TECHNISCHE UNIVERSITÄT MÜNCHEN
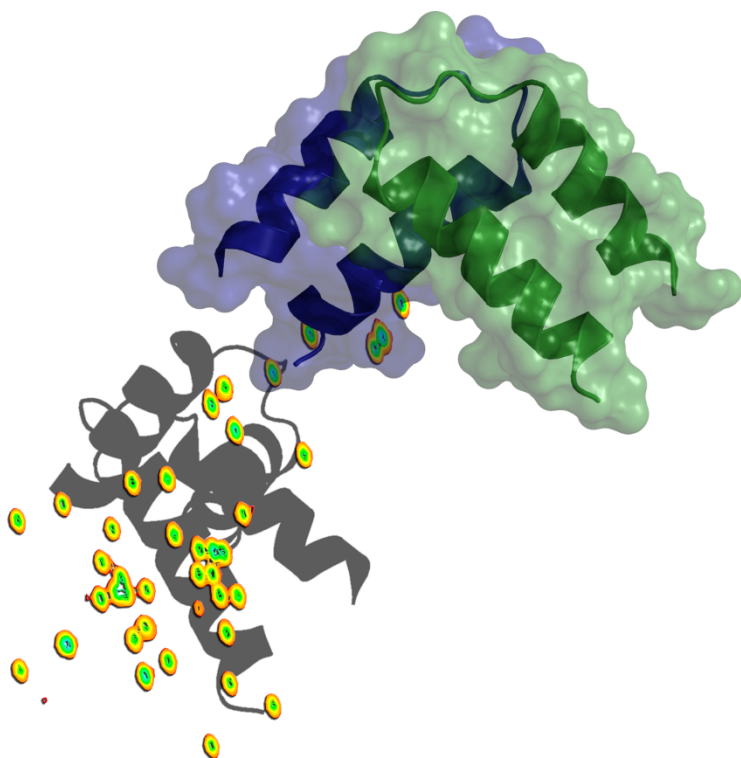
Lehrstuhl für biomolekulare NMR-Spektroskopie, Department Chemie

Structural basis for molecular recognition and function of proteins in alternative mRNA splicing and host-parasite immunobiology

Niels Helge Meyer

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für biomolekulare NMR-Spektroskopie, Department Chemie

Structural basis for molecular recognition and function of proteins in alternative mRNA splicing and host-parasite immunobiology

Niels Helge Meyer

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. Christian F. W. Becker

Prüfer der Dissertation: 1. Univ.-Prof. Dr. Michael Sattler

2. apl. Prof. Dr. Ruth Brack-Werner,

Ludwig-Maximilians-Universität München

Die Dissertation wurde am 20.01.2011 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 03.03.2011 angenommen.

Habe nun, ach! Philosophie, / Juristerei und Medizin,
Und leider auch Theologie / Durchaus studiert, mit heißem Bemühn.
Da steh ich nun, ich armer Tor! / Und bin so klug als wie zuvor;
Heiße Magister, heiße Doktor gar / Und ziehe schon an die zehen Jahr
Herauf, herab und quer und krumm / Meine Schüler an der Nase herum –
Und sehe, dass wir nichts wissen können! / Das will mir schier das Herz verbrennen.

(Faust I, Johann Wolfgang von Goethe)

TABLE OF CONTENTS

SUMMARY

Proteins play a fundamental role in nearly all biological processes by interacting with other proteins, nucleic acids or small molecules. The interplay of proteins and macromolecules can lead to the assembly of structurally complex and highly dynamic molecular machines, with diverse functions in gene expression, cell growth, cell cycle, metabolic pathways, signal transduction, protein folding and transport. The nature of these interactions is governed by the three-dimensional structures of the proteins involved. Therefore, research into the structure and function of proteins (structural biology) has become indispensable in life sciences. Besides X-ray crystallography, Nuclear Magnetic Resonance (NMR) spectroscopy is applied to determine the structure of biological macromolecules. Moreover NMR is essentially the only method to detect dynamics on a wide range of timescales, ranging from ns to days, on sub-molecular level in solution.

The present thesis delivers two conclusive examples, where knowledge of the protein structure has provided critical insight into the function, and accordingly, the rationale behind cell biology assays to validate the structural findings. The first study reveals novel structural aspects in the regulation of alternative mRNA splicing regulation by the protein Sam68. The second study addresses the question of how *Schistosoma mansoni (S. mansoni)*, a parasitic worm, can manipulate and exploit the host immune system by only a single glycoprotein named interleukin-4-inducing principle of *S. mansoni* eggs (IPSE/alpha-1).

Chapter 1 introduces NMR as an important tool in structural biology. Additionally, certain aspects of gene regulation and immunobiology, which are related to the structural projects of the thesis, are reviewed.

Chapter 2 provides a structural and functional characterization of the Src associated during mitosis, 68 kDa protein (Sam68). Sam68 is a member of the signal transducer and acativator of RNA (STAR) domain family of proteins, which regulate certain aspects of RNA metabolism, e.g. alternative mRNA splicing. Typically, the STAR domain has a Qua1-KH-Qua2 domain organisation. The solution structure of its Qua1 homodimerization domain was determined by NMR spectroscopy. The two monomers assemble perpendicular with respect to each other in an unusual arrangement of four helices. This interaction is critical for the function of Sam68 in alternative RNA splicing, as revealed by a cell-based mutational assay. Similar to its sequence homolog human splicing factor 1 (SF1), the RNA-binding K homology (KH) domain is extended by the Qua2 domain, which contacts the target RNA. Unlike SF1, a novel N-terminal extension of the KH-fold, may contribute to the RNA binding of Sam68. These findings could explain the differential RNA binding specificities of the two proteins. Binding of Sam68 to its target RNA guides the constitutive splicing factor U2 auxiliary factor, 65 kDa (U2AF65) to alternative splice sites. NMR titration experiments show, that U2AF65 can bind to the C-terminus of Sam68. This interaction most likely stabilizes the binding of U2AF65 to pre-mRNA regulatory sequences and initiates the formation of the spliceosome at weak splice sites. These data conclusively correlate the structure of Sam68 to its cellular function and reveal how Sam68 regulates the targeting of the spliceosome to alternative splice sites.

Chapter 3 describes the structural analysis of the interleukin-4-inducing principle of *S. mansoni* eggs (IPSE/alpha-1), a major immunogenic component of *S. mansoni*. The stimulation of a T-helper 2 cell (T_H2) response of the host's immune system upon infection with *S. mansoni* involves the specific binding of IPSE/alpha-1 to IgE antibodies on the surface of basophilic granulocytes, which is not mediated by the canonical antigen-antibody interaction. I was able to solve the solution structure of the IPSE/alpha-1 and show that it is a protein of the β/γ-crystallin family. It bears unique structural features which enable the protein to specifically bind to Immunoglobulin (Ig) E. I identified a large positively charged interaction interface on IPSE/alpha-1, presumably binding to a negatively charged patch, provided by the asymmetrically bent IgE molecule. This binding supports a model of basophile activation, which is triggered by the binding of IPSE/alpha-1 to receptor-bound IgE. Presumably, a conformational changes of the IgE molecule leads to subsequent IgE-receptor activation. The implications from this specific protein-protein interaction enhance our understanding on the mechanisms involved in the manipulation of the host's immune system by *S. mansoni*.

ZUSAMMENFASSUNG

Die Interaktion von Proteinen mit anderen Proteinen, Nukleinsäuren und kleinen Molekülen ist die Grundlage nahezu jedes biologischen Prozesses. Das Zusammenspiel von Proteinen und Makromolekülen erlaubt die Bildung komplexer Strukturen bis hin zu dynamischen molekularen Maschinen. Diese üben unterschiedlichste Funktionen in der Genexpression, im Zellzyklus und -wachstum, im Stoffwechsel, in der Signalweiterleitung sowie in der Proteinfaltung und im Proteintransport aus. Die Art der Wechselwirkungen wird maßgeblich von der räumlichen Struktur der beteiligten Proteine bestimmt. Die Aufklärung der Raumstruktur von Proteinen und deren Funktion (Strukturbiologie) ist daher ein unverzichtbarer Bestandteil der Lebenswissenschaften. Neben der Röntgenstrukturanalyse wird die Kernspinresonanz- (engl. *Nuclear Magnetic Resonance, NMR*) Spektroskopie zur Strukturbestimmung von biologischen Makromolekülen verwendet. NMR ist dabei die einzige Methode zur Untersuchung submolekularer, dynamischer Vorgänge auf einer Zeitskala von Nanosekunden bis Tagen in Lösung.

Die vorliegende Doktorarbeit beschreibt anhand zweier Beispiele, dass die Kenntnis der Raumstruktur eines Proteins kritische Einblicke in dessen Funktion geben kann. Auf dieser Grundlage können zellbiologische Experimente entworfen werden, um die gewonnen Erkenntnisse über den Zusammenhang von Struktur und biologischer Funktion zu überprüfen. Die erste Studie zeigt neue, strukturelle Aspekte in der Regulation des alternativen Spleißens von Boten-RNA (mRNA) durch das Protein Sam68 auf. Die zweite Studie befasst sich mit der Frage, wie der parasitäre Wurm *Schistosoma mansoni* das Immunsystem seines Wirtes mittels eines einzigen Proteins, IPSE/alpha-1, manipulieren und zu seinen Zwecken ausnutzen kann.

Kapitel 1 beschreibt die NMR als wichtiges Werkzeug der Strukturbiologie. Darüber hinaus werden verschiedene, für die Projekte dieser Doktorarbeit relevante Aspekte der Genregulation und Immunbiologie erörtert.

Kapitel 2 liefert die strukturelle und funktionale Charakterisierung des Proteins Sam68 *(Src associated during mitosis, 68 kDa protein)*. Sam68 ist ein Mitglied der STAR *(Signal transducer and activator of RNA)*-Domänenfamilie, die mit der Regulierung verschiedener Aspekte des RNA Metabolismus, wie zum Beispiel dem alternative Spleißen von mRNA, assoziiert ist. Typischerweise weist die STAR Domäne eine Qua1-KH-Qua2 Domänen-organisation auf. Die Raumstruktur der Qua1 Homodimerisierungsdomäne von Sam68 in Lösung wurde mittels NMR Spektroskopie bestimmt. Die beiden Monomere lagern sich rechtwinklig zueinander an und bilden so eine ungewöhnliche Anordnung von vier Helices. Durch Mutationsanalyse in einem zellbasierten Assay konnte gezeigt werden, dass diese Interaktion entscheidend für die Aktivität von Sam68 im alternativen mRNA-Spleißen ist. Die RNA-bindende KH *(K homology)* Domäne ist analog zum homologen Protein SF1 *(splicing factor 1)* durch die Qua2-Domäne erweitert. Im Gegensatz zu SF1 ist die KH-Faltung in Sam68 auch N-terminal erweitert. Die Bindung von Sam68 an die Ziel-RNA rekrutiert den konstitutiven Spleißfaktor U2AF65 *(U2 auxiliary factor, 65 kDa)* zur Spleißstelle. NMR Titrationsexperimente zeigen, dass U2AF65 an den C-Terminus von Sam68 binden kann. Diese Interaktion stabilisiert sehr wahrscheinlich die Bindung von

U2AF65 an Regulationssequenzen auf der prä-mRNA und trägt dadurch zur Assemblierung des Spleißosoms an schwachen Speißstellen bei. Die Ergebnisse liefern eine Korrelation der Struktur von Sam68 mit seiner zellulären Aktivität für die Erkennung alternativer Spleißstellen.

Kapitel 3 beschreibt die Strukturanalyse des Proteins IPSE/alpha-1 *(interleukin-4-inducing principle of S. mansoni eggs)*, welches ein Hauptimmunogen von *S. mansoni* ist. Die Stimulierung einer T-Helfer-2-Zellen (T_H2)-Antwort des Wirtimmunsystems in Folge der Infektion mit *S .mansoni* geht mit der spezifischen Bindung von IPSE/alpha-1 an Immunoglobulin (Ig) E Antikörper auf der Oberfläche basophiler Granuloyzten einher. Diese Bindung ist jedoch keine typische Antigen-Antikörper-Reaktion. In der vorliegenden Arbeit konnte ich die Struktur von IPSE/alpha-1 in Lösung aufklären. Meine Ergebnisse zeigen, dass IPSE/alpha-1 eine β/γ-Crystallin-Faltung einnimmt. Die strukturellen Merkmale des Proteins ermöglichen die spezifische Bindung an den IgE Antikörper. Die Kristallstruktur von IgE zeigt einen großflächigen negativ geladenen Bereich, der an eine positiv geladenen Interaktionsfläche auf IPSE/alpha-1 bindet. Dieser Bindungsmodus unterstützt das Modell einer Basophilenaktivierung, die durch eine Interaktion von IPSE/alpha-1 mit Rezeptor-gebundenem IgE ausgelöst wird. Möglicherweise aktiviert eine konformationelle Änderung des asymmetrischen IgE-Moleküls schließlich den IgE-Rezeptor. Die Schlussfolgerungen aus dieser spezifischen Protein-Protein-Wechselwirkung erweitern unser Verständnis über die Mechanismen, die zur Manipulation des Wirtsimmunsystems durch *S. mansoni* führen.

CHAPTER 1:

INTRODUCTION

Introduction to structural biology

In the past decades the increasing number of available genomes has conveyed a virtually complete list of genes and thus their protein products in a given organism. The human genome for example consists of up to 30,000 genes, which may code for over 500,000 proteins [1]. However, when the Human Genome Project delivered its draft of the human genome in 2001 and finally declared the genome to be completed in 2007, it was already obvious that the knowledge of genomic sequence alone could not fulfill the high scientific expectations linked to this project. To understand even the simplest cellular processes a detailed structural knowledge of the involved proteins, their function and moreover their specific interactions is necessary. Proteins rarely act alone; it has been estimated that 80% of all proteins are only active in complex with other proteins [2]. This is the basis of structural biology which aims at understanding the function of a protein by investigating its three dimensional structure at atomic resolution. Instead of focusing on the structure of individual proteins the concept of molecular system biology tries to reveal the properties of an entire biological system, i.e. elucidating the spatial and temporal interactions of the involved macromolecules. This approach involves analysis of the dynamic structure of proteins, the specificity of protein-protein, protein-nucleic acid or protein-metabolite interactions and the pathways of complex networks. Nuclear magnetic resonance (NMR) has turned out to be a key technique, capable of addressing those important questions under near physiological conditions.

NMR Spectroscopy

Nuclear Magnetic Resonance relies on an intrinsic quantum mechanical property of certain nuclei: the nuclear spin. Nuclear spins interact with electromagnetic waves at their resonance frequency, if an external static magnetic field is applied. The response of the excited spins is monitored in the free induction decay (FID). The Fourier Transform of the FID (i.e. the signal intensity as a function of time) provides the NMR spectrum, which is represented as signal intensity as a function of resonance frequency. Nuclei observed by routine NMR methods of biological molecules are typically those, which have a spin quantum number of 1/2, for example ^1H and ^{13}C. An NMR signal is characterized by the position in the spectrum, referred to as chemical shift, the signal intensity, and the line shape. The electronic environment of a nuclear spin influences its chemical shift, which typically results in a broad signal dispersion. ·

Spins which interact with each other are referred to as coupled spins. The terms scalar coupling and J-coupling have been introduced for coupling which is mediated via electrons in chemical bonds. The resonance frequencies of coupled spins can be correlated in two- or more dimensional NMR experiments. Chemical shifts and scalar couplings have been extensively exploited by chemists to characterize organic molecules. Relaxation, which is the return of magnetization to its equilibrium state, has been used to extract structural and dynamical information of biological macromolecules.

## *NMR Spectroscopy of biological macromolecules*

In the past 50 years NMR spectroscopy has become a powerful tool to study biomolecules, such as proteins. Besides X-ray crystallography, it is the only method capable of solving protein structures on an atomic level. In some aspects NMR spectroscopy has an advantage over X-ray crystallography. It does not rely on crystallization of the protein, which can be a tedious process, not seldom failing for a particular target of interest. Especially the co-crystallization of protein complexes is challenging. In cases where the interactions between the proteins are transient, NMR is essentially the only approach to gain structural insight in the complex formation. Moreover NMR spectroscopy is a unique technique to study protein dynamics at atomic resolution.

In its early stage NMR spectroscopy was limited to proteins reasonably smaller than 10 kDa and preferably with a high content of beta sheets. Sequence specific resonance assignment of the backbone and side chain protons of bigger proteins was simply not possible due to increasing overlap in homonuclear proton NMR spectra. ^{15}N- and ^{13}C- isotope enrichment and the development of heteronuclear triple resonance experiments (reviewed by Sattler *et. al.* [3]), solved the problem of overlap and extended the size limitation to proteins up to 20 kDa. Triple resonance experiments correlate the proton resonance frequencies with ^{15}N- or ^{13}C-resonances, in order to obtain a maximum of spectral resolution in a multidimensional NMR experiment. However, for proteins with molecular weights larger than 20 kDa, sensitivity of these experiments declines, due to faster relaxation of magnetization (see next paragraph for details). Perdeuteration of proteins [4, 5] and selective labeling schemes [6] allow the study of even higher molecular weight proteins and protein complexes due to more favorable relaxation properties, which result in sharper NMR signals. Development of modern NMR spectrometers with ever growing field strength in combination with advanced pulse sequences to date allows the analysis of complexes of molecular weights as high as 300 kDa in solution.

NMR relaxation

If an external magnetic field is applied to an ensemble of spin 1/2 nuclei, a macroscopic magnetization is generated according to the Boltzmann distribution. In equilibrium the magnetization is aligned along the magnetic field. By radiofrequency irradiation this magnetization can be converted into magnetization perpendicular to the static magnetic field. Equilibrium is reestablished via two distinct processes: Recovering of magnetization along the static magnetic field (T_1 relaxation) and decay of magnetization in the transverse plane (T_2 relaxation), as depicted in Figure 1.1A.

The major contribution to relaxation arises from the rotational motion of the system under investigation, which is characterized by the correlation time τ_c. Figure 1.1B shows the dependency of T_1 and T_2 relaxation time as a function of the correlation time. The most important relaxation mechanisms in liquids are dipolar interactions (DD) and the chemical shift anisotropy (CSA). T_2 relaxation leads to a broadening of lines in the NMR spectra of macromolecules. Relaxation has been commonly used to characterize dynamical features of proteins and other macromolecules. The relaxation properties of ^{15}N have been found to be suitable to reliably assess molecular dynamics in macromolecules on timescales ranging from ps to s, and to calculate τ_c.

A

B
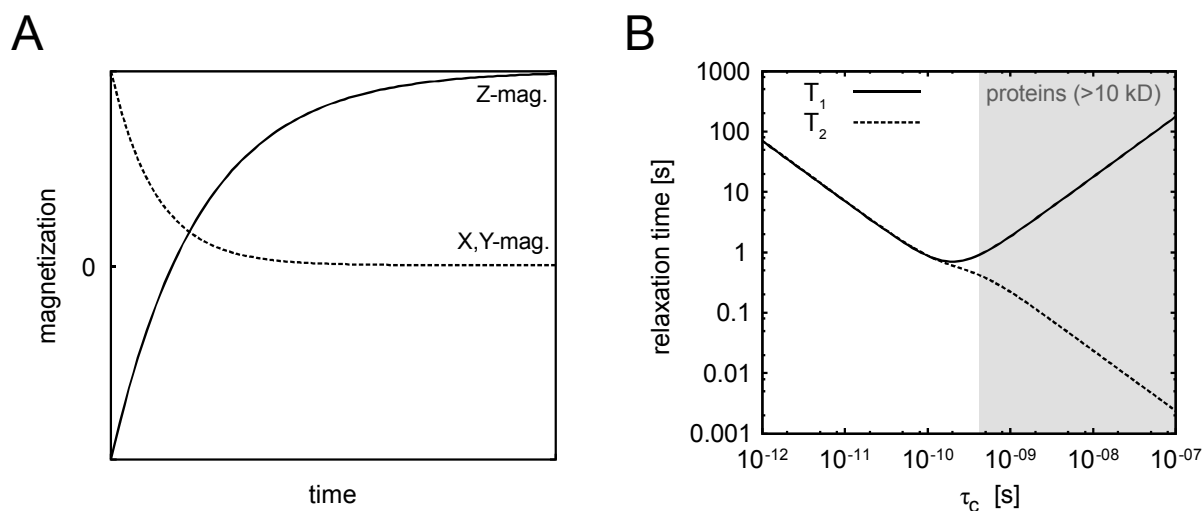
Figure 1.1: Relaxation in NMR Spectroscopy. A) Return of Z- and X,Y-magnetization to its equilibrium state. **B)** Dependency of 1H T1- and T2-relaxation on the correlation time τ_c at a magnetic field strength (B^0) of 19 T, corresponding to a proton lamour frequency of 800 MHz.

NMR structure determination

NMR structure determination in solution mainly relies on an effect known as nuclear Overhauser enhancement (NOE), which is a distance dependent interaction between two NMR active nuclei based on magnetic dipole-dipole interaction. The strength of the NOE is proportional to r^{-6}, with r being the distance of the two nuclei. In a NOESY type of experiment protons which are closer than 6 Å usually give rise to a cross peak [7]. The distance of the protons is extracted from the volume of the respective peak. To reliably determine a solution structure thousands of those cross peaks have to be identified correctly: a cumbersome process which is additionally hampered by a lot of ambiguities in the NOE assignment. The distance information derived from NOEs can be supplemented by dihedral angles, derived from J-couplings [8, 9] or chemical shifts [9] as well as residual dipolar couplings (RDCs) [10]. Also the use of paramagnetic tags or paramagnetic solvents to derive structural information has been described [11]. Parameters extracted from NMR experiments, which can be used in structure calculations are summarized in Table 1.1. The available NMR data is typically used in a restrained molecular dynamics (MD) simulation together with other known parameters, for example bond lengths, bond angles or Van-der-Waals radii. The restraints are translated into energy terms which are used to minimize a target function, containing the potential energy of the system and the restraint energies. Straightforward minimization raises the problem of becoming trapped in a local minimum of the energy

landscape. To avoid this, a set of different starting structures, usually around 100, is optimized in a simulated annealing (SA) protocol. An NMR structure is usually represented by an ensemble of the 10-20 lowest energy structures of the simulation.

The first NMR structure, the 58 aa proteinase inhibitor IIA from bull seminal plasma, was solved in the group of Kurt Wüthrich in 1984 [12]. Back then, structure determination was a process of several years. Resonance assignment, as well as NOE assignment had to be performed manually: A challenging process, since only two dimensional NMR spectra were available at that time. Nowadays higher field strength and more sensitive probeheads allow the routine acquisition of multidimensional NMR spectra in a few days. This has not only speeded up the resonance assignment dramatically but has also allowed the development of automated structure calculation strategies. Commonly used programs for automated structure determination are for example CYANA [13] and ARIA [14], which combine NOE assignment and structure calculation in an iterative process. Preliminary structures based on a limited number of unambiguous NOEs are used to reduce ambiguities in the cross peak assignment of the next iteration. Taken together these advances in NMR based structure determination, the structure of a protein up to 20 kDa can be determined within a few months. However, a thorough manual analysis of the results is still necessary in any case to ensure the correct interpretation of NMR data and to obtain reliable NMR structures. Several algorithms have been used to judge the quality of an NMR structure. Among those PROCHECK [15] and WHATCHECK [16], which compare a given structure with structures in a database, are the most widely used. Recently a new software, CING [17], offers an interface to gather the output of several tools, which not only analyze structural properties but also check the consistency of the restraints used in the structure calculation.

Table 1.1: Experimental restraints used in NMR structure calculations

Measured Parameter	Restraint
NOEs	Distance restraints (<6 Å)
Chemical shift	Backbone dihedral angles
J-coupling	Dihedral angles, H-bonds
RDCs	Orientational restraints
PREs	Long distance restraint (20-30 Å)
Hydrogen exchange rates	H-bonds
Relaxation	Dynamics (correlation time, internal motion)

Protein-Protein interaction

A variety of tools have been successfully established to screen for protein-protein interactions *in vivo* and *in vitro*. Among those, the two hybrid system, chemical cross-linking and the use of affinity tags are the most widely used. The major drawback of these techniques is their high rate of false positives. Another issue is the bias towards high affinity interactions, whereas transient complexes can hardly be identified. The results of the screens have to be confirmed by other methods in order to validate the interaction. Isothermal calorimetry (ITC) and more recently surface plasmon resonance (SPR) have become routinely available not only to validate a potential interaction but also to characterize its thermodynamic and kinetic parameters. Besides, fluorescence and circular dichroism (CD) spectroscopy are used to verify protein-protein interactions. However, spectroscopic techniques usually require a lot of know-how and their success greatly depends on the system under investigation. NMR spectroscopy certainly is the most versatile spectroscopic method to study protein complexes.

Rapid screening for interactions as well as determination of binding affinities belongs to the standard repertoire of NMR applications. Moreover, NMR is the only tool available providing detailed structural information on the binding interface in solution. Classically, the NOE has been used to obtain intermolecular distance restraints, which define the protein interaction interface. In addition, chemical shift perturbation (CSP), residual dipolar coupling (RDC) and paramagnetic relaxation enhancement (PRE), explained in the following paragraphs, have been applied to define protein complexes.

The fingerprint spectrum and chemical shift perturbation

The ^{1}H,^{15}N HSQC experiment correlates the proton resonance frequency with the resonance frequency of its attached nitrogen atom. In a protein ^{1}H,^{15}N HSQC spectrum each backbone amide gives rise to a single peak. Side-chains containing ^{15}N-bound protons (i.e. Glutamine, Asparagine, Lysine, Arginine, Tryptophan and Histidine) can give rise to additional peaks. Proline is the only invisible residue, since it contains no amide group. The number of peaks in a ^{1}H,^{15}N HSQC spectrum is very limited and the peaks are reasonably dispersed. Thus it is characteristic for every protein and therefore referred to as fingerprint spectrum. Moreover the resonance frequency of the amide proton and nitrogen is very sensitive to changes in the chemical micro-environment. Any structural changes of the protein can therefore be probed easily by monitoring the changes in the ^{1}H,^{15}N HSQC spectrum. Likewise binding of the protein to a ligand, such as another protein, nucleic acids or even small molecules causes changes in the spectrum. Those changes are called chemical shift perturbations (CSPs) and are ideally localized in the interaction interface. However, secondary effects, i.e. structural rearrangements induced upon binding, can lead to a blurring of the interface determined by this method.

Intermolecular NOEs in structure determination of protein complexes

Similar to the determination of the structure of an individual protein, high resolution structures of protein complexes can be obtained using NOEs in a restraint molecular dynamics simulation. This approach usually involves protein samples which contain only one isotopically labeled protein whereas the other proteins of the complex are unlabeled. Therefore, the ambiguity of NOE assignment is reduced, which allows the reliable identification of intermolecular contacts by using isotope edited or filtered NOESY experiments [18]. In an edited spectrum only peaks of protons, which are bound to an isotopically labeled heteroatom, usually ^{15}N or ^{13}C, are visible. A filtered spectrum shows only peaks of protons which are bound to ^{14}N or ^{12}C (Figure 1.2).

