universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Characterisation of a double-stranded RNA-binding domain that serves as a nuclear localization signal

angestrebter akademischer Grad

Magister/Magistra der Naturwissenschaften (Mag. rer.nat.)

Verfasserin / Verfasser:	Sabina Dani
Studienrichtung /Studienzweig (lt. Studienblatt):	A 441 Diplomstudium Genetik – Mikrobiologie (Stzw)
Betreuerin / Betreuer:	Ao. Univ-Prof. Dr. Michael Jantsch

Wien, im Dezember 2010

Contents

List of figures

List of tables

Zusammenfassung

Es gibt viele Schritte und Mechanismen die die RNA während und nach der Transkription prozessieren und modifizieren. Ein Beispiel posttranskriptioneller Modifikation ist RNA-editing, welches Deletion, Insertion oder chemische Umwandlung von Nukleotiden umfasst. Adenosin zu Inosin RNA editing wird durch Enzyme katalysiert, die ADARs (adenosine deaminase that act on RNA) genannt werden welche in den meisten Metazoan gefunden werden konnten. Als Folge der Veränderung der RNA Sequenz, können die Eigenschaften der RNA Moleküle und der translatierten Proteine modifiziert werden. ADARs agieren an partiell oder perfekt doppelsträngigen RNAs und sind an einem breiten Spektrum von zellulärer Funktionen beteiligt.

Menschliches ADAR1 hat die Fähigkeit zwischen Nukleus und Cytoplasma zu pendeln und besitzt zwei Promotoren, welche zu zwei Versionen des Proteins führen. Das längere, Interferon induzierte, enthält einen Nuklear Export Signal (NES) und ist im Nukleus sowie im Cytoplasma lokalisiert. Das kürzere hat einen gekürzten Aminoterminus, ist konstitutiv exprimiert und vorwiegend nukleär. Die dritte doppelstrang RNA bindende Domäne von ADAR1 wird von einem untypischen Nuklear Lokalisierung Signal (NLS) überlagert.

Ziel dieser Arbeit war es, die dritte doppelstrang RNA bindende Domäne (dsRBD3) von menschlichen ADAR1 zu charakterisieren, indem ihre dreidimensionale Struktur bestimmt werden sollte. Daher wurden unterschiedliche Konstrukte diese Proteins gemacht, welche sich in der Länge ihrer N- oder C- Termini unterschieden, um hochkonzentriertes, Protein für Kernspinresonanzspektroskopie (NMR), zu reinigen. Zusätzlich wurden verschiedene Puffer, Temperaturen, pH- Werte, Inkubations- und Induktionszeiten getestet um die besten Konditionen für dieses Protein zu finden.

Abstract

There are many steps and mechanisms that process and modify RNA during and after transcription. One example of a post-transcriptional processing event is RNA-editing which includes deletion, insertion or chemical conversion of nucleotides. Adenosine to inosine RNA editing is catalyzed by enzymes named adenosine deaminases that act on RNA (ADARs), which are found in most metazoa. As a consequence of the alteration of the RNA sequence, the properties of the RNA molecules and of the translated proteins can be modified. ADARs act on partially or perfectly double stranded RNA and are involved in a wide range of cellular functions.

The human adenosine deaminase that acts on RNA 1 (ADAR1) has the ability to shuttle between the nucleus and cytoplasm and possesses two promoters which lead to the expression of two versions of the protein. The longer one, interferon inducible, contains a nuclear export signal (NES) and in contrast to the shorter one is located in the nucleus as well as in the cytoplasm. The shorter one, has a truncated amino-terminus, is constitutively expressed and predominantly nuclear. The third double stranded RNA binding domain of ADAR1 is overlapped by an atypical nuclear localization signal (NLS).

The objective of this work was to characterize the third double stranded RNA binding domain (dsRBD3) of the human ADAR1, by determining its three dimensional structure. For this purpose different constructs of this protein which differ from each other in the length of their N- or C-termini, were made to achieve a high concentrated, non aggregated protein for nuclear magnetic resonance spectroscopy (NMR). Additionally different buffers, temperatures, pH values, incubation and induction times, were tested to find the best conditions under which this protein is suitable for NMR analysis.

Introduction

RNA editing

RNA editing is a post-transcriptional processing event, which is widespread in eukaryotes. It can increase the genetic variation because functionally distinct proteins can be produced from a single gene. The molecular mechanisms by which editing reactions operate are different and range from nucleotide insertion or deletion to base modification. In addition to codon changes, this type of modification has been shown to generate both introduction and removal of stop codons as well as change of splice sites (Maas and Rich 2000).

Types of RNA editing

The phenomenon of RNA editing was first discovered more than 20 years ago in kinetoplastid protozoa (Benne, Vandenburg et al. 1986). Here, in the mitochondrial mRNA, many uridine nucleotides are inserted or deleted. These events can produce open reading frames, create translation initiation cordons and correct frame shifts (Feagin, Abraham et al. 1988; Maslov and Simpson 1992; Alfonzo, Thiemann et al. 1997). In protozoa, guide RNAs define the uridine insertion/deletion sites, where editing must occur. The sequence of these guide RNA represent complementary versions of the mature mRNA within the edited region. A hybrid formed between the gRNA and preedited RNA is substrate for multiple cycles of cleavage, addition or deletion of uridylates and religation (Blum, Bakalara et al. 1990).

Another type of RNA editing is base conversion/substitution editing which includes cytidine to uracil and adenosine to inosine deamination.

Cytidine-to-uridine editing is performed in RNA by many cytidine deaminases. The reaction that occurs thereby is a hydrolytic deamination at C4 of cytidine. This particular editing event was found in vertebrate mRNA which codes for apolipoprotein B (apoB), a structural component of plasma lipoproteins and a significant risk factor for the development of atherosclerosis (Chan 1992). The editing of apo-B mRNA involves the site specific deamination of C666 to U, which converts codon 2153 from a glutamine codon CAA, to a premature stop codon UAA (Chen, Habib et al. 1987). In consequence of that , both a truncated apoB48 and the full length variant (apoB100) are produced by a single gene, which have distinct functions in lipoprotein metabolism and atherosclerosis susceptibility (Innerarity, Boren et al. 1996). In most mammals apoB mRNA editing is restricted to the intestine but it may also be found in the livers of some species, including rodents and hepatic expression of apoB mRNA editing influences lipoprotein concentration in plasma (Greeve, Altkemper et al. 1993).

ApoB-mRNA is edited by APOBEC1, a member of CDARs family (cytidine deaminases that act on RNA). APOBEC1 can not edit selectively by itself , only by forming a complex with a factor named APOBEC1 complementation factor ACF, which has been shown to bind to apoB mRNA in vivo and in vitro (Mehta, Kinter et al. 2000).

Cytidine residues are deaminated by members of APOBEC family not only in RNA but also in single stranded DNA. Two members of this family, APOBEC3F and APOBEC3G, affect the effinciency of retroviruses by deaminating cytosines in nascent retroviral cDNA of the human immunodeficiency virus (HIV) which in turn, encodes Vif, a small protein that mediates APOBEC inactivation (Harris and Liddament 2004). Another example of members of APOBEC family is *Activation Induced Deaminase* (AID), which plays a significant role in variegation and expression of immunoglobulins (Neuberger, Harris et al. 2003).

The reverse editing event to C-to-U has also been detected. Uridine to cytidine has been observed mostly in plants but also in RNA in amphibians and mammals (Malek, Lattig et al. 1996).

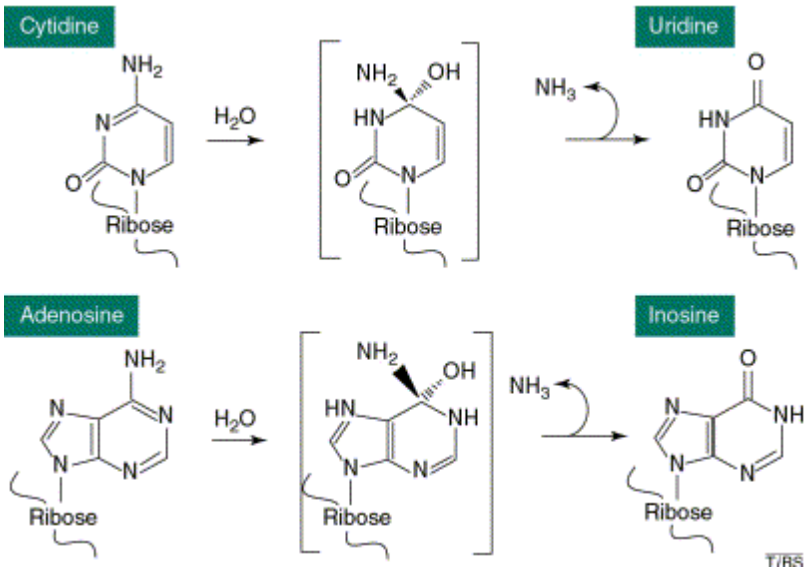


Figure 1: Cytidine and adenosine deamination in RNA
(Gerber and Keller, 2001). During editing cytosine and adenine are converted to uridine and inosine respectively by a hydrolytic deamination of the bases.

Additionally to the above-mentioned editing types, adenosine to inosine is the most widespread editing event in mammals. This phenomenon contributes in transcriptome diversification, it can have an impact on RNA -structure, -localisation, -stability, splicing, and translation. Changes in gene expression are a consequence of the fact that during translation inosine is recognized as a guanosine, thereby causing changes in amino acids and often in protein function.

Adenosine deaminases that act on RNA (ADARs)

Adenosine deaminases that act on RNA (ADARs) can accomplish the reaction of hydrolytic deamination of adenosine to inosine. This enzymes were first indentified in *Xenopus* eggs and it was thought that there is a mysterious dsRNA-unwinding activity (Bass and Weintraub 1987), but further work showed that there occurs an adenosine to inosine deamination.

It has been shown that ADARs are present in all metazoans and were successfully isolated from mammals (Kim, Wang et al. 1994), frogs (Hough and Bass 1997), fish (Slavov, Crnogorac-Jurcevic et al. 2000), birds (Herbert, Lowenhaupt et al. 1995), flies (Palladino, Keegan et al. 2000) and worms (Tonkin, Saccomanno et al. 2002).

All ADARs have some common features; they contain at the C-terminal end a highly conserved catalytic domain and in the region of their N-terminus variable numbers of double stranded RNA binding domains. ADARs act on completely base-paired dsRNA structures, which are recognized by the double stranded RNA binding domains (dsRBDs), and they do not need additional cofactors to be active.

In mammals three ADAR enzymes are known.

ADAR 1 is an omnipresent protein whose expression is regulated. The presence of two promoters leads to expression of two isoforms, a long and a short protein. The longer one, 150 kDa ADAR1i, is mainly cytoplasmic and is expressed from an interferon inducible promoter (Liu, George et al. 1997). In contrast to other ADARs, ADAR1i has a long amino-terminal extension that encloses two Z- DNA binding domains (Herbert, Alfken et al. 1997), which may contribute to the localization of the enzyme near to DNA that is transcribing ADAR substrates (Herbert and Rich 2001). ADAR1c, is 110kDa in size and its expression is conducted by the second promoter, the constitutive one. This form of ADAR1 lacks the first 295 amino acids of the

longer form because of alternative splicing (Patterson and Samuel 1995). Consequently it lacks the Zα subdomain which contains a nuclear export signal (NES) Because of this absence and the presence of a nuclear localization sequence (NLS), which is overlapped by the third RNA binding domain (dsRBD3) of the protein (Eckmann, Neunteufl et al. 2001), ADAR1c accumulates in the nucleus. However, it is believed that both isoforms can edit pre-mRNA site-specifically, independent of their localization. Other features of ADAR1 are the catalytic domain in the C-terminus and three double stranded RNA binding domains; in human the third of these, as mentioned above, overlaps with a nuclear localization signal in the amino terminal region, whose activity does not depend on RNA binding (Eckmann, Neunteufl et al. 2001). ADAR1i has the ability to shuttle between nucleus and the cytoplasm due to the presence of both, nuclear localization and export signals. Additionally, ADAR1i- N-terminus harbors an arginine-glycine (RG) rich domain, whose function is not known (O'Connell, Krause et al. 1995).

Furthermore, even though ADAR1c lacks a classical nuclear export signal (NES), it is also able to shuttle between the nucleus and cytoplasm (Strehblow, Hallegger et al. 2002).

ADAR2 is mainly expressed in the nervous tissue. It is structurally similar to ADAR1, nevertheless it has only two double stranded binding domains, and lacks a Z-domain in the N terminal region. There are two isoforms of human ADAR2, the longer one (ADAR2L) possesses, in contrast to the shorter one (ADAR2S), an Alu-sequence inserted into the deaminase domain (Gerber, O'Connell et al. 1997). An aminoterminal nuclear localization signal (NLS) is also present in ADAR2, which gives this protein the ability to accumulate in the nucleus. Within the nucleolus ADAR1 and ADAR2 co-localize in a novel compartment and they are in constant flux in and out of the nucleolus (Desterro, Keegan et al. 2003).

ADAR3 shows a high (72%) homology to ADAR2. Beside the catalytic domain and the two double stranded RNA binding domains, ADAR3 includes also an arginine/lysine rich domain at its N terminal region, which has the ability to bind single stranded RNA. The presence of this single stranded RNA binding domain and its expression only in specific tissues make it different from ADAR1 (Chen, Cho et al. 2000). Another difference is that ADAR3 shows no editing activity on known substrates in vivo as well as on synthetic dsRNA.

ADATs.

Through searches in yeast and Drosophila with ADAR-type deaminase motifs, a group of enzymes adenosine deaminases that act on tRNA (ADATs), were identified. These enzymes deaminate adenosine to inosine in tRNA at position A37, next to the anticodon or in the wobble position A34. The similar deaminase domain suggests an evolutionary relationship to ADARs (Gerber, Grosjean et al. 1998).

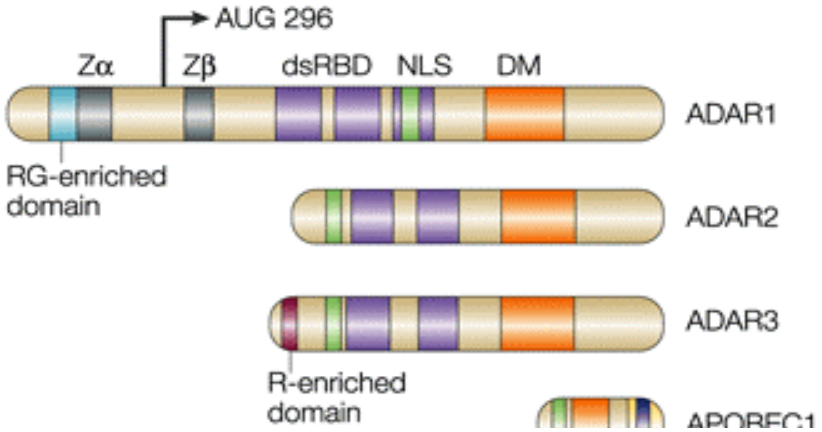


Figure 2: Schematic diagram of human ADARs

(Keegan, Gallo et al. 2001). The enzymatic activity of all ADARs is accomplished through the deaminase domain in the C terminus. An important role in RNA recognition plays the double stranded RNA binding domains in the N terminal region. In this region in ADAR1 are additionally present two Z DNA binding domains and a RG-

enriched domain. A nuclear localization sequence is also contained in all proteins (Keegan, Gallo et al. 2001).

Mechanism and types of A to I editing in vivo

In vivo, there are two types of A to I editing :

Hyper-editing of multiple adenosines, has been almost found within untranslated regions, longer and completely duplexed structures whose functional consequences are still unclear (Athanasiadis, Rich et al. 2004; Levanon, Eisenberg et al. 2004)

The other type is the site selective editing of substrates. Here only a few adenosines within an imperfect RNA fold-back structure are edited. It is supposed that internal mismatches and bulges within an RNA duplex are important for ADAR selectivity. However, it is still not fully understood which properties make an RNA suitable for site selective editing (Kallman, Sahlin et al. 2003; Stephens, Haudenschild et al. 2004).

Binding and catalysis are independent events therefore it could be that ADAR enzymes bind to various transcripts nonproductively (Kallman, Sahlin et al. 2003). It was thought that ADARs binds to editing sites as monomer but some works have suggested that catalytic ADAR is a dimer, and this dimerization is required for RNA editing activity of ADARs (Cho, Yang et al. 2003; Gallo, Keegan et al. 2003). In order that catalysis occurs two monomers must bind, therefore dimerization is probably what contributes in specificity of editing. Additionally, for ADAR editing the double stranded RNA structure and an editing complementary sequence is required in all known substrates (Gerber and Keller 2001). It was shown that editing sites are only coupled within specific distances from each other and the coupled sites of editing are positioned on the same side of a helix, indicating that the three dimensional structure is key in ADAR enzyme substrate recognition (Enstero, Daniel et al. 2009).

Substrates of adenosine deaminases that act on RNA

Editing in non-coding regions

In non-coding regions a great number of human adenosine to inosine RNA editing sites has been identified. These regions consist of inversely oriented repetitive elements, mostly Alu elements and some LINE. It is predicted that more than 85% of pre-mRNA may be edited, with the vast majority being targeted in introns and UTRs (Athanasiadis, Rich et al. 2004; Levanon, Eisenberg et al. 2004). Alu elements are short interspersed elements (SINEs). They are repeat sequences of about 300 base pairs in length, are abundantly found in primates and show a dimeric structure formed by fusion of two monomers derived from the 7SL RNA gene (Kriegs, Churakov et al. 2007). In different tissues there are various editing levels of Alu- containing mRNA, which correlates with the homology between inverted repeats, frequency of the SINEs and their distance (Athanasiadis, Rich et al. 2004; Levanon, Eisenberg et al. 2004).

It was proposed that editing in untranslated regions have a strong impact on the transcriptome and proteome through regulation of splicing, RNA structure, RNA stability and RNA retention (Chen, DeCerbo et al. 2008).

Additionally to repetitive elements, editing can also modify microRNA precursors and microRNA sequences. Editing can change the seed sequence in microRNAs, a small region near the 5`end of the miRNA which is important in defining target specificity, and targeting sequences in mRNAs, leading to inhibition or alteration of processing (Kawahara, Zinshteyn et al. 2007; Kawahara, Zinshteyn et al. 2007; Kawahara, Megraw et al. 2008).

A new function of editing in the control of miRNA biogenesis is the suppression of processing of pri-miR-142 by Drosha, the precursor of miRNA

142. The edited pri-miR-142 was suggested to be degraded by Tudor-SN, a component of RISC and also a ribonuclease specific to inosine-containing dsRNAs (Yang, Chendrimada et al. 2006). Additionally, it has also been shown that as a result of RNA editing of pri-miR-151, its cleavage by Dicer and accumulation of the edited pre-miR-151 RNAs is completely blocked (Kawahara, Zinshteyn et al. 2007).

