Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

Molecular characterization of the neuronal nucleic acid-binding protein Pur-α

Janine Weber

aus

Brandenburg a.d. Havel
Deutschland

2015

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Klaus Förstemann betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 11.05.15

Janine Weber

Dissertation eingereicht am 12.05.15

1. Gutachter: Prof. Klaus Förstemann
2. Gutachter: Prof. Dierk Niessing

Mündliche Prüfung am 08.07.15

Table of Contents

Summary 1

1. Introduction 2

1.1. Purine-rich element binding protein family 2
1.2. Pur-α protein 3
1.3. Pur-α's cellular functions 4
1.3.1. Replication 5
1.3.2. Transcription 5
1.3.3. RNA transport and translational control 6
1.3.4. Pur-α knock out mice 6
1.3.5. Pur-α and cancer 7
1.4. Nucleotide repeat expansion diseases 7
1.4.1. Fragile X syndrome 8
1.4.2. FXTAS 9
1.4.2.1. RNA toxicity – sequestration model 10
1.4.2.2. RAN translation – homopolymeric proteins 12
1.4.2.3. Therapy 12
1.4.3. ALS/FTLD 13
1.4.3.1. Protein toxicity/aggregation 14
1.4.3.2. RNA toxicity 15
1.4.3.3. RAN translation - dipeptide repeat proteins 16
1.4.3.4. Therapy 17
1.5. Pur-α in FXTAS and ALS/FTLD 17
1.6. Objectives 18

2. Results 20

2.1. Structural and functional studies 20
2.1.1. Expression and purification of Drosophila and human Pur-α derivatives 21
2.1.2. EMSA with Pur-α and DNA/RNA 21
2.1.3. NMR with Pur-α and DNA/RNA 22
2.1.4. Crystallization of Pur-α/DNA co-complex 25
2.1.4.1. Structure determination and refinement 25
2.1.4.2. Crystal structure of Pur-α/DNA co-complex 26
2.1.4.3. Structural comparison of Pur-α co-complex and apo-structure 29
2.1.4.4. Sequence alignment of Pur-α from different species 29
2.1.5. EMSA with mutant Pur-α and DNA/RNA 31
2.1.6. EMSA with Pur-α III and DNA/RNA 33
2.1.7. EMSA with Pur-α I-II and FTD/ALS-related repeat RNA 33
2.1.8. Unwinding assay 34
2.1.9. Small angle X-ray scattering (SAXS) 35
2.1.10. Dot blot assay 36
2.1.11. DNase I footprint 37
2.2. Cell culture studies 39
2.2.1. Cellular function of Pur-α 39
2.2.1.1. CHIP-Seq 40
2.2.1.2. Antibody test against human Pur-α 40
2.2.1.3. Generation of vectors for expression tagged human Pur-α 42
2.2.1.4. BAC transfection and cell selection 43
2.2.1.5. Validation of GFP-Pur-α expression 43
2.2.2. Pur-α and FXTAS 44
2.2.2.1. CGG-repeat RNA expression in COS7 cells 45
2.2.2.2. Generation of vectors for CGG-repeat RNA expression 45
2.2.2.3. Test transfection of initial vectors 46

2.2.2.4. Transient transfection and induction of CGG-repeat RNA....47
2.2.2.5. Transient transfection and induction of PP7-tagged CGG-repeat RNA....49
2.2.2.6. Antibody test against αB-crystallin....51

3. Discussion....52
3.1. Interaction of Pur-α with nucleic acids....52
3.1.1. Affinity for DNA and RNA....52
3.1.2. Stoichiometry of the Pur-α/nucleic acid co-complex....53
3.1.3. Interaction sites of Pur-α....54
3.1.4. Unwinding of dsDNA....55
3.1.5. Pur-α mutations cause 5q31.3 microdeletion syndrome....58
3.2. Outlook....61
3.2.1. Doxycycline-inducible CGG-repeat/reporter expression system - Cellular FXTAS model....61
3.2.2. Pur-α's role in transcription....63
3.2.3. Expression of Pur-α mutants in the FXTAS Drosophila model....63