Although those spectra are usually not more complex then the spectra of the individual domains, analysis of NOESY spectra of big protein complexes can be troublesome. More efficient relaxation, due to the increased molecular weight, not only leads to reduced signal intensity, but can also give rise to cross peaks between signals of protons, which are not close in space. This effect is known as spin diffusion. However, even for big complexes it might be still possible to define a protein-protein interface by looking at some unambiguous NOEs, e.g. NOEs between amide protons, which can also be detected in perdeuterated proteins.
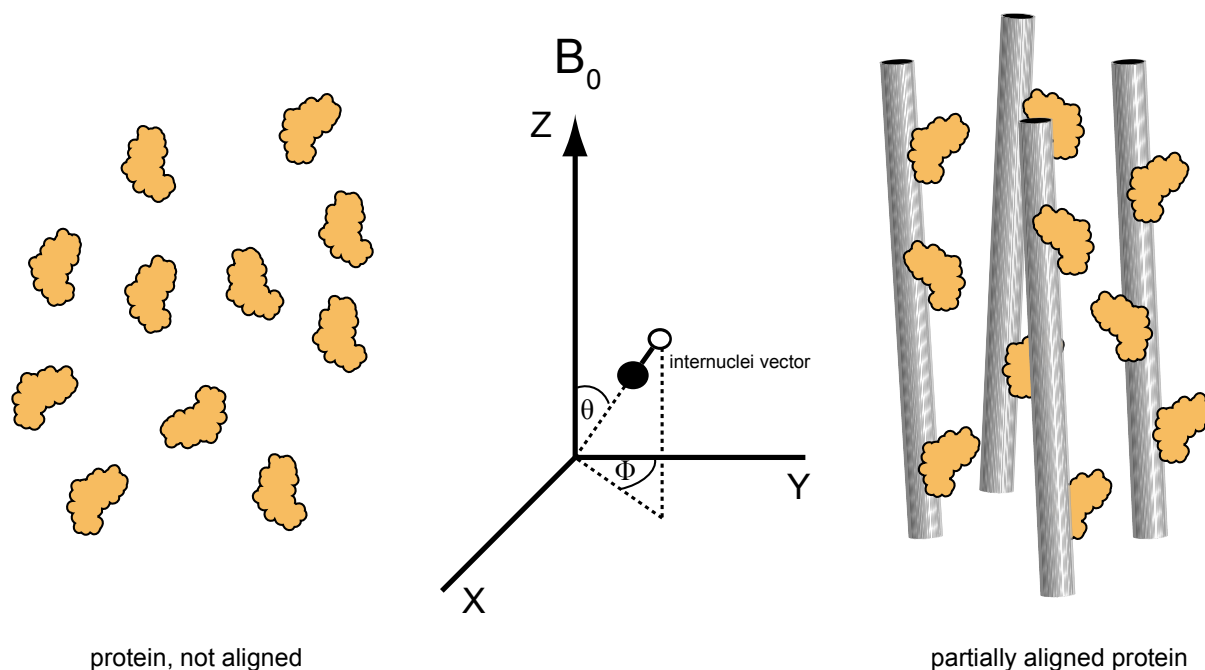
homonuclear NOESY

^{13}C -edited NOESY

^{13}C-filtered and edited NOESY

^{13}C double-filtered NOESY
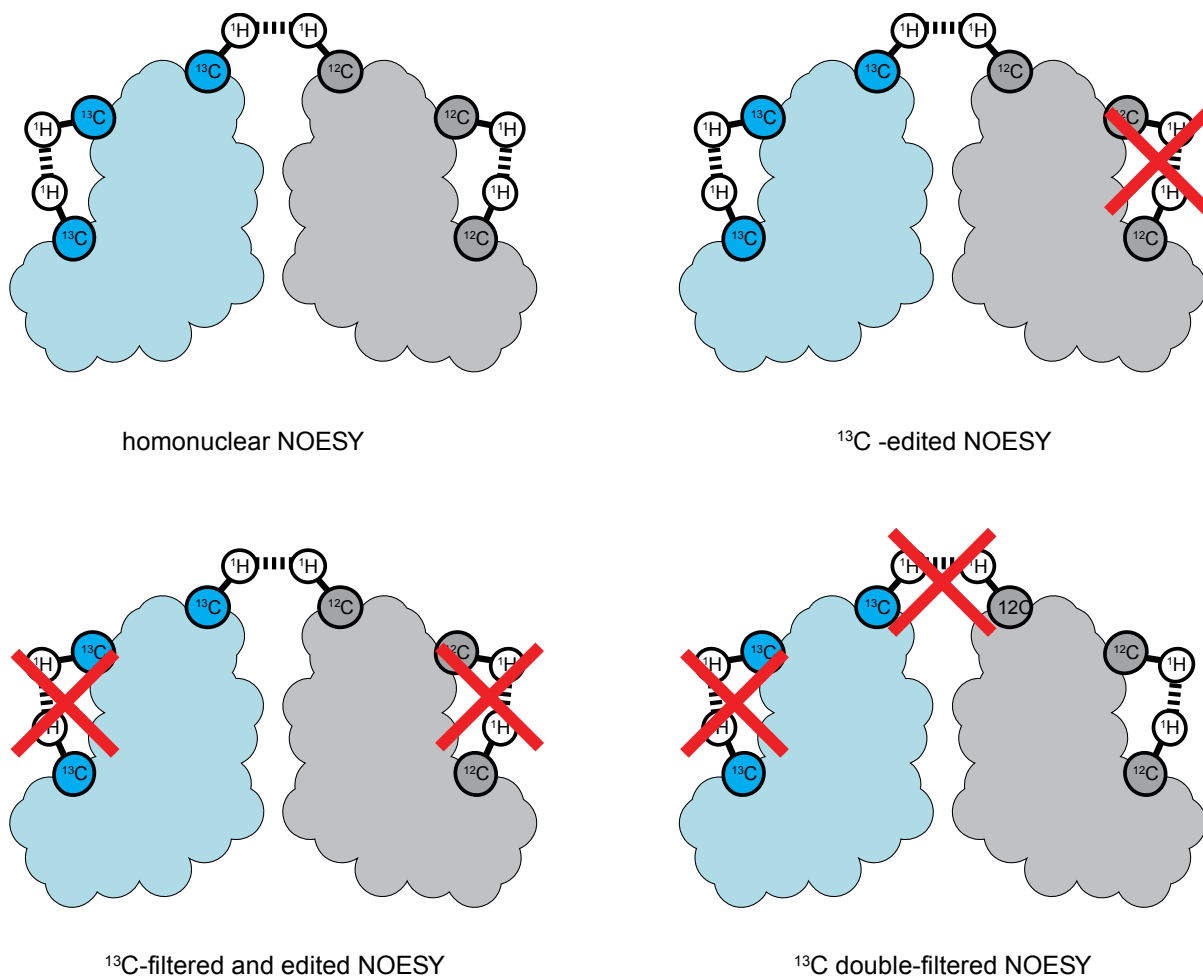
Figure 1.2: Possible magnetization transfer pathways, which give rise to cross peaks in ^{13}C-edited or -filtered NOESY experiments of a protein complex. ^{13}C uniformly labeled and unlabeled protein is depicted in blue and gray, respectively.

Residual dipolar couplings and domain orientation

Dipolar interaction is not only a function of the distance between the involved nuclei, which can be measured as NOE (as described above). It is also influenced by the orientation of the vector connecting a pair of proximal nuclear spins with respect to the external magnetic field. In solution this dependency averages to zero due to free tumbling of the molecules. Nevertheless this effect can be useful to obtain structural information, if the sample is partially aligned in the magnetic field (Figure 1.3). Yet, too strong alignment results in extensive line broadening and therefore loss of signal intensity. Several media have been successfully applied to align protein samples as reviewed by Prestegard and Kishore [19].

$$D_{IS} = -\frac{\gamma_I \gamma_S \mu_0 \hbar}{4\pi^2 r_{ij}^3}\left[A_a \frac{1}{2}(3\cos^2\theta - 1) + \frac{3}{4}A_r \sin^2\theta \cos 2\phi\right]$$

protein, not aligned

partially aligned protein

Figure 1.3: Schematic representation of protein molecules (orange) tumbling free in solution and aligned in an alignment medium (gray cylinders). The equation on top of the illustration defines the dependencies of the RDCs The size of an RDC between two spins I and S (D_{IS}) depends on the angles θ and Φ, which are defined by the orientation of the internuclei vector with respect to the static magnetic field, the axial component A_a and rhombicity A_r of the alignment tensor, defined by the partial alignment of the molecules as well as the distance of the two nuclei r. Furthermore, γ is the gyromagnetic ratio, $\hbar$ is the Planck's constant divided by 2π and μ_0 is the magnetic dipolar momentum.

Dipolar effects in alignment media are referred to as residual dipolar couplings (RDCs) and are used to obtain structural information, mainly on the backbone conformation of proteins. RDCs also have been found to be extremely helpful to analyze protein complexes. If high resolution structures are already available for the individual proteins, RDCs can be used to refine the conformation of those proteins when bound in the complex. Additionally, the orientation of the proteins in the complex with respect to each other can be derived. RDCs, however, do not give information on the protein-protein interface. This has to be extracted by other means, e.g. CSP or a few unambiguous NOEs. Taken together this information allows to obtain high resolution structures, even with only sparse NOE data.

Paramagnetic Tags and solvent PREs

Paramagnetism is a property of compounds, which contain unpaired electrons. This may be chemicals containing a stabilized radical or transition metal ions. Paramagnetism has been used in Electron Paramagnetic Resonance (EPR). Since the magnetic moment of electrons is almost three orders of magnitude larger than that of protons, paramagnetic species also have a pronounced effect on NMR spectra. Even nuclear spins far from the paramagnetic center are affected by unpaired electrons. Various methods have been developed to extract long distance restraints by analyzing the distance dependency of paramagnetic effects on nuclear spins. The

presence of a paramagnetic compound can influence the chemical shift in NMR spectra. This effect is termed pseudo contact shift (PCS). PCS can yield both orientational and distance restraints. Although in practice PCSs might be difficult to obtain, they have been successfully applied in protein NMR [11]. The influence of a paramagnetic center on the relaxation rate of a nuclear spin is referred to as paramagnetic relaxation enhancement (PRE). It describes the enhancement of R_1 and R_2 relaxation rates, which in the case of R_2 leads to a distance dependent broadening of signals in the NMR spectrum (Figure 1.4). Since PREs can be measured much easier than PCSs, they are routinely used in the structural analysis of large protein complexes. Unfortunately biological macromolecules rarely contain paramagnetic centers. Exceptions are for example metal binding proteins. To overcome this limitation several paramagnetic tags for proteins have been developed for EPR and NMR studies. Those tags are usually referred to as spin labels. Most frequently nitroxide containing compounds, which can be coupled to the thiol-group of Cysteine side-chain sulfhydryl groups are used to serve as spin labels. If a modification of the protein is not appreciated or not feasible, PRE can still be useful: Measuring the protein in a paramagnetic solvent allows to define the solvent exposure of NMR active nuclei (e.g. ^{1}H or ^{13}C). For example, this method has been successfully applied to accurately determine protein-protein interaction surfaces, which are excluded from the solvent upon binding [20, 21].
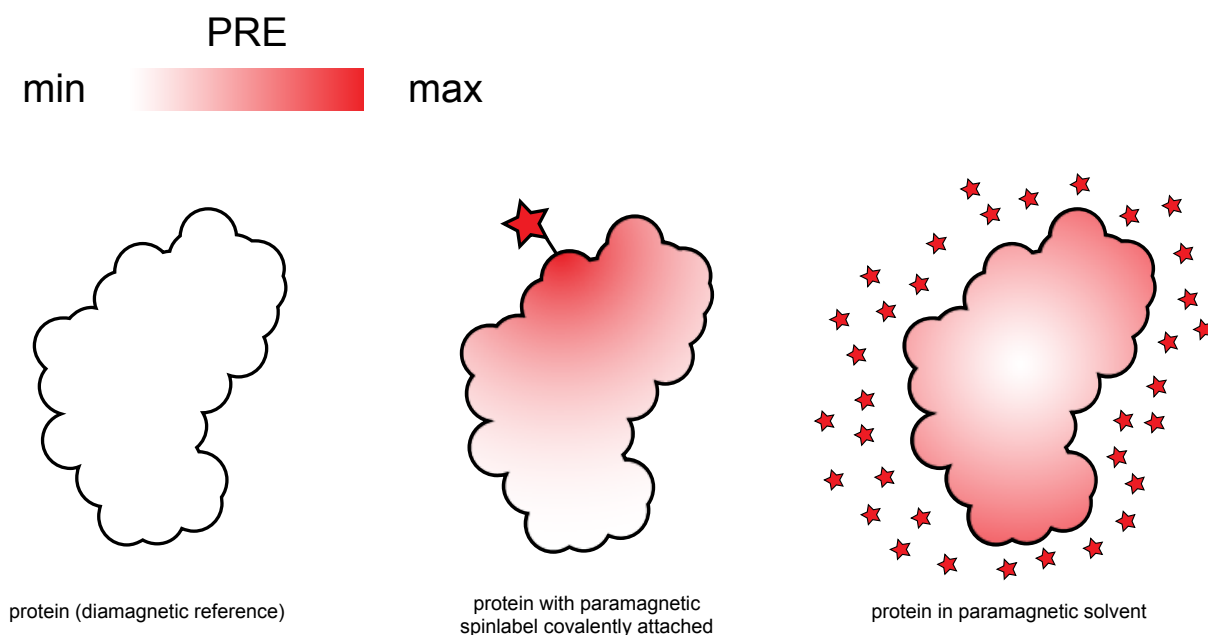
Figure 1.4: Schematic representation of paramagnetic relaxation enhancement in proteins. Spins close the paramagnetic compound (red stars) experience a large PRE.

The PRE is defined as additional contribution to the relaxation rate $\Gamma_{1/2} = R_{1/2}^{para} - R_{1/2}^{dia}$, with $R_{1/2}$ being either the longitudinal or the transverse relaxation rates in the para- and diamagnetic state of the molecule. The PRE and thus the line broadening in the spectrum (R_2) depends on the distance to the spin-label r, whereas the effect decreases with r^{-6}. Thus, distance restraints to the spin label can be extracted from PRE. The use of nitroxide-radicals for example allows to determine distances between 10 and 23 Å from the spin-label.

Introduction to alternative splicing regulation

Processing of genetic information

Genetic information is mostly stored in the form of DNA. DNA is transcribed into messenger RNA (mRNA), the blueprint for translation into proteins. As a cell duplicates, the complete genome is inherited to the daughter cells. Thus, each cell in the organism carries the same genetic information. However, higher organisms require highly specialized cells, which for example can assemble to complex entities such as organs. As a consequence, differentiation into a certain cell type involves regulation of gene expression. This happens on the level of DNA transcription, mRNA translation or, after expression of the protein, via posttranslational modifications. In addition, RNA transport and mRNA degradation are efficient regulatory mechanisms of protein expression. On the one hand this allows the cell to generate an individual proteome, i.e. the set of all proteins expressed from the genome. On the other hand it enables the cell to react instantaneously to extra-cellular triggers and to adapt to the environmental or cell-type specific requirements. It has been shown that entire regions of DNA are packed into highly condensed protein-DNA complexes, the heterochromatin. In general those genes are only poorly expressed. Besides this conclusive silencing of genes, a complex network of protein DNA interaction involving the so-called transcription factors, allow the rapid regulation of gene expression. On the level of mRNA, RNA splicing has been found to generate further diversity of the proteome.

Pre-mRNA splicing

The genome of higher eukaryotic species is organized in a split gene structure: Protein coding sequences (exons) are separated by non coding sequences (introns). This genomic architecture was independently discovered by Sharp and Roberts [22, 23], who both were awarded the noble prize for their pioneering work. Introns have to be removed after transcription of the pre-mRNA to generate the mature mRNA, a process which is termed splicing. Splicing is carried out by the spliceosome, a tightly regulated ribonucleoprotein complex of megadalton size. The spliceosome consists of five small nuclear ribonucleoprotein particles (snRNPs): U1, U2, U4, U5 and U6. Each of them contains a protein associated uridine-rich RNA, and many non-snRNP proteins [24-26]. The catalysis involves two sequential transesterification reactions, which loop out the intron and ligate the two exons [27]. Most human genes contain multiple exons of an average length of 50-250 bp. Introns, however, can be several thousands of bp long [28]. Therefore, splice site recognition takes place at the exon-intron junctions, which share certain sequence characteristics. The 5' splice site contains a GU sequence in a less conserved consensus region. The 3' splice site terminates the intron with AG. The splicing reaction takes place at the so called branch point sequence (BPS) which is preceding a conserved poly pyrimidine (pY) tract. The BPS contains the branch point adenosine, a conserved adenosine base, which is involved in the transesterification. This exon-centered splice site recognition is also referred to as exon definition.
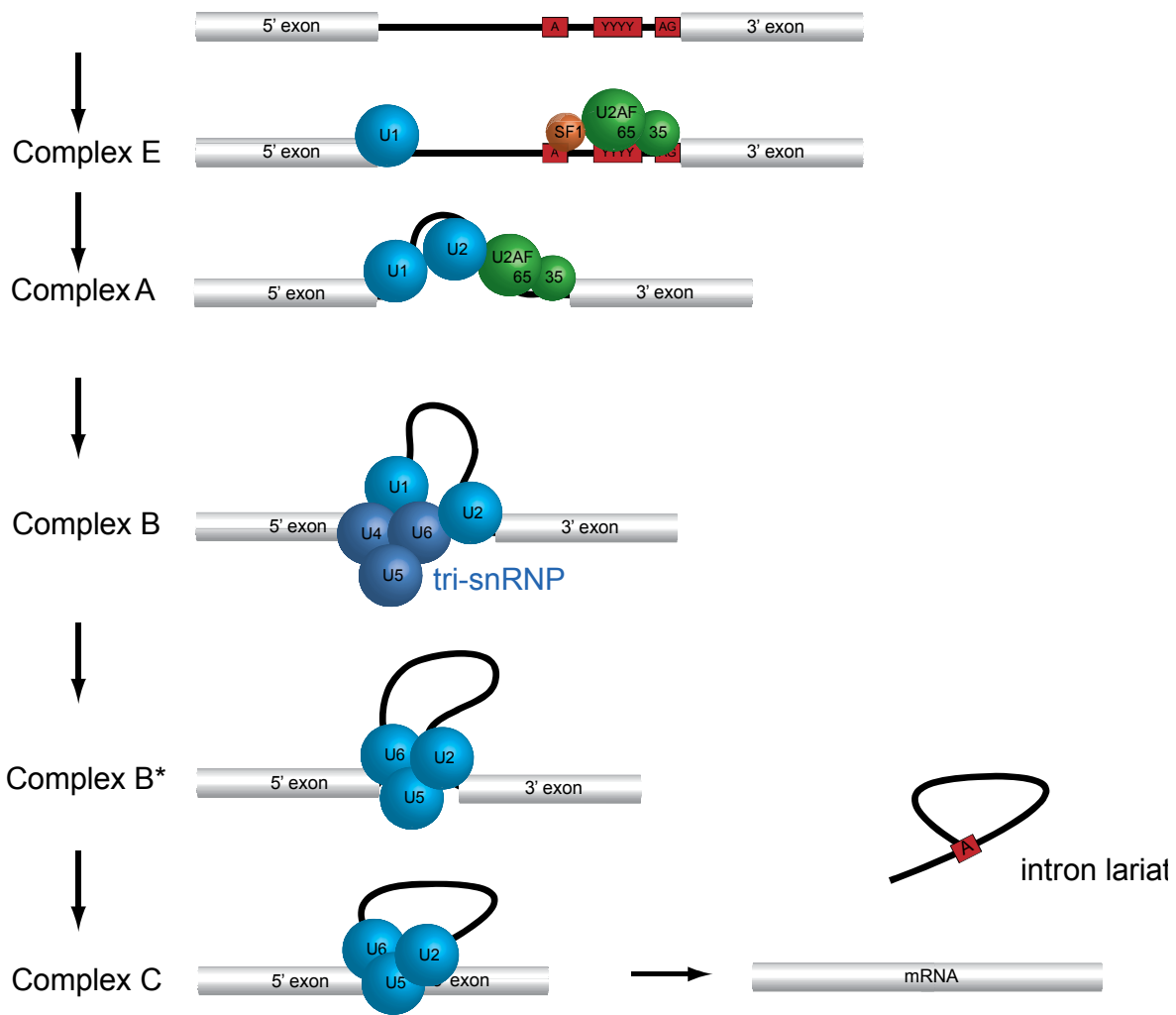
Figure 1.5: Schematic representation of spliceosome assembly. The BPS with the branch point adenosine, the pY tract and the 3' splice site are shown in red. Splicing factors and snRNP subunits of the spliceosome are represented as spheres.

The spliceosome is a sophisticated machinery with a highly dynamic protein/RNA composition. In the canonical pathway, recognition of the 5' splice site by U1 and binding of the heterodimeric U2 auxiliary factor (U2AF) to the 3' splice site initiate spliceosome assembly. Additionally, splicing factor 1 (SF1) targets the branch point sequence. U2AF contains a large subunit, U2AF65, binding to the polypyrimidine tract, and a small subunit, U2AF35, potentially recognizing the 3' splice site. Those events mark the ATP-independent formation of the early (E) complex, which defines the exons. The formation of complex E and the RNA consensus for splice site recognition have been reviewed by Moore [29]. Subsequently the spliceosome undergoes several ATP dependent rearrangements, to form the catalytically active spliceosome. Three distinct complexes A, B and C have been characterized. (Figure 1.5). Conversion of complex E to complex A involves the tight binding of U2 snRNP, which is displacing SF1 from the complex [30, 31]. Thereon the conserved branch point adenosine is bulged out, thus specified as the nucleophile for the first trans-esterification reaction [32]. Association of the U4/U6*U5 tri-snRNP with complex A results in complex B. Cross-intron interactions between U1 and U2 snRNPs convert exon definition into intron definition, since the intron is looped out. Rearrangement of the base pairing interactions between the snRNPs and the 5' splice site displaces U1 and U4 from complex B,

so that an active spliceosome (complex B*) is formed. Complex B* undergoes the first trans-esterification step, which results in the formation of a lariat intron/3' exon intermediate and a free 5' exon [33]. Another step of spliceosomal rearrangement terminates with complex C, the mature spliceosome. In a second catalytic step, the lariat intron is released and the ligation of the two exons is accomplished [34, 35]. The spliceosome is then recycled to undergo another round of splicing activity.

RNA-recognition motif and U2AF homology motif

Although maturation of pre-mRNA requires a variety of different protein factors, the number of involved RNA binding domains is limited. Up to date, the RNA-recognition motif (RRM) is the most frequent domain, which is involved in binding of single stranded RNA [36]. Most RRM domains share an octameric consensus sequence, [RK]-G-[FY]-[GA]-[FY]-[ILV]-X-[FY] (with X being any amino acid), which is called ribonucleoprotein 1 (RNP1) [37, 38]. A second signature sequence, RNP2, with the consensus [ILV]-[FY]-[ILV]-X-N-L, has been identified [39, 40]. Both motifs are embedded in a conserved 80 amino acid domain, which encompasses the RNA-binding function [41-43]. Typically, an RRM domain adopts a $\beta_1\alpha_1\beta_2\beta_3\alpha_2\beta_4$-topology that forms an α/β-sandwich. The four stranded β-sheet contains the RNP1 on strand β_3 and RNP2 on strand β_1. It exposes aromatic residues, which stack with the bases of the RNA. Deviations from this fold have been found in many RRMs, which show N- and C-terminal extensions [44].

Both subunits of U2AF contain atypical sequence variations in the RRM domain. This unusual type of RRM domain has been found in other RRM containing proteins and was grouped in the sub-family of U2AF homology motifs (UHMs). Instead of RNA-binding, this domain mediates protein-protein interaction. The structures of U2AF35-UHM [45] and U2AF65-UHM [46] in complex with their respective ligand revealed several remarkable differences between canonical RRMs and UHMs: They are characterized by the absence of aromatic residues in the RNP motifs, which are critical for RNA recognition Furthermore, an extra C-terminal helix packs against the β-sheet and thus abrogates RNA-binding. Additionally, they contain an extended acidic $\alpha1$-helix as well as a conserved R-X-F motif in the loop connecting α_2 and β_4. These features provide the surface for protein-protein interaction. The UHM ligand motif (ULM) consists of a conserved tryptophan residue which is flanked by positive charged residues upstream and negatively charged residues downstream. Despite the common composition of the ULMs, two-hybrid assays [47] and pull-down experiments suggest, that the interaction between UHM and ULM exhibits a certain degree of specificity [48]. This implies the important role of UHM-ULM interaction in the regulation of RNA metabolism.

Alternative splicing and splice site regulation

Most of the human multi-exon genes can be spliced in multiple ways: Certain exons can be included in or excluded from the mature mRNA. The resulting alternative mRNA products may be translated into protein isoforms with different functions. Figure 1.6 depicts common patterns of alternative splicing. Alternative splicing involves a large number of auxiliary proteins, which assure the correct recognition of diverse splice sites [49].

Splice site sequences in mammals are less conserved than in yeast and are not always recognized by the spliceosome alone [24, 50]. Splicing enhancers, which are specific RNA sequence elements, are mostly targeted by proteins of the SR (Serine-, Arginine-rich) family. These proteins can recruit and stabilize the spliceosome at weakly conserved splice sites. SR proteins are regulated by phosphorylation and dephosphorylation of an Arg-Ser repeat containing (RS) domain [51, 52]. Splicing silencer motifs on the RNA are targeted by proteins to repress splice site selection. Proteins, which bind to splicing silencers mainly belong to the family of heterogeneous nuclear ribonucleoprotein (hnRNP) family. Those proteins can sterically block the access of snRNPs or positive regulatory factors to their RNA target. Since some splicing silencers can bind over 100-200 bp away from the enhancer binding sites, alternative modes of action beyond the 'bind and block' model have been discussed [53]. Combinatorial effects of activators and inhibitors constitute a complex regulatory network for alternative splicing [49] and play an important role in the regulation of gene expression during sex determination [54], nervous system differentiation [55, 56] and apoptotic cell death [57]. However, the limited number of hnRNP and SR proteins, and the resulting limited RNA binding specificity cannot fully explain the vast amount of alternatively spliced pre-mRNAs. Additional proteins are likely required to ensure an adequate specificity for RNA targets and to integrate signal transduction pathways and RNA metabolism. Sam68 has been identified as a candidate protein that regulates alternative splicing in response to environmental cues.
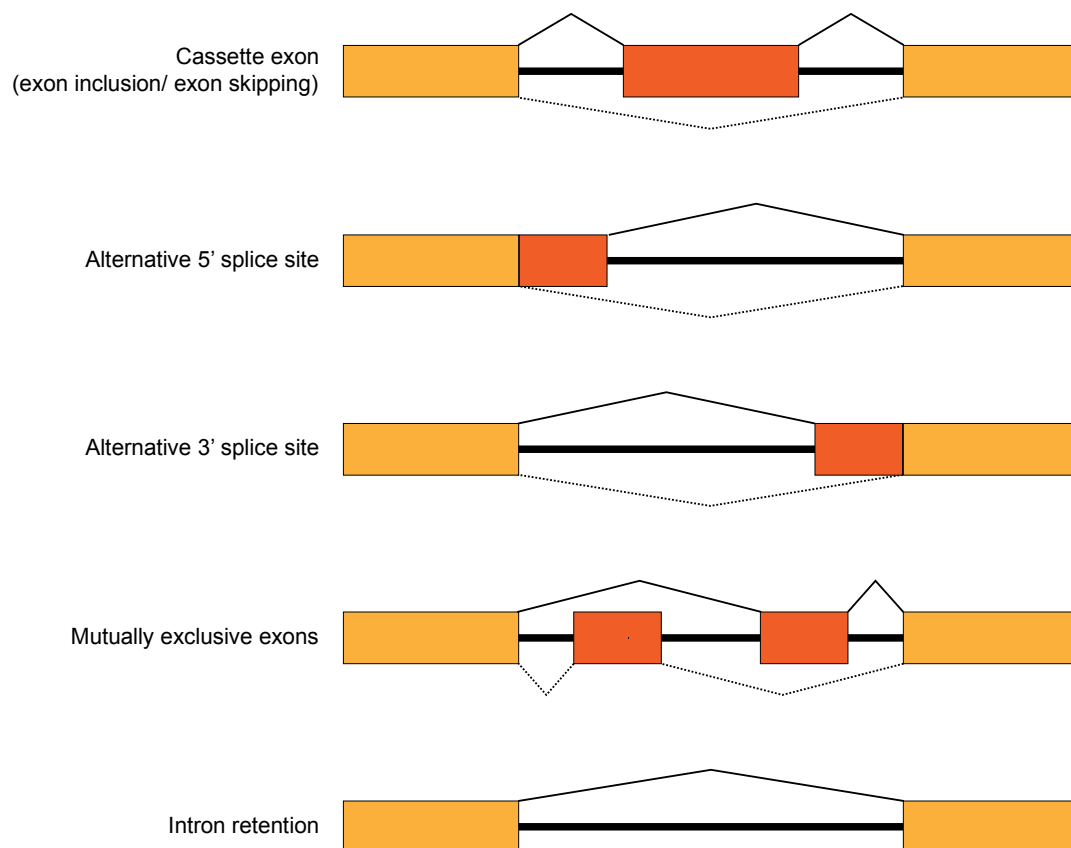

Figure 1.6: Schematic representation of common alternative splicing patterns: exon inclusion or skipping, alternative splice-site selection, mutually exclusive exons, and intron retention. Constitutive exons are coloured in yellow, alternatively spliced exons are shown in orange.
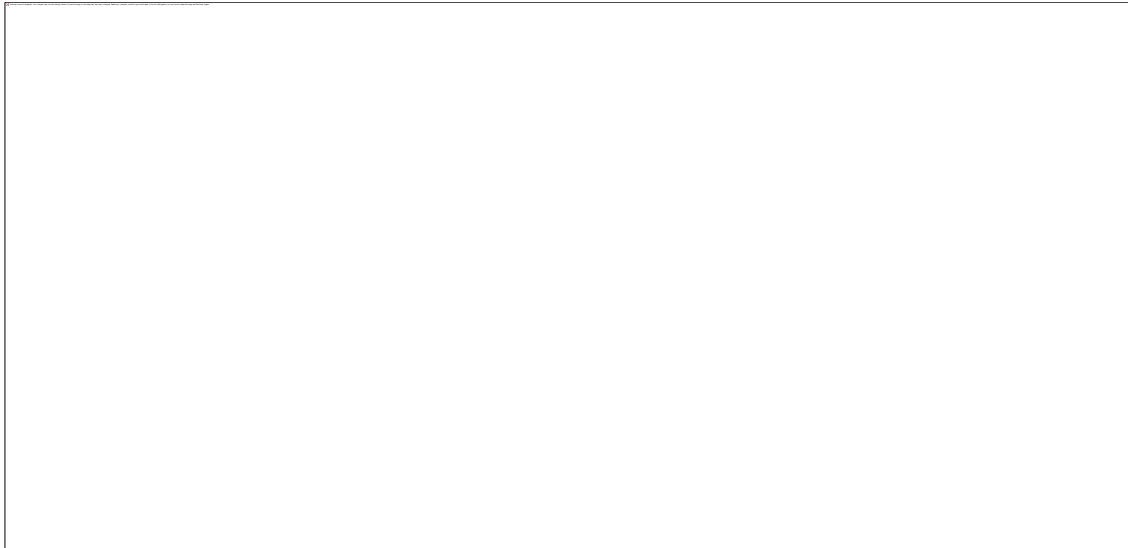
Sam68

The Src-associated during mitosis 68 kDa (Sam 68) protein is a prototypical member of the Signal Transduction and activation of RNA (STAR) family. Members of the STAR-family, which include Quaking 1 (Qk1) Human Splicing factor 1 (SF1) and Germline development defective-1 (Gld-1), link signalling pathways with RNA metabolism, in particular splicing, localization and translation [58].