Editing in coding regions

Mammalian glutamate receptor.

The glutamate receptors are ligand-gated ion channels expressed mainly in the central nervous system that mediate the vast majority of excitatory neurotransmission in the brain. There are three subgroups of the glutamate receptors which show differences in agonist activation and pharmacological characterization : AMPA, NMDA and kainate receptors (Dingledine, Borges et al. 1999).

ADARs can edit five of the gluR subunits: gluR-B, -C, -D, -5, and -6. In gluR-B editing of two exonic sites lead to functional consequences: Q/R site which is located in exon 11 and R/G site located in exon 13. Editing at the Q/R site changes a glutamine codon to an arginine codon leading to a lower Ca+ permeability of the channel (Hume, Dingledine et al. 1991; Burnashev, Monyer et al. 1992). Additionally editing at this site is necessary for an accurate tetramer assembly of AMPA receptor (Greger, Khatri et al. 2003). Editing at the R/G site allows faster recovery of the receptor from desensitization due to a convertion of an arginine codon to a glycine codon (Lomeli, Mosbacher et al. 1994). The Q/R site is edited only by ADAR2 whereas R/G site can be edited by ADAR1 and ADAR2 (Higuchi, Maas et al. 2000).

An increased Ca+ permeability because of underediting of glutamate receptors lead to severe epileptic seizures and death within two weeks of age

(Brusa, Zimmermann et al. 1995). Furthermore after inactivation of the editing enzyme ADAR2, the same phenotype is observed (Higuchi, Maas et al. 2000).

The R/G site shows different editing levels during development and the convertion of arginine to glycine at this position is not essential (Lomeli, Mosbacher et al. 1994). It has been shown that editing at R/G site facilitates alternative splicing events occurring downstream of this site, due to reducing splicing efficiency of the adjacent intron.

Decreased editing at Q/R site has also been shown to lead to accumulation of incompletely spliced gluR-B transcripts and as a result of that they still contain the intron in which the ECS (editing site complementary sequence) for this site is located (Higuchi, Maas et al. 2000; Schoft, Schopoff et al. 2007).

Mammalian serotonin receptor

Serotonin (5-HT) is a neurotransmitter that is involved in the regulation of cognition, mood, aggression, appetite and sexual behavior.

5HT2C serotonin receptors are G-protein coupled receptors, which transmit extracellular signals through this interaction. In the mRNA of the 2C subtype of serotonin receptor, five sites have been shown to be edited from ADAR enzymes (termed as site A to E), leading to five codon exchanges (Burns, Chu et al. 1997). ADAR1 edits the A and B sites while the C, D and E sites are edited by ADAR2 (Higuchi, Maas et al. 2000; Wang, Khillan et al. 2000).

As RNA editing can change the serotonin response, and serotonin has important functions in the central nervous system, some psychiatric disorders such as depression and anxiety are thought to be implicated with altered RNA-editing patterns (Niswender, Herrick-Davis et al. 2001).

Additionally, it was suggested that RNA editing has the ability to regulate serotonergic signal transduction and that this modification may be

important in regulating functions that are interceded by other members of the G-protein coupled receptor superfamily (Burns, Chu et al. 1997; Schmauss 2005).

RNA editing in invertebrates

In *Drosophila melanogaster* only one ADAR gene is known, which is most homologous to mammalian ADAR2 and is expressed in the developing nervous system (Palladino, Keegan et al. 2000). The best characterized dADAR targets are transcripts that encode ion channel subunits as the pore-forming α1 subunit of a voltage gated calcium channel (Kawasaki, Collins et al. 2002), α1 subunit of a voltage gated sodium channel (Hanrahan, Palladino et al. 2000) and glutamate gated chloride channel (Semenov and Pak 1999). *Drosophila* ADAR edits also its own mRNA changing a conserved residue in the catalytic domain, a process which is developmentally regulated (Palladino, Keegan et al. 2000).
Drosophila ADAR mutants lack site selective editing of all known pre-mRNA targets and also nonspecific adenosine deamination activity leading to adult behavioral defects which increase in severity with age and are associated with neurodegeneration (Palladino, Keegan et al. 2000)
Caenorhabditis elegans possesses two ADAR genes adr-1 and adr-2; adr-1 is expressed in the most cells of the nervous system and in the developing vulva. Both of these genes have different roles in RNA editing and are essential for normal function of the nervous system (Tonkin, Saccomanno et al. 2002).

Viral RNA editing

The interferon system is the first line of defense against viral infection in mammals and because of that it was naturally to examine the role of

ADAR1 in viral replication because the long form of ADAR1 is interferon inducible (reviewed in (Sen 2001)).

Modification by an ADAR enzyme can lead to sequence changes also in viral transcripts (Cattaneo 1994). These sequence changes are frequently nonselective or of the hypermutation type of deamination that occurs in substrates that are completely double stranded. In some cases the viral transcripts form intramolecular hairpins (O'Hara, Horodyski et al. 1984; Hajjar and Linial 1995) or interact with an antisense molecule (Kumar and Carmichael 1997), but in some other cases it not clear how dsRNA forms. Formation of aberrant double stranded structures between the plus and minus strands or the minus strands and an mRNA, has been suggested for RNA viruses that replicate through an RNA intermediate (Bass, Weintraub et al. 1989).

One example of the high probability that ADAR1 functions in host defense mechanisms against viral infection and inflammation is the fact that editing of hepatitis C virus RNA genome inhibits its replication. This can explain successful HCV replicon clearance by IFN-α in vitro and my present a promising new therapeutic strategy for HCV as well as other viral infections (Taylor, Puig et al. 2005). Another example is the increase of the inosine-containing mRNAs in T lymphocytes and macrophages stimulated with a variety of inflammatory mediators, including tumor necrosis factor-α (Yang, Luo et al. 2003). Furthermore, the interaction between ADAR1 and nuclear factor 90 family proteins affects the NF90-mediated gene regulation (Nie, Ding et al. 2005). NF90, also known as interleukin enhancer-binding factor 3 (ILF3), is a regulator that stimulates the activation of the cellular antiviral expression cascade and activates interferon-ß (Nie, Ding et al. 2005). ADAR1-NF90 interaction suppresses the interferon activation pathway. This suppression is important for protecting hematopoietic progenitor cells from apoptosis (Hartner, Walkley et al. 2009). Other evidence that supports the function of ADAR1 in DNA-mediated immune

response is its inhibitory effects on interferon-ß induction to the cellular response to DNA treatment (Wang, Choi et al. 2008).

Biological relevance of editing in Human

Dyschromatosis symmetrica hereditaria (DSH), also known as symmetrical dyschromatosis of the extremities, is a pigmenting genodermatosis concentrated on the dorsal extremities with an autosomal dominant inheritance. The locus for this disease was mapped on chromosome 1q21.3 to the ADAR gene. There are 32 ADAR1 mutations identified in DSH patients. H216fs is a frame shift mutation located upstream of the methionine codon 296, which is the initiation codon for translation of the p110 form of ADAR1 (Patterson and Samuel 1995). Besides the knowledge that the dosage of the functional p150 protein is a dominant denominator of DSH, it has also been shown that this protein is involved in apoptotic pathways (Wang, Miyakoda et al. 2004), and seems to regulate cellular RNAi efficacy by sequestering siRNA (Yang, Wang et al. 2005).

It was suggested that RNA editing might have an effect also in Amyotrophic lateral sclerosis (ALS). ALS is a progressive neurodegenerative disease causing muscle weakness, spasticity, muscle atrophy and at a late time paralysis and respiratory failure. In ALS there is an excessive Ca2+ influx through glutamate receptors which contribute to neuronal death (Spreux-Varoquaux, Bensimon et al. 2002). It was found that editing of the GluR-2 Q/R position was decreased in most samples of diseased neurons whereas all control cells showed 100% editing, which indicates that abnormal editing may contribute to neuronal death specifically to sporadic ALS (Kawahara, Ito et al. 2004; Kawahara and Kwak 2005).

Other diseases in which RNA editing might play a role are epilepsy, depression and schizophrenia. It was shown that epileptic phenotypes are most profound in GluR-2 editing deficient mice and the severity of seizures activity correlates directly with the degree of impairment in the editing of

Q/R site (Feldmeyer, Kask et al. 1999). Furthermore there are multiple groups that reported an increased RNA editing efficiency at the A site of serotonin receptor in brains of depressed suicide victims as well as an increased editing at the E site in individuals with major depression (Niswender, Herrick-Davis et al. 2001; Gurevich, Tamir et al. 2002; Iwamoto and Kato 2003).

Double stranded RNA binding proteins (dsRBPs)

The common feature of double stranded RNA binding proteins is their evolutionarily conserved RNA binding domain which has been first identified in the *Xenopus laevis* RNA-binding protein A (XlrbpA) and the *Drosophila* maternal effect protein Staufen. They have the ability to bind to double stranded RNAs and have an important role in some biological processes like for example RNA editing, RNA localization, RNA interference and RNA processing (reviewed in (Doyle and Jantsch 2002; Saunders and Barber 2003)).

Most of the double stranded RNA binding proteins contain multiple dsRBDs but also proteins with single dsRBDs do exist. dsRBDs have different affinities for dsRNA in vitro, furthermore some of them are actually incapable of binding to it (Krovat and Jantsch 1996). They have a length of approximately 70 amino acids shaping a conserved α-ß-ß-ß-α protein topology in which the two α-helices are packed along the face of a three stranded antiparallel ß-sheet (Bass, Hurst et al. 1994; Bycroft, Grunert et al. 1995).

There are some ways how RNA binding occurs. In seeking to establish these mechanisms, the structures of the second dsRBD of XlrbpA (Ryter and Schultz 1998), the third dsRBD of Staufen (Ramos, Grunert et al. 2000) and the dsRBD of Rnt1p, an RNAse III homologue from budding yeast (Wu, Henras et al. 2004) have been determined in complex with RNA.

On the one hand the crystal structure of the dsRBD of XlrbpA revealed that there is no direct base contact in the interaction between dsRBD and the sugar phosphate backbone, therefore proposing a lack of sequence specificity (Ryter and Schultz 1998). On the other hand there are a lot of dsRBDs that have been shown to bind specifically on dsRNA in vivo and in vitro, for example ADAR1 and ADAR2 bind in vivo only a selected group of substrates, or Staufen association with bicoid mRNA, which was one of the first to show the direct interaction between a dsRBD and RNA and the formation of a ribonucleoprotein particle (RNP) (Ferrandon, Elphick et al. 1994; Wagner, Palacios et al. 2001). Additionally a surprising discovery as mentioned above was that not all dsRBDs are able to bind RNA (Krovat and Jantsch 1996).

Through their ability to bind a wide array of different dsRNAs, they can accomplish a large number of different functions.

Drosophila Staufen plays a fundamental role in the formation of the anterior- posterior axis, through the localization of the maternal mRNAs *oskar* and *bicoid*, moving the first one from the anterior to the posterior pole of the oocyte and anchoring the later one at the anterior end of the developing embryo by binding to e stem loop on its 3`-UTR (Ferrandon, Elphick et al. 1994; Wagner, Palacios et al. 2001). The third of the five dsRBD of the Staufen is crucial for the RNA binding and its role afterwards(Ramos, Grunert et al. 2000).

RNA-dependent human protein kinase (PKR) is an important element in the cellular defense mechanism against virus infection. Its activation through the binding of the viral dsRNA to its two dsRBDs, causes it to phosphorylate the translation factor eIF2α, leading to a protein synthesis inhibition in the cell (Meurs, Chong et al. 1990; Green and Mathews 1992).

Additionally to the functions after binding dsRNA, dsRBD proteins show also other functions that are RNA binding independent. The dsRBD2 of Staufen, is responsible for microtubule-dependent localization of *Oscar* mRNA whereas dsRBD5 is needed for its translational control, and both of

them are the only dsRBDs which do not bind dsRNA in vitro (reviewed in Doyle and Jantsch 2002).

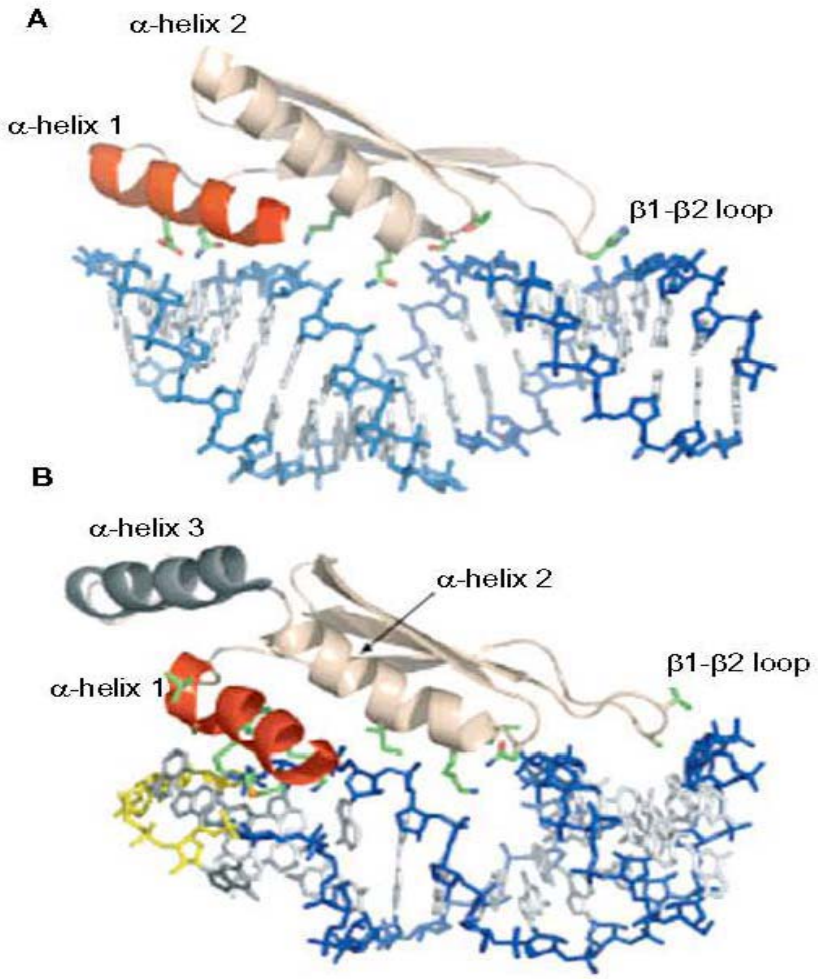


Figure 3: Double-stranded RNA recognition by double-stranded RNA-binding domains

(Stefl, Skrisovska et al. 2005).

(A) The double-stranded RNA-binding domain (dsRBD) of Xlrbpa2 bound to dsRNA (Ryter and Schultz 1998). In red is shown the α-helix 1, which together with β1-β2 loop and the aminoterminal part of α-helix 2, recognize the shape of double stranded RNA. In blue and light blue are coloured the backbones of two coaxial duplexes.

(B) The dsRBM of Rnt1p bound to the stem–loop closed by the AGNN tetraloop (Wu, Henras et al. 2004). Backbone is highlighted in yellow, basis in black whether α-helix 1 in red. The additional carboxy-terminal α-helix 3 of the dsRBM of RNT1p modulates the conformation of α-helix 1 (Leulliot, Quevillon-Cheruel et al. 2004).

Nuclear import and export

The nuclear envelope of all eucaryotes is perforated by structures known as nuclear pore complexes (NPCs), which are composed of aproximately 30 different proteins, *nucleoporins*, present in multiple copies disposing an octagonal symmetry, representing the sites of movement between the nucleus and cytoplasm (Rout and Aitchison 2000).

The nuclear pore complex runs through the nuclear membrane forming a channel and manages the nucleocytoplasmic transport of mebrane proteins (King, Lusk et al. 2006), soluble proteins (Suntharalingam and Wente 2003) and different types of RNAs (Franke and Scheer 1974; Gorlich and Kutay 1999; Vinciguerra and Stutz 2004). The central channel is aproximately 30nm in diameter and allows the passive diffusion of ions and proteins with a molecular mass up to 40 kDa. The major part of the macromolecules perform their transport through the NPC, using an energy-dependent mechanisem mediated by transport receptors that belong to a family of proteins known as ß-karyopherins (Mosammaparast and Pemberton 2004). B-karyopherins accomplish nuclear import and export of membrane proteins, some RNAs and ribosomal subunits (Kohler and Hurt 2007). The ß-karyopherins that mediate nuclear import are called importins and those involved in the nuclear export are called exportins.

The destination of the cargo protein is defined by the interaction of the transport receptors with motifs present in the cargo called nuclear localisation signal (NLS) or nuclear export signal (NES) (Kohler and Hurt 2007).

The GTPase Ran is a monomeric G-protein that mediates the direction of the nucleocytoplasmic transport. It is present in the cytoplasma as well as in the nucleus and exists in two conformational states depending whether bound to GTP or GDP. Because the hydrolysis of GTP through Ran is very slow, the interaction with regulatory proteins regulates its nucleotide binding. External factors like Ran-GTPase activating protein (Ran-GAPs) and Ran binding protein 1 (Ran BP1) mediate the hydrolysis from GTP to GDP. A conformational change is triggered through the binding of RanBP1 to RanGTP, that allows RanGAP to hydrolyze GTP (Conti and Izaurralde 2001).While RanGAP and Ran BP1 are exclusively located in the cytoplasm, guanine nucleotide exchange factors (GEFs) are only located in the nucleus that exchange GTP for GDP. The generation of the gradient over the nuclear membrane through the asymmetrical distribution of Ran regulatory proteins establishes the directionality of the nucleocytoplasmic transport (Gorlich, Seewald et al. 2003).