Sam68 has initially been identified as a 62 kDa protein, p62, which was bound to a tyrosine-phosphorylated GTPase-activating protein (GAP). Later it was characterized as a 68 kDa protein, which was a substrate of Src-Kinase during mitosis, hence the name Sam68 [59]. It has been found to play a role in many critical cellular processes: Beyond cell cylce regulation it is implicated in signal transduction, RNA metabolism, apoptosis and tumorgenesis [60]. Moreover, Sam68 has been controversially discussed to interfere with retroviral infection [61]. The versatility of Sam68 originates from its modular domain composition that is common in the STAR family proteins. It provides regulatory sequences, which allow interaction with signal transduction cascades, a specific RNA binding motif and several protein-protein interaction motifs (Figure 1.7).
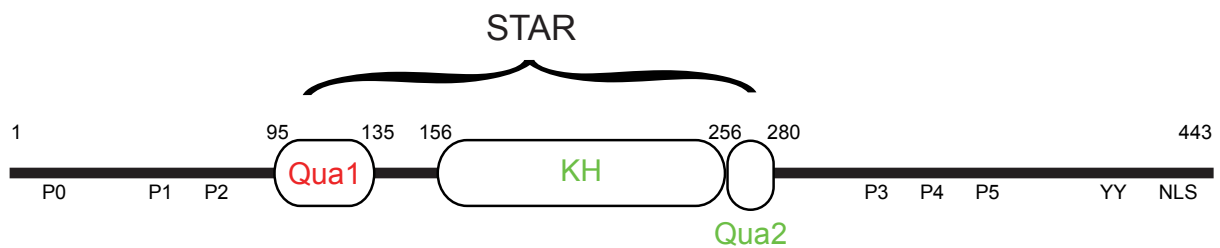
Figure 1.7: Domain organization of Sam68. Domain boundaries of the STAR domain are indicated as residue numbers. Relative positions of Proline rich motifs (P0-P5), a C-terminal tyrosine rich domain (YY) and the nuclear localization signal (NLS) are depicted.

RNA binding of Sam68 is facilitated by an hnRNP K [62] homology (KH) domain, which is one of the most prevalent RNA binding motifs [63]. In proteins of the STAR family, the KH domain is usually flanked by a conserved C- and N-terminal domain referred to as NK- and CK- or Qua1 and Qua2-domain, respectively [64-66]. This domain arrangement is termed GSG (named after the proteins GRP33, Sam68, GLD-1) domain or STAR domain. The only known exception from this characteristic domain composition in STAR-domain proteins is SF1 which lacks a Qua1 domain [67]. The amino acid sequence within this domain is highly conserved (Figure 1.8). Structural studies on SF1 revealed that the Qua2 domain extends the RNA binding surface of the KH-domain, which is a general RNA binding motif.

Figure 1.8: Sequence alignment of members of the STAR family. Residue numbers are given for Sam68. Secondary structure of SF1 is depicted above the alignment. Only the sequence of the conserved STAR domain was aligned. Note, that SF1 does not include a Qua1 domain (top section of the alignment).

The KH domain of SF1 adopts a β1-α1-α2-β2-β3-α3 topology, which is extended by a fourth helix α4 from the Qua2 domain. In the RNA bound form α4 packs against the KH domain [68]. In the free form of the KH-Qua2 domain, which has been determined for the homologous protein Quaking 1 (Qk1), α4 is highly dynamic and not packed to the KH domain [69]. The RNA is bound in a hydrophobic groove between the Qua2 domain, a conserved GPRG loop, and the variable loop of the KH domain (Figure 1.9). The consensus sequence for RNA binding of SF1 is YNCURAY. Conversely, RNA binding of Sam68 is much less specific. SELEX and in vivo cross-linking studies have shown that Sam68 generally binds to U- and A-rich sequences with high affinity to RNA containing either a UAAA or UUUA motif [64, 66, 70, 71]. Since no high resolution structure is available for the STAR domain of Sam68, the structural basis in the differential RNA recognition of SF-1 and Sam68 remains elusive.
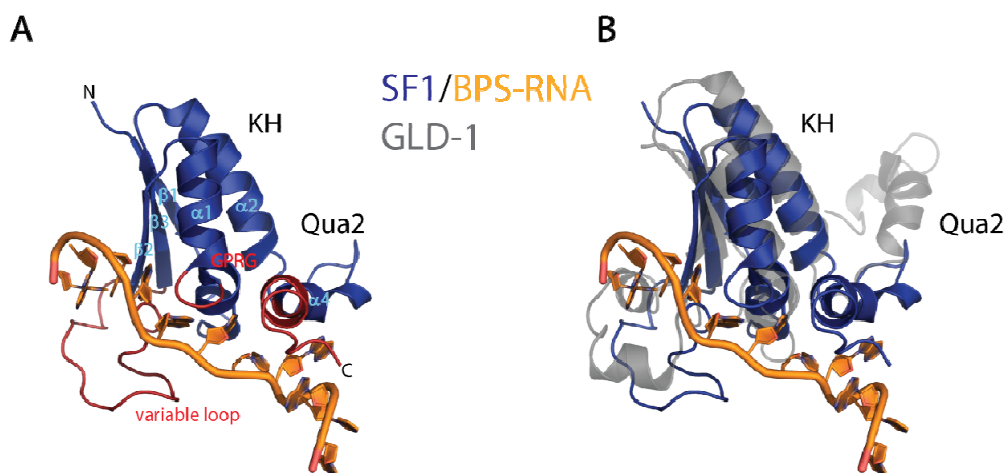
Figure 1.9: Structure of the KH-Qua2 domain. A) Ribbon representation of the structure of the complex of SF1 with the branch point sequence RNA. Regions critical for RNA binding are coloured in red. **B)** Overlay of the structure of the KH-Qua2 domains of SF1 (blue) in complex with its target RNA (orange) and the RNA-free form of GLD-1 (gray).

28

Several groups could show that the Qua1 domain of STAR proteins forms homodimers *in vitro* and *in vivo* [64, 71-73]. This is consistent with the bipartite nature of the RNA recognition motifs of many STAR domain proteins [74]. Until recently no structural information was available on the homodimerization of STAR domain proteins and its physiological role was obscure.

The regions N- and C-terminal of the STAR domain are unstructured. They harbour various protein-protein interaction motifs and are targets for post-translational modifications: Six Proline rich domains in the C- and N-terminus facilitate binding of Src homology (SH3) and WW domains. A Tyrosine rich (YY) domain in the C-terminus is recognized by SH2 domain containing proteins, after Tyrosine phosphorylation [60]. Phosphorylation of the N-terminus takes place at three specific Erk-phosphorylation sites (S58, T71 and 84) [75]. A C-terminal RGG box, is the target of Arginine-methylation [76]. Acetylation of Sam68 has rarely been detected in the N- and C-terminus but is a frequent modification of Lysine residues in the STAR domain [77].

The effect of posttranslational modification is manifold. On one hand, phosphorylation has been shown to decrease binding of Sam68 to its RNA targets. Band shift assays have revealed that phosphorylation of Sam68 with Erk leads to an impairment of RNA binding *in vitro*. Ribonucleoprotein immunoprecipitation followed by RT-PCR was used to validate this effect *in vivo*. [78] On the other hand phosphorylation of Sam68 results in an increased association with SH2-domain containing proteins, which has been reviewed in detail [58]. Acetylation of Sam68 at specific lysine residues by histone acetyltransferases has been detected in several tumor cell lines and enhances RNA binding [77]. Sam68 is mono- and di-methylated in a number of residues in the C-terminus and in the N-terminal RGG-box. A C-terminal nuclear location signal preferentially directs Sam68 to the nucleus. However, methylated Sam68 locates in the cytoplasm [76]. Moreover, methylation of Sam68 decreases its affinity to poly(U) RNA [79]. This plurality of various posttranslational modifications embeds Sam68 in a complex regulatory network.

Functional regulation of Sam68 via posphorylation by various kinases links Sam68 to signal transduction pathways. Sam68 has been implicated in T cell receptor, leptin receptor and insulin receptor signalling. How Sam68 integrates the stimuli of different receptor pathways to regulate alternative splicing remains a compelling issue.

Sam68 in alternative splicing regulation

A direct correlation between Sam68 and alternative splicing has been established by Stamm and co-workers, when they demonstrated the co-localization of Sam68 with the splicing factor YT521-B. The interaction of Sam68 with a Glutamic acid/Arginine-rich region on YT521-B is negatively regulated by phosphorylation, suggesting that Sam68 may be part of a signal transduction pathway that can influence splice site selection. Further evidence for the role of Sam68 in alternative splicing results from its binding to spliceosome-associated protein FBP21 [80] and the testis-specific splicing factor RBM [81]. A first hint that Sam68 can directly interact with alternatively spliced RNA aroused, when Grossmann *et al.* cross-linked Sam68 to an intronic regulatory RNA sequence of the tropomysin pre-mRNA [82].

More recently the pivotal role of Sam68 in the alternative splicing of diverse RNA targets, like CD 44 and Bcl-x, which are linked to apoptosis and cancer, was revealed [83]. The Bcl-x

transcript can be spliced to produce antiapoptotic Bcl-xL or proapoptotic Bcl-xs. Cancer cells often show an up-regulation of the antiapoptotic Bcl-xL isoform, which is associated with increased risk of metastasis. Paronetto *et al.* showed that Sam68 increases the level of proapoptotic Bcl-xs in a phosphorylation dependent manner. The group of König was able to clearly connect the Sam68 depended alternative splicing of CD44 to the mitogen activated protein kinase (MAPK) pathway and established Sam68 as a prototype regulator of alternative splicing in response to signalling events [75].

The binding of Sam68 to U2AF65 links the function of Sam68 to spliceosome formation, by supporting the binding of U2AF65 to pre-mRNA. U2AF65 contains three RNA recognition motifs (RRMs). Only two of them, RRM1 and RRM2 exhibit the canonical RRM fold and bind to RNA *in vitro*. The third RRM domain is a UHM-type RRM and binds to the N-terminus of SF-1. However, the Sam68 sequence does not contain a canonical ULM and no structural details on the interaction between Sam68 and U2AF65 is available. However, König and co-workers proposed a model of how Sam68 dependent alternative splicing of CD44 pre-mRNA might occur: First, Sam68 binds to a binding motif in the intronic and the exonic RNA respectively. U2AF65 is then targeted to the pre-mRNA by interaction with Sam68, which stabilizes association of the complex. Phosphorylation of Sam68 results in decreased affinity of Sam68 to its target RNA, so that it is released from the complex and the next step in spliceosome formation can be facilitated [84]. The structural details of how phosphorylation triggers the release of RNA are still under investigation.

Introduction to immune response after *Schistosoma mansoni* infection

The human immune system

The integrity of the human organism is constantly challenged by a vast amount of pathogens, which populate the environment. Pathogens include bacteria, viruses, fungi and other parasitic eukaryotes, and can cause severe disease when entering the body. To counteract this persistent invasion a complex interconnection of mechanisms, the immune system, evolved to identify and neutralize pathogens. The cells with affiliation to this defence system are termed white blood cells or leukocytes. In general the immune response is categorized into two principal approaches: The innate immune system and the adaptive immune system. The innate immune response is an immediate and universal defence mechanism, which can defeat most of the pathogens. For example, it efficiently prevents free bacterial growth throughout the body. The adaptive immune response is activated by the innate immune system and specifically recognizes and remembers a certain pathogen. If the body is encountered again with the same pathogen the adaptive immune system provides a stronger and more efficient defence mechanism. This process is called immunological memory.

The innate immunity

The major functions of the innate immune system include the recruitment of leukocytes to the location of infection, the phagocytosis of pathogenic particles and the release of anti-bacterial or anti-fungal substances. Additionally, the destruction of compromised host cells, such as virus infected cells, is accomplished. One of the first reactions to an infection is the inflammation which causes the commonly known symptoms: redness, heat, swelling and pain. During an inflammation, so called cytokines are released by injured cells, which initially summon leukocytes to the focus of infection. Also certain leukocytes, e.g. granulocytes, release cytokines, to further organize an efficient defence. The pattern of chemokine release is pathogen specific and determines which immune cells are recruited to oppose the infection. Basophilic granulocytes for instance release histamines which orchestrate the defence against parasites. Typically the first cells which counter the infection are macrophages. They clear pathogens by phagocytosis. Such incorporated intruders are efficiently destroyed by reactive oxygen species (ROS). Their remains are presented on the surface of macrophages as so called antigens. Antigens can be recognized by antibodies, a class of molecules produced by cells of the adaptive immune system.

The adaptive immunity

The key function of the adaptive immune system is the discrimination between self and foreign or non-self components. Everything which is recognized as non-self is attacked in an adaptive immune response. This function is essentially mediated by T- and B-cells.

T-cells specifically recognize antigens presented on the surface of macrophages or other antigen presenting cells (APCs) via the T-cell receptor (TCR). Dendritic cells for example are specialized to present antigens to lymphocytes. Several subsets of T-cells with divergent function have been discovered. The T-helper cells play a central role in the adaptive immunity. Their major function is the direction of the immune response. Stimulated by cytokines they differentiate into different subtypes, e.g. T_H1 cells or T_H2 cells, with diverse effector function. A simplistic model characterizes an immune response as a balance between a cellular T_H1 mediated response and a humoral T_H2 mediated response. The cellular immune response specifically targets virus infected or tumour cells and results in an acute inflammation reaction. The humoral immune response is directed against extra-cellular pathogens. It is essentially mediated by antibodies, which are proteins mainly produced by specialized B cells. Antibodies can be released from the cells and bind to invaders. Additionally, they occur bound to receptors on the surface of leukocytes to regulate their function.

Antibodies/Immunoglobulins

Antibodies, also termed Immunglobolins (Ig's), consists of two identical heavy and two identical light chains. The chains are linked by disulphate bonds and arranged in a dimeric structure, as depicted in Figure 1.10. Historically, antibodies are subdivided into three parts. The constant fragment (F_C), which facilitates receptor binding on cell surfaces, and the two antigen binding fragments (F_{AB}), which recognize the antigen. Five isoforms of Immunoglobulin (IgA, IgE, IgD, IgE and IgM) with different effector functions have been found in humans. They generally differ in the composition of the heavy chain, and are involved into different immunological functions. The predominant isoform in human is IgG. Its heavy chain includes three constant Ig-domains ($C_\gamma1$- $C_\gamma3$) and one variable domain (V_H). The light chain contains one constant (C_L) and one variable (V_L) Ig-domain. The flexibility of the antigen binding F_{AB} moieties is assured by a hinge region between $C_\gamma1$ and $C_\gamma2$.
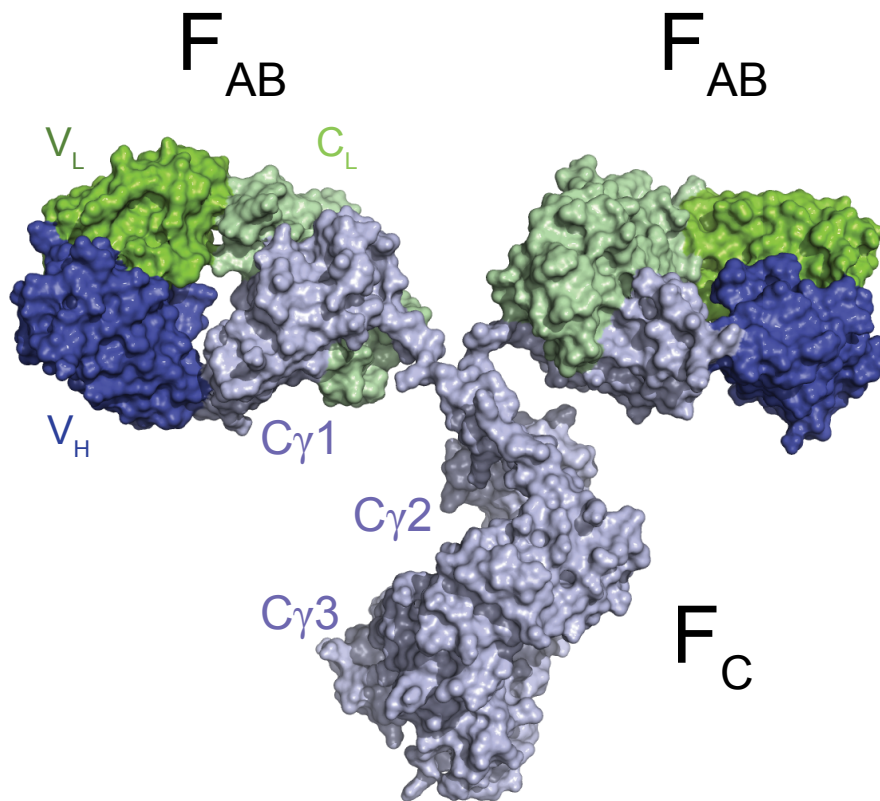
Figure 1.10: Surface representation of the structure of an IgG antibody. The heavy chain is shown in blue, the light chain is coloured green. The variable antigen binding domains (V_H and V_L) are represented in dark colours [85].

Immunoglobulin E

IgE is the least abundant immunoglobulin in the blood serum. It only reaches 0.05% of the typical IgG concentration. Nevertheless, it can mediate a potent immune reaction. IgE might have evolved as the first line of defence against pathogens. In particular, it triggers the immune reaction upon encounter with parasites [86]. With the steady improvement of sanitation in industrial countries, IgE lost its importance in host defence. Moreover IgE has been identified to be the major cause of allergies, i.e. the hypersensitive reactions to harmless antigens [87]. Although the general architecture of IgE is similar to IgG it shows some remarkable structural differences. Its heavy chain contains an additional Ig-domain replacing the hinge region between F_{AB} and F_C. The flexibility of the two antigen binding domains is therefore widely constricted. Structural analysis of the F_C moiety including small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) [88] as well as Förster resonance energy transfer (FRET) [89] show, that IgE does not resemble the Y-shaped structure of IgG in solution. Rather than that, the F_C moiety undergoes an asymmetrical bent. This was confirmed when a high resolution structure of IgE F_C was released [90]. It has been shown that the asymmetry of IgE plays a key role in the interaction with its high affinity receptor FcεRI [91].

Schistosomiasis and Schistosoma mansoni infection

Schistosomiasis is a parasitic disease, which in most cases leads to chronical manifestation. It is accompanied by a heavy parasitic load, leading to severe morbidity or death. With an estimate of over 200 million cases worldwide, it is the second most prevalent tropical disease. Schistosomiasis is predominantly endemic in rural and developing countries. In particular the disease is found in Africa, the Middle East, the Caribbean and South America. During 2008 less than 18 million infected people received a treatment in 17 of the 76 endemic countries (World Health Organisation).

Schistosomiasis is caused by trematode worms of the family *Schistosoma*. Among them, *S. mansoni* is the major cause of intestinal Schistosomiasis [92]. During the life cycle of *S. mansoni* its eggs are transmitted via faeces. Larvae hatch from their eggs in contaminated fresh water and infect water snails, the intermediate host. Here the larvae can multiply and mature into Cercaria, which are infective to humans. Large amounts are released into water and start the active search for the human host. Cercaria are capable of penetrating the human skin where they transform into the schistosomulum state. Schistosomules enter the circulatory system and eventually accumulate in the liver, where they mature into adult worms. Adult schistosomes pair and relocate into mesenteric or rectal veins to start production of eggs. Via the intestinal tract eggs can leave the human host to complete the life cycle.

While the worm resides in the human body it has to outmanoeuvre the host's immune system. One efficient mechanism is the down-regulation of the immune system by immunomodulatory and anti-inflammatory proteins from the eggs. Exposure to juvenile or adult worms induces a strong T_H1 reaction. A sudden down-regulation of this reaction and the stimulation of a T_H2 response coincides with the onset of egg production [93]. This response being significantly weaker than the initial T_H1 response is inappropriate to clear the heavy worm and egg load. It has been postulated that a glycoprotein, secreted from the eggs, is a major factor which primes the T_H2 response. This protein has been termed interleukin-4-inducing principle of *S. mansoni* eggs (IPSE) [94]. IPSE is identical with *S. mansoni* egg antigen alpha-1, which had been characterized 25 years earlier. However its function was obscure at that time [95].

IPSE/alpha-1

IPSE/alpha-1 is a 114 residue protein without any known sequence homolog. Secondary structure prediction indicates that IPSE/alpha-1 forms a compact, globular domain with a high content of β-sheets. A Y/FXXXXY/FXG signature sequence foretells its structural similarity to the β/γ-crystallin family. As revealed by SDS-PAGE under reducing and non-reducing conditions, IPSE forms disulfide-linked homodimers. Out of the 7 Cysteines, the C-terminal one facilitates this dimerization. A shorter protein variant, IPSEΔNLS, is missing a potential nuclear location signal and also lacks the C-terminal Cysteine residue. Consistently, this variant is incapable of homodimerization [96]. The remaining 6 Cysteines form three intramolecular disulfide bridges [97]. N-glycosylation has been shown for residues N61 and N80, which are in regions outside of the predicted β-sheets.
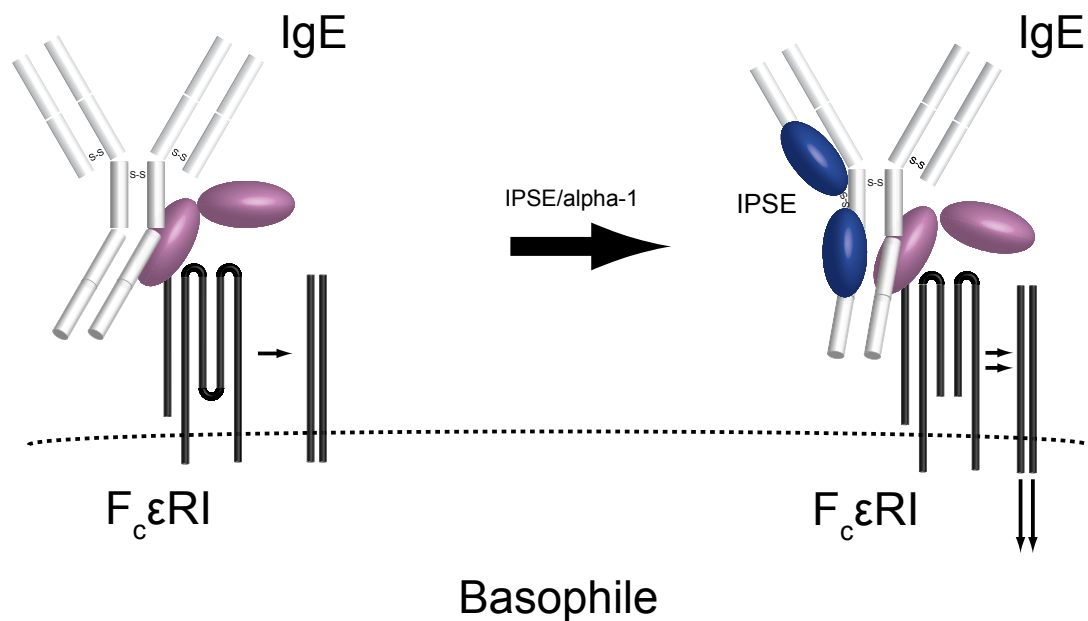
Figure 1.11: Schematic represantion of IgE receptor activation on the surface of basophils. IPSE potentially induces a structural rearrangment in the IgE molecule. This triggers the activation of the receptor signalling cascade.

IPSE is exclusively produced in *S. mansoni* eggs and secreted into the surrounding. It triggers the release of interleukin-4 from basophilic granulocytes [94]: A process clearly implicated in the T_H2 shift of the host immune system, which is observed upon egg deposition [98]. The release of interleukin-4 depends on IgE but is independent of antigen binding, since it occurs prior to the generation of *S. mansoni* specific IgE. Glycan epitopes or egg glycolipids and glycoproteins are the predominant target for antigen-antibody interaction in schistosomiasis. However, glycosylation does not seem to be a prerequisite for the function of IPSE/alpha-1 in activation of a T_H2 response. Recombinant IPSE from *E.Coli*, missing these posttranslational modifications, still binds to IgE an antigen unspecific manner. A speculative model proposes, that IPSE/alpha-1 jump-starts interleukin-4 production thereby initiating an early T_H2 response [98]. In this model IPSE/alpha-1 binds to IgE bound to the FcεRI receptor on the surface of basophilic granulocytes. Binding likely involves the constant regions of IgE, rather than the variable antigen binding domains. Conventional receptor signalling requires the cross-linking of at least two IgE molecules via their antigen binding domains. This brings several receptor molecules in close proximity. Receptor aggregation leads to phosphorylation of immunoreceptor tyrosine-based activation motifs (ITAM). The signal is then transduced via phospholipase C pathway, the Calcium pathway or the MAPK pathway [99]. Despite the dimeric nature of IPSE/alpha-1, and the fact that monomeric IPSEΔNLS is not functional, IPSE/alpha-1 does not seem to crosslink IgE molecules *in vitro*. Thus, a different mode of receptor activation seems applicable. Structural rearrangements of IgE upon IPSE binding might provide a novel cue to initiate receptor activation (Figure 1.11).

Scope of the thesis

The increasing numbers of protein structures in the Protein Data Bank (PDB) by now allows a general understanding of the function of proteins. However, to understand more complex cellular mechanism, a detailed knowledge of the complex structure and the dynamical interactions between several proteins and other macromolecules is necessary. NMR provides excellent tools for these studies. This thesis presents two examples of how the structure and interactions of proteins elevates the understanding of their function. Moreover the extracted knowledge was used to generate conceived functional models.

The first part of the thesis covers particular aspects of the regulation of alternative mRNA splicing by the protein Sam68. Sam68 and STAR domain proteins in general have been intensively investigated over the past two decades and a lot of cell biological knowledge about those proteins has been gathered. However, no high resolution structure has been deposited for Sam68 and structural data on other STAR domain proteins is very limited. When this project was started in 2006, only two high resolution structures of the KH-Qua2 moieties of STAR proteins had been reported. This part of my thesis aims at closing the gap between structural and cell biological studies, to deliver a broad understanding on the molecular function of Sam68. In particular, I focus on the structural properties of Sam68 which account for the differential RNA binding specificity. RNA binding is directly correlated to the function of Sam68. To elucidate the influence of Sam68 homodimerization on the recognition of the bipartite target RNA, I investigate the solution structure of Sam68 Qua1 domain. Finally, the analysis of its interaction with the splicing factor U2AF65 establishes a structural context to alternative splicing.

The second part of the thesis describes the structural analysis of IPSE/alpha-1 and its interaction with IgE. It has been proposed that IPSE/alpha-1 activates the FcεRI signalling without cross-linking of IgE molecules. The solution structure of IPSE/alpha-1 is the first step towards the understanding of this novel mechanism of receptor activation. Structural analysis of the complex of IPSE/alpha-1 and IgE is a challenging process due to its high molecular weight and the flexibility of the immunoglobulins. Use of high field NMR spectrometers in combination with modern pulse sequences are used to get an insight into complex formation.

CHAPTER II:

MOLECULAR BASIS FOR ALTERNATIVE SPLICING REGULATION BY THE PROTEIN SAM68

Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain.

Summary

Sam68 is an RNA-binding protein, which belongs to the STAR (signal transducer and activator of RNA) family. STAR family proteins are adaptors between signal transduction and RNA metabolism, thus regulating critical cellular functions such as cell cycle regulation and tissue development. In particular Sam68 modulates the alternative splicing of the pre-mRNAs of CD44 and Bcl-xL, which are related to apoptosis and tumor progression. Sam68 and other members of the STAR family recognize bipartite RNA sequences. Consistently, the protein forms homodimers *in vitro* and *in vivo*. However, the structural and functional roles of the self-association remain to be elucidated.

Here, I present the structure of the Sam68 Qua1 homodimer. Each Qua1 monomer comprises two α-helices, which are aligned in an antiparallel fashion. Extensive contacts between hydrophobic residues lead to a tight packing of the two helices of each monomer. The Qua1 dimer is formed by perpendicular stacking of the two monomers. This interaction is mainly driven by a largely hydrophobic contact area, provided by conserved hydrophobic residues. Additionally, a network of intermolecular hydrogen bonds and electrostatic interaction stabilize the homodimerization.

Based on the solution structure of Sam68 point mutants were designed to interfere with dimer formation. Single point mutations of critical residues are sufficient to destabilize the homodimer. The almost invariant F118 in the loop connecting the two alpha helices has been identified as a key determinant of homodimerization.

Analysis of Sam68 variants in a cell-based assay revealed that regulation of alternative splicing by Sam68 requires homodimerization of the protein. Furthermore, the conserved Y103 is crucial for Sam68 function.

Sam68 dimerization via the Qua1 domain may support the correct recognition of its target RNAs. This step is necessary to recruit splicing factors to the corresponding pre-mRNA. Phosphorylation of Y103 might trigger the release of Sam68 from the RNA to initiate spliceosome assembly.

Structural basis for homodimerization of the Src-associated during mitosis, 68 kD protein (Sam68) Qua1 domain

N. Helge Meyer [1,2], Konstantinos Tripsianes[1,2], Michelle Vincendeau[3], Tobias Madl[1,2], Fatiha Kateb[1,2], Ruth Brack-Werner[3], Michael Sattler[1,2]

[1] Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
[2] Munich Center for Integrated Protein Science (CiPS[M]) at Department Chemie, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany
[3] Institute of Virology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany

Sam68 (Src-associated in mitosis, 68 kDa) is a prototypical member of the STAR (signal transducer and activator of RNA) family of RNA-binding proteins. STAR proteins bind mRNA targets and modulate cellular processes such as cell cycle regulation and tissue development in response to extracellular signals. Sam68 has been shown to modulate alternative splicing of the pre-mRNAs of CD44 and Bcl-xL, which are linked to tumor progression and apoptosis. Sam68 and other STAR proteins recognize bipartite RNA sequences and are thought to function as homodimers. However, the structural and functional roles of the self-association are not known.

Here, we present the solution structure of the Sam68 Qua1 homodimerization domain. Each monomer consists of two antiparallel α-helices connected by a short loop. The two subunits are arranged perpendicular to each other in an unusual four-helical topology. Mutational analysis of Sam68 in vitro and in a cell-based assay revealed that the Qua1 domain and residues within the dimerization interface are essential for alternative splicing of a CD44 minigene. Together, our results indicate that the Qua1 homodimerization domain is required for regulation of alternative splicing by Sam68.

Introduction

Sam68 (Src associated during mitosis, 68 kDa) [59] belongs to the STAR (signal transducer and activator of RNA) family of RNA-binding proteins, which also includes Qk1 (Quaking 1), SF1 (Splicing factor 1) and Gld-1 (Germline development defective-1) [60]. STAR family proteins link signaling pathways and many aspects of RNA metabolism (splicing, localization, and translation). They are regulated by post-translational modifications such as phosphorylation, acetylation and arginine methylation [60].

Sam68 acts in post-transcriptional regulation of pre-mRNA splicing in response to extracellular signals [100]. It is involved in a variety of pathways, including insulin and T-cell receptor signaling [58] and plays a key role in cell cycle regulation [101]. Sam68 exhibits binding specificity for homopolymeric poly(U) RNA and has been shown to recognize UAAA or UUUA sequences with high affinity as determined by SELEX and *in vivo* cross-linking [66, 70]. Posttranslational modifications can regulate Sam68 function by critically

affecting the accessibility to RNA [84, 102]. Tyrosine phosphorylation by Src kinase during mitosis enhances the interaction of Sam68 with Ras-GAP [103] but prevents its association with RNA. On the other hand, acetylation of lysine residues by histone acetyltransferases enhances RNA binding [77]. Finally, overexpression of Sam68 has been shown to support prostate cancer, cell proliferation and survival [104].

Sam68 has been identified as a key determinant in the alternative splicing of various important RNA targets, like CD44 [75] and Bcl-XL [83], which are linked to apoptosis and cancer. In particular, alternative splicing of CD44 impacts embryonic development and immune response [105-108]. Up to 10 variant exon sequences can be included in the mature CD44 mRNA. Among them, variable exon 5 (exon v5) inclusion is associated with tumor progression and T-Cell activation [107, 109]. Sam68 is a target for phosphorylation by ERK (extracellular-signal-regulated kinase), which promotes inclusion of CD44 exon v5 in response to Ras activation by phorbol ester stimulation. Sam68 binds two RNA sequences within exon v5 and the preceding intron, respectively [75, 84]. Recent studies indicate that Sam68 interacts with the heterodimeric splicing factor U2AF (U2 small nuclear ribonucleoprotein particle auxiliary factor) [84]. Recognition of the 3' splice site by U2AF is a key step in spliceosome formation. The interaction with Sam68 is thought to stabilize U2AF binding to its cognate RNA elements. Interestingly, phosphorylation of Sam68 by ERK interferes with RNA binding and reduces pre-mRNA occupancy by U2AF. Altogether, these data indicate that Sam68 can regulate alternative splicing in a signal-dependent manner.

Sam68 shares high similarity with Splicing Factor 1 (SF1), another binding partner of U2AF. The so-called STAR domain, also referred to as GRP33/SAM68/GLD1 (GSG) domain, consists of an hnRNP K homology (KH) domain, flanked by two domains N- and C-terminal of the KH domain, referred to as Qua1 and Qua2, respectively [60, 100]. Interestingly, SF1 is the only known member devoid of the Qua1 subdomain and functions as a monomer [67]. The KH domain is one of the major RNA binding motifs in eukaryotic cells [63]. The structural basis for the recognition of single-stranded RNA by the KH-Qua2 domain of SF1 has been described, indicating that the Qua2 domain extends the KH fold and the KH-Qua2 tandem domain is essential for sequence-specific RNA recognition [68]. The solution structure of the free form of the KH-Qua2 region of Quaking showed that the Qua2 helix does not contact the KH domain in the absence of RNA [69]. Biochemical data and *in vivo* studies suggest that the Qua1 domain can oligomerize [64, 71-73]. However, the structural basis for oligomerization and the role for the *in vivo* function of STAR proteins are unknown.

Here, we present the NMR solution structure of the Sam68 Qua1 domain. Qua1 forms a homodimer, comprised of a perpendicular interaction of two helical hairpins. A network of hydrophobic and electrostatic interactions stabilizes the dimer interface. Based on NMR and biophysical data we show that the Qua1 domain is sufficient for homodimerization of Sam68 *in vitro*. Cell-based splicing assays identify critical residues in the dimer interface and reveal that Qua1 is necessary for the function of Sam68 in alternative splicing.

Experimental procedures

Sample Preparation. H. sapiens Sam68 Qua1 (95-135), Qua1 (95-156), KH-Qua2 (147-280) and Qua1-KH-Qua2 (95-280), which were derived from pcDNA 3.1 HsSam68 described in [75], were cloned in a modified pETM-11 vector (European Molecular Biology Lab) containing an additional N-terminal double Z-tag and expressed in *E.coli* BL21(DE3). Point

mutations were introduced in pcDNA 3.1 HsSam68 and pETM-11 ZZ Qua1 (95-135) using the Quick-Change site directed mutagenesis (Invitrogen) protocol. The corresponding proteins contain two additional residues at the N-terminal tobacco etch virus (TEV) protease cleavage site, which are derived from the expression vector.

Bacteria were grown in LB medium for preparation of unlabeled sample or M9 medium supplemented with ^{13}C-labeled glucose and/or ^{15}NH$_4$Cl for uniformly labeled samples, respectively. The protein was expressed at 20° C for 16 hours after induction with 250 mM IPTG.

Proteins were purified using Ni^{2+}-affinity chromatography (Ni-NTA, Qiagen). After cleavage of the His- and the Z-tag with TEV protease the tags were separated from the protein by a second Ni^{2+}-affinity step. Qua1 samples were heat shocked for 5 min at 85° C. The samples were further purified by size exclusion chromatography on a HiLoad 16/60 Superdex 75 (GE Healthcare Biosciences) and kept in 10 mM phosphate pH 6.5, 100 mM NaCl.

An asymmetrically labeled sample to record intermolecular NOEs was prepared as follows: Equimolar amounts of unlabeled and ^{15}N/^{13}C uniformly labeled Qua1 (residues 95-135) were mixed and incubated with 1% SDS at 85° C for 10 minutes. After slowly cooling down, SDS was removed by dilution and ultra filtration using a 15 ml Amicon (Millipore) with 5 kDa molecular weight cutoff.

NMR Spectroscopy. NMR measurements were carried out at 298 K on a Bruker Avance III 750 MHz spectrometer equipped with a TXI probehead, a 600 MHz spectrometer equipped with a TCI cryoprobehead or on an Avance 900 instrument equipped with a TXI cryoprobehead. Spectra where processed with NMRPipe [110] and analyzed with Sparky[111]. Backbone assignment was done semi-automatically using MARS [112]. For backbone and side chain assignment CBCA(CO)NH, CBCANH and (H)CCH-TOCSY spectra were recorded [3]. Distance information was obtained from ^{15}N- and ^{13}C-edited NOESY spectra with a mixing time of 70 ms. To distinguish inter- and intramolecular NOEs a set of isotope-edited and filtered NOESY spectra was recorded [3, 18]. Experiments were carried out using a 1 mM uniformly ^{15}N/^{13}C-labeled or 2 mM asymmetrically ^{15}N/^{13}C-labeled Qua1 (95-156) sample, respectively. ^{15}N R_1 and R_2 relaxations rates and {^{1}H}-^{15}N heteronuclear NOE data were measured at 750 MHz proton Larmor frequency and 298 K as described [113]. H$^\alpha$-C$^\alpha$, N-C' and H^N-N residual dipolar couplings (RDCs) were recorded using HNCO-based NMR experiments [114] with a 1 mM Qua1 (95-135) sample that was aligned in a medium containing 15 mg/ml Pf1 phage (Profos AG, Regensburg, Germany) as described [115]. Paramagnetic relaxation enhancements (PREs) were measured from saturation-recovery ^{1}H,^{15}N HSQC experiments (recovery times between 0.01 and 4.0 s) at concentrations of 0, 1, 2, 3, 4, 5, 7, 10 mM of the soluble paramagnetic agent Gd(DTPA-BMA). Back-calculation and data analysis were carried out according to [21].

Structure calculation. Automated NOE cross-peak assignment was performed using the software CYANA 3.0 [13]. Automatically assigned NOEs and completeness of the NOE cross peaks were manually inspected. Homodimer symmetry is explicitly taken into account for network anchoring, and identical conformation of the two monomers is ensured by dihedral angle difference restraints for all corresponding torsion angles. Additionally, a symmetric relative orientation of the two monomers is maintained by distance difference restraints between symmetry related intermolecular C$^\alpha$-C$^\alpha$ distances [116]. Distance restraints from the CYANA calculation and TALOS+ [9] derived torsion angles and RDC restraints

were used in a water refinement calculation [117] with Aria 2.2 [14]. Quality of the structure ensemble was validated using the iCING [17] web server as well as PROCHECK [15] and WHATCHECK [16]. Molecular images were generated using Pymol [118].

Circular dichroism (CD) spectroscopy. Temperature series of Far-UV (190-250 nm) CD spectra were recorded on 100 μM Qua1 wildtype or mutant proteins in 20 mM phosphate, pH 6.5, 50 mM NaCl using a JASCO J-715 spectropolarimeter.

Cell culture and transfections. U138MG cells were cultured as previously described [119]. All overexpression experiments were performed in 6-well plates. Cells were seeded at a density of 2×10^5 cells per well one day prior to transfection and cultured for 30 hours after transfection. Transfections were performed using FuGENE[TM]HD Transfection Reagent (Roche Diagnostic) according to the manufacturer's instructions. For cotransfection experiments 2 μg of CD44 v5 minigene-plasmid DNA [75] was cotransfected with 300 ng plasmids expressing Sam68 or wildtype or mutants. As a control, the CD44 v5 minigene-plasmid DNA was cotransfected with 300 ng of the pcDNA 3.1 expression plasmid. Expression of all proteins was checked by Western-Blot analysis (data not shown).

RT-PCR analysis. Cytoplasmic RNA was prepared using the PARIS Kit (Ambion) according to the manufacturer's protocol. 1 U/μg RNA was transcribed with Superscript II (Invitrogen) according to the manufacturer's protocol, using random hexamers. RT-PCR analyses were carried out as described in [75] using 25 PCR cycles. RT-PCR bands were quantified densitometrically using ImageJ Software. Three independent experiments were performed.

Results

Qua1 is sufficient for Sam68 homodimerization. It has been shown that the STAR domain of Sam68 (Figure 2.1A) and other members of the STAR family dimerize [64, 71-73]. The apparent molecular weight estimated from size exclusion chromatography of ~40 kDa is consistent with a dimer of the Sam68 STAR domain (not shown). To identify the region that mediates dimerization we expressed proteins comprising the Qua1, KH-Qua2 and Qua1-KH-Qua2 domains in *E.coli*. An overlay of ^{1}H,^{15}N correlation NMR spectra of the Qua1 (residues 95-135) and KH-Qua2 (residues 147-280) subdomains with that of the STAR domain (residues 95-280) shows that chemical shifts seen in the smaller domains are very similar to those of common residues in the STAR domain. This indicates that there are no strong contacts between Qua1 and KH-Qua2 and that these two regions are structurally autonomous (Figure 2.1B).

The ^{1}H NMR frequencies in the ^{1}H,^{15}N HSQC spectrum of Qua1 (residues 95-156) are mainly found within a small region centered around 8 ppm indicating the presence of mostly helical and random coil conformation. Analysis of the ^{13}C$^{\alpha/\beta}$ secondary chemical shifts reveals that residues 95-135 comprise two α-helices that are interconnected by a short loop. The 20 C-terminal amino acids (residues 136-156) do not exhibit any secondary structure. ^{15}N R_1 and R_2 relaxation rates as well as {^{1}H}-^{15}N heteronuclear NOE data show that the C-terminus is highly flexible, while the loop connecting the two helices has a slightly increased flexibility compared to the helical segments (Figure 2.2). Thus, only residues 95-135 define the globular fold of the Qua1 domain. A tumbling correlation time of $\tau_c = 9.5$ ns was estimated from the ^{15}N R_2/R_1 relaxation rates ratio for Qua1, consistent with the molecular weight expected for a homodimer (14 kDa). Even at concentrations of 10 μM, there is no

significant change in the ^{15}N relaxation times and thus τ_c (data not shown), indicating that the dimerization constant is at least in the low micromolar to nanomolar range. Additionally the apparent molecular weight determined by size exclusion chromatography matches well with a homodimer (Suppl. Fig. 2.1). These observations suggest that the Qua1 domain is sufficient for homodimerization.

For structural analysis and to further confirm that Qua1 is a homodimer we attempted to identify intermolecular NOEs. In the case of a symmetric homodimer the NOESY spectrum contains both intramolecular and intermolecular cross-peaks. To unambiguously distinguish intermolecular from intramolecular NOEs isotope filtered experiments were recorded on an asymmetrically ^{15}N/^{13}C-labeled sample [3, 18]. To obtain such a sample, equimolar amounts of uniformly ^{15}N/^{13}C-labeled and unlabeled Qua1 are mixed. Random exchange of the subunits yields 50% of asymmetrically isotope labeled dimers. Isotope edited/filtered NMR experiments will detect exclusively intermolecular NOEs for this population, while no signals are observed for symmetrically labeled or unlabeled homodimers or monomeric species. Thermal denaturation in presence of 1% SDS and refolding of equal amounts of labeled and unlabeled Qua1 resulted in a fingerprint spectrum identical to that of the native protein, indicating that the fold was restored (data not shown). After refolding, more than 100 intermolecular cross peaks were observed in an isotope-edited/filtered NOESY experiment (Figure 2.1C). Taken together, these data establish that Qua1 is a stable homodimer.

Structure of the Sam68 Qua1 homodimer. We determined the solution structure of the Sam68 Qua1 homodimer (comprising residues 95-135) by heteronuclear triple-resonance NMR spectroscopy using uniformly as well as asymmetrically labeled ^{15}N/^{13}C-labeled protein. The structure is defined based on 2226 experimental distance restraints, including 2*228 intermolecular distance restraints, derived from ^{13}C- and ^{15}N-edited NOESY HSQC spectra as well as ^{12}C/^{14}N-filtered, ^{15}N- or ^{13}C-edited NOESY-HSQC spectra. A summary of the structural and restraint statistics is given in Table 1. The ensemble of the 20 lowest energy structures obtained after water refinement is shown in Figure 2.3A. Quality of the structures was further evaluated by comparison of measured and back-calculated relaxation rate enhancements upon addition of the paramagnetic co-solvent Gd(DTPA-BMA) [21] and showed perfect agreement (Suppl. Fig. 2.2).

Table 2.1: Structural statistics of the Sam68 Qua1 homodimer.

NOE-based distance restraints[1]	
Intraresidual, sequential	950
Medium range ($2 \leq |i-j| \leq 4$)	504
Longe range ($|i-j| \geq 5$)	316
Intermolecular	456
Total	2226
Other restrains	
$\phi + \psi$ dihedral angle restraints	128
Residual dipolar coupling restraints (H^N-N, N-C', H^α-C^α)	128
Coordinate Precision RMSD	
Backbone (Å)	0.32±0.09
Heavy atom (Å)	0.97±0.16
Consistency (structure vs. restraints)	
RMSD (Å) from experimental distance restraints[1]	0.019±0.002
RMSD (°) from experimental torsion angle restraints[2]	0.911±0.104
RDC Q-Factor[3]	0.230±0.005
WHATCHECK[4]	**Structure Z Scores**
First generation packing quality	0.497
Second generation packing quality	-0.028
Ramachandran plot appearance	0.010
χ_1/χ_2 rotamer normality	-2.057
Backbone conformation	1.325
Ramchandran plot[4]	
Most favoured regions	96.0%
Allowed regions	4.0%
Generously allowed regions	0.0%
Disallowed regions	0.0%

Statistics are given for the 20 lowest energy structures after water refinement out of 100 calculated. The CNS E_{repel} function was used to simulate van der Waals interactions with an energy constant of 25 kcal mol^{-1} Å^{-4} using "PROLSQ" van der Waals radii [117]. RMSD and PROCHECK values apply for residues 101-135.

[1] Distance restraints were employed with a soft square-well potential using an energy constant of 50 kcal mol^{-1} Å^{-2}. No distance restraint was violated by more than 0.5 Å.

[2] Torsion angle restraints derived from TALOS [9] were applied to ϕ, ψ backbone angles using energy constants of 200 kcal mol^{-1} rad^{-2}. No dihedral angle restraint was violated by more than 5°.

[3] RDCs were employed with a harmonic potential using an energy constant of 0.5 kcal mol^{-1} Hz^{-2}. Q-factor as defined in [120].

[4] PROCHECK [15] and WHATCHECK [16] were used to determine the quality of the structure. Positive WHATCHECK Z-scores indicate that structure is better than average.
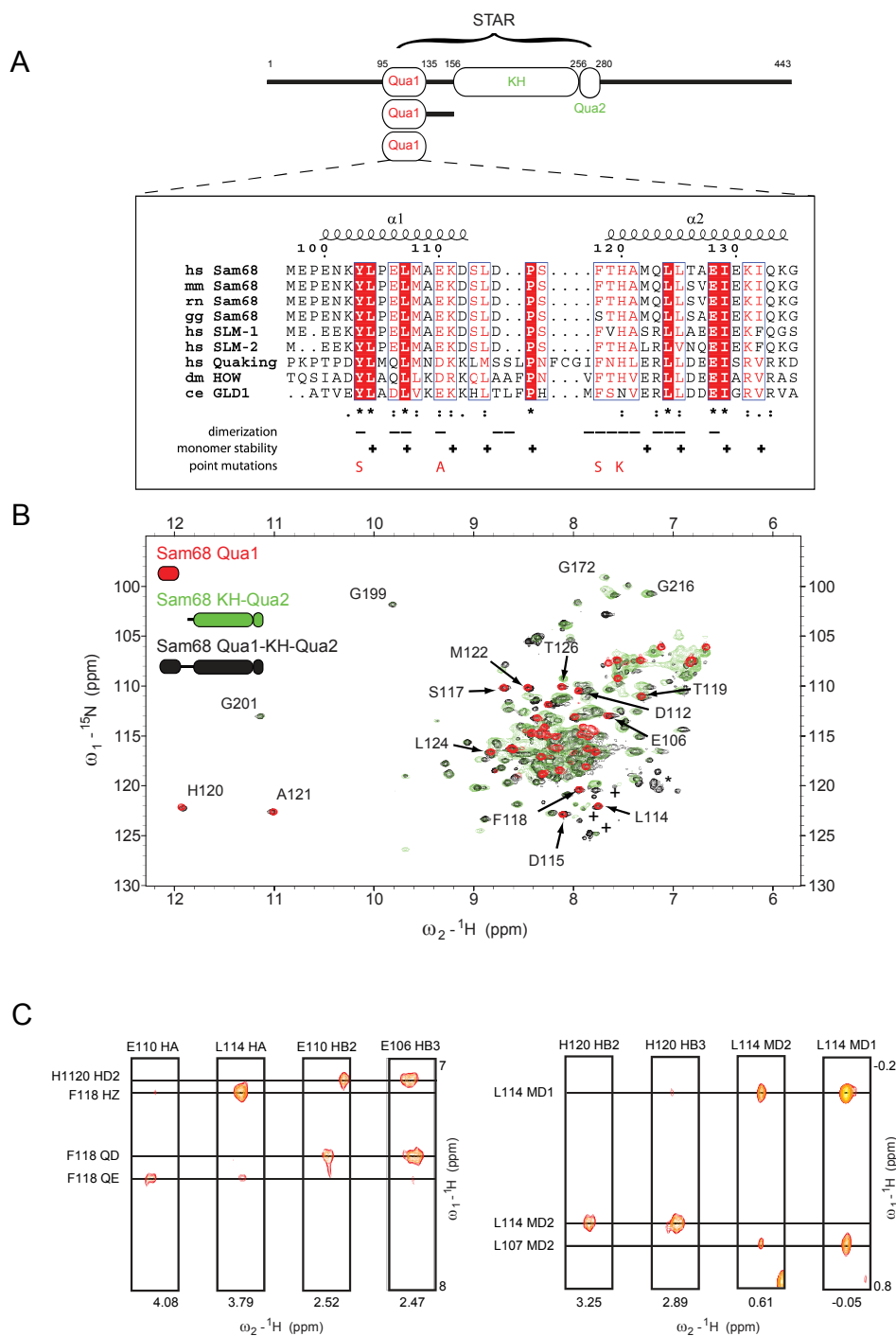
Figure 2.1: **Topology and NMR spectra of the Sam68 STAR domain. A)** Domain organization of Sam68. The Qua1 proteins used for NMR are indicated. Sequence alignment of the Qua1 domains of different members in the STAR family, residue numbers are given for Sam68 Qua1. Secondary structure of Sam68 Qua1 is depicted above the alignment. Residues important for monomer and dimer stability are indicated below, by "_" and "+", respectively. Conserved residues that were mutated are indicated. Sequences were aligned using the program ClustalW [121] and analyzed using the program ESpript [122]. **B)** Overlay of a ^{1}H,^{15}N TROSY spectrum of perdeuterated Qua1-KH-Qua2 (residues 95-280) (blue) and ^{1}H,^{15}N HSQC spectra of Qua1 (residues 95-135) (red) and KH-Qua2 (residues 147-280) (green). Note, that the linker connecting the Qua1 and KH domains (135-156) is not included in the Qua1 protein. Signals arising from arginine side chains are marked with an asterisk and some small peaks of degradation products are marked with "+". **C)** Intermolecular NOEs recorded on an asymmetrically ^{15}N/^{13}C-labeled sample.

Each Qua1 monomer comprises two α-helices which are aligned in an antiparallel fashion with a tilt angle of ~30°, connected by a short loop. Hydrophobic residues are spaced every three to four residues apart on each helix and interact with residues arrayed on the other helix of the same monomer, i.e. L104, L107, K111 and L114 in helix α1 contact I132, I129, L125 and M122 in helix α2, respectively (Suppl. Fig. 2.3). This arrangement is reminiscent of a short "zipper" where the extensive hydrophobic contacts lead to a tight packing of the two helices of each monomer. A potential side-chain hydrogen bond between the invariant Y103 (helix α1) and E128 (helix α2) further stabilizes the helical hairpin. This hydrogen bond may be consistent with a lethal phenotype of the E48G mutation in the mouse Qk1 paralog (E48 corresponds to Sam68 E128). Another important feature of this topology is the highly conserved P116 in the loop connecting the two helices, which allows the reversal of the peptide backbone.
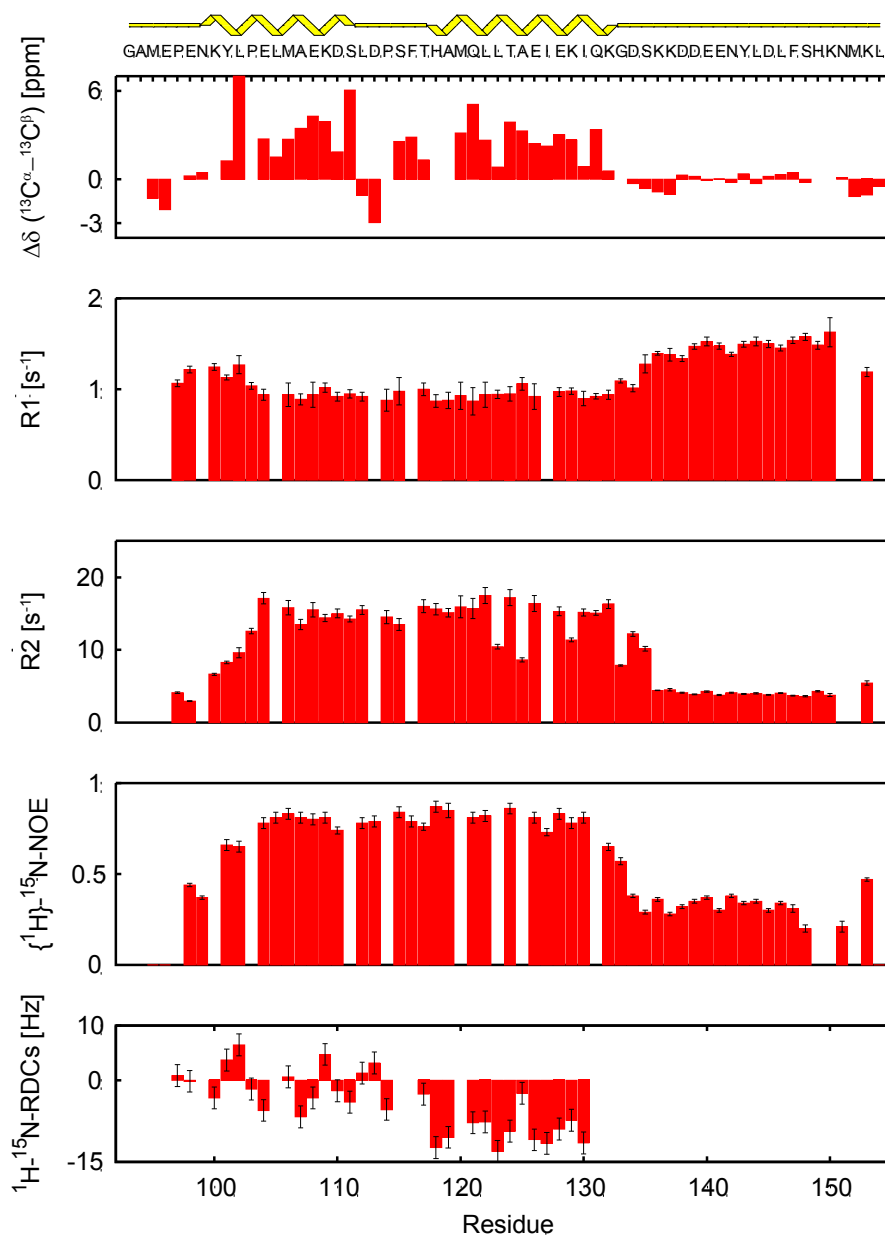
Figure 2.2: NMR characterization of the Sam68 Qua1 domain. [13]C secondary chemical shifts, [15]N R1, R2 relaxation rates, {[1]H}-^{15}N heteronuclear NOE and [1]H-^{15}N residual dipolar couplings are plotted *versus* Sam68 Qua1 residue numbers. The secondary structure is indicated at the top.

The Qua1 dimer is formed by perpendicular stacking of the two monomers with a C2 symmetry (Figure 2.3A and B). Numerous hydrophobic contacts, which mainly involve the loop region connecting the two helices, stabilize the dimer interface. Below this largely hydrophobic contact area, cross subunit hydrogen bonds and electrostatic contacts between the N-terminal end of helix α2 and the α1'/α2' helices of the other monomer provide additional inter-subunit interactions (Figure 2.3). The interface area covers 624 $\mathring{A}^2$ per monomer, which accounts for more than 18% of the total surface area of one monomer (Figure 2.3C).
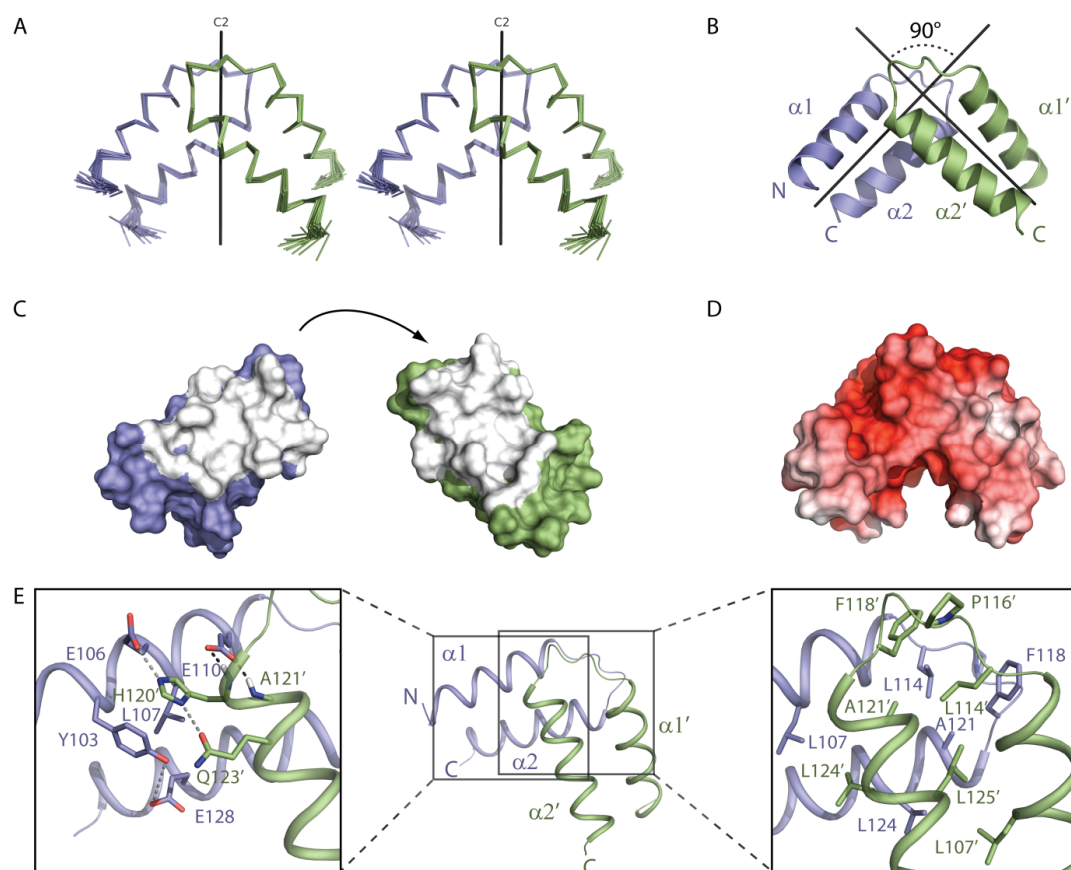
Figure 2.3: NMR structure of the Sam68 Qua1 homodimer. A) Stereo view of the ensemble of the 20 lowest energy structures and **B)** ribbon representation of the Sam68 Qua1 dimerization domain. **C)** Surface representation of two separated Qua1 monomers, showing a view onto the dimer interface (white). **D)** Surface representation colored according to electrostatic surface potential at ±7 kB T/e for positive (blue) or negative (red) charge potential using the program APBS [123]. **E)** Close-up view of the interface. Side chains of key residues in the dimer interface are shown as sticks. Hydrogen bonds between the E110 side chain and the H120 and A121 amides are indicated by black dashed lines. Potential hydrogen bonds and electrostatic interactions are shown by gray dashed lines.