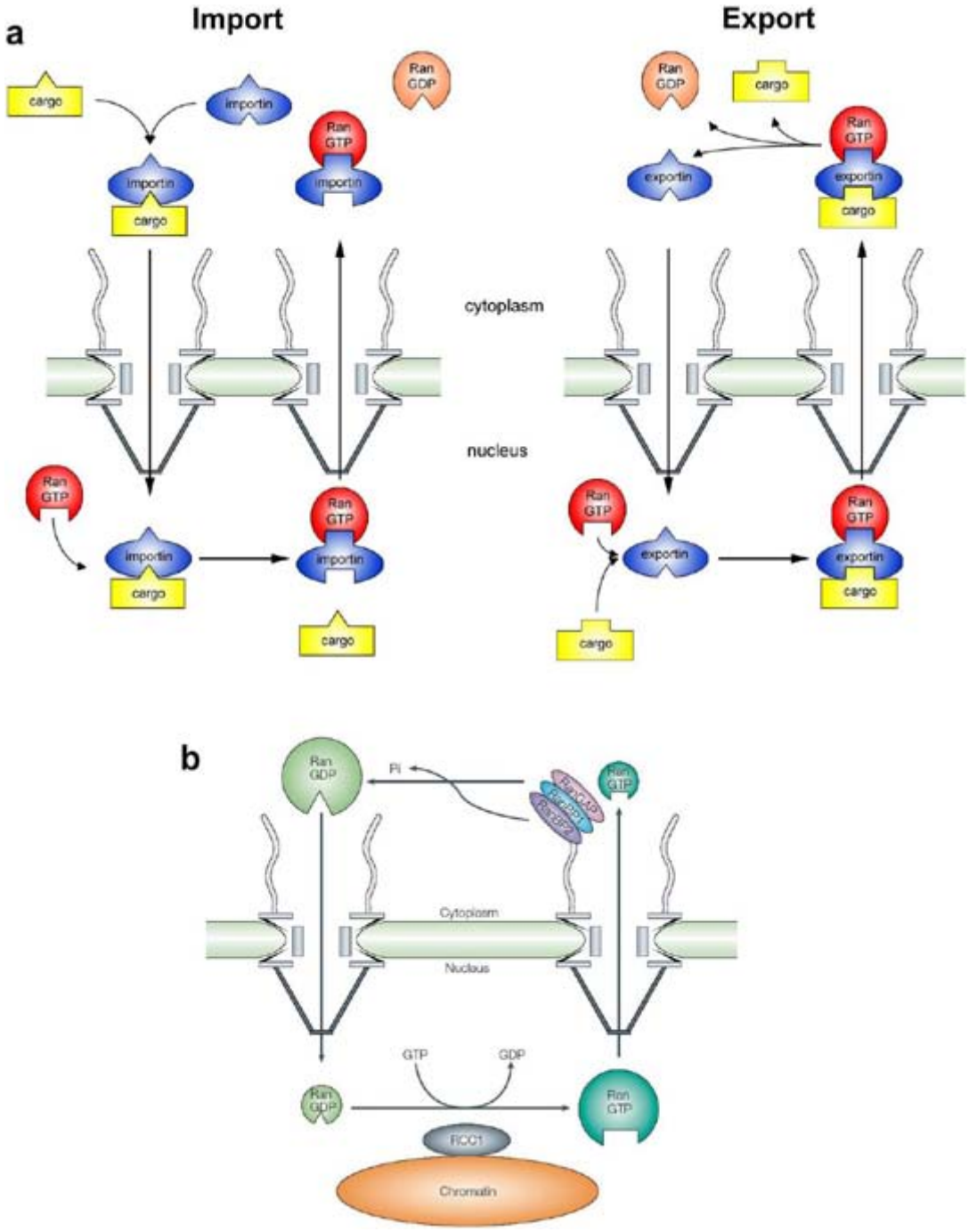


Figure 4: Nucleocytoplasmic transport

(a) Represented are nucleocytoplasmic import (left) and export (right). (b) Ran mediated directionality of the nucleocytoplasmic transport (Atlas of Genetics and Cytogenetics in Oncology and Haematology (atlasgeneticsoncology.org)).

(a) After binding of importin to the nuclear localization signal of the cargo protein in the cytoplasm follows the interaction with NPC to mediate the translocation of the complex. The importin Ran-GTP complex in the nucleus lead to the release of cargo. The newly formed complex is further recycled to the cytoplasm. The translocation of the cargo protein from the nucleus to the cytoplasm takes place after binding of its nuclear export signal to the exportin, which is stimulated by RanGTP. In the cytoplasm in consequence of removal of GTP from the complex by GTP hydrolyses, exportin dissociates from the cargo and is back recycled to the nucleus.

(b) In the nucleus a high concentration of Ran GTP results after exchange from RanGDP to RanGTP by chromatin bound RCC1, which is a nucleotide exchange factor for Ran (RanGEF). Maintenance of a low concentration of Ran GTP after importin- RanGTP translocation is mediated by RanGTP-binding proteins and RanGAP.

Nuclear transport of hsADAR1

Beside protein cargos, different types of RNA represent another group of export cargos (Lei and Silver 2002). Two exportins bind RNA directly, exportin-t which recognize a nuclear export signal in tRNAs mediate its export from the nucleus and exportin-5 which recognize an RNA hairpin structure with a 3´overhang mediate mRNA precursors export (Arts, Kuersten et al. 1998; Kutay, Lipowsky et al. 1998; Kim 2004). Some factors involved in the mRNA export are also involved in mRNA transcription, processing and splicing indicating that these processes are strictly linked (Lei and Silver 2002).

A number of pri-miRNAs have been shown to be edited either by ADAR1 or ADAR2 (Yang, Chendrimada et al. 2006). ADAR1 which shuttles between nucleus and cytoplasm could be involved in the transport of the RNA substrates over the nuclear membrane (Strehblow, Hallegger et al. 2002). The control of the nuclear concentration of this enzyme could regulate its

nuclear transport, and as a consequence of that hyperediting might be prevented.

As was mentioned above the human RNA-editing enzyme ADAR1 is expressed in the longer 150kDa, interferon inducible version and in the amino-terminally truncated 110kDa version. The first one ADAR1i is nuclear and cytoplasmic while the second one ADAR1c is predominantly nuclear (Patterson and Samuel 1995). Even though ADAR1c lacks a NES it is able to shuttle between nucleus and cytoplasm (Strehblow, Hallegger et al. 2002). It was shown that the NLS overlaps the third dsRBD of ADAR1c, and that transportin-1 is required for the nuclear import of ADAR1 (Fritz, Strehblow et al. 2009). Different dsRBD constructs were used to test the influence of RNA binding on the interaction between them and TRN1, what leads to the conclusion that dsRNA binding inhibited interaction between TRN1 and dsRBD3 of ADAR1 (Fritz, Strehblow et al. 2009). Furthermore it was shown that exportin-5 also interacts with the dsRBDs interfering with TRN1 for dsRBD binding and suggesting a dual role for dsRBD as a nuclear import and export signal (Fritz, Strehblow et al. 2009).

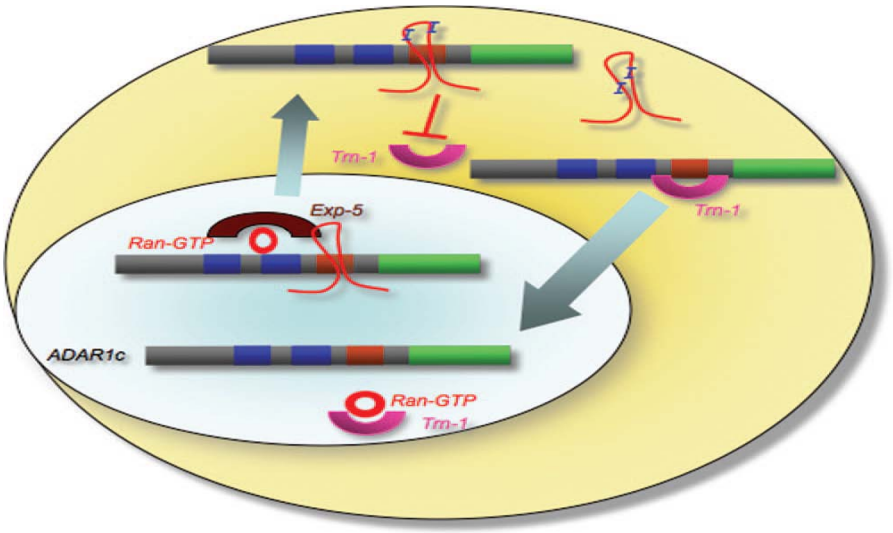


Figure 5: A model of ADAR1 nucleocytoplasmic shuttling in a dsRNA-dependent way

(Fritz, Strehblow et al. 2009). TRN1 binds ADAR1 in the cytoplasm and transports it to the nucleus, where it is released by RanGTP binding. Nuclear export is then mediated through the interaction between Exp-5 and one or multiple dsRBDs in a Ran-GTP-dependent manner. Once arrived in the cytoplasm RanGTP is hydrolysed to RanGDP, releasing the complex. The presence of dsRNA precludes the interaction of dsRBD3 with TRN1 preventing the reimport of ADAR1 bound to RNA.

Aim of the thesis

The aim of the thesis was to characterize, determining the three dimensional structure, the third double stranded RNA binding domain of the human adenosine deaminases that act on RNA (hADAR1c).

It has been shown that the third dsRBD of ADAR1 is essential for both RNA binding as well as for enzyme activity. While the longer version hADAR1i has the ability to shuttle between nucleus and cytoplasm very likely because of the presence of both nuclear localization signal and nuclear export signal, the shorter N-terminal truncated version ADAR1c lacking the NES is predominantly nuclear. Additionally it has been shown, that the NLS of ADAR1c overlaps its third dsRBD, and that transportin-1 binds ADAR1 through this domain and is able to import hADAR1-dsRBD3 to the nucleus in a RanGTP dependent manner without the need of additional factors. Furthermore, dsRBD1 and the presence of dsRNA show an inhibitory effect, because in the first case TRN1 cannot mediate the nuclear import of dsRBD1+3 and in the second case TRN1 and dsRBD3 are able to interact only if dsRNA is absent.

A three-dimensional structure of dsRBD3 of ADAR1c and further on a 3D structure of this domain bound by RNA could probably provide insights, on how this interaction exactly occurs and help by finding out control mechanisms for its transport.

Nuclear magnetic resonance spectroscopy (NMR) is a method used to analyze three dimensional structures. For this purpose a high protein concentration is needed. From previous experiments, we know that the third dsRBD of ADAR1c is hardly soluble or more specifically only a low concentration is achieved in solution.

In trying to increase the solubility and consequently the concentration of this protein, we made forty constructs that differed in C- and N-termini

length. These constructs were tested using different buffers, pH, temperatures, and culture growing conditions to find out the most promising candidate and also the best conditions in which this construct is highly soluble.

Materials and Methods

Competent bacterial strains

For cloning, Escherichia coli XL1Blue competent cells were used. These cells are endonuclease (endA) and recombination (recA) deficient, consequently greatly improving the quality of miniprep DNA and insert stability.

Escherichia coli BL21 derivatives BL21 (DE3) and in later experiments Rosetta C41 were used for protein expression. They allow high efficiency protein expression of any gene that is under the control of a T7 promotor.

Media

The bacteria are grown in a rich medium (2xTY) and to label the protein with 15N a minimal medium (M9) was used.

TY Medium (pH=7,4)	
Chemicals	Weight
Tryptone	16g
Yeast extract	10g
NaCl	5g

M9 Medium (1l)	
Chemicals (concentration)	Volume
M9 Salt (5x)	200ml
MgSO4 (1M)	2ml
Glucose (20%)	20ml
CaCl2 (1M)	0.1ml
Water	778ml

M9 Salt	
Chemicals	Weight
Na2PO4 . 2H2O	47g
KH2PO4	15g
NH4Cl (15N-labeled)	5g

Table 1: TY-medium, M9-medium and M9-salt.

Forty-two different constructs of dsRBD3- hADAR1

To generate different lengths of dsRBD3, different primers were designed, containing BamHI (GGA TCC) and HindIII (AAG CTT) restriction sites which are also present in the multiple cloning sites (MCS), downstream of the T7 promotor in cloning vector pRSETA (MJ109) and pET30(MJ..). Vector pRSETA contains a His-Tag and pET30 contains a GB1- domain (56 amino acid protein) as a solubility-enhancement tags. Further steps and methods (shown below) are necessary to get the recombinant DNA, the consequent protein expression and purification.

To amply the specific sequences, polymerase chain reaction was performed.

Polymerase Chain Reaction (PCR)

Polymerase Chain Reaction (PCR)	
Solution (concentration)	Volume
Template DNA (hADAR1GFP; MJ7)	100ng (1µl of 1:10 dilution)
Primer 5`	100ng (1µl of 1:10 dilution)
Primer 3`	100ng (1µl of 1:10 dilution)
dNTP (10mM)	1.5µl
Buffer 10x ((NH4)2SO4; -MgCl2)	5µl
MgCl2 (25mM)	2.5µl
Taq Polymerase	0.5µl
Water dist.	37.5µl
	25µl total volume

Table 2: Solutions for the polymerase chain reaction.

Steps	Time	Temperature	Number of cycles
Melting	4`30``	95°C	1
Annealing	30``	50°C	30
	30``	50°C	
Elongation	45``	72°C	
	5`	72°C	1

Table 3: PCR-conditions

For primer sequence and combination see the tables below.

Primer 5‘	**Primer 3‘**
a. MJ2527 5‘-CAGGATCCATGATGCCCAACAAGGTCA-3‘	**g**. MJ2533 5‘-GTGGTACCCTAGCGTTCTGCCTTCTCGTTCTC-3‘
b. MJ2528 5‘-CAGGATCCGAATCCATGATGCCCAACAAG-3‘	**h**. MJ2534 5‘-GTGGTACCCTATGTGAAACCCATGCGTTCTGC-3‘
c. MJ2529 5‘-CAGGATCCAACTTGGAATCCATGATGCCCAAC-3‘	i. MJ2535 5‘-GTGGTACCCTAGGTTACCTCTGTGAAACCCAT-3‘
d. MJ2530 5‘-CAGGATCCCTTGATAACTTGGAATCCATGATG-3‘	**j**. MJ2536 5‘-GTGGTACCCTATGTCACTTGGGTTACCTCTGT-3‘
e. MJ2531 5‘-CAGGATCCGAGTCACTTGATAACTTGGAATCC-3‘	**k**. MJ2537 5‘-GTGGTACCCTAACTGGCCCCTGTCACTGGGGT-3‘
f. MJ2532 5‘-CAGGATCCTCAGAGTCACTTGATAACTTGGAA-3‘	**l**. MJ2538 5‘-GTGGTACCCTATCTGAGACTGGCCCCTGTCAC-3‘

Table 4: Primer designed for the different constructs.

Constructs 1-36 (in pRSETA) - Primer combination					
1 - ag	7 - bg	13 - cg	19 - dg	25 - eg	31 - fg
2 - ah	8 - bh	14 - ch	20 - dh	26 - eh	32 - fh
3 - ai	9 - bi	15 - ci	21 - di	27 - ei	33 - fi
4 - aj	10 - bj	16 - cj	22 - dj	28 - ej	34 - fj
5 - ak	11 - bk	17 - ck	23 - dk	29 - ek	35 - fk
6 - al	12 - bl	18 - cl	24 - dl	30 - el	36 - fl

Construct	**Primer combination**	
	5‘	**3‘**
A	MJ2695 5‘-CTGGGATCCGAGGCGACCAACTCCATG-3‘	MJ2683 5‘-GCTAAGCTTCTCAGGACCTTGAGAGGAGGAGC-3‘
B	MJ2527 5‘-CAGGATCCATGATGCCCAACAAGGTCA-3‘	MJ2683 5‘-GCTAAGCTTCTCAGGACCTTGAGAGGAGGAGC-3‘
C	MJ2695 5‘-CTGGGATCCGAGGCGACCAACTCCATG-3‘	MJ2533 5‘-GTGGTACCCTAGCGTTCTGCCTTCTCGTTCTC-3‘
E	MJ2527 5‘-CAGGATCCATGATGCCCAACAAGGTCA-3‘	MJ2763 5‘-GGGTTCCGCGCACATTTCCCCG-3‘

Construct	Primer combination	
	5‘	3‘
N13	MJ2529 5‘-CAGGATCCAACTTGGAATCCATGATGCCCAAC-3‘	MJ2735 5‘-GGTGGATCCGCGTTCTGCCTTCTCGTTCTC-3‘
NB	MJ2527 5‘-GGTGGATCCGCGTTCTGCCTTCTCGTTCTC-3‘	MJ2737 5‘-GCTGGGATCCTGGGGACCTTGAGAGGAGGAG-3‘

Table 5 : Primer combination

Restriction Digest

Restriction digestion was performed in a total volume of 20µl. Master Mix, including restriction enzyme and corresponding buffer, was prepared on ice. DNA was added to the Master Mix aliquots and incubated usually for 1,5 hour at 37°C.

Solution	Volume
DNA	5µl
Restriction enzyme	1µl
Buffer	2µl
RNase/BSA	2µl
Water dist.	10µl
	20µl total

.

Table 6: Solutions for the restriction digest.

Gel elektrophoresis

To test if PCR, restriction, ligation, or minipreps have properly functioned, DNA was separated on agarose gels of 0.5-2% depending on fragment sizes.

Agarose was solubilised in 1x TAE buffer (2M Tris, 1M NaAc, 50mM EDTA, pH=7.4). Samples were mixed with 6x loading buffer and loaded onto the gel. As a running buffer was used 1x TAE and gels were run at constant voltage of 70V until the desired separation was achieved.

Ligation and Transformation

For the ligation, a ratio of 1 vector to 3 inserts was used. Vector, insert, ligase, appropriate ligase buffer and distilled water were mixed at a total volume of 20µl. After that, the tubes were put usually in the 16°C incubator over night.

For transformation, aliquots of frozen cells were thawed on ice. After addition of plasmid DNA, the mixture remained on ice for at least 15 minutes. A subsequent 90-second heat shock at 42°C was followed by the addition of SOB medium. Cells were incubated in a 37°C shaker for 1 hour, than plated on Luria Bertani (LB) - Agar with the appropriate antibiotic to maintain plasmid selection, and afterwards the plates were incubated overnight at 37°C.

Isolation of Plasmid DNA

To extract and purify plasmid DNA mini- and midi-preparation`s protocols were used.

For minipreps, were used cells, which were span 10 minutes at 3.000 rpm after an overnight inoculation of colonies in 3ml TY medium (one colony/ 3ml Medium) with appropriate antibiotic. The supernatant was discarded, cell pellet resuspended in 200µl GTE (Glucose/Tris/EDTA), transferred in fresh eppendorf tube and incubated 5 minutes at room temperature. After adding 400µl alkaline SDS (1% SDS/ 0.2M NaOH), the tubes were inverted several times and incubated 5 minutes on ice. The next steps after adding 300µl KAc were to mix them, to leave them 5 minutes on ice and to spin 5 minutes at 14.000 rpm. Supernatant was transferred in fresh eppendorf tube and 700µl 2-propanol was added. A subsequently 10-20 minutes incubation at -20°C was followed by 5-10 minutes spinning at 14.000 rpm. The pellet was dissolved in 200µl water and 200µl 5M NH4Ac was added. After shaking 5-10 minutes and spinning 5 minutes at 14.000 rpm, the supernatant was transferred again in fresh eppendorf tube and 1ml 96% EtOH was added. The incubation at -80°C for 30 minutes (or overnight at -20°C) is followed by spinning for 5 minutes at 14.000 rpm. The pellet was washed with 50µl 70% EtOH, span 5 minutes at 14.000, dried and dissolved in 50µl distilled water.