A key determinant of the dimer formation is the almost invariant F118 in the α1-α2 loop. Its aromatic side chain stacks with the L114' side chain from the other monomer (Figure 2.3E, right). The side chains of L107, A121, L124 and L125 mediate additional hydrophobic interactions stabilizing the dimer interface. Near the intersection of the α2/α2' helices the H120' side chain contacts the aromatic ring of Y103 in the other monomer. Apart from this hydrophobic network cross-subunit hydrogen bonds between E110 and the backbone amide protons of A121' and H120' are observed (Figure 2.3E, left). This is reflected by large downfield shifts of the ^{1}H NMR frequencies of the backbone amides of A121 and H120 (Figure 2.4).
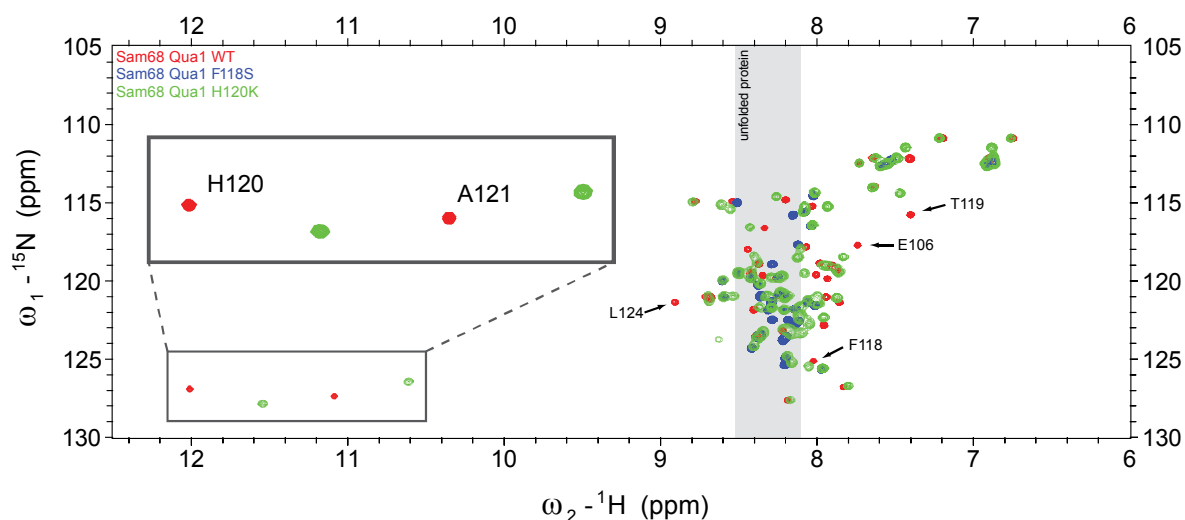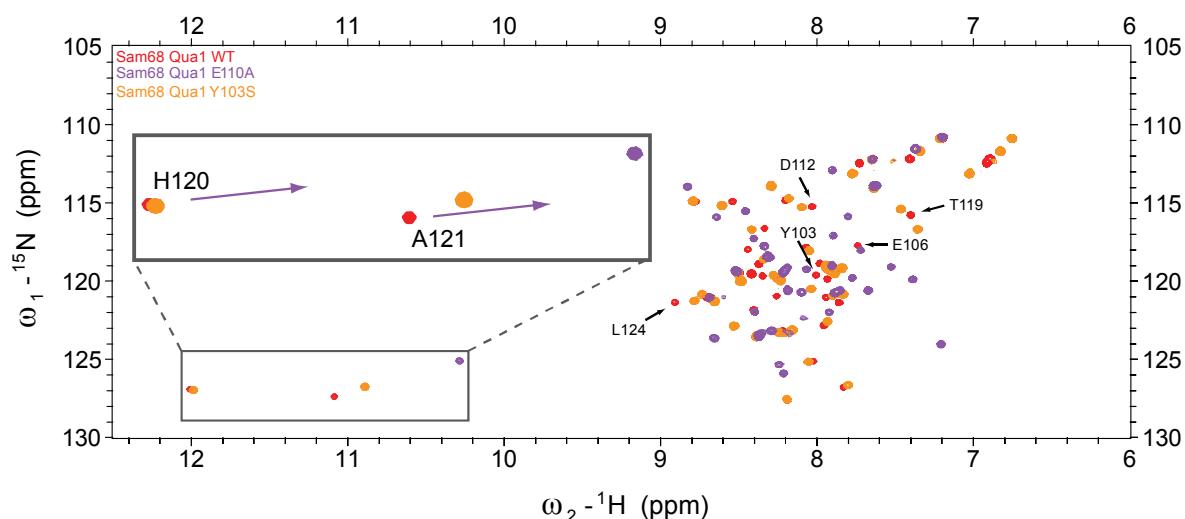
Figure 2.4: NMR spectra of wildtype and mutant Sam68 Qua1. Overlay of the ^{1}H,^{15}N HSQC spectra of the different Qua1 mutants that disrupt the overall fold **(A)** or that have mainly local structural effects **(B)**. A) The F118S mutant is unfolded as shown by a drastically reduced chemical shift dispersion. The typical chemical shift range for unfolded proteins is indicated by a gray box. The H120K mutant shows two sets of NMR signals, which account for folded and unfolded moieties of the protein (see Suppl. Fig 4). **B)** The mutants Y103S and E110A are folded. However, the intermolecular hydrogen bond between the E110 side chain and the backbone amides of H120 and A121 cannot form in the E110A mutant, resulting in an upfield shift of the backbone amide resonances for those residues (indicated by arrows). The E110A mutant may lead to a local rearrangement of the aromatic side chains of H120 and indirectly of Y103, which is reflected by the overall chemical shift differences compared to the wildtype protein. Similarly, the relatively large changes of the NMR chemical shifts in the Y103S mutant may be attributed to the removal of aromatic ring current effects.

Furthermore, the side chain of H120' is in close proximity to the side chains of E106 and Q123', implying electrostatic stabilization (Figure 2.3E, left). Charged residues are distributed at the periphery and at the opposite side of the interface giving an overall negative charge to the Qua1 homodimer (Figure 2.3D).

An arrangement of two antiparallel helices is a common feature in a variety of different proteins and protein folds. However, structural similarity searches [124] did not reveal any structure in the PDB database with a similar perpendicular arrangement of two helical hairpins, indicating that the Qua1 homodimer adopts a novel four-helical dimer topology.

Mutational analysis of the dimer interface. We confirmed the dimer contacts seen in the structure by mutational analysis. Various single point mutations were designed and ^{1}H,^{15}N HSQC experiments of the corresponding proteins were recorded to monitor the influence of the mutation on the overall fold. The capability to form homodimers was determined by analysis of apparent local correlation times, τ_c^{app}, estimated from the ^{15}N relaxation data. The τ_c^{app} values reflect the size of the protein in solution and the distribution of τ_c^{app} therefore allows estimating the ratio between the monomeric and the dimeric forms (Suppl. Fig. 2.4).

Mutation of F118 (F118S) leads to a virtually complete loss of structure and dimerization (Figure 2.4A). Remarkably, at concentrations above 1 mM a small portion of the protein can still dimerize as indicated by weak NMR signals in the ^{1}H,^{15}N HSQC, which reflect the folded dimer (data not shown). The H120K variant is in slow exchange between unfolded protein and structurally intact dimer. Accordingly, two sets of signals with distinct ^{15}N relaxation properties are observed in the ^{1}H,^{15}N HSQC spectra (Figure 2.4A). The Y103S and E110A variants maintain the overall fold and the protein still forms dimers (Figure 2.4B). However, the distribution of local correlation times, τ_c^{app}, derived from the NMR relaxation data (Suppl. Fig. 2.4) is slightly decreased in both mutants compared to the wildtype, consistent with a fast exchange between monomeric and dimeric species. The effect of the E110A mutation on dimerization is more pronounced, as the average τ_c^{app} measured for the E110A variant is lower than for the Y103S mutant. The E110A Qua1 domain cannot form intermolecular hydrogen bonds to the backbone amides of A121' and H120'. This results in an upfield shift of the respective resonances seen in the ^{1}H,^{15}N HSQC spectrum of the E110A mutant (Figure 2.4B). Thermal denaturation of the Qua1 mutants followed by CD spectroscopy demonstrates that the thermal stability of the Y103S and E110A variants is decreased compared to the wildtype protein (Suppl. Fig. 2.5), even though both mutants are still mainly dimeric. In contrast, the F118S and H120K mutants are highly unstable or already unfolded at room temperature (data not shown).

In summary, the structure-guided mutagenesis shows a good correlation between the importance of contacts in the interface and dimer formation. Specifically, the F118S mutation completely disrupts the dimer, while the other mutations appear to have more local effects only destabilizing the Qua1 dimer.

Functional Qua1 domain is required for splicing regulation by Sam68. Sam68 has previously been shown to increase the inclusion of the CD44 exon v5 in mRNAs produced from a CD44v5 minigene (Figure 2.5A) [75]. Here, we used this assay to investigate whether the Qua1 domain is required for regulation of splicing by Sam68. These experiments were performed in the human astrocytic cell line U138MG, since human astrocytes were previously reported to express relatively low levels of endogenous Sam68 [125] (Suppl. Fig. 2.6). Overexpression of wildtype full-length Sam68 led to a more than three-fold increase of

exon v5 inclusion (Figure 2.5B,C). As expected, a mutant lacking the KH region, which mediates RNA binding, failed to increase exon v5 inclusion. Interestingly, the mutant lacking the Qua1 domain was similarly ineffective as the ΔKH mutant in this assay. This demonstrates that the Qua1 domain is essential for Sam68 activity *in vivo*. The same effect is observed for the F118S variant, which *in vitro*, in the context of the Qua1 domain, does not dimerize. H120K, which is disturbed in dimerization *in vitro* (Suppl. Fig. 2.4) has significantly reduced splicing activity. In contrast, the E110A mutant is only slightly destabilized *in vitro*, and, consistently, supports alternative splicing almost as the wild type protein. Although mutation of Y103 has the least effect on dimerization *in vitro* it seems to be crucial for Sam68 activity, as the Y103S variant is virtually inactive in the splicing assay.
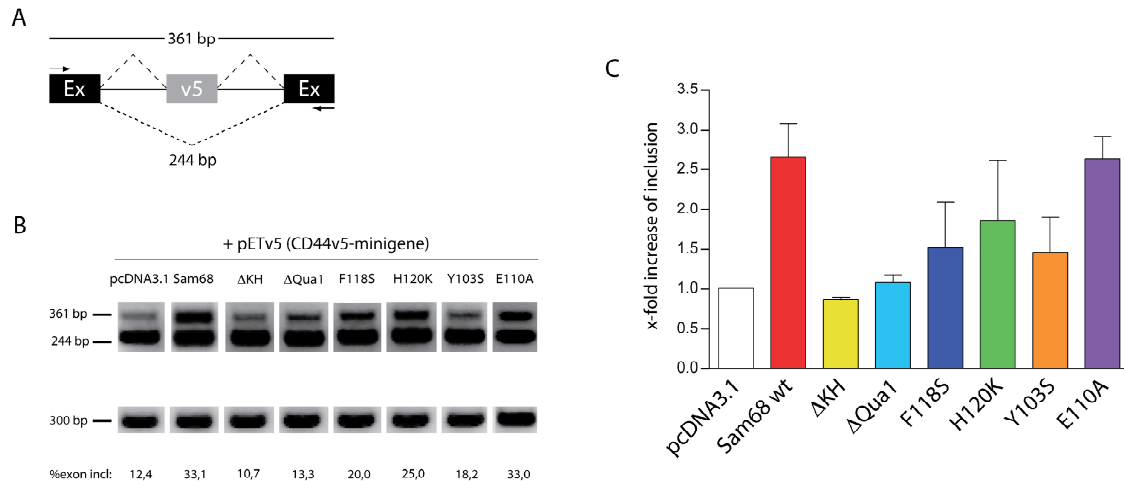
Figure 2.5: Mutational analysis of the Sam68 Qua1 domain reduces CD44 exon v5 inclusion *in vivo*. A) Schematic presentation of the CD44 v5 minigene organization: black boxes display the insulin exons; the gray box represents CD44 exon v5. PCR primer positions are indicated as arrows. Inclusion of CD44 v5 leads to a PCR product with a predicted size of 361 bp whereas exclusion of exon v5 results in a 244 bp PCR product. **B)** RT-PCR analysis of minigene splicing patterns from cytoplasmic RNA of U138MG cells co-transfected with the CD44 v5 minigene construct pETv5 [75] and with plasmids for expression of wildtype or mutant Sam68 proteins. Cells were co-transfected with the minigene construct and the empty vector plasmid (pcDNA3.1) as control. The expression levels of a house keeping gene (RNA polymerase II) are shown for comparison. Values for exon inclusion (expressed as percentage of variant exon-containing RT-PCR products relative to total RT-PCR products) are displayed. **C)** RT-PCR bands were quantified densitometrically using ImageJ Software. Each Sam68 protein was assayed in three independent co-transfection experiments. The graph shows the x-fold increase of % exon inclusion for the RT-PCR products in U138MG cells, relative to the control. Columns represent the mean results of three independent experiments and error bars the standard error of the mean.

Discussion

We show that the Sam68 Qua1 domain adopts a unique four helical dimer fold with the two monomers stacking perpendicularly. The dimer interface is stabilized by a combination of hydrophobic interactions and cross-subunit hydrogen bonds. Residues that mediate these interactions are highly conserved (Figure 2.1A). We demonstrated that mutations of critical residues that build the dimer interface impair the quaternary arrangement. The Qua1 domain is highly conserved within the STAR family, with the exception of the loop region, implying that the Qua1 dimeric structures of these proteins are similar.

While this manuscript was in preparation the crystal structure of the Gld-1 Qua1 dimer was reported [126], with the same topology described here for Sam68. The two structures superimpose very well (backbone coordinate RMSD 1.1 Å; Suppl. Fig. 2.7), although local differences are seen for intermolecular contacts in the loop region, presumably linked to its different composition.

Our results demonstrate that the Qua1 domain alone is sufficient for dimerization of the Sam68 STAR domain. Although we cannot strictly exclude additional contributions from the KH-Qua2 domain to the dimer interface [64], it seems likely that the Qua1 domain is the main determinant for the dimerization of Sam68. In any case, our results unequivocally show that the Qua1 domain is required for the functional activity in splicing regulation of Sam68 *in vivo*. Thus, dimerization of Sam68 is at least as important for splicing as the binding to the target RNA itself, which is mediated by the KH-Qua2 domain. As the surface of Qua1 is negatively charged (Figure 2.3D), it is unlikely that the Qua1 domain contributes directly to RNA binding. Instead, Qua1 homodimerization brings together two KH-Qua2 domains for recognition of bipartite RNA sequences (Figure 2.6), a conserved feature within the STAR family.

With respect to alternative splicing of the CD44 v5 exon, Qua1 dimerization may serve two functions: Firstly, it could stabilize binding to two independent RNA binding sites, one within exon v5 and one in the preceding intron, thus promoting splice site definition by other factors. For instance, Sam68 stabilizes binding of U2AF65 to the CD44 pre-mRNA (Figure 2.6), which has no canonical U2AF65 binding site downstream of the branch point RNA sequence [84]. Secondly, Sam68 appears to promote rearrangement of the spliceosome by leaving the protein-RNA complex. It has been postulated that this is facilitated by a reduced RNA binding affinity linked to phosphorylation [84].

Phosphorylation of Y103 might act as a possible switch in disassembling the Sam68 dimer and thus releasing the protein from the CD44 pre-mRNA (Figure 2.6). The Y103S mutation in Sam68 only partially affects the structural integrity *in vitro* (Figure 2.4B, Suppl. Fig 5), while our *in vivo* splicing assays indicate that Y103 is critical for Sam68 function as it strongly reduces exon v5 inclusion on the CD44 mRNA. The loss of splicing activity is comparable to the F118S mutation, which completely destabilizes the dimeric fold. This cannot be explained exclusively by the modest effect of the Y103S mutation on dimerization of Qua1 in vitro. One possible explanation for this paradox might be the potential phosphorylation of Y103 by a tyrosine kinase [127]. Since Y103 is located close to the edge of the dimer interface, it is in principle accessible to phosphorylation by a kinase. Phosphorylation could sterically interfere with the dimerization. Additionally the intermolecular hydrogen bond network, which we have shown to be crucial for dimerization, could be disturbed by additional hydrogen bond acceptors and donors provided by the phosphate group. Finally, the additional negative charge of the phosphate might lead to repulsion of the two monomers and thereby destabilize the dimer. Phosphorylation of Y103 might thus play a role for the Sam68 splicing activity, beyond this residue being important for the stability of the Qua1 dimer.

Future studies should focus on understanding the molecular mechanisms of the function of the Sam68 STAR domain and its modulation by phosphorylation.
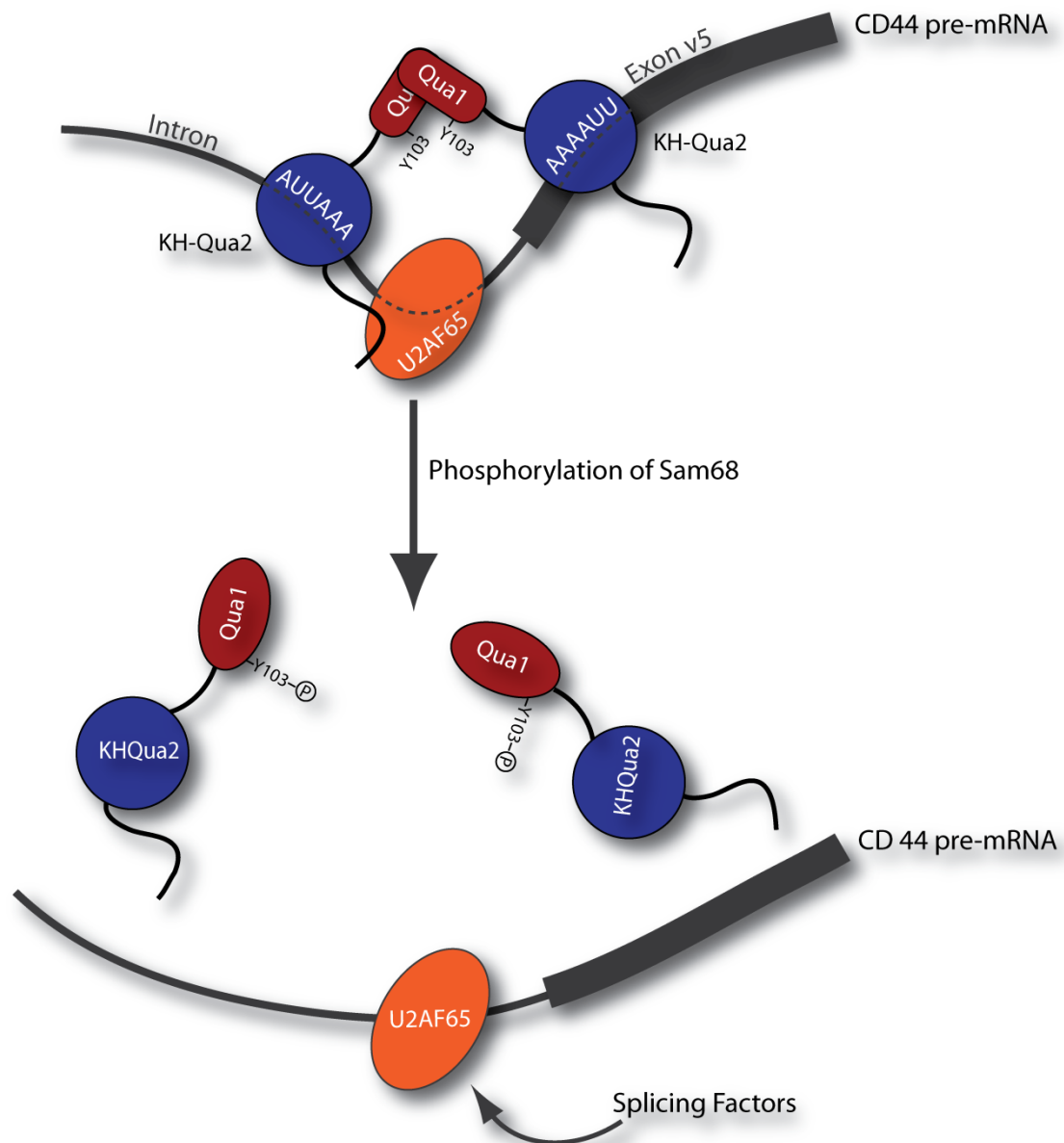
Figure 2.6: Model of the Sam68 interaction with CD44 pre-mRNA and U2AF65. Sam68 dimerization via the Qua1 domain may allow simultaneous binding to cognate RNA elements in the CD44 exon v5 and the preceding intron. Phosphorylation of Sam68 has been shown to reduce the RNA binding affinity and thus releases Sam68 from the pre-mRNA. This is required for recruiting constitutive splicing factors and spliceosome assembly. Phosphorylation of Y103 might additionally contribute to the release of Sam68 from the pre-mRNA by destabilizing the dimerization via Qua1.

PDB Accession Codes

The coordinates and NMR restraint files for the Qua1 homodimer are deposited in the PDB with accession number 2XA6. Chemical shifts are deposited in the BMRB, accession 16969.

Acknowledgements.

Supplemental Figures
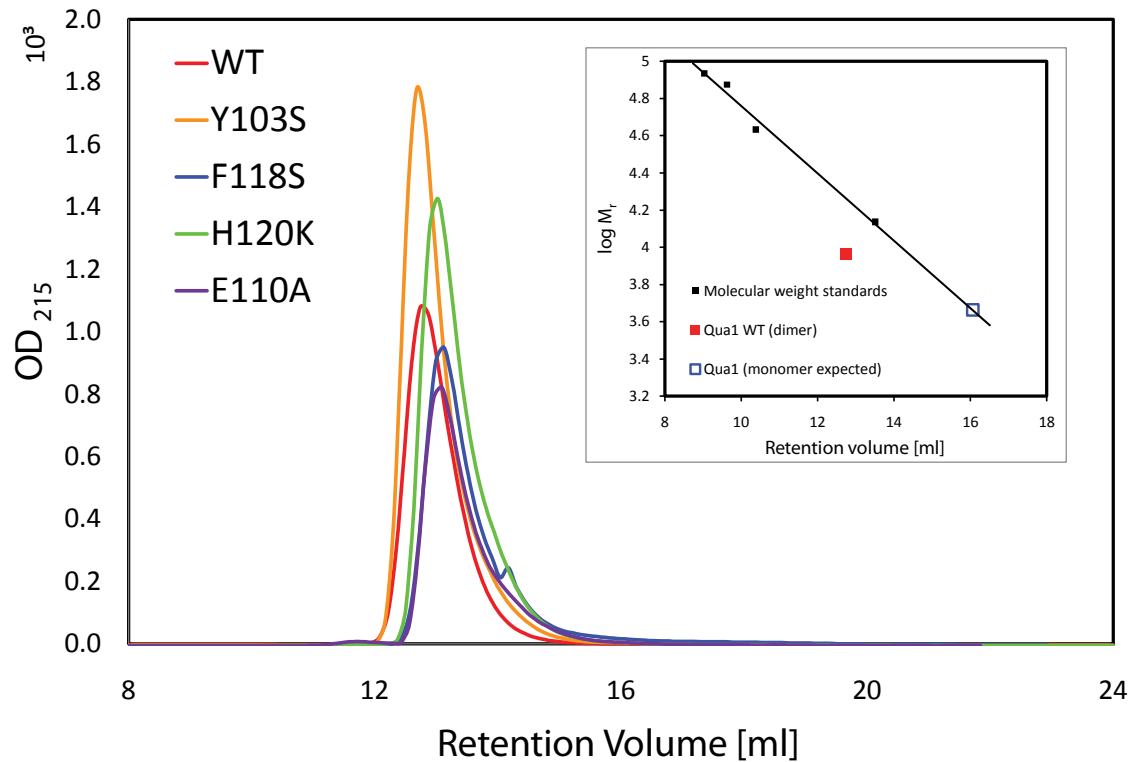
Supplementary Figure 2.1: Size exclusion chromatography of Sam68 wildtype as well as Y103S, F118S, H120K and E110A variants. Proteins were loaded onto a Superdex 75 10/300 gel filtration column (GE Healthcare). Molecular weight was calibrated with standard proteins: Ribonuclease A (13.7 kDa), Ovalbumine (43.0 kDa monomer, 86 kDa dimer) and Conalbumine (75 kDa) (GE Healthcare). The retention volume of wt and Y103S and mutant protein corresponds well to that expected a dimer. F118S, H120K and E110A have a slightly increased retention volume compared to the wildtype. This however does not match the retention volume expected for the monomeric form of the protein. This might be due to the fact that the mutants are destabilized or unfolded resulting in an increased hydrodynamic radius.
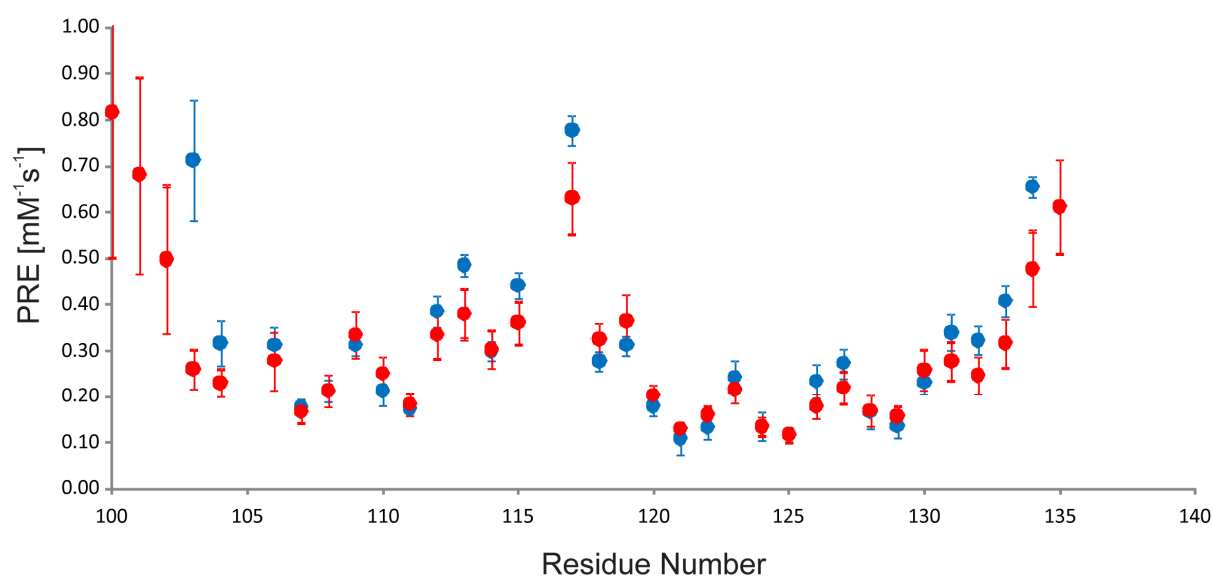
Supplementary Figure 2.2: Paramagnetic relaxation enhancements (PREs) for the backbone amide protons of Qua1. Measured PREs (blue) and PREs back-calculated for the ensemble of NMR structures (red). For some surface protons – particularly protons located in loops and the flexible termini– the measured PRE is higher than calculated. This is due to chemical exchange between these protons and water protons. Water is coordinated to gadolinium in Gd(DTPA-BMA) and thus experiences large relaxation enhancements. As this water ligand exchanges with free water and then to amide protons the large relaxation enhancement is partly transferred.
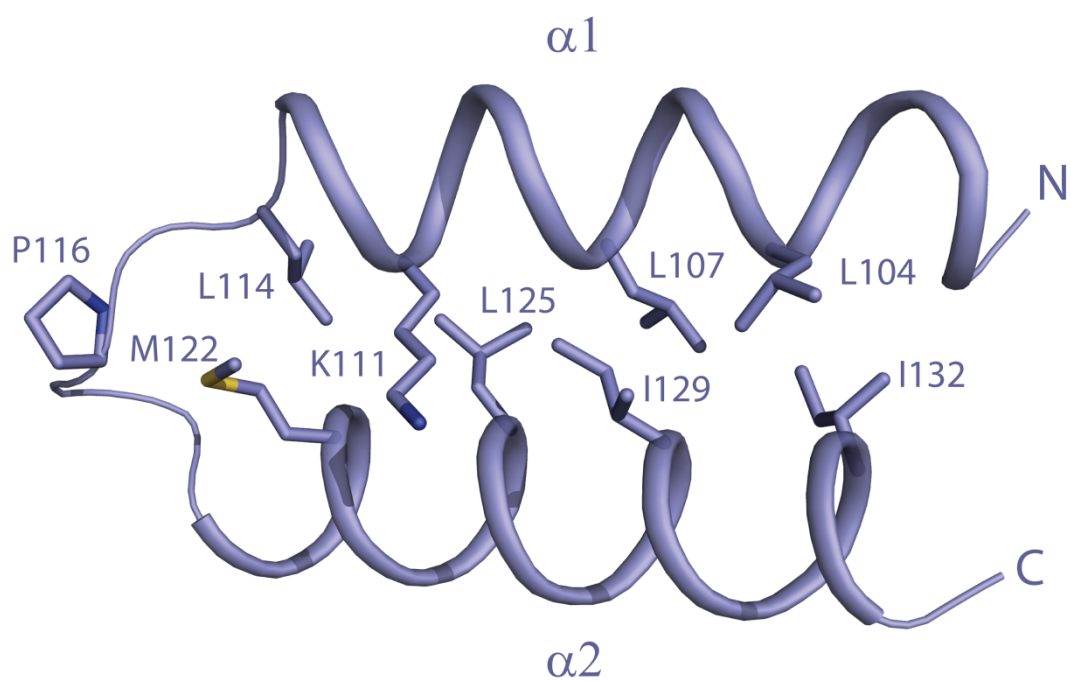
Supplementary Figure 2.3: Ribbon representation of a Qua1 monomer. Side chains of residues, which stabilize the helical hairpin, are shown as sticks.
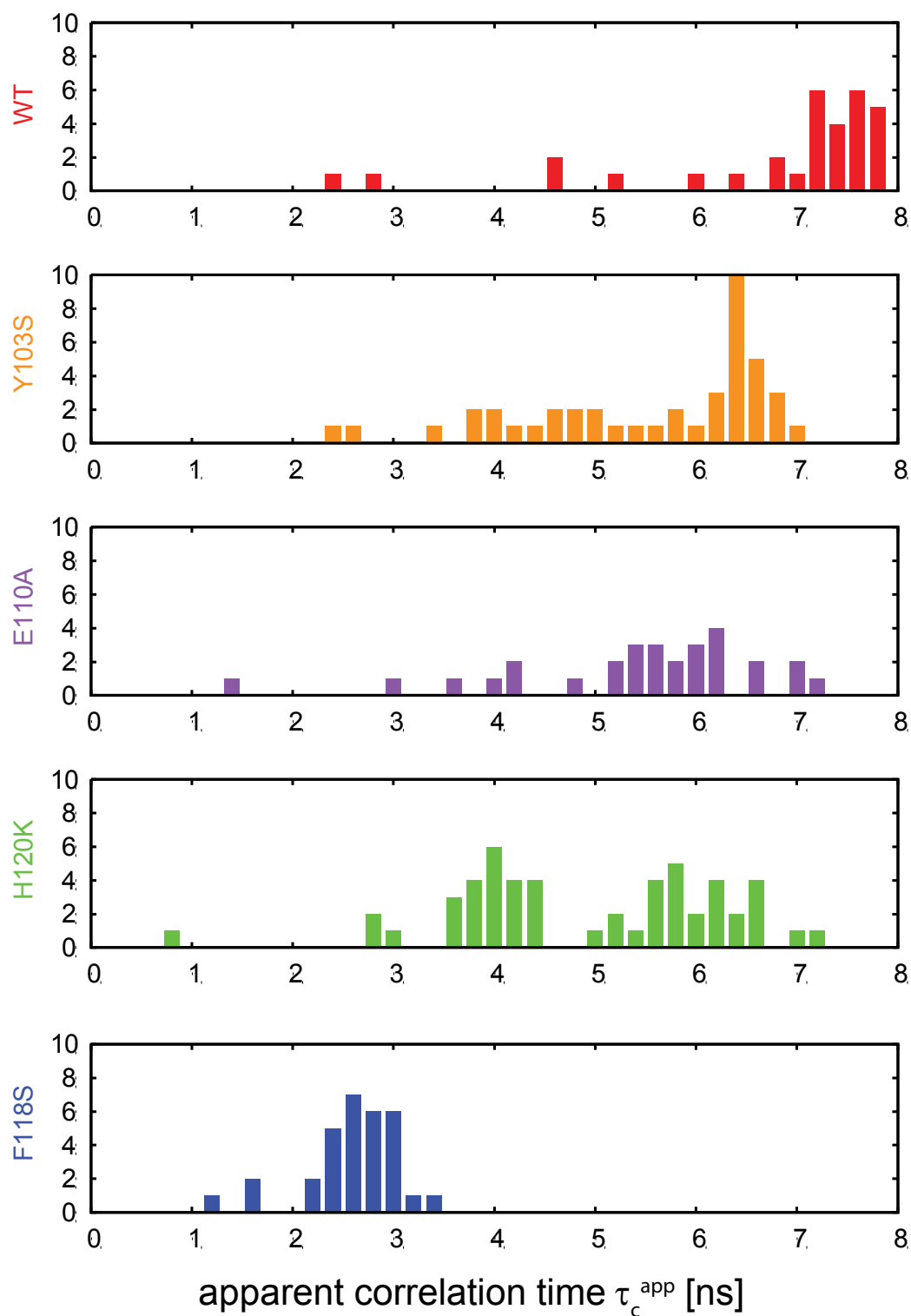
Supplementary Figure 2.4: Histograms of apparent correlation times (τ_c^{app}) derived from ^{15}N relaxation data for Sam68 Qua1 (residues 95-135). Apparent local correlation times were calculated from the ^{15}N R_2/R_1 relaxation rate ratios. The average τ_c^{app} for the wild type Qua1 is consistent with a homodimer, whereas the reduced values for the mutants indicate that dimerization is disturbed. F118S, which is unfolded and monomeric, has the lowest average correlation time. All other mutants show correlation times in between those of F118S and wild type. Note, that the very small correlation times for wild type Qua1 arise from flexible regions of the protein.
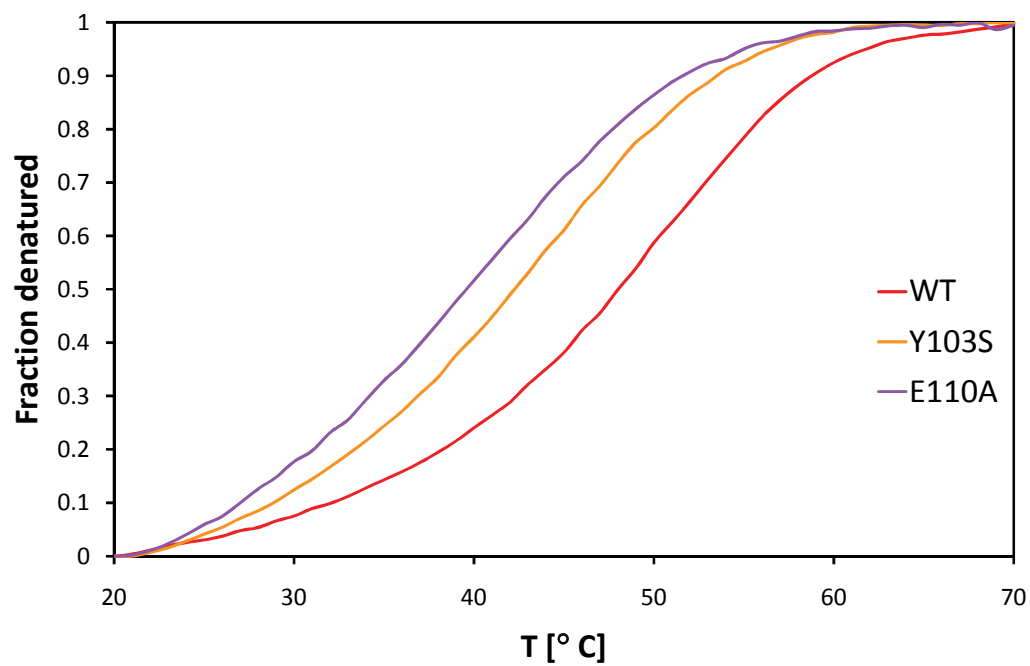
Supplementary Figure 2.5: Thermal denaturation of Sam68 Qua1 WT and the E110A and Y103S variants monitored by CD spectroscopy. The fraction of denatured protein was determined by measuring molar ellipticity θ at 222 nm at steps of 1° C (heating rate 2° C min^{-1}).

Supplementary Figure 2.6: Sam68 expression in U138MG and HeLa cells. Western Blot analysis of U138MG and HeLa cell lysates show significant more endogenous Sam68 expression in HeLa cells compared to the astrocytic U138MG cells. GAPDH western blot served as a loading control.
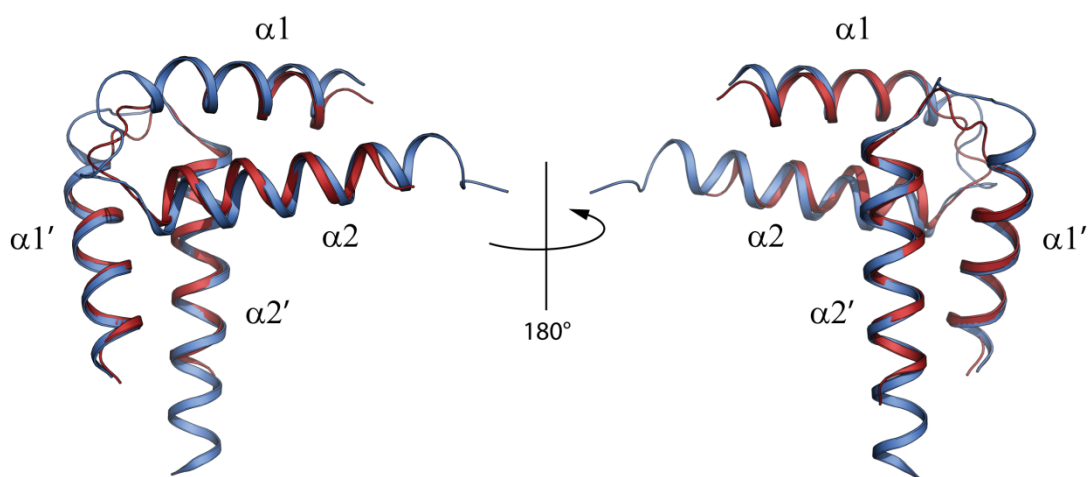
Supplementary Figure 2.7: Superposition of the Sam68 Qua1 (red) and Gld-1 Qua1 (blue) structures. Backbone atoms of residues 100-114, 115-116 and 117-134 of Sam68 Qua1 are aligned with the respective residues 146-160, 163-164 and 166-183 of Gld-1 Qua1. The two structures superimpose with a backbone coordinate RMSD of 1.1 Å. The sequence identity of the Qua1 domains of the two proteins is 23%. Note that in the linker region only P116 is conserved.

Unpublished results

The study of large proteins and protein complexes by NMR spectroscopy can be challenging. Although perdeuteration of proteins allows the acquisition of standard triple resonance experiments, overlap of peaks severely interferes with resonance assignment. Especially unfolded and random coil regions of proteins give rise to peaks with hardly any dispersion in the spectrum. To study the STAR domain of Sam68, protein variants of variable length have been recombinantly expressed and purified.

The STAR domain boundaries have been defined according to a sequence alignment with other members of the STAR family. The conserved STAR domain, ranging from aa 135 to aa 280, expresses with high yield in *E. Coli*. Estimated from size exclusion chromatography, the molecular weight of the Sam68 STAR domain is approximately 40 kDa, which is consistent with the mass of a homodimer. This finding was validated by analytical ultracentrifugation (data not shown). The high overlap in the ^{1}H,^{15}N HSQC spectrum is consistent with the presence of several loop regions in the RNA binding domain. Additionally, the flexible linker between the Qua1 and the KH-Qua2 moiety of the STAR domain contributes to signal overlap. To support the resonance assignment shorter protein variants, containing only the Qua1 domain with and without the linker domain or the KH-Qua2 domain have been produced (Figure 2.7).
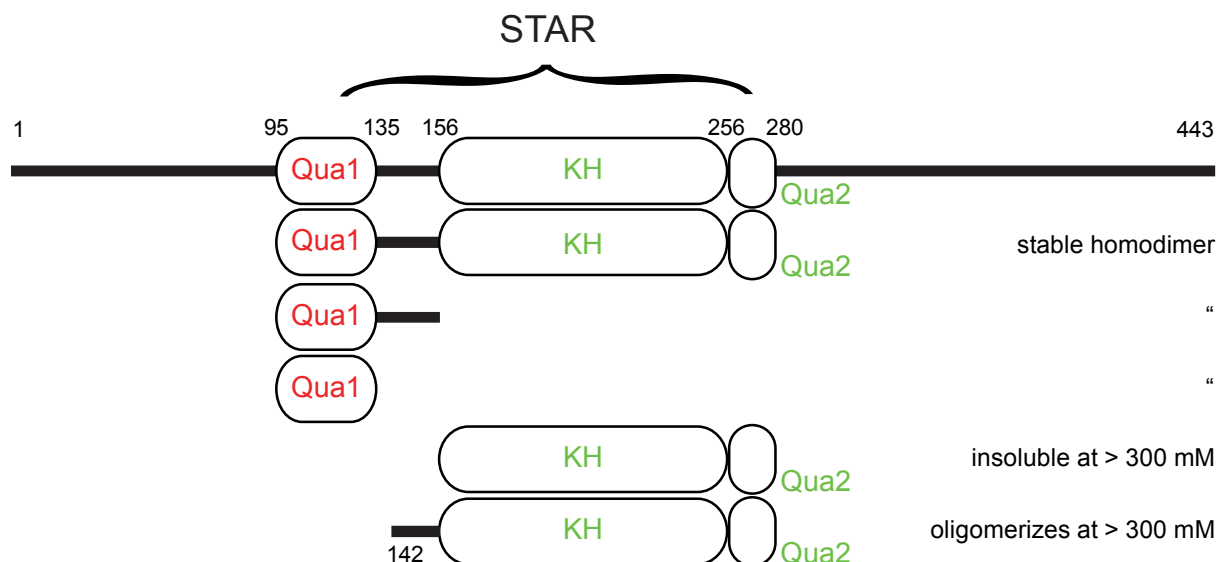
Figure 2.7: Protein variants used for characterization by NMR spectroscopy.

Resonance assignments from the Qua1 domain could be directly transferred to the STAR domain since hardly any difference was observed between the amide chemical shifts of the individual Qua1 domain and the Qua1 domain in the context of the STAR domain. However,

the last ten residues of the flexible linker between the Qua1 and the KH-Qua2 domain experience large chemical shift perturbations (Figure 2.7). This might indicate that the KH-Qua2 region has an N-terminal extension which has not been described before. Consistently, the KH-Qua2 variant, missing those residues, is not stable at high concentrations. Secondary chemical shift analysis shows that the extended N-terminus adopts a β-sheet conformation. However, N-terminally elongated KH-Qua2 (aa 142-280) still tends to aggregate at protein concentrations above 300 µM, as judged from the quality of the NMR spectra recorded on this protein variant. Thus, the Qua1 domain can be regarded as solubility tag for the less soluble KH-Qua2 moiety.
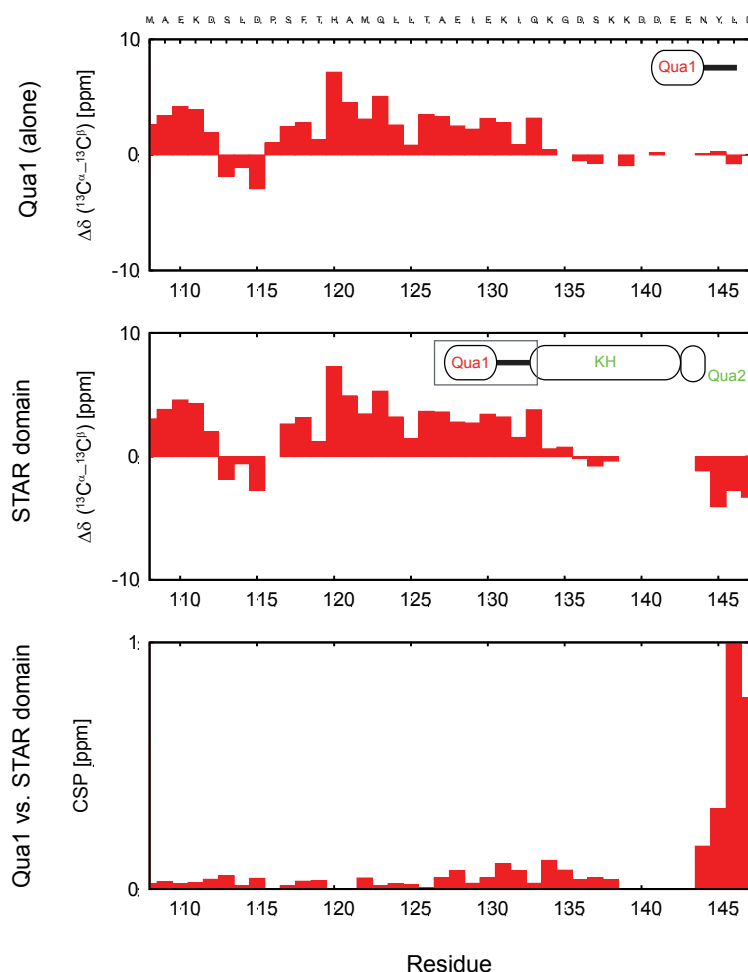
Figure 2.8: Analysis of the chemical shift of residues of the Qua1 domain alone and in context of the STAR domain. The last five amino acids of the linker region form a β-sheet, which elongates the KH fold.

Using different protein variants of the Sam68 STAR domain barely 50% of the backbone amide resonances could be sequence-specifically assigned. For a detailed analysis more than 90% assigned residues would be desirable. Improvements of the solubility of the KH-Qua2 (aa 142-280) variant in combination with perdeuteration might further enhance the quality of the NMR spectra for this domain. Assignments of the KH-Qua2 domain may serve as a basis for completion of the backbone assignment of the STAR domain. Nevertheless, even such a scarce set of assignments can be useful. Since the assigned amides are distributed over the whole amino acid sequence, they can be used to probe interactions between certain regions of the protein and for example RNA.

To characterize the binding of Sam68 to its target RNA, I performed an NMR titration. In this titration an A- and U-rich oligonucleotide was used to bind to Sam68 (Figure 2.9). Free and bound state of the protein are in fast exchange on a 900 MHz spectrometer. Therefore shifts of several peaks in the ^{1}H,^{15}N-TROSY spectrum can be observed. Fast exchange at high magnetic field strength is rather common for weak binding in the µM range. Note, that interactions of both KH-Qua2 domains of the homodimer to two distinct binding sites on one RNA molecule might support the binding in a cooperative manner. Therefore, longer RNA oligonucleotides containing two binding sites for Sam68 should be analyzed to assess the RNA binding affinity. Most interestingly, residues in the N-terminal elongation of the KH-domain are affected in the NMR titration as well. These region might thus provide additional contacts to RNA, which alter the binding specificity of Sam68 with respect to other members of the STAR domain family.
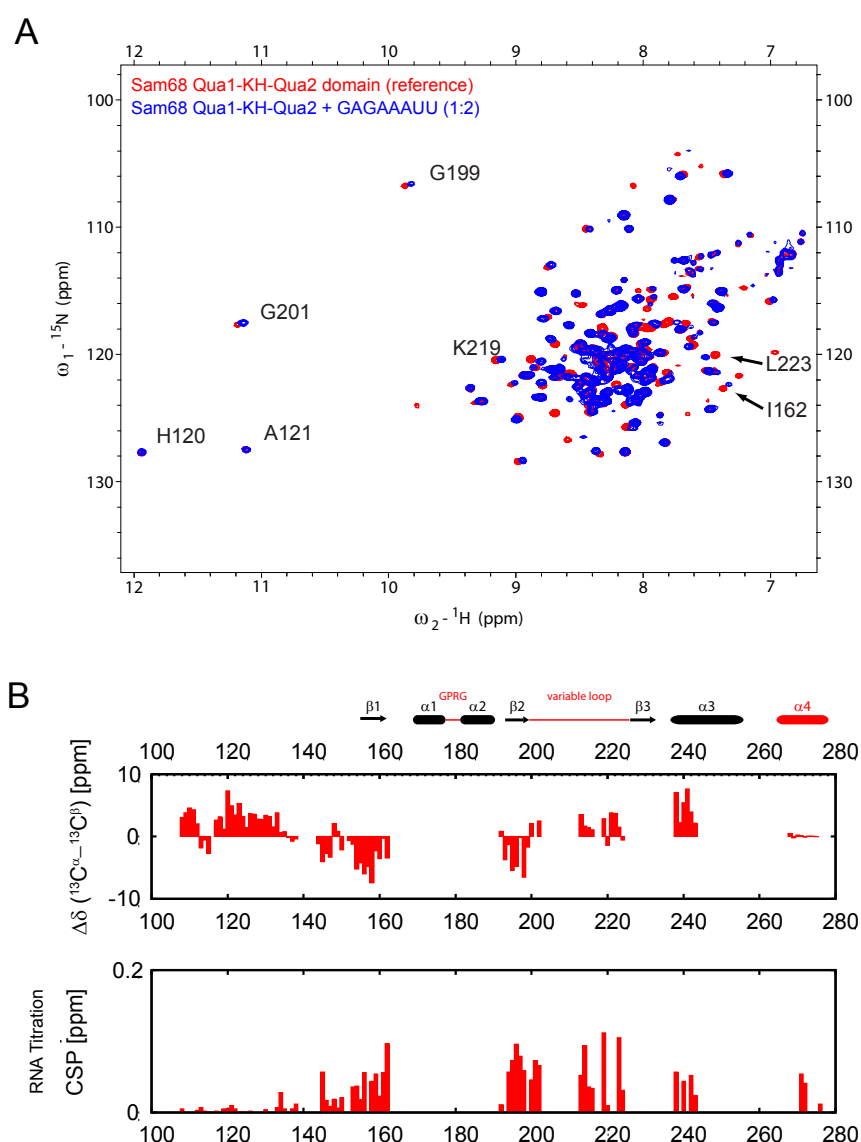
Figure 2.9: NMR Titration of Sam68 STAR domain with an RNA oligonucleotide. A) ^{1}H,^{15}N-TROSY spectra of Sam68 STAR domain alone (red) and in complex with a 2-fold excess of GAGAAAUU-RNA (blue). Spectra were recorded at a static magnetic field of 21 Tesla corresponding to 900 MHz proton lamour frequency. **B)** ^{13}C$^\alpha$-^{13}C$^\beta$ secondary chemical shift of assigned residues of Sam68 STAR domain and chemical shift perturbation upon RNA binding. The secondary structure of the homologues SF1 is indicated at the top. Elements which are important for specific RNA recognition are highlighted in red.

Binding of Sam68 to U2AF65 is a critical step in the regulation of alternative mRNA splicing. However, the molecular basis for the interaction of these proteins is unclear. NMR titrations revealed that the unstructured C-terminal part of Sam68 can bind to the RRM domains of U2AF65, whereas the N-terminus does not bind. This binding might be mediated via a potential ULM in Sam68 recognized by one of the RRM domains. However, chemical shift perturbations observed are not restricted to one of the RRM domains but rather distributed throughout the whole molecule (Figure 2.10). The C-terminus of Sam68 is rich in Tyrosine residues. Therefore it cannot be excluded that additional unspecific hydrophobic interactions stabilize the interaction between the Sam68 C-terminus and the RRM domains of U2AF65 *in vitro*, as they expose aromatic residues on their surface. To validate the physiological relevance of this interaction, several C-terminal point mutations of Sam68, especially in the region of the potential ULM, should be characterized in a cell-based splicing assay as described in [128].
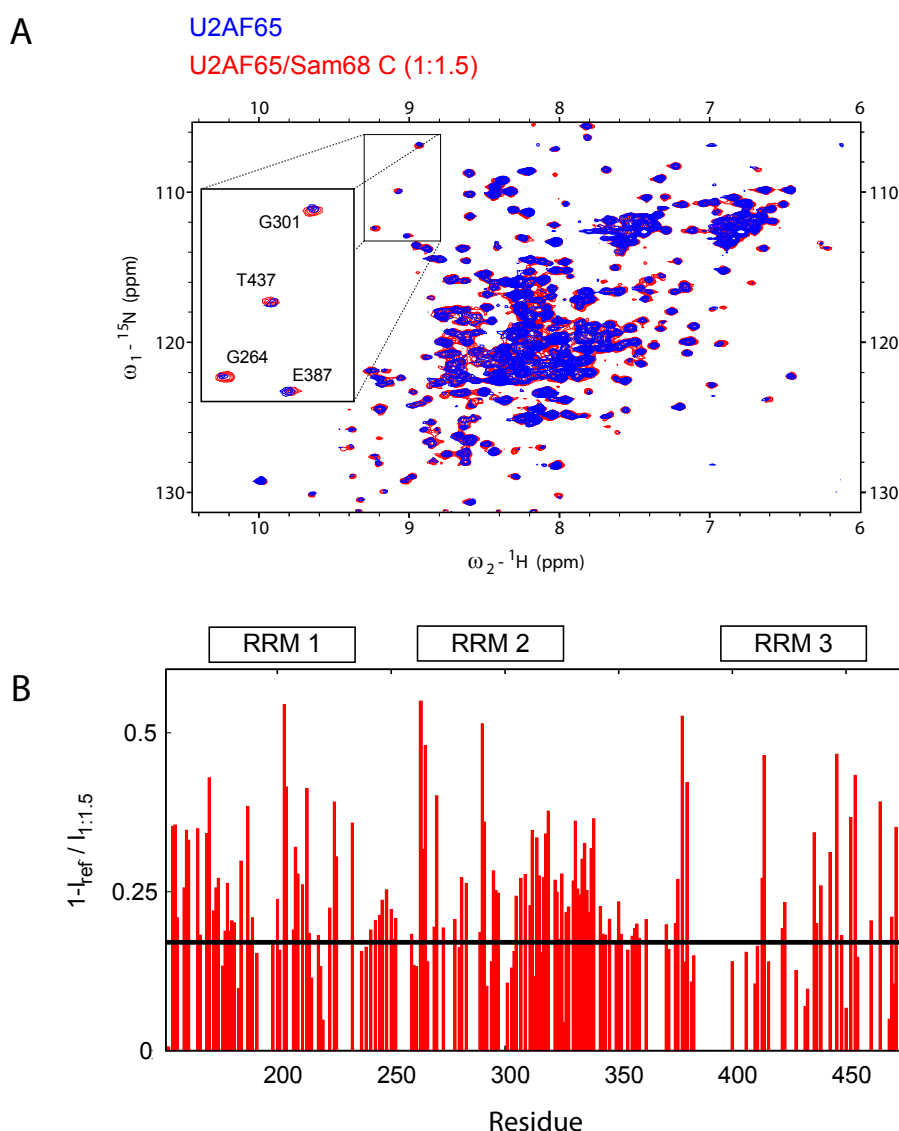
Figure 2.10: Titration of U2AF65 RRM 1,2,3 with the C-terminus of Sam68 A) ^{1}H,^{15}N-HSQC spectrum of U2AF65 RRM 1,2,3 domains alone (red) and in complex with a 1.5-fold excess of Sam68 C-terminus. Spectra were recorded on a 750 MHz spectrometer (17.6 T). B) Reduction of U2AF65 HSQC cross peak intensity (I_{ref}) upon the addition of 1.5-fold excess of Sam68 C-terminus ($I_{1:1.5}$).

CHAPTER 3:

IPSE/ALPHA-1, AN IgE-BINDING CRYSTALLIN

^{1}H, ^{13}C and ^{15}N chemical shift assignments of IPSEΔNLS

Summary

During infection with *Schistosoma mansoni*, the deposition of eggs modulates the initial T_H1 response into a T_H2 response. IPSE/alpha-1 from *Schistosoma mansoni* eggs has a potent IL-4 inducing effect on human basophils. Therefore it might be the major component triggering the switch in the immune response.

Chemical shifts for NMR-active nuclei provide detailed information on molecular conformation and dynamics of proteins. Here, we present the sequence-specific resonance assignment for the core domain of IPSE/alpha-1. Approximately 98 % of all resonances have been assigned. Chemical shift based structure prediction provides first evidence that IPSE/alpha-1 is a member of the crystallin family.

^{1}H, ^{13}C and ^{15}N chemical shift assignments of IPSEΔNLS

N. Helge Meyer [1,2], Gabriele Schramm[3], Michael Sattler[1,2]

[1] Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany,
[2] Biomolecular NMR and Munich Center for Integrated Protein Science (CiPS^M) at Department Chemie, Technische Universität München, Garching, Germany
[3] Forschungszentrum Borstel, Germany

Correspondence: Michael Sattler, E-mail: sattler@helmholtz-muenchen.de

Abstract

The interleukin-4-inducing principle from *Schistosoma mansoni* eggs (IPSE/alpha-1) is a major immunogenic component of schistosomes. It potently triggers the release of interleukin-4 from basophilic granulocytes in an IgE-dependent manner, suggesting a key function in the modulation of the host's immune response to *Schistosoma mansoni* infection. Here we present the near complete assignment of an IPSE/alpha-1 variant, IPSEΔNLS, which comprises the core domain of the protein.