DNA Extraction

To the isolated DNA was added 1µl RNAse A and was incubated for 15 minutes at 37°C. Addition of 50µl 1x TE (minimal total volume: 100µl) and 50µl phenol was followed by a vigorous vortex and 5 minutes spinning at 14.000 rpm. The supernatant was transferred in a fresh eppendorf tube, 50µl PCI (Phenol/ Chloroform/ Isoamylalcohol) was added and again vortex

and spinning 5 minutes at 14.000 rpm are done. This step was repeated with CI (Chloroform/ Isoamylalcohol) and after transferring the supernatant in a fresh eppendorf tube, 10µl NaOAc (1/10 Vol.) and 300µl 96% EtOH (3 Vol.) was added and mixed. The last two steps include the incubation at -80°C for 2 hours (or at -20°C overnight) and the DNA precipitation.

Cycle Sequencing

To check if the recombinant DNA contains the proper sequence, Biotech PCR-Machine was used to perform cycle sequencing.

Solution (concentration)	Volume
DNA (500ng/µl)	1µl
Primer (0.8pMol/µl)	4 µl
Buffer (2.5*)	1 µl
Big dye	1 µl
Water dist.	3 µl
	10 µl total

PCR Program (cycle sequencing)	
1.	Rapid thermal ramp to 96°C
2.	96°C for 10"
3. for 25 cycles	Rapid thermal ramp to 50°C
4.	50°C for 5"
5.	Rapid thermal ramp to 60°C
6.	60°C for 4`

Table 7: Solutions and PCR-program used for the cycle sequencing

SDS-polyacrylamid gel electrophoresis (SDS PAGE)

Before loading the samples onto the gel to separate them according to their size, the proteins were denatured mixing them at first with 2x SDS sample buffer (220mM Tris-HCl pH 6.8, 1.8% SDS, 29% glycerol, 0.03% bromophenol-blue, 0.03% ß-mercaptoethanol) and then by boiling them for 5 minutes at 96°C.

After preparing the minigel apparatus (1cm spacer), the separation gel was mixed (values for 13% gel):

Solutions (concentration)	Volume
30% Acrylamide	3ml
Water distilled	1.3ml
Tris pH=8.8 (1M)	2.6ml
SDS 10%	50µl
APS 10%	50µl
Temed	15µl

Table 8: Solutions used for the separation gel (13%)

The gel was cast into the apparatus leaving circa 1.5 cm empty, overlaid with isopropanol and allowed to polymerize. Next the loading gel is mixed, poured and a comb is placed to create the wells.

Loading gel:

Solutions	Volume
Acrylamide (Stock)	4ml
APS 10%	35µl
Temed	10µl

Table 9: Solutions used for the loading gel.

After the loading gel is polymerized, the apparatus was placed into electrophoresis tank, filled with 1x SDS running buffer (25mM Tris-HCl, 190mM glycine, 0.1% SDS), and then the comb was removed. After loading the protein extracts (10µl), the gel was run with 25mA/gel until the sample buffer dye reached the bottom.

Western Blotting

To perform western blotting a polyvinylidene fluoride membrane (GE Water & Process Technologies), activated in methanol and equilibrated in transfer buffer (20mM Tris-HCl, 150mM glycine), was placed on the gel on which the proteins are separated after electrophoresis and all air bubbles were removed. The gel and membrane were placed between two Whatman papers and two sponges, and that sandwich was then put into the tank, which was filled up with transfer buffer. The transfer was carried out at room temperature with icepacks in the apparatus at 360-450mA for 80 minutes. Next, the membrane was washed with 0.5% milk in TBST (25mM Tris, pH 7.4, 3.0mM KCl, 140mM NaCl and 0.05% Tween 20) and then blocked with 5% dry skim milk in TBST for 10 minutes at room temperature. The only one antibody used (AP + Ni-NTA-AB) was diluted in 0.5% milk in TBST (1: 1000), applied to the membrane and incubated for one hour at room temperature. Next, it was washed three times each 5 minutes with 0.5% milk in TBST, and after mixing the APB with 33µl BCIP (5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt) and 66µl NBT (Nitro-Blue Tetrazolium Chloride) it was incubated for 30 minutes at room temperature. Then the membrane was washed two times with water and let dry on paper overnight. The protein on the blot was detected by NBT/BCIP chromogen precipitation. NBT/BCIP together, when reacted with Alkaline Phosphatase yield an intense, insoluble black-purple precipitate.

Expression and purification of recombinant protein

Escherichia coli BL21 (DE3) or Rosetta (C41) strain harboring either pRSETA or pET30 vector encoding for His- or GB1-fusion protein were grown in 50-100ml cultures with the appropriate antibiotic at 37°C. At an OD600 (optic density) of about 0.7 the expression of the protein dsRBD3 was induced by addition of isopropyl-ß-D-thiogalactopyranoside (IPTG) to a final concentration of 1mM. After 3 hours induction the cells were harvested by centrifugation at 14.000rpm for 10 minutes. Cell pellet was resuspended in lysis buffers and further lysed and sonicated as shown in the three steps below:

Step one: lysis with 10mM KI, 20mM KH2PO4 + K2HPO4, 0.1% triton, 1mM DTT, sonication for 5 minutes, centrifugation for 10 minutes at 14000rpm.

Step two: lysis with 150mM KI, 20mM KH2PO4 + K2HPO4, 0.1% triton, 1mM DTT, sonication for 5 minutes, centrifugation for 10 minutes at 14000rpm.

Step three: lysis with 1M KI, 20mM KH2PO4 + K2HPO4, 0.1% triton, 1mM DTT, sonication for 5 minutes, centrifugation for 10 minutes at 14000rpm.

The supernatants were loaded into a Ni-NTA column at room temperature. The column was washed with a buffer containing 10mM imidazol, 150mM KI, 0.1% triton and 1mM DTT at the first step, whether at the second step the washing buffer contained 50mM imidazol, 150mM KI, 0.1% triton and 1mM DTT. The protein was then eluted in fractions of about 500µl volume, with a buffer containing 500mM imidazol, 150mM KI, 0.1% triton and 1mM DTT. The protein was dialysed in four steps, each for 2 hours. Thereby it was also denatured and slowly renatured by stepwise dilution of guanidine hydrochloride, using the buffers below:

First step: 6M guanidine hydrochloride, 150mM KI, 20mM K2HPO4 + KH2PO4 (pH=7.2), 1mM DTT.

Second step: 3M guanidine hydrochloride, 150mM KI, 20mM K2HPO4 + KH2PO4 (pH=7.2), 1mM DTT.

Third step: 1mM guanidine hydrochloride, 150mM KI, 20mM K2HPO4 + KH2PO4 (pH=7.2), 1mM DTT.

Fourth step: 150mM KI, 20mM K2HPO4 + KH2PO4 (pH=7.2), 1mM DTT.

The protein concentration was then measured using Bradford assay.

Bradford assay

Bradford dye concentrate was diluted with distilled water at a ratio 1:5.

To determine the protein concentration a standard curve was created using 5 cuvettes within 1ml Bradford solution and the following amounts of BSA (bovine serum albumin): 0 (baseline), 1.25, 2.5, 5, 10 micrograms. In an additional cuvette 1ml Bradford solution was mixed with 1µl of the protein, of which concentration was going to be determined. Finally a spectrophotometer was used to measure the OD (optical density) at 595nm.

Results

In this work, nuclear magnetic resonance spectroscopy (NMR) (implemented by Pierre Barraud, in the laboratory of Prof.Dr. Frederic Allain, from the institute for molecular biology and biophysics in Zurich) was used to obtain the resolution of the three dimensional structure of the protein which in contrast to X- Ray, is applied mainly for small proteins, usually smaller than 35kDa. The application of NMR requires, protein sample to be well-behaved and non-aggregated sample to be at an adequate high protein concentration. Since the whole dsRBD3 protein is only partially soluble and not nearly sufficient for applying NMR, constructs with different lengths were made trying to include in the protein sequence as few as possible nonpolar (hydrophobic) aminoacids. Additionally different buffers, pH values, temperatures and incubations times were tested to find out the best conditions on which the protein behaves at best for NMR purpose.

Purification of His-tagged dsRBD3

The first 36 constructs (1-36) and the four following ones termed A, B, C, and E of dsRBD3 were recombinantly expressed in pRSETA with a His-Tag which was fused to the protein C-terminal. A Ni-NTA column was used for affinity purification. The constructs were made so that the first six have the same N-Terminus but with a progressively extending C-Terminus. The next six have a longer N-Terminus compared to the first 6 constructs but the C-Terminus has the same progressive extension, and so on. The constructs 1-36 are made the way shown schematically below.

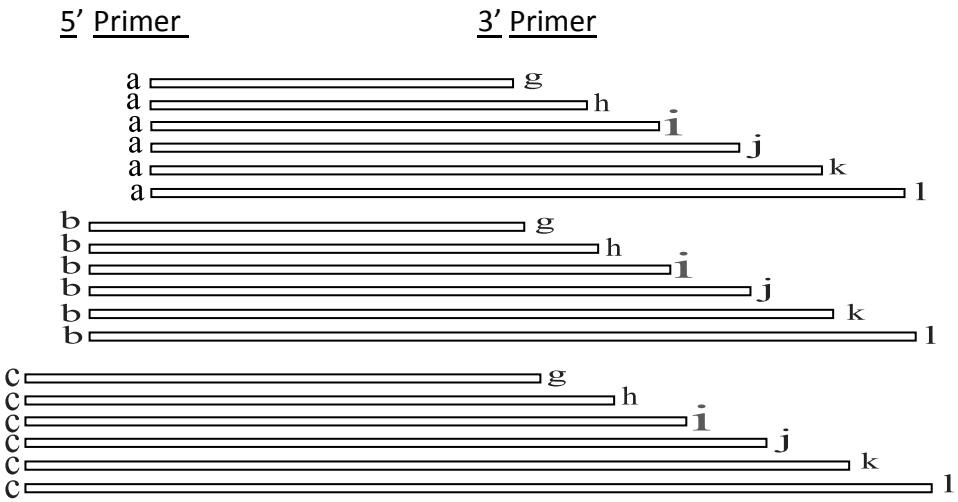


Figure 6: Constructs 1-36 of dsRBD3 done extending either the N- or the C-Terminus

Additionally, the length of the constructs termed A, B, C and E show a relation as in the scheme below.

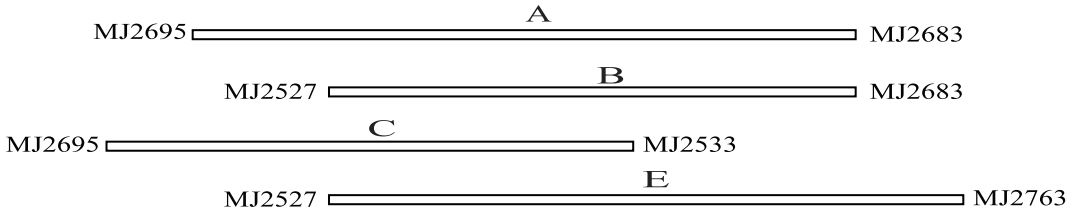


Figure 7: Schema of dsRBD3 constructs termed A, B, C and E

.

After protein expression as described above, the cell pellet was lysed in three steps with lyses buffer containing first 10mM NaCl, hepes pH=7.4, 0.1% triton, then 100mM NaCl, hepes (pH=7.4), 0.1% Triton, and finally 1M NaCl, hepes (pH=7.4), 0.1% Triton. Between the steps the lysate was sonicated for 4 minutes and centrifuged for 5 minutes at 14000rpm to obtain the pellet for the next lyses step. The supernatant gained at the end was washed and eluted from the Ni-NTA column, using for the washing 100mM imidazol, 1XPBS (1.37M NaCl, 27mM KCl, 43mM Na2PO4, 14mM KH2PO4, pH=7.4), 1%triton, and for elution 500mM imidazol, 1XPBS and

1% triton. Initially Flow pressure liquid chromatography (FPLC) at 4°C was performed. Trying to improve the conditions, the lysate was in the further experiments manually loaded onto the Ni-NTA column at room temperature showing that experiments carried out in this temperature gives better results as at a lower temperature of 4°C.

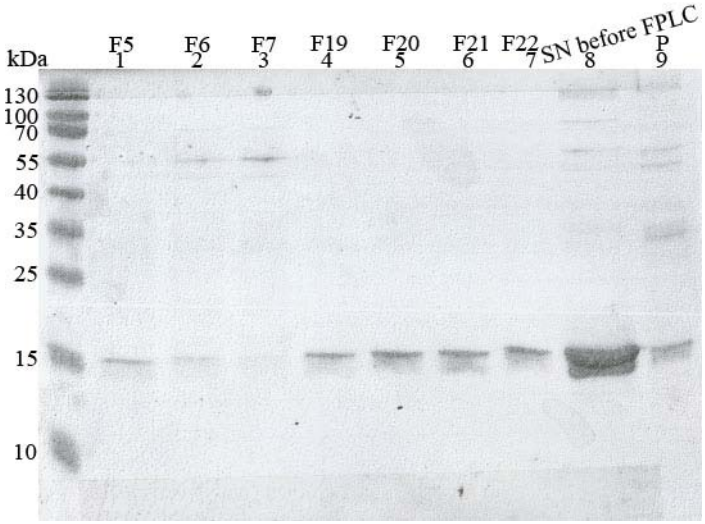


Figure 8: Purification of His- tagged construct number 13 using FPLC at 4°C

Shown are FPLC- fractions (lane 1-7), supernatant before loading into FPLC (lane 9) and pellet (lane 10) before purification.

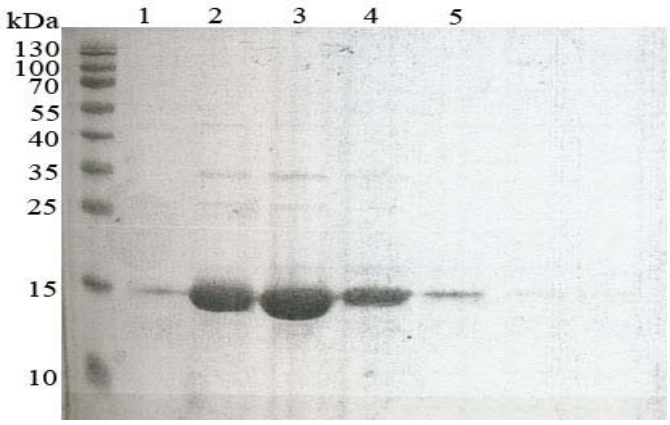


Figure 9: Purification of His-tagged construct number 13 using Ni-NTA column manually at room temperature

After elution with 500mM Imidazol 10µl of each fraction are loaded.

To optimize the expression level of the protein, expression-time and induction-period were among others also tested (constructs lettered A, B und C).

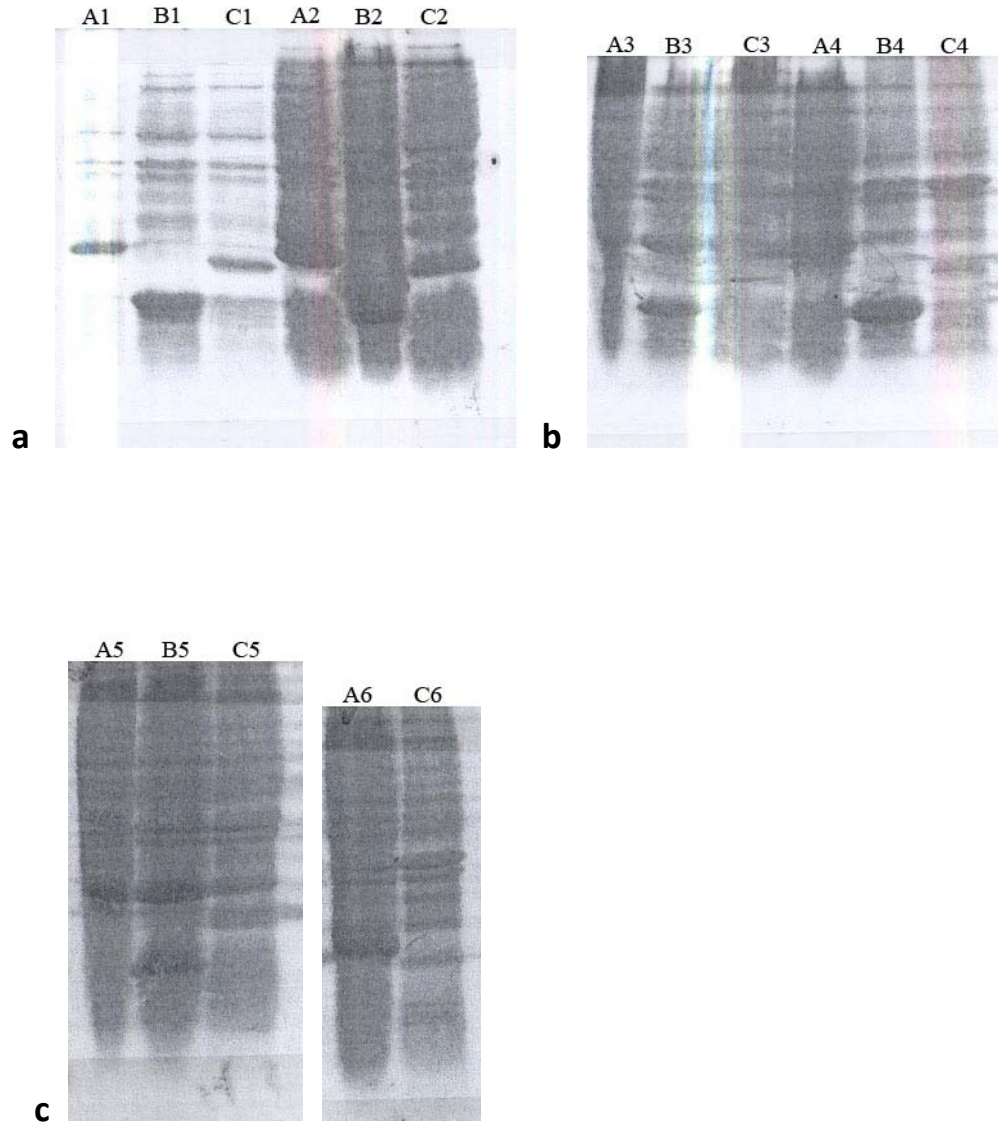


Figure 10: Different incubation and induction times tested on construct A, B and C.

(**a**) A1, B1, C1 cultures were incubated for 4 hours and inducted for 2 hours, whether A2, B2, C2 cultures were incubated for 6 hours and inducted for 2 hours. (**b**) A3, B3, C3 cultures were inducted overnight after 4 hours incubation, whether A4, B4, C4 were also inducted overnight but incubated for 6 hours. (**c**) A5, B5, C5 and A6, C6 were

incubated overnight, but the first three were inducted for 2 hours whether the last two for 4 hours.

We performed the cell growth and protein induction almost always for about 7 hours incubation and 3 hours induction when cells were grown in TY medium. Cells grown in minimal medium (M9) need a much longer incubation time of about 24 hours to reach an optical density of about 0.7.

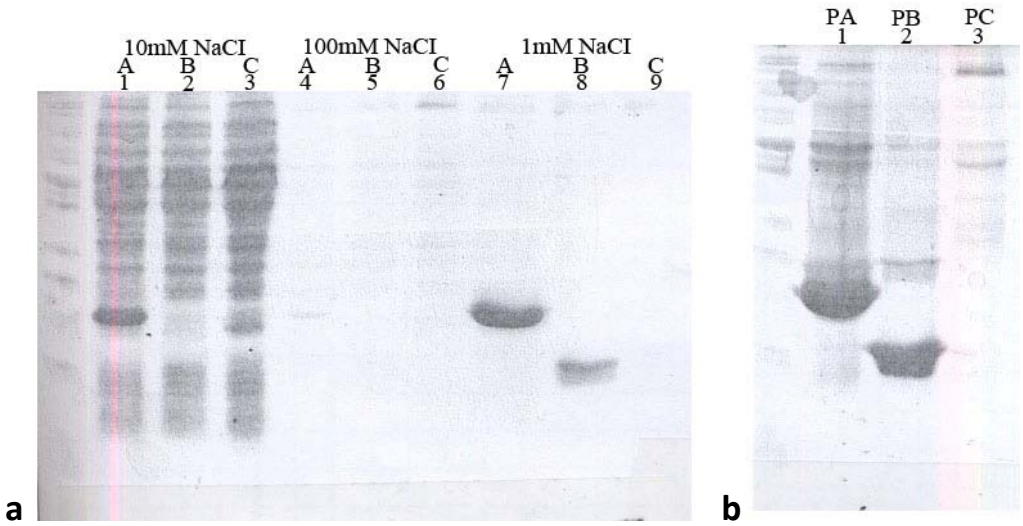


Figure 11: Solubility test of A, B, C- constructs.

(a) Lanes 1-3 show the supernatant of constructs A,B,C after lysis with 10mM NaCl, hepes pH=7.4, 0.1% triton whether in the lanes 4-6 are loaded supernatants after lysis with 100mM NaCl, hepes pH=7.4, 0.1% triton and in lanes 7-9 after lysis with 1mM NaCl, hepes pH=7.4, 0.1% triton. (b) Lanes 1-3 show the pellet remained after lysis with 1mM NaCl, hepes pH=7.4, 0.1% triton. As shown in lanes 7-9 (a) the best result is obtained from the lysis with the buffer containing 1mM NaCl.

Purification of GB1-tagged dsRBD3

GB1 (immunoglobulin G binding domain 1 of Streptococcal protein G) is a solubility enhancement tag (SET), a 56 amino acid protein highly stable and soluble.

We apply GB1- tag because the most known protein tags, such as MBP (Maltose Binding Protein) or GST (Gluthation-S-Transferase) have a large size, which hinders the structural characterization of a fusion protein by NMR. Additionally cleavage of the fusion proteins often re-introduces problems with solubility and stability (Zhou, Lugovskoy et al. 2001).

The two best dsRBD3 constructs among the forty described above, construct 13 and B were fused with GB1-tag, and termed N13 and NB.

Cell-pellets containing this constructs were lysed with a different buffer as the constructs described above, which contains KI instead of NaCl. The washing buffer contains also KI instead of NaCl. After lysis with 150mM KI, 10mM hepes pH=7.4, 0,1% triton and centrifugation, supernatant was manually loaded into the Ni-NTA column and after two steps washing with 10mM imidazol, 150mM KI, 0.1% Triton and 50mM imidazol, 150mM KI, 0.1% triton, was eluted with 500mM imidazol, 150mM KI, 0.1% triton.

After that, dialysis was performed for 4 hours. Dialysis solution contained: 150mM KI, 20mM K2HPO4 + KH2PO4 (pH=7,2), 1mM DTT (Dithiothreitol).

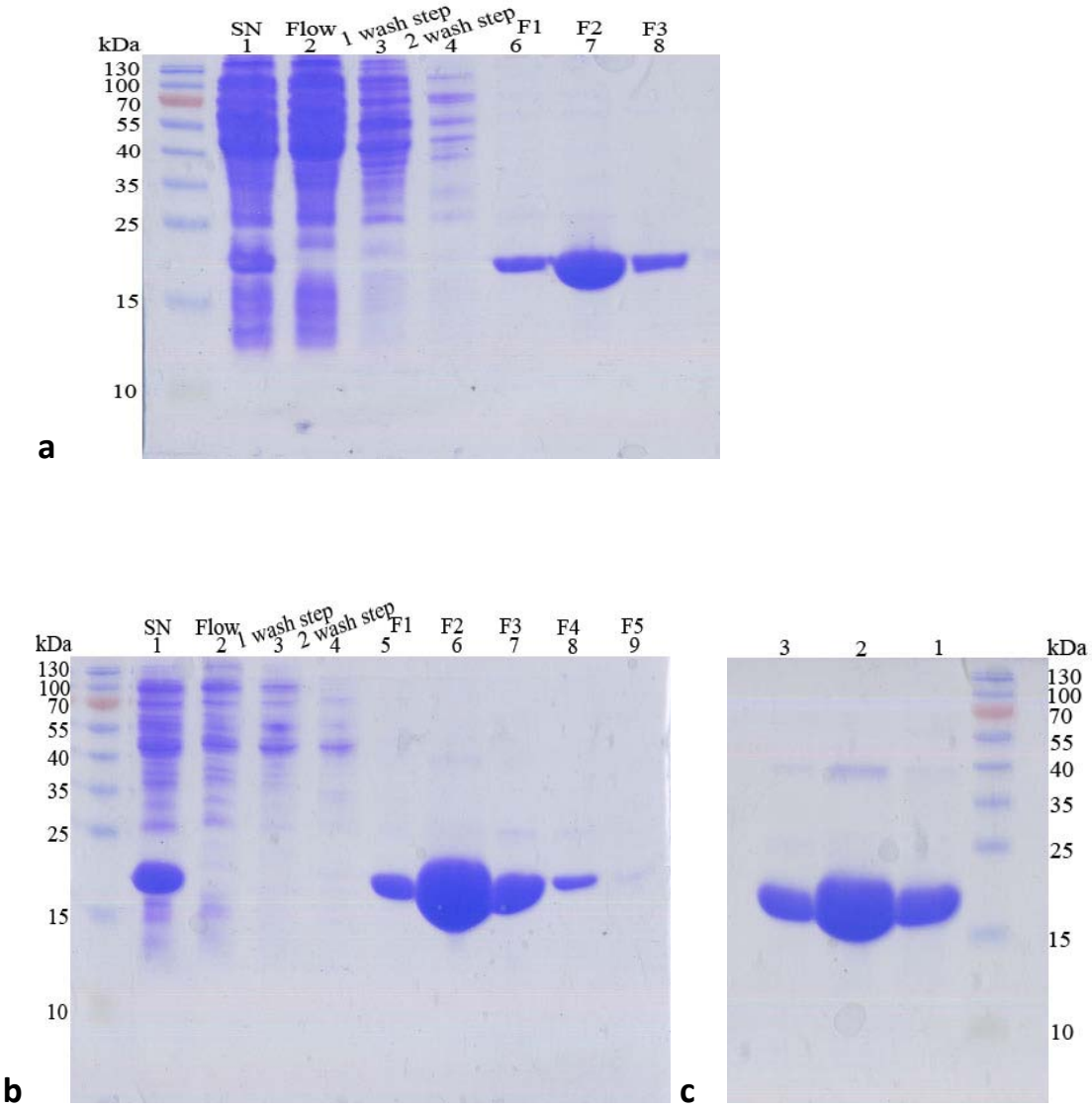


Figure 12: Purification of clones grown in different media and dialysis.
(a) Purification of clone N13, grown in TY rich medium. Shown are: supernatant before loading into the column (lane 1), flow (lane2), first wash step with 10mM imidazol (lane 3), second wash step with 100mM imidazol (lane4), N13 fractions (lane 6-8). **(b) Purification of clone N13, grown in M9 minimal medium**. Shown are like in figure (a), supernatant, flow, washing steps and the fraction after elution. **(c) Fractions after dialysis**

The last experiments were performed using the lysis buffer mentioned in the materials and method part. This buffer contained 20mM KH2PO4 +

K2HPO4 (pH=7,2) instead of hepes and 1mM DTT was added. Dialysis was carried out with a gradient of guanidine hydrochloride in addition to 150mM KI, 20mM KH2PO4 + K2HPO4 (pH=7,2), 1mM DTT.

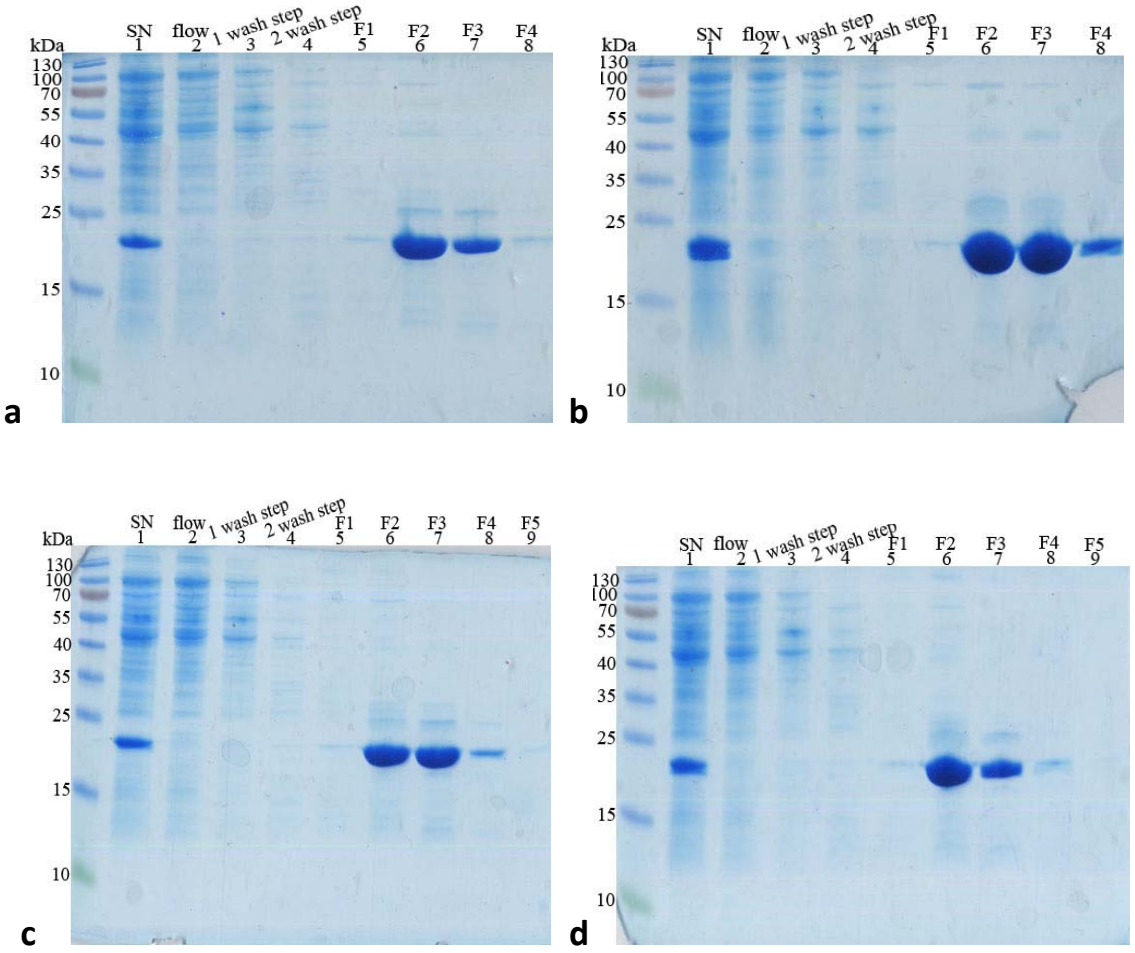


Figure 13: Purificaton of different clones of construst NB.

In all pictures is shown: supernatant before purification (lane 1), flow (lane2), first and second wash step (lane 3 and 4), eluted fractions (lane 5-9)

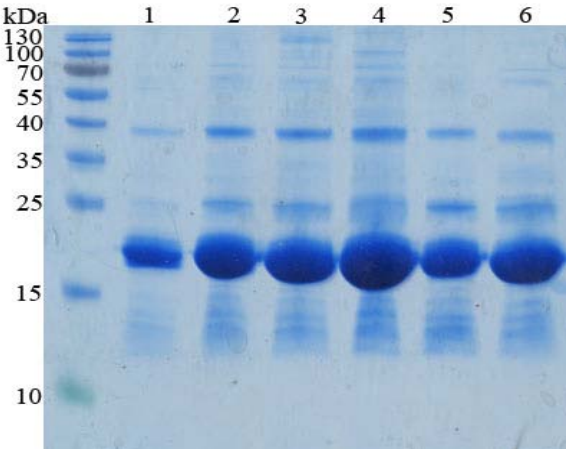
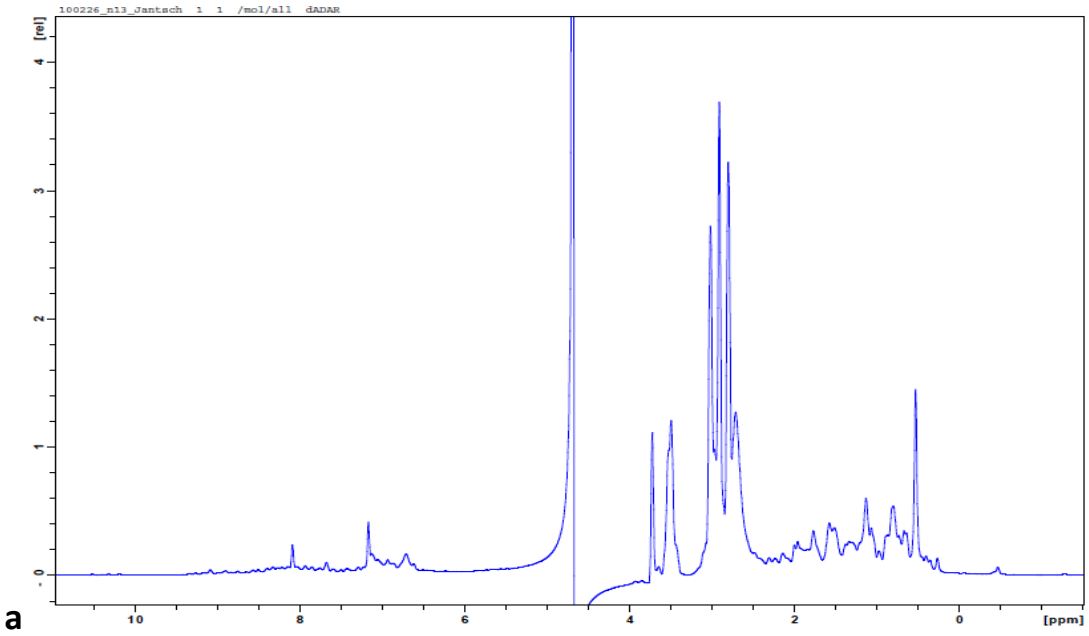


Figure 14: Clone N13 after dialysis.

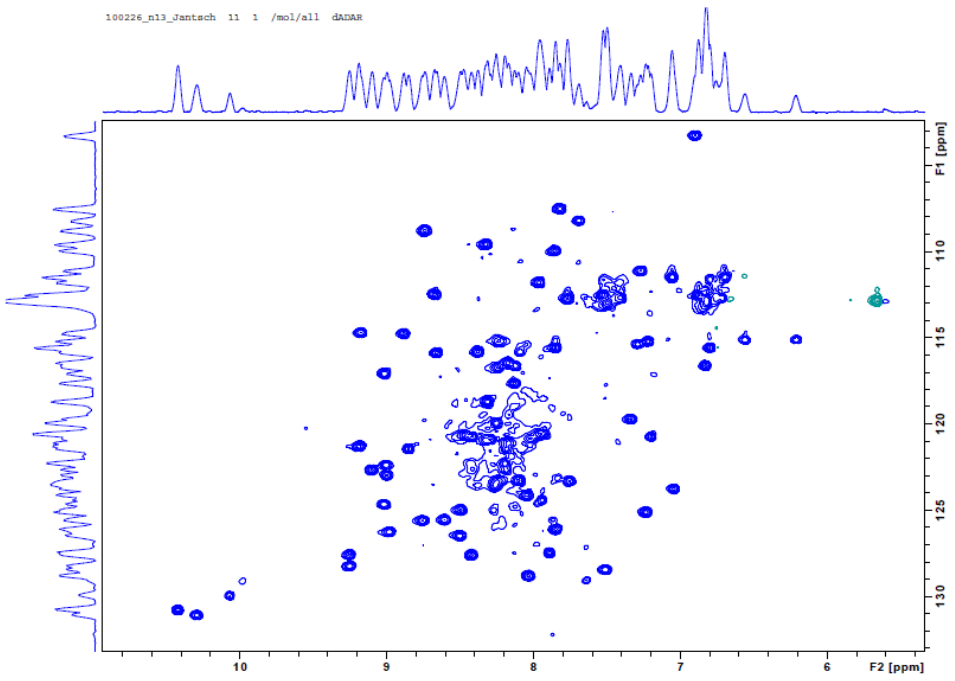
Fractions with a large amount of protein were singly dialysed.

Nuclear magnetic resonance (NMR) spectroscopy

After dialysis the protein concentration was measured: the highest protein concentration reached was 6,4 µg/µl. Further on the protein-samples were send to institute for molecular biology and biophysics (Prof.Dr. Frederic Allain, Pierre Barraud) and nuclear magnetic resonance spectroscopy was applied.