Biological context

Schistosomiasis is a chronic disease, which is caused by parasitic trematode worms of the genus *Schistosoma.* Approximately 700 million people in 74 mostly rural countries are at risk of being infected. More than 207 million people, with 85% of the infections occurring in Africa, are suffering from Schistosomiasis and 200.000 death cases per year have been reported for sub-Saharan Africa (World Health Organisation, Schistosomiasis Facts, 2010). *Schistosoma mansoni*, which is the best characterized member of *Schistosoma* family, is the major cause of intestinal Schistosomiasis [92].

A major part of the pathology of a chronic *S. mansoni* infection results from the immune response of the host to the parasite eggs, which induce a T helper type 2 (T_H2) reaction. During chronic infection, *S. mansoni* eggs can accumulate in the intestine or liver resulting in granulomatous inflammation and subsequently fibrosis and organ failure [129]. IPSE/alpha-1 has been identified to be a major immunogenic component secreted from *S. mansoni* eggs [95]. It has been shown that IPSE efficiently mediates the release of the signature T_H2 cytokines IL-4 and IL-13 and the degranulation of basophils. This step is supposed to be critical for the initiation/amplification of the T_H2 response of the host. IPSE/alpha-1 activates basophils *via* an antigen-independent interaction with IgE bound to the surface of basophils [98]. However, details on this interaction and the subsequent receptor activation are obscure.

To study the structural basis for the interaction between IPSE/alpha-1 and IgE we assigned the resonances of the core domain of IPSE/alpha-1.

Methods and experiments

A variant of IPSE/alpha-1, named IPSEΔNLS, which is lacking the last 10 amino acids was expressed from the expression vector pProExHtB IPSEΔNLS as a His-tagged fusion protein [96]. To obtain uniformly ^{15}N, ^{13}C-labeled protein freshly transformed *E. coli* BL21(DE3) cells were cultivated in minimal (M9) medium supplemented with ^{13}C-D-Glucose and ^{15}NH$_4$Cl. IPSEΔNLS was purified from inclusion bodies as described elsewhere [94]. After removal of the His-tag using tobacco etch virus (TEV) protease, the protein was concentrated to a final concentration of 1 mM and kept in phosphate buffered saline (PBS, 137 mM NaCl, 2,7 mM KCl, 10 mM Na$_2$HPO$_4$.) at a pH of 7.0.

All NMR measurements were carried out at 298 K on a Bruker Avance III 750 MHz spectrometer equipped with a TXI probehead. Spectra where processed with NMRPipe [110] and analyzed with Sparky [111]. For assignment CBCA(CO)NH, CBCANH and spectra were recorded [3]. Some of the missing amide proton and ^{13}C^α, ^{13}C^β resonance assignments could be obtained from ^{15}N-edited NOESY spectra with a mixing time of 70 ms. Sidechain ^{13}C and ^{1}H resonances were assigned using an H(C)CH-TOCSY spectrum.

Assignments and data deposition

Near complete assignments were obtained for IPSEΔNLS. Approximately 98% of all ^{15}N, ^{1}H^N, ^{13}C^α and ^{13}C^β, including all resonances of 5 Prolines could be assigned. Due to severe line-broadening in the ^{15}N-edited experiments, ^{15}N and ^{1}H^N resonances of D21, C23, K115 and N123 could not be assigned. Due to spectral overlap the resonance assignments are missing for D21 and E120 ^{1}H^α and K118 ^{1}H^ε. The chemical shifts of V63 ^{15}N and ^{1}H^N are unambiguously assigned and have unusually high chemical shift values. T82 and R36 have high ^{1}H^N shifts. The atypically low chemical shifts of N80 and Y81 ^{1}H^α are likely induced by ring current effects of Y81. The ^{13}C^β of all cysteine residues are consistent with the oxidized state.

Blast search in the sequence database does not reveal any similar sequences to IPSE/alpha-1, besides proteins originating from other Schistosomes. However, secondary-structure predictions indicate a mainly β-sheet fold. Due to the presence of Y/FXXXXY/FXG signature sequence, IPSE/alpha-1 was predicted to be a member of the crystallin family. Secondary chemical shift analysis indeed shows, that IPSEΔNLS is an all β-sheet protein and give the first experimental clue that IPSE adopts a crystallin fold.

A

B

**Figure 3.1: A) ^{1}H,^{15}N-HSQC spectrum of IPSEΔNLS and B) ^{13}C secondary chemical shifts.**

Characterization of IPSE by NMR spectroscopy

IPSE/alpha-1 has recently been identified to be one of the major immunogenic component in *Shistosoma mansoni* infection [95]. However, blast search in the sequence database does not reveal any similar sequences, besides proteins originating from other Schistosomes. A Y/FXXXXY/FXG motif in the sequence indicates the presence of a Greek key motif. Besides that, little is known about the structure and the molecular function of this protein. A detailed biophysical characterization of IPSE by NMR spectroscopy was carried out to understand the findings from recent biochemical studies [94, 97, 98].
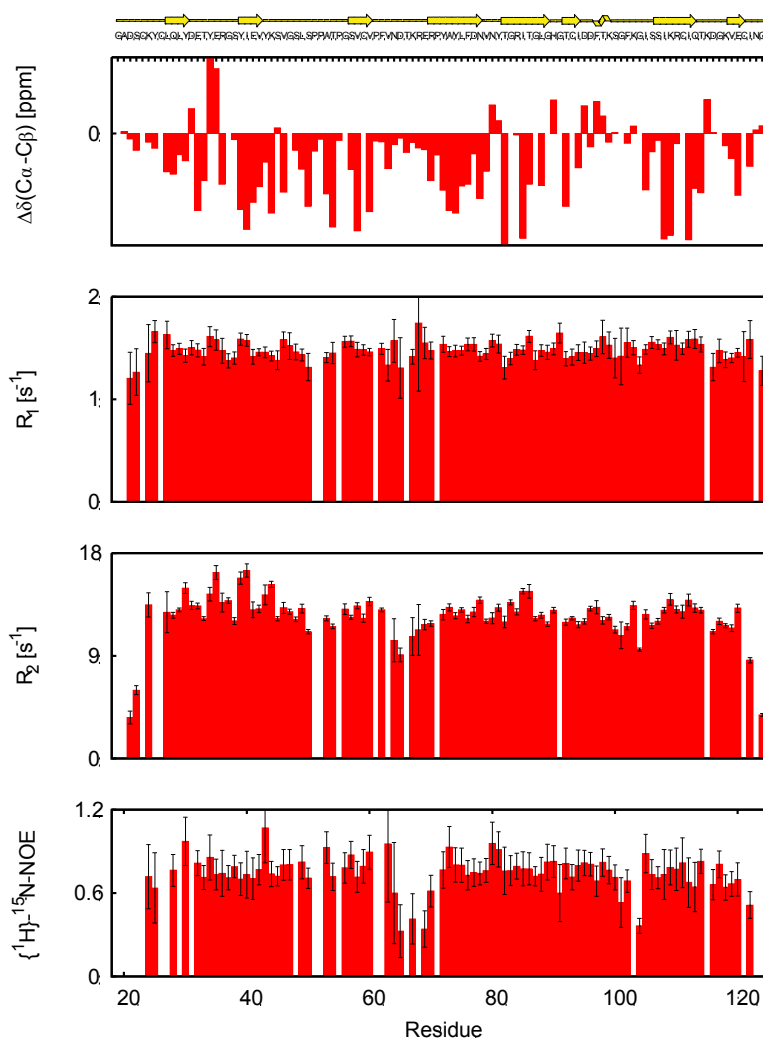
Figure 3.2: NMR characterization of IPSEΔNLS. [13]C secondary chemical shifts, [15]N R_1, R_2 relaxation rates and {[1]H}-^{15}N heteronuclear NOE are plotted *versus* residue numbers. The secondary structure is indicated at the top.

Secondary chemical shift analysis shows that IPSEΔNLS (aa 19-121) is an all β-fold protein apart from a small one turn α-helix (Figure 3.2). This observation was confirmed by CD Spectroscopy (data not shown). IPSEΔNLS contains 8 beta strands connected by loops of variable lengths. NMR relaxation analysis indicates that the protein is rigid with the exception of one loop (res 62-69) which experiences an increased degree of flexibility (Figure 3.2). According to ^{15}N R_1/R_2 ratio IPSEΔNLS is a monomer in solution. This is consistent with a missing C-terminal cysteine, which is known to be involved in covalent dimerization of the molecule [94].

Solution and crystal structure of IPSE/alpha-1

To elucidate the molecular basis for the function of IPSE we determined the solution structure of IPSEΔNLS (aa 19-121) using hetero-nuclear double- and triple resonance NMR spectroscopy. 2650 experimental distance restraints and 161 dihedral angle restraints, which are based on TALOS+ [9] predictions (Table 3.1), were used to calculate the solution structure (Figure 3.3A).

In its monomeric form, the IPSE scaffold consists of an 8 anti-parallel beta strand (β1-β8) sandwich arranged in two intertwined Greek key motifs with a pseudo 2-fold symmetry. A characteristic β hairpin motif forms a loop between the first two β strands of each motif. This arrangement clearly resembles the crystallin fold, in which the third β strand of each motif is exchanged with the fourth strand in the other motif. The structures of the C-terminal domain of human γ-B-crystallin and the IPSE/alpha-1 monomer align with a backbone RMSD of 1.718 Å.
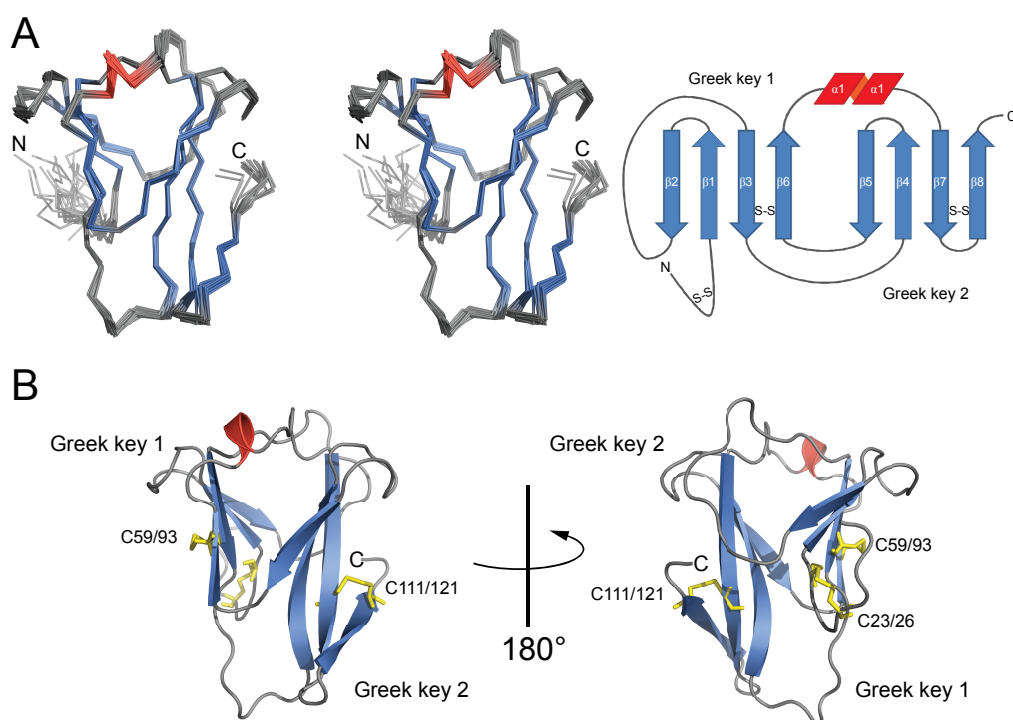
Figure 3.3: Structure of IPSE/alpha-1. A) Stereo view of the ensemble of the 20 lowest energy structures solved by NMR (left) and topology diagram (right). B) Ribbon representation of IPSEΔNLS solution structure.

Table 3.1: Structural statistics of the Sam68 Qua1 homodimer.

NOE-based distance restraints[1]	
Intraresidual, sequential	1151
Medium range ($2 \leq |i\text{-}j| \leq 4$)	277
Longe range ($|i\text{-}j| \geq 5$)	1202
Total	2630

Other restrains	
$\phi+\psi$ dihedral angle restraints	162

Coordinate Precision RMSD	
Backbone (Å)	0.44±0.09
Heavy atom (Å)	1.06±0.11

Consistency (structure vs. restraints)	
RMSD (Å) from experimental distance restraints[1]	0.011±0.001
RMSD (°) from experimental torsion angle restraints[2]	0.919±0.052

WHATCHECK[3]	**Structure Z Scores**
First generation packing quality	-0.403
Second generation packing quality	-1.436
Ramachandran plot appearance	-0.562
χ_1/χ_2 rotamer normality	-2.807
Backbone conformation	-0.491

Ramchandran plot[3]	
Most favoured regions	91.7%
Allowed regions	8.2%
Generously allowed regions	0.1%
Disallowed regions	0.0%

Statistics are given for the 20 lowest energy structures after water refinement out of 100 calculated. The CNS E_{repel} function was used to simulate van der Waals interactions with an energy constant of 25 kcal mol^{-1} Å^{-4} using "PROLSQ" van der Waals radii [117]. RMSD and PROCHECK values apply for residues 27-120.

[1] Distance restraints were employed with a soft square-well potential using an energy constant of 50 kcal mol^{-1} Å^{-2}. No distance restraint was violated by more than 0.3 Å.

[2] Torsion angle restraints derived from TALOS [9] were applied to φ, ψ backbone angles using energy constants of 200 kcal mol^{-1} rad^{-2}.

[3] PROCHECK [15] and WHATCHECK [16] were used to determine the quality of the structure. Positive WHATCHECK Z-scores indicate that structure is better than average.

Similarity and differences of IPSE with other crystallins

Despite the strong structural homology, IPSE/alpha-1 shows only moderate sequence homology with gamma B crystallin (Figure 3.4B). Sequence identity of IPSE/alpha-1 with the C-or N-terminal domain of γ-B-crystallin is less than 15%. This has been reported for several other members of the γ/β crystallin superfamily found in lower organisms as reviewed in [130]. One of these proteins is Spherulin 3A from *Physarum polycephalum*. Its role as a link between a potentially monomeric ancestor and the dimeric crystallin fold found in higher eukaryotes is controversely discussed [131, 132]. Similar to IPSE, Spherulin 3A dimerizes via a disulfide-bridge. However, the respective cysteins are located at the N- rather than at the C-terminus (Figure 3.4A). The sequence similarities of IPSE/alpha-1 and Spherulin 3A to γ-B-crystallin are comparable.
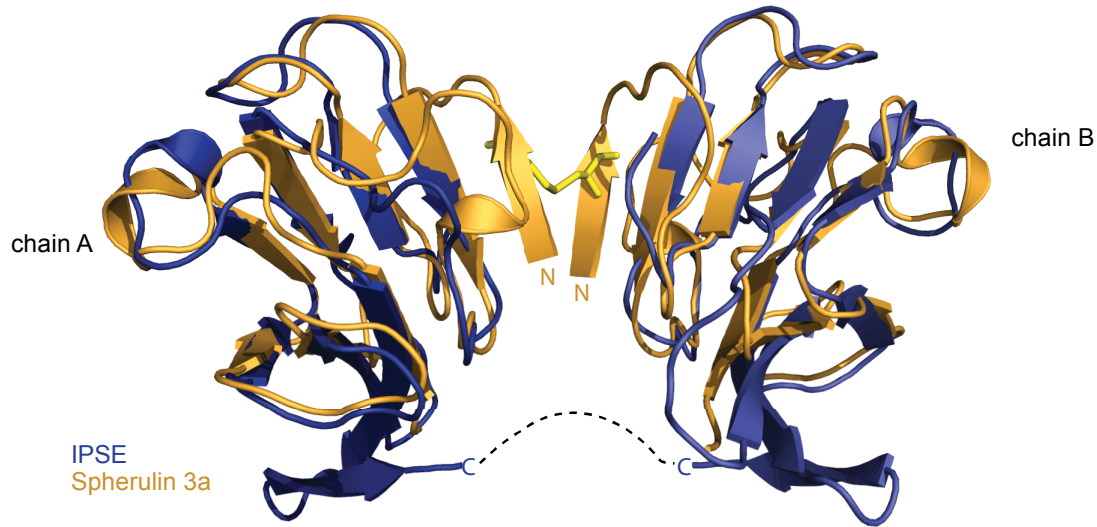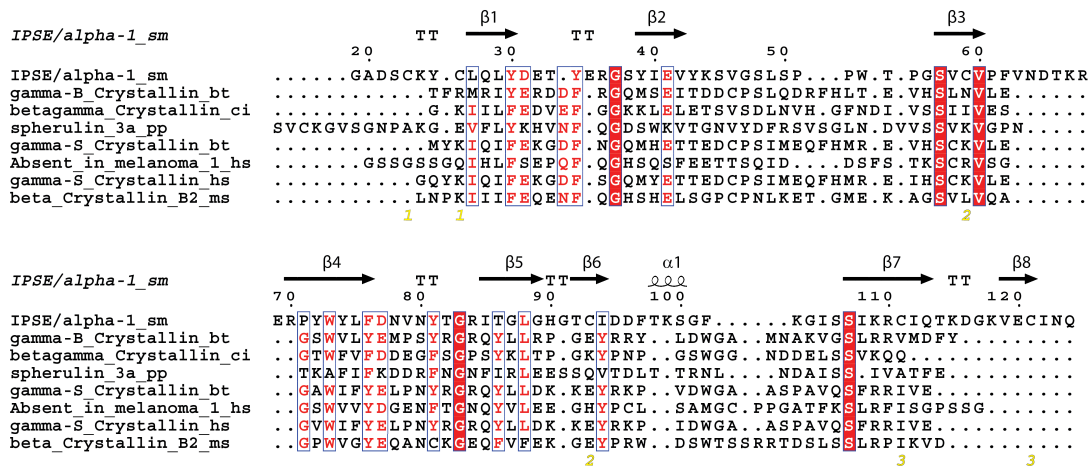
Figure 3.4: Structural homology of IPSE with with proteins of the crystallin family. A) Structure superposition of two IPSE monomers with dimeric spherulin 3A. Structures align with a backbone RMSD of 1.8 Å. **B)** Structure based alignment of proteins of the crystallin superfamily. Secondary structure elements of IPSE are depicted at the top. Disulfide bridges are indicated as yellow numbers.

A few prominent motifs in the amino acid sequence seem to be adequate to maintain the stable crystallin fold. One of those is the highly conserved signature sequence Y/FXXXXY/FXG interacting with a widely conserved Serine on the fourth beta strand of each Greek key motif. Interestingly this sequence is mutated to YDETYERG, therefore shifting the second Y by one amino acid, in the first Greek key motif of IPSE, without disturbing the formation of the hairpin loop.

Whilst the stability of the γ-B-crystallin mainly depends on the strong hydrophobic core, the IPSE fold is additionally stabilized via three disulfide bonds (23/26, 59/93 and 111/121), which have been identified by NMR spectroscopy (Figure 3.4B). Those disulfide bonds presumably compensate the lack of a Tyrosine corner, which is present in most γ-crystallins, so that the protein is stable over a wide pH range and in presence of 2M Urea. Consequently,

addition of reducing agents, such as DTT, dramatically destabilizes the protein and leads to protein precipitation (data not shown).

Different from other proteins of the crystallin family the second and the third beta strand of the C-terminal Greek key motif are extended. Stacking interaction of Y74 and Y72 in combination with additional hydrophobic contacts leads to a curvature of the C-terminal β-sheet. The flexible loop between the two Greek key motifs is extended as well. Both, the extended β sheets as well as the flexible loop could in principle provide a novel interface for protein-protein interaction in the crystallin family.

Molecular interaction with human Immunoglobulin E
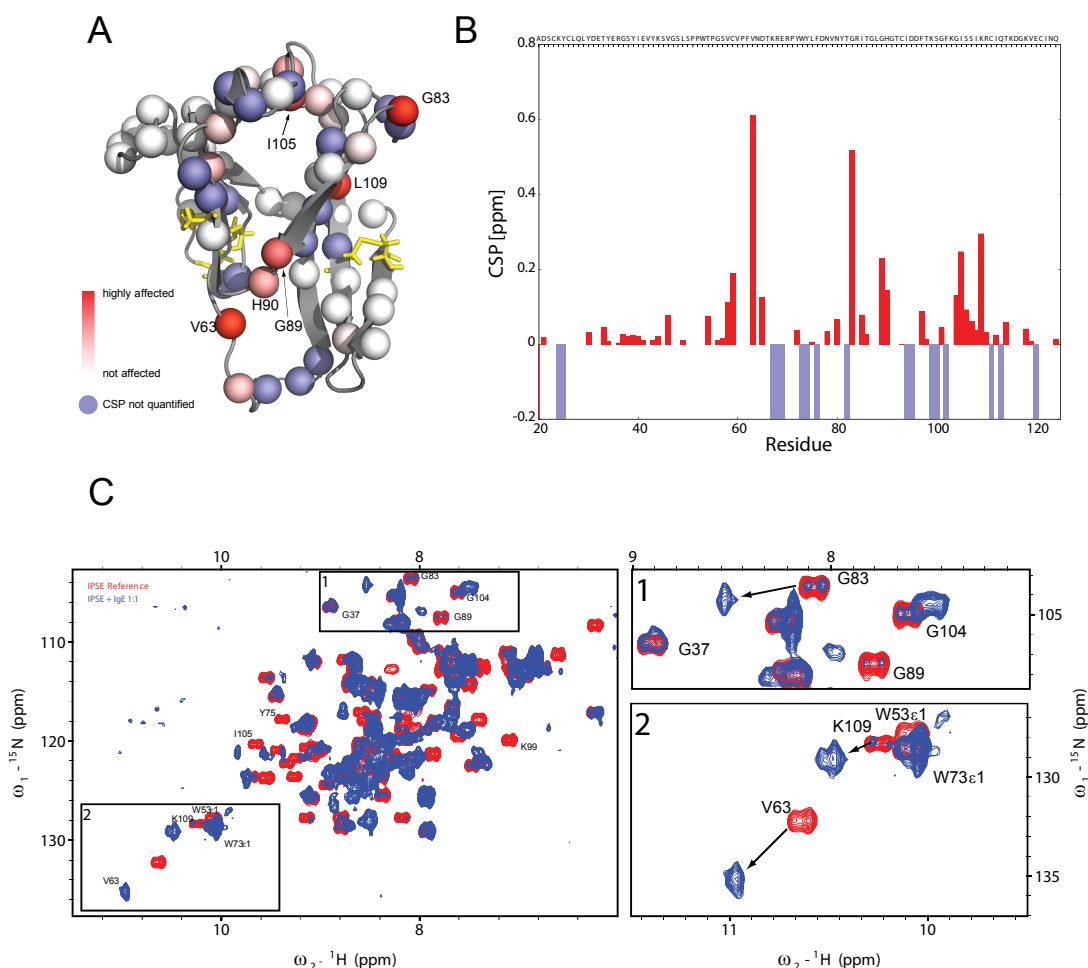
Figure 3.5: Chemical shift perturbation of IPSE upon titration with IgE. A) Mapping of the IgE-binding interface on the structure of IPSE/alpha-1. Nitrogen atoms of the backbone amide groups are represented as spheres and coloured according to their CSP in the NMR titration experiment. **B)** CSP perturbation per residue. Negative values indicate, that the respective peak is broadened beyond detection limit. **C)** ^{15}N,^{1}H CRINEPT spectra of IPSE/alpha-1 (red) and IPSE/alpha-1 in a 1:1 complex with IgE (blue).

It has been previously reported that IPSE binds IgE. To confirm and to characterize this binding on a molecular level, an NMR titration has been performed. Perdeuterated and uniformly ^{15}N/ ^{13}C-labeled IPSEΔNLS was titrated to unlabeled human IgE to analyze chemical shift perturbations. Upon binding to the 190 kDa IgE molecule, IPSEΔNLS signals experience extensive line broadening. This is due to the low molecular tumbling rate and the therefore strongly increased longitudinal relaxation rates of the high molecular weight complex. Hardly any signal can be observed in conventional HSQC or TROSY experiments. Therefore cross relaxation-enhanced polarization transfer (CRINEPT) [133] was used to detect IPSEΔNLS bound to IgE. Signals of amide protons close to the interaction site can still be broadened and might disappear because protons of the unlabeled IgE provide an additional source of dipolar relaxation. Amide protons in more flexible regions of the protein might tumble faster than the entire complex. Therefore, those signals are typically less broadened. However, they might still experience chemical shift perturbation due to a change in chemical environment upon binding of IPSE to IgE.

The NMR titration revealed a large interface covering the whole front of the molecule, including the flexible loop, which links the two Greek key motifs and β3-6. The backside, essentially β1 and β2 with the hairpin loop in between on the N-terminal Greek key as well as β7 and β8 on the C-terminal Greek key motif are left unaffected upon binding to IgE (Figure 3.5). The amide group of V63, which is in close proximity to the solvent exposed F62, experiences the strongest chemical shift perturbation. Another patch of strongly affected residues is closely located to the highly conserved Y81. It is likely that the interaction interface involves these aromatic side chains. Besides this, most of the interface is located in overall positively charged regions of the protein, especially in the flexible loop. This implies an essential role of electrostatic and charged interactions.

Model of the IPSE-IgE complex

I used a rigid body docking approach to gain structural insight into the mechanisms which underlie the activation of IgE F_C receptor upon binding of IPSE to IgE. It has been recently shown, that IPSE can bind to both the F_C and the F_{AB} moieties of IgE. Besides electrostatic and Van-der-Waals interactions, chemical shift perturbation was used to drive the docking process of IPSE to IgE. (Figure 3.6).

IPSE/alpha-1 adopts the crystallin fold, providing a novel interaction interface for IgE-binding. Furthermore, we could demonstrate that IPSE is capable of activating basophils and that mutations in the indentified interface abolish this function. This is the first report of a protein resembling the crystallin-fold, which major function is the modulation of the immune response. We propose a structure based model of how *Shistosoma mansoni* might stimulate the release of the signature T_H2 cytokines (IL-4 and IL-13) and the degranulation of basophils to initiate a T_H2 shift of the host immune system. This manipulation protects the parasite from annihilation, consequently leading to severe shistosomiasis in the human host.

In the past 30 years the IgE receptor on mast cells served as a prototype to characterize the receptor-mediated response to foreign antigens. Cross-linking of two or more IgE-receptor complexes on the cell surface stimulates degranulation of the cells subsequently priming a T_H2 response [134]. Note, that in spite of forming homodimers, IPSE/alpha-1 cannot crosslink IgE (data not shown). Thus, the general model of IgE-receptor activation does not seem to be applicable for the investigated system.

X-ray crystallography as well as small-angle X-ray scattering in solution revealed that the IgE F_C domain is an asymmetrical homodimer [88, 90]. NMR titration of IPSEΔNLS against IgE in combination with rigid body docking suggest, that monomeric IPSEΔNLS presumably binds to a grove which is provided by the asymmetrical bending of the F_C-region. However, a second IPSE monomer subunit would not fit into this grove, which is entirely covered by one IPSE monomer. The second subunit possibly binds to the F_{AB} moiety. One can speculate that binding of both subunits of the IPSE/alpha-1 dimer might involve a rearrangement of the structure of IgE which subsequently could lead to receptor activation.
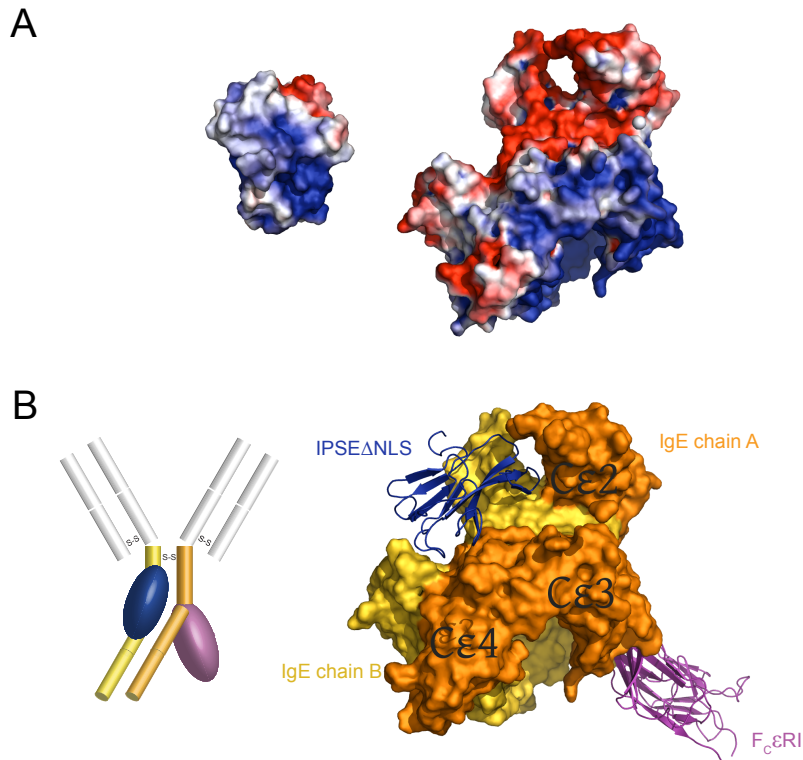
Figure 3.6: Structural model of the interaction of IPSE/alpha-1 with IgE. A) Surface representation of IPSE and IgE F_C colored according to electrostatic surface potential at ±3 kB T/e for positive (blue) or negative (red) charge potential using the program APBS [123] **B)** IPSE docked to the crystal structure of IgE (Cε2-Cε4) using ClusPro 2.0. High affinity receptor FcεRI was placed according to the complex structure of IgE (Cε3-Cε4) and FcεRI.