100226_n13_Jantsch 1 1 /mol/all dADAR
[rel]
4
3
2
1
0
10
8
6
4
2
0
[ppm]

a


100226_n13_Jantsch 11 1 /mol/all dADAR
F1 [ppm]
110
115
120
125
130
10
9
8
7
6
F2 [ppm]

b

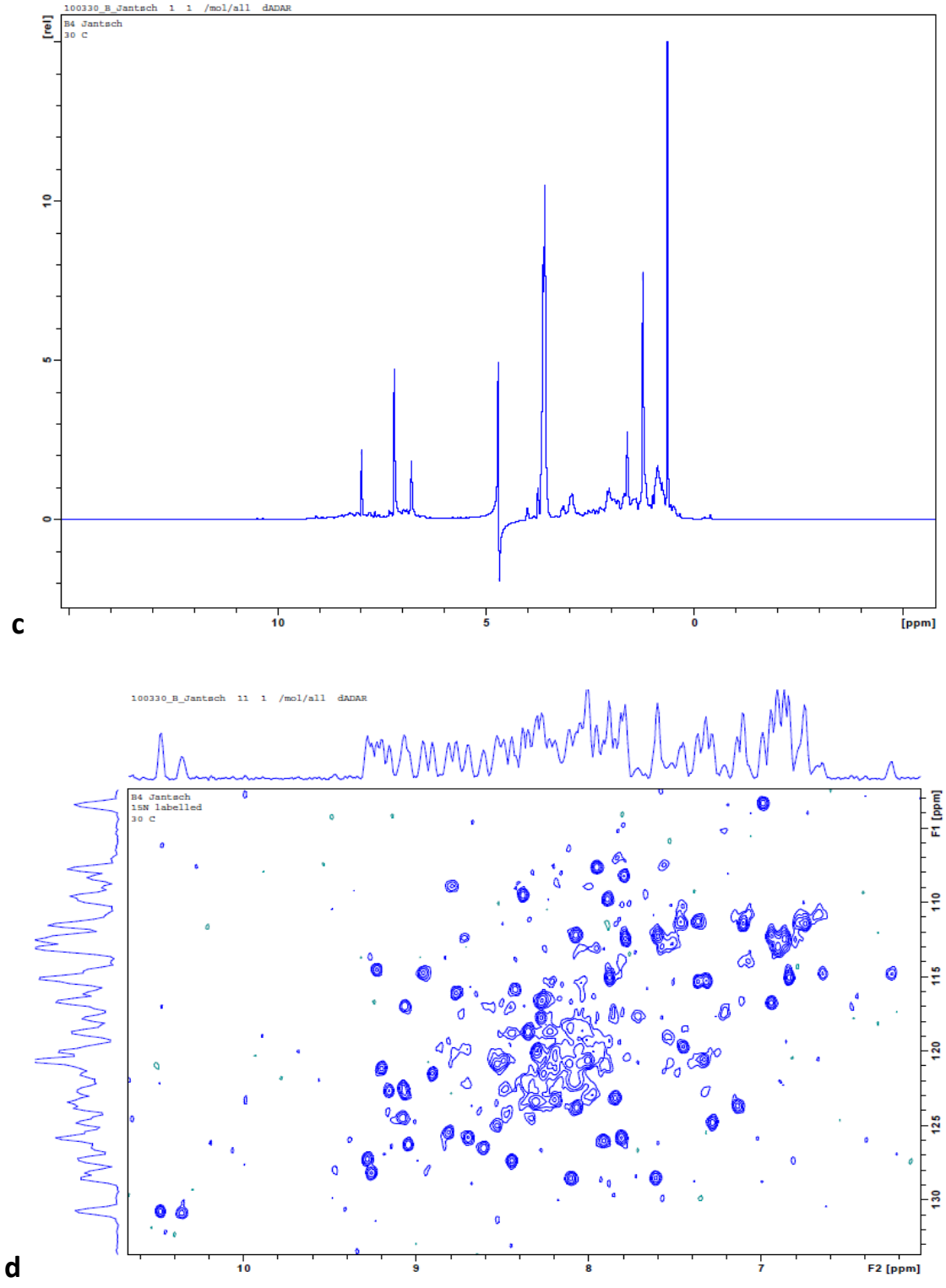


Figure 15: NMR-spectroscopy spectra.

(**a**) 1D spectrum of clone N13 (**b**) 2D 15N-1H-HSQC spectrum of clone N13 (**c**) 1D spectrum of clone NB (**d**) 2D 15N-1H-HSQC spectrum of clone NB

On the first 1D spectrum we can see buffer lines and therefore we changed the buffer conditions and switched from hepes to potassium phosphate-buffer. On the 2D spectra there are about 60 amide signals coming from the backbone of GB1 tag, and additional peaks coming from the GB1 side chain. From the domain come only some additional weak peaks. It could be that there was an aggregation problem or the domain was not folded, or both, not folded and aggregated. Therefore further experiments are necessary to obtain the 3D structure of this protein.

Nucleotide sequence of construct N13 (495bp)

ATGCAGTACAAACTGATCCTGAACGGTAAAACCCTGAAAGGTGAAAC
CACCACCGAAGCTGTTGACGCTGCTACCGCGGAAAAAGTTTTCAAAC
AGTACGCTAACGACAACGGTGTTGACGGTGAATGGACCTACGACGA
CGCTACCAAAACCTTCACCGTTACCGAAGGATCCAACTTGGAATCCA
TGATGCCCAACAAGGTCAGGAAGATTGGCGAACTCGTGAGATACCTG
AACACCAACCCTGTGGGTGGCCTTTTGGAGTACGCCCGCTCCCATGG
CTTTGCTGCTGAATTCAAGTTGGTCGACCAGTCCGGACCTCCCCACG
AGCCCAAGTTCGTTTACCAAGCAAAAGTTGGGGGTCGCTGGTTCCCA
GCCGTCTGCGCACACAGCAAGAAGCAAGGCAAGCAGGAAGCAGCAG
ATGCGGCTCTCCGTGTCTTGATTGGGGAGAACGAGAAGGCAGAACG
CGGATCCCACCACCACCACCACCACTAA

Amino acid sequence of construct N13

MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYD
DATKTFTVTEGSNLESMMPNKVRKIGELVRYLNTNPVGGLLEYARSH
GFAAEFKLVDQSGPPHEPKFVYQAKVGGRWFPAVCAHSKKQGKQEA
ADAALRVLIGENEKAERGSHHHHHH*

Nucleotide sequence of construct NB (384bp)

ATGATGCCCAACAAGGTCAGGAAGATTTGCGAACTCGTGAGATACCTG
AACCCCAACCCTGTGGGTGGCCTTTTGGAGTACGCCCGCTCCCATGGC

TTTGGTGCTGAATTCAAGTTGGTCGACCAGTCCGGACCTCCTCACGAGC
CCAAGTTCGTTTACCAAGCAAAAGTTGGGGGTCGCTGGTTCCCAGCCG
TCTGCGCACACAGCAAGAAGCAAGGCAAGCAGGAAGCAGCAGATGCGG
CTCTCCGTGTCTTGATTGGGGAGAACGAGAAGGCAGAACGCATGGGTT
TCACAGAGGTAACCCCAGTGACAGGGGCCAGTCTCAGAAGAACTATGC
TCCTCCTCTCAAGGTCCCCAGGATCCCACCACCACCACCACCACTAA

Amino acid sequence of construct NB

MMPNKVRKICELVRYLNPNPVGGLLEYARSHGFGAEFKLVDQSGPPHE
PKFVYQAKVGGRWFPAVCAHSKKQGKQEAADAALRVLIGENEKAERMG
FTEVTPVTGASLRRTMLLLSRSPGSHHHHHH*

Discussion

A range of organisms from E.coli to human have proteins which contain double stranded RNA binding domains (dsRBDs). They are involved in different cellular processes such as RNA-processing, RNA-editing or RNA-localisation. On the one hand they bind specifically to A-helices formed by dsRNAs or RNA-DNA hybrids, on the other hand they can also interact with different cellular factors in an RNA-independent manner (St Johnston, Brown et al. 1992; Bevilacqua and Cech 1996). It has been shown that dsRBDs can influence translational control and also the cellular distribution of their respective protein (Doyle and Jantsch 2002; Saunders and Barber 2003)

The third double stranded RNA binding domain (dsRBD3) of human ADAR1 is overlapped by a nuclear localization signal and the NLS activity displayed by this domain does not depend on RNA binding indicating a dual function for this domain (Eckmann, Neunteufl et al. 2001). ADAR1c has the ability to shuttle between the nucleus and cytoplasm even though it lacks a nuclear export signal. Transportin-1 (TRN1) was identified as the import receptor for ADAR1 and it was shown that TRN1 binds hADAR1 via its third dsRBD and that the dsRNA binding interfere with transportin-1 binding (Fritz, Strehblow et al. 2009). The interaction between export factor exportin-5 and each dsRBD in ADAR1 is stimulated by RNA-binding and might control nuclear export while dsRBD3 can mediate nuclear import (Fritz, Strehblow et al. 2009). It seems that after ADAR1 export to the cytoplasm, the nuclear import can occur only after the dissociation of the bound RNA which allows TRN1 to bind to the NLS in dsRBD3.

Finding out the three dimensional structure of the dsRBD3 of hADAR1, and furthermore that of this domain bound to RNA could provide more information about their interaction and the nuclear transport. Information about the biological function and the mechanisms linked to the functions as well as information on relationships between macromolecules could also be acquired.

Appendix

Sequences of the constructs 1-36 and A, B, C, E

Construct 1 (287bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S

 91/31                                                               121/41
151/51

 |                                                                    |
|

cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R

 181/61                                                              211/71
241/81

 |                                                                    |
|

tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E

271/91

 |

 aag gca gaa cgc atg gg

 K   A   E   R   M
```

Construct 2 (297bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S



 91/31                                                                 121/41
151/51

 |                                                                      |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R



 181/61                                                                211/71
241/81

 |                                                                      |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E



 271/91

 |

 aag gca gaa cgc atg ggt ttc aca gag

 K   A   E   R   M   G   F   T   E
```

Construct 3 (306bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

 M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S

91/31                                                                121/41
151/51

 |                                                                    |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R


 181/61                                                               211/71
241/81

 |                                                                    |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E


 271/91                                 301/101

 |                                      |

 aag gca gaa cgc atg ggt ttc aca gag gta acc cca

 K   A   E   R   M   G   F   T   E   V   T   P
```

Construct 4 (314)

```
1/1                                                                 31/11
61/21

 |                                                                  |
|

 atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

 M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S


 91/31                                                              121/41
151/51

 |                                                                  |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R


 181/61                                                             211/71
241/81

 |                                                                  |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E


 271/91                                 301/101

 |                                      |

 aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca g

 K   A   E   R   M   G   F   T   E   V   T   P   V   T
```

Construct 5 (324bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

 M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S


 91/31                                                               121/41
151/51

 |                                                                    |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R


 181/61                                                              211/71
241/81

 |                                                                    |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E


 271/91                                 301/101

 |                                      |

 aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc agt ctc

 K   A   E   R   M   G   F   T   E   V   T   P   V   T   G   A   S   L
```

Construct 6 (331bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

 atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

 M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S


 91/31                                                                121/41
151/51

 |                                                                      |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R


 181/61                                                               211/71
241/81

 |                                                                      |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E


 271/91                                 301/101

 |                                      |

 aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc agt ctc aga
aga a

 K   A   E   R   M   G   F   T   E   V   T   P   V   T   G   A   S   L   R   R
```

Construct 7 (291bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac
acc aac cct gtg ggt ggc ctt ttg gag tac gcc

 E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T
N   P   V   G   G   L   L   E   Y   A



 91/31                                                               121/41
151/51

 |                                                                    |
|

 cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac
gag ccc aag ttc gtt tac caa gca aaa gtt ggg

 R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E
P   K   F   V   Y   Q   A   K   V   G



 181/61                                                              211/71
241/81

 |                                                                    |
|

 ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca
gca gat gcg gct ctc cgt gtc ttg att ggg gag

 G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A
D   A   A   L   R   V   L   I   G   E



 271/91

 |

 aac gag aag gca gaa cgc atg

 N   E   K   A   E   R   M
```

Construct 8 (295bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

 M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S



 91/31                                                               121/41
151/51

 |                                                                    |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R



 181/61                                                              211/71
241/81

 |                                                                    |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E



 271/91

 |

 aag gca gaa cgc atg ggt ttc aca

 K   A   E   R   M   G   F   T
```

Construct 9 (313bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac
acc aac cct gtg ggt ggc ctt ttg gag tac gcc

 E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T
N   P   V   G   G   L   L   E   Y   A



 91/31                                                               121/41
151/51

 |                                                                    |
|

 cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac
gag ccc aag ttc gtt tac caa gca aaa gtt ggg

 R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E
P   K   F   V   Y   Q   A   K   V   G



 181/61                                                              211/71
241/81

 |                                                                    |
|

 ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca
gca gat gcg gct ctc cgt gtc ttg att ggg gag

 G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A
D   A   A   L   R   V   L   I   G   E



 271/91                                 301/101

 |                                      |

 aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca

 N   E   K   A   E   R   M   G   F   T   E   V   T   P
```

Construct 10 (325bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac
acc aac cct gtg ggt ggc ctt ttg gag tac gcc

 E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T
N   P   V   G   G   L   L   E   Y   A



 91/31                                                               121/41
151/51

 |                                                                    |
|

 cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac
gag ccc aag ttc gtt tac caa gca aaa gtt ggg

 R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E
P   K   F   V   Y   Q   A   K   V   G



 181/61                                                              211/71
241/81

 |                                                                    |
|

 ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca
gca gat gcg gct ctc cgt gtc ttg att ggg gag

 G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A
D   A   A   L   R   V   L   I   G   E



 271/91                                 301/101

 |                                      |

 aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc

 N   E   K   A   E   R   M   G   F   T   E   V   T   P   V   T   G   A
```

Construct 11 (331bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac
acc aac cct gtg ggt ggc ctt ttg gag tac gcc

 E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T
N   P   V   G   G   L   L   E   Y   A



 91/31                                                               121/41
151/51

 |                                                                    |
|

 cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac
gag ccc aag ttc gtt tac caa gca aaa gtt ggg

 R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E
P   K   F   V   Y   Q   A   K   V   G



 181/61                                                              211/71
241/81

 |                                                                    |
|

 ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca
gca gat gcg gct ctc cgt gtc ttg att ggg gag

 G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A
D   A   A   L   R   V   L   I   G   E



 271/91                              301/101

 |                                   |

 aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc agt
ctc

 N   E   K   A   E   R   M   G   F   T   E   V   T   P   V   T   G   A   S   L
```

Construct 12 (338bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac
acc aac cct gtg ggt ggc ctt ttg gag tac gcc

 E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T
N   P   V   G   G   L   L   E   Y   A



 91/31                                                               121/41
151/51

 |                                                                    |
|

 cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac
gag ccc aag ttc gtt tac caa gca aaa gtt ggg

 R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E
P   K   F   V   Y   Q   A   K   V   G



 181/61                                                              211/71
241/81

 |                                                                    |
|

 ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca
gca gat gcg gct ctc cgt gtc ttg att ggg gag

 G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A
D   A   A   L   R   V   L   I   G   E



 271/91                                                              301/101
331/111

 |                                                                    |
|

 aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc agt
ctc aga aga a

 N   E   K   A   E   R   M   G   F   T   E   V   T   P   V   T   G   A   S   L
R   R
```

Construct 13 (298bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac
ctg aac acc aac cct gtg ggt ggc ctt ttg gag

 N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L
N   T   N   P   V   G   G   L   L   E



 91/31                                                                121/41
151/51

 |                                                                    |
|

 tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct
cct cac gag ccc aag ttc gtt tac caa gca aaa

 Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P
H   E   P   K   F   V   Y   Q   A   K



 181/61                                                               211/71
241/81

 |                                                                    |
|

 gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag
gaa gca gca gat gcg gct ctc cgt gtc ttg att

 V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E
A   A   D   A   A   L   R   V   L   I



 271/91

 |

 ggg gag aac gag aag gca gaa cgc atg

 G   E   N   E   K   A   E   R   M
```

Construct 14 (307bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac
ctg aac acc aac cct gtg ggt ggc ctt ttg gag

 N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L
N   T   N   P   V   G   G   L   L   E



 91/31                                                                121/41
151/51

 |                                                                    |
|

 tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct
cct cac gag ccc aag ttc gtt tac caa gca aaa

 Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P
H   E   P   K   F   V   Y   Q   A   K



 181/61                                                               211/71
241/81

 |                                                                    |
|

 gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag
gaa gca gca gat gcg gct ctc cgt gtc ttg att

 V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E
A   A   D   A   A   L   R   V   L   I



 271/91                              301/101

 |                                   |

 ggg gag aac gag aag gca gaa cgc atg ggt ttc aca

 G   E   N   E   K   A   E   R   M   G   F   T
```

Construct 15 (316bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac
ctg aac acc aac cct gtg ggt ggc ctt ttg gag

 N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L
N   T   N   P   V   G   G   L   L   E



 91/31                                                                121/41
151/51

 |                                                                    |
|

 tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct
cct cac gag ccc aag ttc gtt tac caa gca aaa

 Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P
H   E   P   K   F   V   Y   Q   A   K



 181/61                                                               211/71
241/81

 |                                                                    |
|

 gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag
gaa gca gca gat gcg gct ctc cgt gtc ttg att

 V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E
A   A   D   A   A   L   R   V   L   I



 271/91                                 301/101

 |                                      |

 ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc

 G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
```

Construct 16 (325bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

 aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac
ctg aac acc aac cct gtg ggt ggc ctt ttg gag

 N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L
N   T   N   P   V   G   G   L   L   E


 91/31                                                                 121/41
151/51

 |                                                                      |
|

 tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct
cct cac gag ccc aag ttc gtt tac caa gca aaa

 Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P
H   E   P   K   F   V   Y   Q   A   K


 181/61                                                                211/71
241/81

 |                                                                      |
|

 gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag
gaa gca gca gat gcg gct ctc cgt gtc ttg att

 V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E
A   A   D   A   A   L   R   V   L   I


 271/91                                 301/101

 |                                      |

 ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca

 G   E   N   E   K   A   E   R   M   G   F   T   E   V   T   P   V   T
```

Construct 17 (334)

aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac cct gtg ggt ggc ctt ttg gag

N L E S M M P N K V R K I G E L V R Y L N T N P V G G L L E

tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc aag ttc gtt tac caa gca aaa

Y A R S H G F A A E F K L V D Q S G P P H E P K F V Y Q A K

gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat gcg gct ctc cgt gtc ttg att

V G G R W F P A V C A H S K K Q G K Q E A A D A A L R V L I

ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc agt

G E N E K A E R M G F T E V T P V T G A S

Construct 18 (340bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac
ctg aac acc aac cct gtg ggt ggc ctt ttg gag

 N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L
N   T   N   P   V   G   G   L   L   E


 91/31                                                                121/41
151/51

 |                                                                    |
|

 tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct
cct cac gag ccc aag ttc gtt tac caa gca aaa

 Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P
H   E   P   K   F   V   Y   Q   A   K


 181/61                                                               211/71
241/81

 |                                                                    |
|

 gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag
gaa gca gca gat gcg gct ctc cgt gtc ttg att

 V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E
A   A   D   A   A   L   R   V   L   I


 271/91                                                               301/101
331/111

 |                                                                    |
|

 ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg
gcc agt ctc aga

 G   E   N   E   K   A   E   R   M   G   F   T   E   V   T   P   V   T   G   A
S   L   R
```

Construct 19 (301bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg
aga tac ctg aac acc aac cct gtg ggt ggc ctt

 L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R
Y   L   N   T   N   P   V   G   G   L


 91/31                                                                121/41
151/51

 |                                                                    |
|

 ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc
gga cct cct cac gag ccc aag ttc gtt tac caa

 L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G
P   P   H   E   P   K   F   V   Y   Q


 181/61                                                               211/71
241/81

 |                                                                    |
|

 gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc
aag cag gaa gca gca gat gcg gct ctc cgt gtc