A role of monomeric IgE in the initiation of cell signaling events upon binding to FcεRI has been controversially discussed. A high concentration of IgE partially activates the FcεRI, thereby triggering cellular events such as enhancing mast cell survival [135], degranulation and cytokine production [136], cell adheasion [137] and migration [138]. Although different IgE molecules generate different levels of basophil activation, the structural differences that account for functional divergence are not known [139]. However, Schweitzer-Stenner and Pecht consider non-physiological experimental conditions to cause the effects of monomeric IgE [140]. Nevertheless IPSE/alpha-1 enhances the effect of monomeric IgE on the activation state of basohiles and leads to an immediate degranulation. To understand the structural principles underlying this mechanism, future studies should focus on a potential structural rearrangement of IgE upon IPSE/alpha-1 binding.

REFERENCES

1.   International Human Genome Sequencing Consortium., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.

2.   Berggard, T., S. Linse, and P. James, *Methods for the detection and analysis of protein-protein interactions.* Proteomics, 2007. **7**(16): p. 2833-42.

3.   Sattler, M., J. Schleucher, and C. Griesinger, *Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients.* Prog. Nucl. Magn. Reson. Spectrosc., 1999. **34**: p. 93-158.

4.   Shan, X., et al., *Assignment of N-15, C-13(alpha), C-13(beta), and HN resonances in an N-15, C-13, H-2 labeled 64 kDa trp repressor-operator complex using triple-resonance NMR spectroscopy and H-2-decoupling.* Journal of the American Chemical Society, 1996. **118**(28): p. 6570-6579.

5.   Sattler, M. and S.W. Fesik, *Use of deuterium labeling in NMR: overcoming a sizeable problem.* Structure, 1996. **4**(11): p. 1245-9.

6.   Staunton, D., et al., *Cell-free expression and selective isotope labelling in protein NMR.* Magn Reson Chem, 2006. **44 Spec No**: p. S2-9.

7.   Kumar, A., R.R. Ernst, and K. Wuthrich, *A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules.* Biochem Biophys Res Commun, 1980. **95**(1): p. 1-6.

8.   Karplus, M., *VICINAL PROTON COUPLING IN NUCLEAR MAGNETIC RESONANCE.* Journal of the American Chemical Society, 1963. **85**(18): p. 2870-&.

9.   Shen, Y., et al., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR, 2009. **44**(4): p. 213-23.

10.   Bax, A., *Weak alignment offers new NMR opportunities to study protein structure and dynamics.* Protein Sci, 2003. **12**(1): p. 1-16.

11.   Otting, G., *Protein NMR using paramagnetic ions.* Annu Rev Biophys, 2010. **39**: p. 387-405.

12.   Williamson, M.P., T.F. Havel, and K. Wuthrich, *Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry.* J Mol Biol, 1985. **182**(2): p. 295-315.

13.   Güntert, P., *Automated structure determination from NMR spectra.* Eur Biophys J, 2009. **38**(2): p. 129-43.

14.   Rieping, W., et al., *ARIA2: automated NOE assignment and data integration in NMR structure calculation.* Bioinformatics, 2007. **23**(3): p. 381-2.

15.   Laskowski, R.A., et al., *AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR.* J Biomol NMR, 1996. **8**(4): p. 477-86.

16.   Hooft, R.W., et al., *Errors in protein structures.* Nature, 1996. **381**(6580): p. 272.

17.   Vuister, *e.a., CING software.* to be published.

18.   Otting, G. and K. Wüthrich, *Heteronuclear filters in two-dimensional [1H,1H]-NMR spectroscopy: combined use with isotope labelling for studies of macromolecular conformation and intermolecular interactions.* Q Rev Biophys, 1990. **23**(1): p. 39-96.

19.   Prestegard, J.H. and A.I. Kishore, *Partial alignment of biomolecules: an aid to NMR characterization.* Curr Opin Chem Biol, 2001. **5**(5): p. 584-90.

20.   Madl, T., et al., *Structural analysis of protein interfaces from 13C direct-detected paramagnetic relaxation enhancements.* J Am Chem Soc, 2010. **132**(21): p. 7285-7.

21. Madl, T., W. Bermel, and K. Zangger, *Use of relaxation enhancements in a paramagnetic environment for the structure determination of proteins using NMR spectroscopy.* Angew Chem Int Ed Engl, 2009. **48**(44): p. 8259-62.

22. Berget, S.M., C. Moore, and P.A. Sharp, *Spliced segments at the 5' terminus of adenovirus 2 late mRNA.* Proc Natl Acad Sci U S A, 1977. **74**(8): p. 3171-5.

23. Chow, L.T., et al., *An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA.* Cell, 1977. **12**(1): p. 1-8.

24. Reed, R., *Mechanisms of fidelity in pre-mRNA splicing.* Curr Opin Cell Biol, 2000. **12**(3): p. 340-5.

25. Stevens, S.W., et al., *Composition and functional characterization of the yeast spliceosomal penta-snRNP.* Mol Cell, 2002. **9**(1): p. 31-44.

26. Zhou, Z., et al., *Comprehensive proteomic analysis of the human spliceosome.* Nature, 2002. **419**(6903): p. 182-5.

27. Tseng, C.K. and S.C. Cheng, *Both catalytic steps of nuclear pre-mRNA splicing are reversible.* Science, 2008. **320**(5884): p. 1782-4.

28. Hawkins, J.D., *A survey on intron and exon lengths.* Nucleic Acids Res, 1988. **16**(21): p. 9893-908.

29. Moore, M.J., *Intron recognition comes of AGe.* Nat Struct Biol, 2000. **7**(1): p. 14-6.

30. Das, R., Z. Zhou, and R. Reed, *Functional association of U2 snRNP with the ATP-independent spliceosomal complex E.* Mol Cell, 2000. **5**(5): p. 779-87.

31. Hong, W., et al., *Association of U2 snRNP with the spliceosomal complex E.* Nucleic Acids Res, 1997. **25**(2): p. 354-61.

32. Query, C.C., M.J. Moore, and P.A. Sharp, *Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model.* Genes Dev, 1994. **8**(5): p. 587-97.

33. Makarov, E.M., et al., *Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome.* Science, 2002. **298**(5601): p. 2205-8.

34. Valadkhan, S. and J.L. Manley, *Splicing-related catalysis by protein-free snRNAs.* Nature, 2001. **413**(6857): p. 701-7.

35. Collins, C.A. and C. Guthrie, *The question remains: is the spliceosome a ribozyme?* Nat Struct Biol, 2000. **7**(10): p. 850-4.

36. Hall, K.B., *RNA-protein interactions.* Curr Opin Struct Biol, 2002. **12**(3): p. 283-8.

37. Sachs, A.B., M.W. Bond, and R.D. Kornberg, *A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression.* Cell, 1986. **45**(6): p. 827-35.

38. Adam, S.A., et al., *mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence.* Mol Cell Biol, 1986. **6**(8): p. 2932-43.

39. Lahiri, D.K. and J.O. Thomas, *A cDNA clone of the hnRNP C proteins and its homology with the single-stranded DNA binding protein UP2.* Nucleic Acids Res, 1986. **14**(10): p. 4077-94.

40. Dreyfuss, G., M.S. Swanson, and S. Pinol-Roma, *Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation.* Trends Biochem Sci, 1988. **13**(3): p. 86-91.

41. Query, C.C., R.C. Bentley, and J.D. Keene, *A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein.* Cell, 1989. **57**(1): p. 89-101.

42. Scherly, D., et al., *Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA.* EMBO J, 1989. **8**(13): p. 4163-70.

43. Birney, E., S. Kumar, and A.R. Krainer, *Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors.* Nucleic Acids Res, 1993. **21**(25): p. 5803-16.

44. Maris, C., C. Dominguez, and F.H. Allain, *The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression.* FEBS J, 2005. **272**(9): p. 2118-31.

45. Kielkopf, C.L., et al., *A novel peptide recognition mode revealed by the X-ray structure of a core U2AF35/U2AF65 heterodimer.* Cell, 2001. **106**(5): p. 595-605.

46. Selenko, P., et al., *Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP.* Mol Cell, 2003. **11**(4): p. 965-76.

47. Tronchere, H., J. Wang, and X.D. Fu, *A protein related to splicing factor U2AF35 that interacts with U2AF65 and SR proteins in splicing of pre-mRNA.* Nature, 1997. **388**(6640): p. 397-400.

48. Kielkopf, C.L., S. Lucke, and M.R. Green, *U2AF homology motifs: protein recognition in the RRM world.* Genes Dev, 2004. **18**(13): p. 1513-26.

49. Chen, M. and J.L. Manley, *Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches.* Nat Rev Mol Cell Biol, 2009. **10**(11): p. 741-54.

50. Berget, S.M., *Exon recognition in vertebrate splicing.* J Biol Chem, 1995. **270**(6): p. 2411-4.

51. Graveley, B.R., *Sorting out the complexity of SR protein functions.* RNA, 2000. **6**(9): p. 1197-211.

52. Tacke, R. and J.L. Manley, *Determinants of SR protein specificity.* Curr Opin Cell Biol, 1999. **11**(3): p. 358-62.

53. Spellman, R. and C.W. Smith, *Novel modes of splicing repression by PTB.* Trends Biochem Sci, 2006. **31**(2): p. 73-6.

54. Graveley, B.R., *Sex, AGility, and the regulation of alternative splicing.* Cell, 2002. **109**(4): p. 409-12.

55. Grabowski, P.J., *Splicing regulation in neurons: tinkering with cell-specific control.* Cell, 1998. **92**(6): p. 709-12.

56. Black, D.L. and P.J. Grabowski, *Alternative pre-mRNA splicing and neuronal function.* Prog Mol Subcell Biol, 2003. **31**: p. 187-216.

57. Jiang, Z.H. and J.Y. Wu, *Alternative splicing and programmed cell death.* Proc Soc Exp Biol Med, 1999. **220**(2): p. 64-72.

58. Najib, S., et al., *Role of Sam68 as an adaptor protein in signal transduction.* Cell Mol Life Sci, 2005. **62**(1): p. 36-43.

59. Taylor, S.J. and D. Shalloway, *An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis.* Nature, 1994. **368**(6474): p. 867-71.

60. Lukong, K.E. and S. Richard, *Sam68, the KH domain-containing superSTAR.* Biochim Biophys Acta, 2003. **1653**(2): p. 73-86.

61. Suhasini, M. and T.R. Reddy, *Cellular proteins and HIV-1 Rev function.* Curr HIV Res, 2009. **7**(1): p. 91-100.

62. Siomi, H., et al., *The pre-mRNA binding K protein contains a novel evolutionarily conserved motif.* Nucleic Acids Res, 1993. **21**(5): p. 1193-8.

63. Burd, C.G. and G. Dreyfuss, *Conserved structures and diversity of functions of RNA-binding proteins.* Science, 1994. **265**(5172): p. 615-21.

64. Chen, T., et al., *Self-association of the single-KH-domain family members Sam68, GRP33, GLD-1, and Qk1: role of the KH domain.* Mol Cell Biol, 1997. **17**(10): p. 5707-18.

65. Chen, T. and S. Richard, *Structure-function analysis of Qk1: a lethal point mutation in mouse quaking prevents homodimerization.* Mol Cell Biol, 1998. **18**(8): p. 4863-71.

66. Lin, Q., S.J. Taylor, and D. Shalloway, *Specificity and determinants of Sam68 RNA binding. Implications for the biological function of K homology domains.* J Biol Chem, 1997. **272**(43): p. 27274-80.

67. Berglund, J.A., et al., *The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC.* Cell, 1997. **89**(5): p. 781-7.

68. Liu, Z., et al., *Structural basis for recognition of the intron branch site RNA by splicing factor 1.* Science, 2001. **294**(5544): p. 1098-102.

69. Maguire, M.L., et al., *Solution structure and backbone dynamics of the KH-QUA2 region of the Xenopus STAR/GSG quaking protein.* J Mol Biol, 2005. **348**(2): p. 265-79.

70.   Itoh, M., et al., *Identification of cellular mRNA targets for RNA-binding protein Sam68.* Nucleic Acids Res, 2002. **30**(24): p. 5452-64.

71.   Zorn, A.M. and P.A. Krieg, *The KH domain protein encoded by quaking functions as a dimer and is essential for notochord development in Xenopus embryos.* Genes Dev, 1997. **11**(17): p. 2176-90.

72.   Di Fruscio, M., T. Chen, and S. Richard, *Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis.* Proc Natl Acad Sci U S A, 1999. **96**(6): p. 2710-5.

73.   Wu, J., et al., *The quaking I-5 protein (QKI-5) has a novel nuclear localization signal and shuttles between the nucleus and the cytoplasm.* J Biol Chem, 1999. **274**(41): p. 29202-10.

74.   Galarneau, A. and S. Richard, *The STAR RNA binding proteins GLD-1, QKI, SAM68 and SLM-2 bind bipartite RNA motifs.* BMC Mol Biol, 2009. **10**: p. 47.

75.   Matter, N., P. Herrlich, and H. König, *Signal-dependent regulation of splicing via phosphorylation of Sam68.* Nature, 2002. **420**(6916): p. 691-5.

76.   Cote, J., et al., *Sam68 RNA binding protein is an in vivo substrate for protein arginine N-methyltransferase 1.* Mol Biol Cell, 2003. **14**(1): p. 274-87.

77.   Babic, I., A. Jakymiw, and D.J. Fujita, *The RNA binding protein Sam68 is acetylated in tumor cell lines, and its acetylation correlates with enhanced RNA binding activity.* Oncogene, 2004. **23**(21): p. 3781-9.

78.   Tisserant, A. and H. Konig, *Signal-regulated Pre-mRNA occupancy by the general splicing factor U2AF.* PLoS One, 2008. **3**(1): p. e1418.

79.   Rho, J., et al., *Arginine methylation of Sam68 and SLM proteins negatively regulates their poly(U) RNA binding activity.* Arch Biochem Biophys, 2007. **466**(1): p. 49-57.

80.   Bedford, M.T., et al., *Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains.* J Biol Chem, 2000. **275**(21): p. 16030-6.

81.   Venables, J.P., et al., *T-STAR/ETOILE: a novel relative of SAM68 that interacts with an RNA-binding protein implicated in spermatogenesis.* Hum Mol Genet, 1999. **8**(6): p. 959-69.

82.   Grossman, J.S., et al., *The use of antibodies to the polypyrimidine tract binding protein (PTB) to analyze the protein components that assemble on alternatively spliced pre-mRNAs that use distant branch points.* RNA, 1998. **4**(6): p. 613-25.

83.   Paronetto, M.P., et al., *The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x.* J Cell Biol, 2007. **176**(7): p. 929-39.

84.   Tisserant, A. and H. König, *Signal-regulated Pre-mRNA occupancy by the general splicing factor U2AF.* PLoS ONE, 2008. **3**(1): p. e1418.

85.   Harris, L.J., et al., *Refined structure of an intact IgG2a monoclonal antibody.* Biochemistry, 1997. **36**(7): p. 1581-97.

86.   Thornton, C.A., et al., *Fetal exposure to intact immunoglobulin E occurs via the gastrointestinal tract.* Clin Exp Allergy, 2003. **33**(3): p. 306-11.

87.   Stanworth, D.R., *The discovery of IgE.* Allergy, 1993. **48**(2): p. 67-71.

88.   Beavil, A.J., et al., *Bent domain structure of recombinant human IgE-Fc in solution by X-ray and neutron scattering in conjunction with an automated curve fitting procedure.* Biochemistry, 1995. **34**(44): p. 14449-61.

89.   Zheng, Y., et al., *Conformations of IgE bound to its receptor Fc epsilon RI and in solution.* Biochemistry, 1991. **30**(38): p. 9125-32.

90.   Wan, T., et al., *The crystal structure of IgE Fc reveals an asymmetrically bent conformation.* Nat Immunol, 2002. **3**(7): p. 681-6.

91.   Garman, S.C., et al., *Structure of the Fc fragment of human IgE bound to its high-affinity receptor Fc epsilonRI alpha.* Nature, 2000. **406**(6793): p. 259-66.

92.   Hotez, P.J., et al., *Developing vaccines to combat hookworm infection and intestinal schistosomiasis.* Nat Rev Microbiol, 2010. **8**(11): p. 814-26.

93.   Pearce, E.J., et al., *Th2 response polarization during infection with the helminth parasite Schistosoma mansoni.* Immunol Rev, 2004. **201**: p. 117-26.

94.   Schramm, G., et al., *Molecular characterization of an interleukin-4-inducing factor from Schistosoma mansoni eggs.* J Biol Chem, 2003. **278**(20): p. 18384-92.

95.   Schramm, G., et al., *IPSE/alpha-1: a major immunogenic component secreted from Schistosoma mansoni eggs.* Mol Biochem Parasitol, 2006. **147**(1): p. 9-19.

96.   Mayerhofer, H., et al., *Cloning, expression, purification, crystallization and preliminary X-ray crystallographic analysis of interleukin-4-inducing principle from Schistosoma mansoni eggs (IPSE/alpha-1).* Acta Crystallogr Sect F Struct Biol Cryst Commun, 2009. **65**(Pt 6): p. 594-6.

97.   Wuhrer, M., et al., *IPSE/alpha-1, a major secretory glycoprotein antigen from schistosome eggs, expresses the Lewis X motif on core-difucosylated N-glycans.* FEBS J, 2006. **273**(10): p. 2276-92.

98.   Schramm, G., et al., *Cutting edge: IPSE/alpha-1, a glycoprotein from Schistosoma mansoni eggs, induces IgE-dependent, antigen-independent IL-4 production by murine basophils in vivo.* J Immunol, 2007. **178**(10): p. 6023-7.

99.   Nadler, M.J., et al., *Signal transduction by the high-affinity immunoglobulin E receptor Fc epsilon RI: coupling form to function.* Adv Immunol, 2000. **76**: p. 325-55.

100.   Vernet, C. and K. Artzt, *STAR, a gene family involved in signal transduction and activation of RNA.* Trends Genet, 1997. **13**(12): p. 479-84.

101.   Pillay, I., H. Nakano, and S.V. Sharma, *Radicicol inhibits tyrosine phosphorylation of the mitotic Src substrate Sam68 and retards subsequent exit from mitosis of Src-transformed cells.* Cell Growth Differ, 1996. **7**(11): p. 1487-99.

102.   Wang, L.L., S. Richard, and A.S. Shaw, *P62 association with RNA is regulated by tyrosine phosphorylation.* J Biol Chem, 1995. **270**(5): p. 2010-3.

103.   Guitard, E., et al., *Sam68 is a Ras-GAP-associated protein in mitosis.* Biochem Biophys Res Commun, 1998. **245**(2): p. 562-6.

104.   Busa, R., et al., *The RNA-binding protein Sam68 contributes to proliferation and survival of human prostate cancer cells.* Oncogene, 2007. **26**(30): p. 4372-82.

105.   Arch, R., et al., *Participation in normal immune responses of a metastasis-inducing splice variant of CD44.* Science, 1992. **257**(5070): p. 682-5.

106.   Cooper, D.L. and G.J. Dougherty, *To metastasize or not? Selection of CD44 splice sites.* Nat Med, 1995. **1**(7): p. 635-7.

107.   Sherman, L., et al., *The CD44 proteins in embryonic development and in cancer.* Curr Top Microbiol Immunol, 1996. **213 (Pt 1)**: p. 249-69.

108.   Sherman, L., et al., *A splice variant of CD44 expressed in the apical ectodermal ridge presents fibroblast growth factors to limb mesenchyme and is required for limb outgrowth.* Genes Dev, 1998. **12**(7): p. 1058-71.

109.   Weg-Remers, S., et al., *Regulation of alternative pre-mRNA splicing by the ERK MAP-kinase pathway.* EMBO J, 2001. **20**(15): p. 4194-203.

110.   Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.

111.   Goddard, T. and D. Kneller, *SPARKY 3 (www.cgl.ucsf.edu/home/sparky).* 2004.

112.   Jung, Y.S. and M. Zweckstetter, *Mars -- robust automatic backbone assignment of proteins.* J Biomol NMR, 2004. **30**(1): p. 11-23.

113.   Farrow, N.A., et al., *Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation.* Biochemistry, 1994. **33**(19): p. 5984-6003.

114.   Yang, D., et al., *TROSY-based HNCO pulse sequences for the measurement of ^1HN-^{15}N, ^{15}N-^{13}CO, ^1HN-^{13}CO, ^{13}CO-$^{13}C\alpha$ and ^1HN-$^{13}C\alpha$ dipolar couplings in ^{15}N,^{13}C, ^2H-labeled proteins.* J. Biomol. NMR, 1999. **14**: p. 333-343.

115. Rückert, M. and G. Otting, *Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments.* J. Am. Chem. Soc., 2000. **122**: p. 7793-7797.

116. Nilges, M., *A calculation strategy for the structure determination of symmetric dimers by 1H NMR.* Proteins, 1993. **17**(3): p. 297-309.

117. Linge, J.P., et al., *Refinement of protein structures in explicit solvent.* Proteins, 2003. **50**(3): p. 496-506.

118. DeLano, W.L., *The PyMOL Molecular Graphics System.* DeLano Scientific, Paolo Alto, CA, USA., 2002: p. http://www.pymol.org.

119. Hadian, K., et al., *Identification of an hnRNP-recognition region in the HIV Rev protein.* J Biol Chem, 2009.

120. Lipsitz, R.S. and N. Tjandra, *Residual dipolar couplings in NMR structure analysis.* Annu Rev Biophys Biomol Struct, 2004. **33**: p. 387-413.

121. Thompson, J.D., T.J. Gibson, and D.G. Higgins, *Multiple sequence alignment using ClustalW and ClustalX.* Curr Protoc Bioinformatics, 2002. **Chapter 2**: p. Unit 2 3.

122. Gouet, P., et al., *ESPript: analysis of multiple sequence alignments in PostScript.* Bioinformatics, 1999. **15**(4): p. 305-8.

123. Baker, N.A., et al., *Electrostatics of nanosystems: application to microtubules and the ribosome.* Proc Natl Acad Sci U S A, 2001. **98**(18): p. 10037-41.

124. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-68.

125. Li, J., et al., *Expression of exogenous Sam68, the 68-kilodalton SRC-associated protein in mitosis, is able to alleviate impaired Rev function in astrocytes.* J Virol, 2002. **76**(9): p. 4526-35.

126. Beuck, C., et al., *Structure of the GLD-1 Homodimerization Domain: Insights into STAR Protein-Mediated Translational Regulation.* Structure, 2010. **18**(3): p. 377-389.

127. Richard, S., et al., *Association of p62, a multifunctional SH2- and SH3-domain-binding protein, with src family tyrosine kinases, Grb2, and phospholipase C gamma-1.* Mol Cell Biol, 1995. **15**(1): p. 186-97.

128. Meyer, N.H., et al., *Structural basis for homodimerization of the Src-associated during mitosis, 68-kDa protein (Sam68) Qua1 domain.* J Biol Chem, 2010. **285**(37): p. 28893-901.

129. Wilson, M.S., et al., *Immunopathology of schistosomiasis.* Immunol Cell Biol, 2007. **85**(2): p. 148-54.

130. Jaenicke, R. and C. Slingsby, *Lens crystallins and their microbial homologs: structure, stability, and function.* Crit Rev Biochem Mol Biol, 2001. **36**(5): p. 435-99.

131. Rosinke, B., et al., *Ca2+-loaded spherulin 3a from Physarum polycephalum adopts the prototype gamma-crystallin fold in aqueous solution.* J Mol Biol, 1997. **271**(4): p. 645-55.

132. Clout, N.J., et al., *Crystal structure of the calcium-loaded spherulin 3a dimer sheds light on the evolution of the eye lens betagamma-crystallin domain fold.* Structure, 2001. **9**(2): p. 115-24.

133. Riek, R., et al., *Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules.* Proc Natl Acad Sci U S A, 1999. **96**(9): p. 4918-23.

134. Holowka, D., et al., *Insights into immunoglobulin E receptor signaling from structurally defined ligands.* Immunol Rev, 2007. **217**: p. 269-79.

135. Kawakami, T. and S.J. Galli, *Regulation of mast-cell and basophil function and survival by IgE.* Nat Rev Immunol, 2002. **2**(10): p. 773-86.

136. Kitaura, J., et al., *Early divergence of Fc epsilon receptor I signals for receptor up-regulation and internalization from degranulation, cytokine production, and survival.* J Immunol, 2004. **173**(7): p. 4317-23.

137.	Lam, V., et al., *IgE alone stimulates mast cell adhesion to fibronectin via pathways similar to those used by IgE + antigen but distinct from those used by Steel factor.* Blood, 2003. **102**(4): p. 1405-13.

138.	Kitaura, J., et al., *IgE- and IgE+Ag-mediated mast cell migration in an autocrine/paracrine fashion.* Blood, 2005. **105**(8): p. 3222-9.

139.	Kitaura, J., et al., *Evidence that IgE molecules mediate a spectrum of effects on mast cell survival and activation via aggregation of the FcepsilonRI.* Proc Natl Acad Sci U S A, 2003. **100**(22): p. 12911-6.

140.	Schweitzer-Stenner, R. and I. Pecht, *Death of a dogma or enforcing the artificial: monomeric IgE binding may initiate mast cell response by inducing its receptor aggregation.* J Immunol, 2005. **174**(8): p. 4461-4.

TABLE OF FIGURES

ABBREVIATIONS

1D, 2D, 3D   one-, two-, three-dimensional
APC   antigen presenting cell
ATP   adenosine triphosphate
BMRB   Biological Magnetic Resonance Bank
BPS   branch point sequence
CD   circular dichroism
CRINEPT   cross relaxation-enhanced polarization transfer
CSA   chemical shift anisotropy
CSP   chemical shift perturbation
Dalton   Da
DD   dipolar interactions
DNA   deoxyribonucleic acid
E. Coli   Escherichia Coli
EPR   Electron Paramagnetic Resonance
ERK   extracellular-signal-regulated kinase
FID   free induction decay
FRET   Förster resonance energy transfer
FT   Fourier Transformation
GAP   GTPase-activating protein
Gld-1   Germline development defective-1
GSG   GRP33, Sam68, GLD-1
GTP   guanosine triposphate
hnRNP   heterogeneous nuclear ribonucleoprotein
HSQC   heteronuclear single quantum correlation
Ig   Immunoglobulin
IL   interleukin
IPSE   interleukin-4-inducing principle of S. mansoni eggs
ITAM   immunoreceptor tyrosine-based activation motif
KH   K homology
MAPK   mitogen activated protein kinase
MD   molecular dynamics
mRNA   messanger RNA
MWCO   molecular weight cut-off
NLS   nuclear location signal
NMR   nuclear magnetic resonance

NOE   nuclear Overhauser enhancement
NOESY   NOE spectroscopy
PAGE   polyacrylamide gel electrophorese
PCR   polymerase chain reaction
PCS   pseudo contact shift
PDB   Protein Data Bank
ppm   parts per million
PRE   relaxation enhancement
pY   poly-pyrimidine
Qk   Quaking
RDC   residual dipolar coupling
RMSD   root mean square deviation
RNA   ribonucleic acid
RNP   ribonucleoprotein
ROS   reactive oxygen species
RRM   RNA-recognition motif
RS   Arginine-Serine repeat
Sam68   Src associated during mitosis, 68 kDa protein
SANS   small angle neutron scattering
SA   simulated annealing
SAXS   small angle X-ray scattering
SDS   sodium dodecyl sulfate
SELEX   systematic evolution of ligands by exponential enrichment
SF1   human splicing factor 1
SH   Src homology
S. mansoni   *Schistosoma mansoni*
snRNP   small nuclear ribonucleoprotein particles
STAR   signal transducer and acativator of RNA
TCR   T-cell receptor
TEV   tobacco etch virus
TH2   T-helper type 2
TOCSY   total correlation spectroscopy
TROSY   transverse relaxation optimized
U2AF   U2 auxiliary factor
UHM   U2AF homology motif
ULM   UHM ligand motif

Acknowledgments

CURRICULUM VITAE

Personal Information

Name	Helge Meyer
Date of Birth	18. December 1981
Nationality	German
Gender	Male

Education

Date	Since June 2007
Intended degree	Dr. rer nat., Technische Universität München, AK Sattler

Date	October 2004 - July 2006
Degree	Master of Science (Molecular Life Science, Universität zu Lübeck)

Date	October 01- July 2004
Degree	Bachelor of Science

Work Experience

Date	September 2006 - Mai 2007
Occupation	research fellow
Main activities	protein purification, NMR
Employer	European Molecular Biology Lab/ Biochemie-Zentrum Heidleberg

Date	September - November 2005
Occupation	Internship
Main activities	molecular biological and biochemical methods
Employer	Institute for biochemistry and molecular biology, Charité Berlin

Date	October 2003
Occupation	Internship
Main activities	fluorescence microscopy
Employer	Institute for immunology, Universität zu Lübeck

Date	July 2003
Occupation	Internship
Main activities	Analysis of blood cultures, Purification of cells (FACS, MACS etc.)
Employer	Institute for microbiology, University of Lübeck

Date	2002-2005
Occupation	Assistant scientist
Main activities	Internship mentor (biophysical chemistry, inorganic and organic chemistry)
Employer	Institute for Chemistry, Universität zu Lübeck