 A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K
Q   E   A   A   D   A   A   L   R   V


 271/91

 |

 ttg att ggg gag aac gag aag gca gaa cgc

 L   I   G   E   N   E   K   A   E   R
```

Construct 20 (313bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg
aga tac ctg aac acc aac cct gtg ggt ggc ctt

 L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R
Y   L   N   T   N   P   V   G   G   L


 91/31                                                               121/41
151/51

 |                                                                    |
|

 ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc
gga cct cct cac gag ccc aag ttc gtt tac caa

 L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G
P   P   H   E   P   K   F   V   Y   Q


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc
aag cag gaa gca gca gat gcg gct ctc cgt gtc

 A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K
Q   E   A   A   D   A   A   L   R   V


 271/91                                 301/101

 |                                      |

 ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca

 L   I   G   E   N   E   K   A   E   R   M   G   F   T
```

Construct 21 (322bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

 ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg
aga tac ctg aac acc aac cct gtg ggt ggc ctt

 L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R
Y   L   N   T   N   P   V   G   G   L


 91/31                                                                 121/41
151/51

 |                                                                      |
|

 ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc
gga cct cct cac gag ccc aag ttc gtt tac caa

 L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G
P   P   H   E   P   K   F   V   Y   Q


 181/61                                                                211/71
241/81

 |                                                                      |
|

 gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc
aag cag gaa gca gca gat gcg gct ctc cgt gtc

 A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K
Q   E   A   A   D   A   A   L   R   V


 271/91                                 301/101

 |                                      |

 ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc

 L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
```

Construct 22 (330bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

 ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg
aga tac ctg aac acc aac cct gtg ggt ggc ctt

 L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R
Y   L   N   T   N   P   V   G   G   L


 91/31                                                                121/41
151/51

 |                                                                      |
|

 ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc
gga cct cct cac gag ccc aag ttc gtt tac caa

 L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G
P   P   H   E   P   K   F   V   Y   Q


 181/61                                                               211/71
241/81

 |                                                                      |
|

 gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc
aag cag gaa gca gca gat gcg gct ctc cgt gtc

 A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K
Q   E   A   A   D   A   A   L   R   V


 271/91                                 301/101

 |                                      |

 ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg
ac

 L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T   P   V
```

Construct 23 (340bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

 ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg
aga tac ctg aac acc aac cct gtg ggt ggc ctt

 L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R
Y   L   N   T   N   P   V   G   G   L


 91/31                                                                121/41
151/51

 |                                                                      |
|

 ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc
gga cct cct cac gag ccc aag ttc gtt tac caa

 L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G
P   P   H   E   P   K   F   V   Y   Q


 181/61                                                               211/71
241/81

 |                                                                      |
|

 gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc
aag cag gaa gca gca gat gcg gct ctc cgt gtc

 A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K
Q   E   A   A   D   A   A   L   R   V


 271/91                                                               301/101
331/111

 |                                                                      |
|

 ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg
aca ggg gcc agt

 L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T   P   V   T
G   A   S
```

Construct 24 (346bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg
aga tac ctg aac acc aac cct gtg ggt ggc ctt

 L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L   V   R
Y   L   N   T   N   P   V   G   G   L


 91/31                                                               121/41
151/51

 |                                                                    |
|

 ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc
gga cct cct cac gag ccc aag ttc gtt tac caa

 L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q   S   G
P   P   H   E   P   K   F   V   Y   Q


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc
aag cag gaa gca gca gat gcg gct ctc cgt gtc

 A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q   G   K
Q   E   A   A   D   A   A   L   R   V


 271/91                                                              301/101
331/111

 |                                                                    |
|

 ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg
aca ggg gcc agt ctc aga

 L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T   P   V   T
G   A   S   L   R
```

Construct 25 (307bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag
ctc gtg aga tac ctg aac acc aac cct gtg ggt

 E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L
V   R   Y   L   N   T   N   P   V   G



 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac
cag tcc gga cct cct cac gag ccc aag ttc gtt

 G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q
S   G   P   P   H   E   P   K   F   V



 181/61                                                              211/71
241/81

 |                                                                    |
|

 tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag
caa ggc aag cag gaa gca gca gat gcg gct ctc

 Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q
G   K   Q   E   A   A   D   A   A   L



 271/91                                 301/101

 |                                      |

 cgt gtc ttg att ggg gag aac gag aag gca gaa cgc

 R   V   L   I   G   E   N   E   K   A   E   R
```

Construct 26 (319bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag
ctc gtg aga tac ctg aac acc aac cct gtg ggt

 E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L
V   R   Y   L   N   T   N   P   V   G



 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac
cag tcc gga cct cct cac gag ccc aag ttc gtt

 G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q
S   G   P   P   H   E   P   K   F   V



 181/61                                                              211/71
241/81

 |                                                                    |
|

 tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag
caa ggc aag cag gaa gca gca gat gcg gct ctc

 Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q
G   K   Q   E   A   A   D   A   A   L



 271/91                                 301/101

 |                                      |

 cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca

 R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T
```

Construct 27 (328bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag
ctc gtg aga tac ctg aac acc aac cct gtg ggt

 E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L
V   R   Y   L   N   T   N   P   V   G



 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac
cag tcc gga cct cct cac gag ccc aag ttc gtt

 G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q
S   G   P   P   H   E   P   K   F   V



 181/61                                                              211/71
241/81

 |                                                                    |
|

 tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag
caa ggc aag cag gaa gca gca gat gcg gct ctc

 Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q
G   K   Q   E   A   A   D   A   A   L



 271/91                                 301/101

 |                                      |

 cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc

 R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
```

Construct 28 (329bp)

```
1/1                                                                 31/11
61/21

 |                                                                  |
|

 gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag
ctc gtg aga tac ctg aac acc aac cct gtg ggt

 E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L
V   R   Y   L   N   T   N   P   V   G


 91/31                                                              121/41
151/51

 |                                                                  |
|

 ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac
cag tcc gga cct cct cac gag ccc aag ttc gtt

 G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q
S   G   P   P   H   E   P   K   F   V


 181/61                                                             211/71
241/81

 |                                                                  |
|

 tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag
caa ggc aag cag gaa gca gca gat gcg gct ctc

 Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q
G   K   Q   E   A   A   D   A   A   L


 271/91                                 301/101

 |                                      |

 cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc c

 R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
```

Construct 29 (346bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

 gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag
ctc gtg aga tac ctg aac acc aac cct gtg ggt

 E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L
V   R   Y   L   N   T   N   P   V   G



 91/31                                                                121/41
151/51

 |                                                                      |
|

 ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac
cag tcc gga cct cct cac gag ccc aag ttc gtt

 G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q
S   G   P   P   H   E   P   K   F   V



 181/61                                                               211/71
241/81

 |                                                                      |
|

 tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag
caa ggc aag cag gaa gca gca gat gcg gct ctc

 Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q
G   K   Q   E   A   A   D   A   A   L



 271/91                                                               301/101
331/111

 |                                                                      |
|

 cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc
cca gtg aca ggg gcc agt

 R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T   P
V   T   G   A   S
```

Construct 30 (352bp)

```
1/1                                                                 31/11
61/21

 |                                                                  |
|

 gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc gag
ctc gtg aga tac ctg aac acc aac cct gtg ggt

 E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E   L
V   R   Y   L   N   T   N   P   V   G



 91/31                                                              121/41
151/51

 |                                                                  |
|

 ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc gac
cag tcc gga cct cct cac gag ccc aag ttc gtt

 G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D   Q
S   G   P   P   H   E   P   K   F   V



 181/61                                                             211/71
241/81

 |                                                                  |
|

 tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag aag
caa ggc aag cag gaa gca gca gat gcg gct ctc

 Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K   Q
G   K   Q   E   A   A   D   A   A   L



 271/91                                                             301/101
331/111

 |                                                                  |
|

 cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta acc
cca gtg aca ggg gcc agt ctc aga

 R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T   P
V   T   G   A   S   L   R
```

Construct 31 (312bp)

```
1/1                                                                  31/11
61/21

 |                                                                    |
|

 tca gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc
gag ctc gtg aga tac ctg aac acc aac cct gtg

 S   E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E
L   V   R   Y   L   N   T   N   P   V


 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggt ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc
gac cag tcc gga cct cct cac gag ccc aag ttc

 G   G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D
Q   S   G   P   P   H   E   P   K   F


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gtt tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag
aag caa ggc aag cag gaa gca gca gat gcg gct

 V   Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K
Q   G   K   Q   E   A   A   D   A   A


 271/91                                 301/101

 |                                      |

 ctc cgt gtc ttg att ggg gag aac gag aag gca gaa cgc at

 L   R   V   L   I   G   E   N   E   K   A   E   R
```

Construct 32 (322bp)

```
1/1                                                                    31/11
61/21

 |                                                                      |
|

 tca gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc
gag ctc gtg aga tac ctg aac acc aac cct gtg

 S   E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E
L   V   R   Y   L   N   T   N   P   V



 91/31                                                                 121/41
151/51

 |                                                                      |
|

 ggt ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc
gac cag tcc gga cct cct cac gag ccc aag ttc

 G   G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D
Q   S   G   P   P   H   E   P   K   F



 181/61                                                                211/71
241/81

 |                                                                      |
|

 gtt tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag
aag caa ggc aag cag gaa gca gca gat gcg gct

 V   Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K
Q   G   K   Q   E   A   A   D   A   A



 271/91                                 301/101

 |                                      |

 ctc cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca

 L   R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T
```

Construct 33 (331bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 tca gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc
gag ctc gtg aga tac ctg aac acc aac cct gtg

 S   E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E
L   V   R   Y   L   N   T   N   P   V


 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggt ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc
gac cag tcc gga cct cct cac gag ccc aag ttc

 G   G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D
Q   S   G   P   P   H   E   P   K   F


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gtt tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag
aag caa ggc aag cag gaa gca gca gat gcg gct

 V   Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K
Q   G   K   Q   E   A   A   D   A   A


 271/91                                 301/101

 |                                      |

 ctc cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta
acc

 L   R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
```

Construct 34 (332bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 tca gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc
gag ctc gtg aga tac ctg aac acc aac cct gtg

 S   E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E
L   V   R   Y   L   N   T   N   P   V


 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggt ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc
gac cag tcc gga cct cct cac gag ccc aag ttc

 G   G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D
Q   S   G   P   P   H   E   P   K   F


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gtt tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag
aag caa ggc aag cag gaa gca gca gat gcg gct

 V   Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K
Q   G   K   Q   E   A   A   D   A   A


 271/91                                 301/101

 |                                      |

 ctc cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta
acc c

 L   R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
```

Construct 35 (349bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 tca gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc
gag ctc gtg aga tac ctg aac acc aac cct gtg

 S   E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E
L   V   R   Y   L   N   T   N   P   V


 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggt ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc
gac cag tcc gga cct cct cac gag ccc aag ttc

 G   G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D
Q   S   G   P   P   H   E   P   K   F


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gtt tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag
aag caa ggc aag cag gaa gca gca gat gcg gct

 V   Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K
Q   G   K   Q   E   A   A   D   A   A


 271/91                                                              301/101
331/111

 |                                                                    |
|

 ctc cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta
acc cca gtg aca ggg gcc agt

 L   R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
P   V   T   G   A   S
```

Construct 36 (355bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 tca gag tca ctt gat aac ttg gaa tcc atg atg ccc aac aag gtc agg aag att ggc
gag ctc gtg aga tac ctg aac acc aac cct gtg

 S   E   S   L   D   N   L   E   S   M   M   P   N   K   V   R   K   I   G   E
L   V   R   Y   L   N   T   N   P   V


 91/31                                                               121/41
151/51

 |                                                                    |
|

 ggt ggc ctt ttg gag tac gcc cgc tcc cat ggc ttt gct gct gaa ttc aag ttg gtc
gac cag tcc gga cct cct cac gag ccc aag ttc

 G   G   L   L   E   Y   A   R   S   H   G   F   A   A   E   F   K   L   V   D
Q   S   G   P   P   H   E   P   K   F


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gtt tac caa gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc tgc gca cac agc aag
aag caa ggc aag cag gaa gca gca gat gcg gct

 V   Y   Q   A   K   V   G   G   R   W   F   P   A   V   C   A   H   S   K   K
Q   G   K   Q   E   A   A   D   A   A


 271/91                                                              301/101
331/111

 |                                                                    |
|

 ctc cgt gtc ttg att ggg gag aac gag aag gca gaa cgc atg ggt ttc aca gag gta
acc cca gtg aca ggg gcc agt ctc aga

 L   R   V   L   I   G   E   N   E   K   A   E   R   M   G   F   T   E   V   T
P   V   T   G   A   S   L   R
```

Construct A (430bp)

```
 1/1                                                                  31/11
61/21

 |                                                                    |
|

 gag gcg acc aac tcc atg gct tct gat aac cag cct gaa ggt atg atc tca gag tca
ctt gat aac ttg gaa tcc atg atg ccc aac aag

 E   A   T   N   S   M   A   S   D   N   Q   P   E   G   M   I   S   E   S   L
D   N   L   E   S   M   M   P   N   K


 91/31                                                               121/41
151/51

 |                                                                    |
|

 gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac cct gtg ggt ggc ctt
ttg gag tac gcc cgc tcc cat ggc ttt gct gct

 V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P   V   G   G   L   L
E   Y   A   R   S   H   G   F   A   A


 181/61                                                              211/71
241/81

 |                                                                    |
|

 gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc aag ttc gtt tac caa
gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc

 E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K   F   V   Y   Q   A
K   V   G   G   R   W   F   P   A   V


 271/91                                                              301/101
331/111

 |                                                                    |
|

 tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat gcg gct ctc cgt gtc
ttg att ggg gag aac gag aag gca gaa cgc atg

 C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A   A   L   R   V   L
I   G   E   N   E   K   A   E   R   M
```

```
 361/121                                                                391/131
421/141

 |                                                                      |
|

 ggt ttc aca gag gta acc cca gtg aca ggg gcc agt ctc aga aga act atg ctc ctc
ctc tca agg tcc

 G   F   T   E   V   T   P   V   T   G   A   S   L   R   R   T   M   L   L   L
S   R   S
```

Construct B (344bp)

```
1/1                                                                     31/11
61/21

 |                                                                      |
|

 atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

 M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S

 91/31                                                                  121/41
151/51

 |                                                                      |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R


 181/61                                                                 211/71
241/81

 |                                                                      |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag

 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E
```

271/91 301/101
331/111

| |
|

aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc agt ctc aga
aga act atg ctc ctc c

K A E R M G F T E V T P V T G A S L R R
T M L L

Construct C (351bp)

1/1 31/11
61/21

| |
|

gag gcg acc aac tcc atg gct tct gat aac cag cct gaa ggt atg atc tca gag tca
ctt gat aac ttg gaa tcc atg atg ccc aac aag

E A T N S M A S D N Q P E G M I S E S L
D N L E S M M P N K

91/31 121/41
151/51

| |
|

gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac cct gtg ggt ggc ctt
ttg gag tac gcc cgc tcc cat ggc ttt gct gct

V R K I G E L V R Y L N T N P V G G L L
E Y A R S H G F A A

181/61 211/71
241/81

| |
|

gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc aag ttc gtt tac caa
gca aaa gtt ggg ggt cgc tgg ttc cca gcc gtc

E F K L V D Q S G P P H E P K F V Y Q A
K V G G R W F P A V

```
 271/91                                                                301/101
331/111

 |                                                                     |
|

 tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat gcg gct ctc cgt gtc
ttg att ggg gag aac gag aag gc

 C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A   A   L   R   V   L
I   G   E   N   E   K
```

Construct E (401bp)

```
 1/1                                                                    31/11
61/21

 |                                                                     |
|

 atg atg ccc aac aag gtc agg aag att ggc gag ctc gtg aga tac ctg aac acc aac
cct gtg ggt ggc ctt ttg gag tac gcc cgc tcc

 M   M   P   N   K   V   R   K   I   G   E   L   V   R   Y   L   N   T   N   P
V   G   G   L   L   E   Y   A   R   S



 91/31                                                                 121/41
151/51

 |                                                                     |
|

 cat ggc ttt gct gct gaa ttc aag ttg gtc gac cag tcc gga cct cct cac gag ccc
aag ttc gtt tac caa gca aaa gtt ggg ggt cgc

 H   G   F   A   A   E   F   K   L   V   D   Q   S   G   P   P   H   E   P   K
F   V   Y   Q   A   K   V   G   G   R



 181/61                                                                211/71
241/81

 |                                                                     |
|

 tgg ttc cca gcc gtc tgc gca cac agc aag aag caa ggc aag cag gaa gca gca gat
gcg gct ctc cgt gtc ttg att ggg gag aac gag
```

```
 W   F   P   A   V   C   A   H   S   K   K   Q   G   K   Q   E   A   A   D   A
A   L   R   V   L   I   G   E   N   E


 271/91                                                                  301/101
331/111

 |                                                                       |
|

 aag gca gaa cgc atg ggt ttc aca gag gta acc cca gtg aca ggg gcc agt ctc aga
aga act atg ctc ctc ctc tca agg tcc cca gaa

 K   A   E   R   M   G   F   T   E   V   T   P   V   T   G   A   S   L   R   R
T   M   L   L   L   S   R   S   P   E


 361/121                                 391/131

 |                                       |

 gca cag cca aag aca ctc cct ctc act ggc agc acc ttc c

 A   Q   P   K   T   L   P   L   T   G   S   T   F
```

References

Alfonzo, J. D., O. Thiemann, et al. (1997). "The mechanism of U insertion/deletion RNA editing in kinetoplastid mitochondria." Nucleic Acids Res **25**(19): 3751-3759.

Arts, G. J., S. Kuersten, et al. (1998). "The role of exportin-t in selective nuclear export of mature tRNAs." EMBO J **17**(24): 7430-7441.

Athanasiadis, A., A. Rich, et al. (2004). "Widespread A-to-I RNA editing of Alu-containing mRNAs in the human transcriptome." PLoS Biol **2**(12): e391.

Bass, B. L., S. R. Hurst, et al. (1994). "Binding properties of newly identified Xenopus proteins containing dsRNA-binding motifs." Curr Biol **4**(4): 301-314.

Bass, B. L. and H. Weintraub (1987). "A developmentally regulated activity that unwinds RNA duplexes." Cell **48**(4): 607-613.

Bass, B. L., H. Weintraub, et al. (1989). "Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA." Cell **56**(3): 331.

Benne, R., J. Vandenburg, et al. (1986). "Major Transcript of the Frameshifted Coxll Gene from Trypanosome Mitochondria Contains 4 Nucleotides That Are Not Encoded in the DNA." Cell **46**(6): 819-826.

Bevilacqua, P. C. and T. R. Cech (1996). "Minor-groove recognition of double-stranded RNA by the double-stranded RNA-binding domain from the RNA-activated protein kinase PKR." Biochemistry **35**(31): 9983-9994.

Blum, B., N. Bakalara, et al. (1990). "A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information." Cell **60**(2): 189-198.

Brusa, R., F. Zimmermann, et al. (1995). "Early-onset epilepsy and postnatal lethality associated with an editing-deficient GluR-B allele in mice." Science **270**(5242): 1677-1680.

Burnashev, N., H. Monyer, et al. (1992). "Divalent ion permeability of AMPA receptor channels is dominated by the edited form of a single subunit." Neuron **8**(1): 189-198.

Burns, C. M., H. Chu, et al. (1997). "Regulation of serotonin-2C receptor G-protein coupling by RNA editing." Nature **387**(6630): 303-308.

Bycroft, M., S. Grunert, et al. (1995). "NMR solution structure of a dsRNA binding domain from Drosophila staufen protein reveals homology to the N-terminal domain of ribosomal protein S5." EMBO J **14**(14): 3563-3571.

Cattaneo, R. (1994). "Biased (A-->I) hypermutation of animal RNA virus genomes." Curr Opin Genet Dev **4**(6): 895-900.

Chan, L. (1992). "Apolipoprotein B, the major protein component of triglyceride-rich and low density lipoproteins." J Biol Chem **267**(36): 25621-25624.

Chen, C. X., D. S. Cho, et al. (2000). "A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains." RNA **6**(5): 755-767.

Chen, L. L., J. N. DeCerbo, et al. (2008). "Alu element-mediated gene silencing." EMBO J **27**(12): 1694-1705.

Chen, S. H., G. Habib, et al. (1987). "Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon." Science **238**(4825): 363-366.

Cho, D. S., W. Yang, et al. (2003). "Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA." J Biol Chem **278**(19): 17093-17102.

Conti, E. and E. Izaurralde (2001). "Nucleocytoplasmic transport enters the atomic age." Curr Opin Cell Biol **13**(3): 310-319.

Desterro, J. M., L. P. Keegan, et al. (2003). "Dynamic association of RNA-editing enzymes with the nucleolus." J Cell Sci **116**(Pt 9): 1805-1818.

Dingledine, R., K. Borges, et al. (1999). "The glutamate receptor ion channels." Pharmacol Rev **51**(1): 7-61.

Doyle, M. and M. F. Jantsch (2002). "New and old roles of the double-stranded RNA-binding domain." J Struct Biol **140**(1-3): 147-153.

Eckmann, C. R., A. Neunteufl, et al. (2001). "The human but not the Xenopus RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein." Mol Biol Cell **12**(7): 1911-1924.

Enstero, M., C. Daniel, et al. (2009). "Recognition and coupling of A-to-I edited sites are determined by the tertiary structure of the RNA." Nucleic Acids Res **37**(20): 6916-6926.

Feagin, J. E., J. M. Abraham, et al. (1988). "Extensive editing of the cytochrome c oxidase III transcript in Trypanosoma brucei." Cell **53**(3): 413-422.

Feldmeyer, D., K. Kask, et al. (1999). "Neurological dysfunctions in mice expressing different levels of the Q/R site-unedited AMPAR subunit GluR-B." Nat Neurosci **2**(1): 57-64.

Ferrandon, D., L. Elphick, et al. (1994). "Staufen protein associates with the 3'UTR of bicoid mRNA to form particles that move in a microtubule-dependent manner." Cell **79**(7): 1221-1232.

Franke, W. W. and U. Scheer (1974). "Pathways of nucleocytoplasmic translocation of ribonucleoproteins." Symp Soc Exp Biol(28): 249-282.

Fritz, J., A. Strehblow, et al. (2009). "RNA-regulated interaction of transportin-1 and exportin-5 with the double-stranded RNA-binding domain regulates nucleocytoplasmic shuttling of ADAR1." Mol Cell Biol **29**(6): 1487-1497.

Gallo, A., L. P. Keegan, et al. (2003). "An ADAR that edits transcripts encoding ion channel subunits functions as a dimer." EMBO J **22**(13): 3421-3430.

Gerber, A., H. Grosjean, et al. (1998). "Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2." EMBO J **17**(16): 4780-4789.

Gerber, A., M. A. O'Connell, et al. (1997). "Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette." RNA **3**(5): 453-463.

Gerber, A. P. and W. Keller (2001). "RNA editing by base deamination: more enzymes, more targets, new mysteries." Trends Biochem Sci **26**(6): 376-384.

Gorlich, D. and U. Kutay (1999). "Transport between the cell nucleus and the cytoplasm." Annu Rev Cell Dev Biol **15**: 607-660.

Gorlich, D., M. J. Seewald, et al. (2003). "Characterization of Ran-driven cargo transport and the RanGTPase system by kinetic measurements and computer simulation." EMBO J **22**(5): 1088-1100.

Green, S. R. and M. B. Mathews (1992). "Two RNA-binding motifs in the double-stranded RNA-activated protein kinase, DAI." Genes Dev **6**(12B): 2478-2490.

Greeve, J., I. Altkemper, et al. (1993). "Apolipoprotein B mRNA editing in 12 different mammalian species: hepatic expression is reflected in low concentrations of apoB-containing plasma lipoproteins." J Lipid Res **34**(8): 1367-1383.

Greger, I. H., L. Khatri, et al. (2003). "AMPA receptor tetramerization is mediated by Q/R editing." Neuron **40**(4): 763-774.

Gurevich, I., H. Tamir, et al. (2002). "Altered editing of serotonin 2C receptor pre-mRNA in the prefrontal cortex of depressed suicide victims." Neuron **34**(3): 349-356.

Hajjar, A. M. and M. L. Linial (1995). "Modification of retroviral RNA by double-stranded RNA adenosine deaminase." J Virol **69**(9): 5878-5882.

Hanrahan, C. J., M. J. Palladino, et al. (2000). "RNA editing of the Drosophila para Na(+) channel transcript. Evolutionary conservation and developmental regulation." Genetics **155**(3): 1149-1160.

Harris, R. S. and M. T. Liddament (2004). "Retroviral restriction by APOBEC proteins." Nat Rev Immunol **4**(11): 868-877.

Hartner, J. C., C. R. Walkley, et al. (2009). "ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling." Nat Immunol **10**(1): 109-115.

Herbert, A., J. Alfken, et al. (1997). "A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase." Proc Natl Acad Sci U S A **94**(16): 8421-8426.

Herbert, A., K. Lowenhaupt, et al. (1995). "Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA." Proc Natl Acad Sci U S A **92**(16): 7550-7554.

Herbert, A. and A. Rich (2001). "The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1." Proc Natl Acad Sci U S A **98**(21): 12132-12137.

Higuchi, M., S. Maas, et al. (2000). "Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2." Nature **406**(6791): 78-81.

Hough, R. F. and B. L. Bass (1997). "Analysis of Xenopus dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases." RNA **3**(4): 356-370.

Hume, R. I., R. Dingledine, et al. (1991). "Identification of a site in glutamate receptor subunits that controls calcium permeability." Science **253**(5023): 1028-1031.

Innerarity, T. L., J. Boren, et al. (1996). "Biosynthesis of apolipoprotein B48-containing lipoproteins. Regulation by novel post-transcriptional mechanisms." J Biol Chem **271**(5): 2353-2356.

Iwamoto, K. and T. Kato (2003). "RNA editing of serotonin 2C receptor in human postmortem brains of major mental disorders." Neurosci Lett **346**(3): 169-172.

Kallman, A. M., M. Sahlin, et al. (2003). "ADAR2 A-->I editing: site selectivity and editing efficiency are separate events." Nucleic Acids Res **31**(16): 4874-4881.

Kawahara, Y., K. Ito, et al. (2004). "Glutamate receptors: RNA editing and death of motor neurons." Nature **427**(6977): 801.

Kawahara, Y. and S. Kwak (2005). "Excitotoxicity and ALS: what is unique about the AMPA receptors expressed on spinal motor neurons?" Amyotroph Lateral Scler Other Motor Neuron Disord **6**(3): 131-144.

Kawahara, Y., M. Megraw, et al. (2008). "Frequency and fate of microRNA editing in human brain." Nucleic Acids Res **36**(16): 5270-5280.

Kawahara, Y., B. Zinshteyn, et al. (2007). "RNA editing of the microRNA-151 precursor blocks cleavage by the Dicer-TRBP complex." EMBO Rep **8**(8): 763-769.

Kawahara, Y., B. Zinshteyn, et al. (2007). "Redirection of silencing targets by adenosine-to-inosine editing of miRNAs." Science **315**(5815): 1137-1140.

Kawasaki, F., S. C. Collins, et al. (2002). "Synaptic calcium-channel function in Drosophila: analysis and transformation rescue of temperature-sensitive paralytic and lethal mutations of cacophony." J Neurosci **22**(14): 5856-5864.

Keegan, L. P., A. Gallo, et al. (2001). "The many roles of an RNA editor." Nat Rev Genet **2**(11): 869-878.

Kim, U., Y. Wang, et al. (1994). "Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing." Proc Natl Acad Sci U S A **91**(24): 11457-11461.

Kim, V. N. (2004). "MicroRNA precursors in motion: exportin-5 mediates their nuclear export." Trends Cell Biol **14**(4): 156-159.

King, M. C., C. P. Lusk, et al. (2006). "Karyopherin-mediated import of integral inner nuclear membrane proteins." Nature **442**(7106): 1003-1007.

Kohler, A. and E. Hurt (2007). "Exporting RNA from the nucleus to the cytoplasm." Nat Rev Mol Cell Biol **8**(10): 761-773.

Kriegs, J. O., G. Churakov, et al. (2007). "Evolutionary history of 7SL RNA-derived SINEs in Supraprimates." Trends Genet **23**(4): 158-161.

Krovat, B. C. and M. F. Jantsch (1996). "Comparative mutational analysis of the double-stranded RNA binding domains of Xenopus laevis RNA-binding protein A." J Biol Chem **271**(45): 28112-28119.

Kumar, M. and G. G. Carmichael (1997). "Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts." Proc Natl Acad Sci U S A **94**(8): 3542-3547.

Kutay, U., G. Lipowsky, et al. (1998). "Identification of a tRNA-specific nuclear export receptor." Mol Cell **1**(3): 359-369.

Lei, E. P. and P. A. Silver (2002). "Protein and RNA export from the nucleus." Dev Cell **2**(3): 261-272.

Leulliot, N., S. Quevillon-Cheruel, et al. (2004). "A new alpha-helical extension promotes RNA binding by the dsRBD of Rnt1p RNAse III." EMBO J **23**(13): 2468-2477.

Levanon, E. Y., E. Eisenberg, et al. (2004). "Systematic identification of abundant A-to-I editing sites in the human transcriptome." Nat Biotechnol **22**(8): 1001-1005.

Liu, Y., C. X. George, et al. (1997). "Functionally distinct double-stranded RNA-binding domains associated with alternative splice site variants of the interferon-inducible double-stranded RNA-specific adenosine deaminase." J Biol Chem **272**(7): 4419-4428.

Lomeli, H., J. Mosbacher, et al. (1994). "Control of kinetic properties of AMPA receptor channels by nuclear RNA editing." Science **266**(5191): 1709-1713.

Maas, S. and A. Rich (2000). "Changing genetic information through RNA editing." Bioessays **22**(9): 790-802.

Malek, O., K. Lattig, et al. (1996). "RNA editing in bryophytes and a molecular phylogeny of land plants." EMBO J **15**(6): 1403-1411.

Maslov, D. A. and L. Simpson (1992). "The polarity of editing within a multiple gRNA-mediated domain is due to formation of anchors for upstream gRNAs by downstream editing." Cell **70**(3): 459-467.

Mehta, A., M. T. Kinter, et al. (2000). "Molecular cloning of apobec-1 complementation factor, a novel RNA-binding protein involved in the editing of apolipoprotein B mRNA." Mol Cell Biol **20**(5): 1846-1854.

Meurs, E., K. Chong, et al. (1990). "Molecular cloning and characterization of the human double-stranded RNA-activated protein kinase induced by interferon." Cell **62**(2): 379-390.

Mosammaparast, N. and L. F. Pemberton (2004). "Karyopherins: from nuclear-transport mediators to nuclear-function regulators." Trends Cell Biol **14**(10): 547-556.

Neuberger, M. S., R. S. Harris, et al. (2003). "Immunity through DNA deamination." Trends Biochem Sci **28**(6): 305-312.

Nie, Y., L. Ding, et al. (2005). "ADAR1 interacts with NF90 through double-stranded RNA and regulates NF90-mediated gene expression independently of RNA editing." Mol Cell Biol **25**(16): 6956-6963.

Niswender, C. M., K. Herrick-Davis, et al. (2001). "RNA editing of the human serotonin 5-HT2C receptor. alterations in suicide and implications for serotonergic pharmacotherapy." Neuropsychopharmacology **24**(5): 478-491.

O'Connell, M. A., S. Krause, et al. (1995). "Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase." Mol Cell Biol **15**(3): 1389-1397.

O'Hara, P. J., F. M. Horodyski, et al. (1984). "Vesicular stomatitis virus mutants resistant to defective-interfering particles accumulate stable 5'-terminal and fewer 3'-terminal mutations in a stepwise manner." J Virol **49**(3): 793-798.

Palladino, M. J., L. P. Keegan, et al. (2000). "dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing." RNA **6**(7): 1004-1018.

Palladino, M. J., L. P. Keegan, et al. (2000). "A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity." Cell **102**(4): 437-449.

Patterson, J. B. and C. E. Samuel (1995). "Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase." Mol Cell Biol **15**(10): 5376-5388.

Ramos, A., S. Grunert, et al. (2000). "RNA recognition by a Staufen double-stranded RNA-binding domain." EMBO J **19**(5): 997-1009.

Rout, M. P. and J. D. Aitchison (2000). "Pore relations: nuclear pore complexes and nucleocytoplasmic exchange." Essays Biochem **36**: 75-88.

Ryter, J. M. and S. C. Schultz (1998). "Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA." EMBO J **17**(24): 7505-7513.

Saunders, L. R. and G. N. Barber (2003). "The dsRNA binding protein family: critical roles, diverse cellular functions." FASEB J **17**(9): 961-983.

Schmauss, C. (2005). "Regulation of serotonin 2C receptor pre-mRNA editing by serotonin." Int Rev Neurobiol **63**: 83-100.

Schoft, V. K., S. Schopoff, et al. (2007). "Regulation of glutamate receptor B pre-mRNA splicing by RNA editing." Nucleic Acids Res **35**(11): 3723-3732.

Semenov, E. P. and W. L. Pak (1999). "Diversification of Drosophila chloride channel gene by multiple posttranscriptional mRNA modifications." J Neurochem **72**(1): 66-72.

Sen, G. C. (2001). "Viruses and interferons." Annu Rev Microbiol **55**: 255-281.

Slavov, D., T. Crnogorac-Jurcevic, et al. (2000). "Comparative analysis of the DRADA A-to-I RNA editing gene from mammals, pufferfish and zebrafish." Gene **250**(1-2): 53-60.

Spreux-Varoquaux, O., G. Bensimon, et al. (2002). "Glutamate levels in cerebrospinal fluid in amyotrophic lateral sclerosis: a reappraisal using a new HPLC method with coulometric detection in a large cohort of patients." J Neurol Sci **193**(2): 73-78.

St Johnston, D., N. H. Brown, et al. (1992). "A conserved double-stranded RNA-binding domain." Proc Natl Acad Sci U S A **89**(22): 10979-10983.

Stefl, R., L. Skrisovska, et al. (2005). "RNA sequence- and shape-dependent recognition by proteins in the ribonucleoprotein particle." EMBO Rep **6**(1): 33-38.

Stephens, O. M., B. L. Haudenschild, et al. (2004). "The binding selectivity of ADAR2's dsRBMs contributes to RNA-editing selectivity." Chem Biol **11**(9): 1239-1250.

Strehblow, A., M. Hallegger, et al. (2002). "Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain." Mol Biol Cell **13**(11): 3822-3835.

Suntharalingam, M. and S. R. Wente (2003). "Peering through the pore: nuclear pore complex structure, assembly, and function." Dev Cell **4**(6): 775-789.

Taylor, D. R., M. Puig, et al. (2005). "New antiviral pathway that mediates hepatitis C virus replicon interferon sensitivity through ADAR1." J Virol **79**(10): 6291-6298.

Tonkin, L. A., L. Saccomanno, et al. (2002). "RNA editing by ADARs is important for normal behavior in Caenorhabditis elegans." EMBO J **21**(22): 6025-6035.

Vinciguerra, P. and F. Stutz (2004). "mRNA export: an assembly line from genes to nuclear pores." Curr Opin Cell Biol **16**(3): 285-292.

Wagner, C., I. Palacios, et al. (2001). "Dimerization of the 3'UTR of bicoid mRNA involves a two-step mechanism." J Mol Biol **313**(3): 511-524.

Wang, Q., J. Khillan, et al. (2000). "Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis." Science **290**(5497): 1765-1768.

Wang, Q., M. Miyakoda, et al. (2004). "Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene." J Biol Chem **279**(6): 4952-4961.

Wang, Z., M. K. Choi, et al. (2008). "Regulation of innate immune responses by DAI (DLM-1/ZBP1) and other DNA-sensing molecules." Proc Natl Acad Sci U S A **105**(14): 5477-5482.

Wu, H., A. Henras, et al. (2004). "Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III." Proc Natl Acad Sci U S A **101**(22): 8307-8312.

Yang, J. H., X. Luo, et al. (2003). "Widespread inosine-containing mRNA in lymphocytes regulated by ADAR1 in response to inflammation." Immunology **109**(1): 15-23.

Yang, W., T. P. Chendrimada, et al. (2006). "Modulation of microRNA processing and expression through RNA editing by ADAR deaminases." Nat Struct Mol Biol **13**(1): 13-21.

Yang, W., Q. Wang, et al. (2005). "ADAR1 RNA deaminase limits short interfering RNA efficacy in mammalian cells." J Biol Chem **280**(5): 3946-3953.

Zhou, P., A. A. Lugovskoy, et al. (2001). "A solubility-enhancement tag (SET) for NMR studies of poorly behaving proteins." J Biomol NMR **20**(1): 11-14.

Curriculum Vitae

Sabina Dani

Personal Data

Date of Birth:	26.06.1984
Place of Birth:	Shkoder, Albania
Nationality:	Albania

Education

	After the basic education:
1998 - 2002	Academic high school, "28- Nentori" (Albania)
10.2003 - 12.2010	Studies of "Genetics and Microbiology" at the University of Vienna, Austria
05.2009 - 04.2010	Diploma thesis: Laboratory of Ao.Univ. Prof. Dr. Michael Jantsch, Department of Chromosome Biology Max F. Perutz Laboratories, University of Vienna.

Research Interests

Genomics, human genetic diseases, molecular medicine, and the progress from research to therapeutic approaches.

Languages:

German - fluently spoken and written
English - fluently spoken and written
Albanian - native language
Italian - fluently spoken and written