4. Materials and Methods....65
4.1. Consumables and chemicals....65
4.2. Oligonucleotides....65
4.2.1. DNA oligonucleotides for cloning....65
4.2.2. DNA oligonucleotides for interaction studies and crystallization....66
4.2.3. RNA oligonucleotides for interaction studies and crystallization....66
4.3. Plasmids....66
4.3.1. Commercial plasmids....66
4.3.2. Plasmids for recombinant protein expression in *E. coli*....67
4.3.3. Plasmids for RNA/protein expression in mammalian cell lines....67
4.4. BAC (Bacterial artificial chromosome) clone, tagging cassettes, tagging and verification oligonucleotides....68
4.5. *E. coli* strains....68
4.6. Mammalian cell lines....68
4.7. Media and supplements for bacterial cell culture....69
4.8. Medium, supplements and reagents for mammalian cell culture....69
4.9. Antibodies....70
4.10. General buffers and stock solutions....70
4.11. Molecular biology....70
4.11.1. Cloning....70
4.11.2. Transformation of E. coli and isolation plasmid DNA....71
4.12. Bioinformatics....71
4.12.1. Protein parameters....71
4.12.2. Nucleic acid parameters and primer design....72
4.12.3. Sequence alignment....72
4.13. Protein expression and purification....72
4.13.1. Protein Expression....72
4.13.2. Protein Purification....73
4.14. Methods for protein analysis....73
4.14.1. SDS PAGE....73
4.14.2. Circular dichroism (CD) spectroscopy....74
4.15. Structural biology....74
4.15.1. Co-crystallization of Pur-α and nucleic acid and structure determination....74
4.16. Protein-DNA/RNA interaction studies....75
4.16.1. RNase-free water....75
4.16.2. NMR experiments....75
4.16.3. Isotopic labeling of oligonucleotides and primer....75
4.16.4. Electrophoretic mobility shift assay (EMSA)....76
4.16.5. SAXS (small angle X-ray scattering)....76
4.16.6. Dot blot assay....77
4.16.7. Unwinding assay....77

4.16.8. DNase I footprint 78
4.17. Cell culture 79
4.17.1. Cell cultivation 79
4.17.2. Transfection 79
4.17.3. Induction of protein/RNA expression 79
4.17.4. Establishment of a stable cell line expressing tagged human Pur-α 80
4.17.5. Establishment of a stable cell line expressing CGG-repeat RNA 80
4.18. Western Blot 81
4.18.1. Cell harvest and lysis 81
4.18.2. Determination of total protein concentration 81
4.18.3. Blotting and immuno-staining of proteins 81
4.18.4. Antibody testing (αB-crystallin, Pur-α) 82

5. Appendix 83
5.1. Abbreviations 83
5.2. Index of figures 84
5.3. Index of tables 85

6. References 86
Acknowledgements 94

Summary

Pur-α (purine-rich element binding protein A) is a multifunctional protein binding to ss/dsDNA and RNA. It is involved in replication, transcription, mRNA transport and translation in neurons. Homozygous *Pur-α* mutant mice die within 4 weeks after birth, suffering from severe neurological defects. Pur-α unwinds dsDNA in an ATP-independent manner, thereby providing access for replication and transcriptional regulators. Still, Pur-α's role in cellular functions is not well understood.

Pur-α has also been implicated in the pathomechanism of heritable, neurodegenerative diseases like ALS/FTLD (amyotrophic lateral sclerosis / frontotemporal lobar degeneration) and FXTAS (fragile X-associated tremor/ataxia syndrome). FXTAS is caused by premutation expansions (55-200 CGG repeats) in the 5'UTR of the *fmr1* gene. ALS/FTLD can be triggered by hexanucleotide (G4C2) repeat expansions in the first intron of the *C9orf72* gene. The pathological hallmark for both diseases is the formation of neuronal, intranuclear and cytoplasmic inclusions. It is thought that these repeat-RNA containing inclusions sequester RNA-binding proteins, leading to altered transcription, RNA processing and trafficking. Pur-α binds to both types of RNA repeats and accumulates in these pathogenic inclusions.

The first goal of this study was to gain insights into the molecular principles of Pur-α's binding to nucleic acids and its cellular functions. For this, structural analysis were combined with various biochemical *in vitro* and cellular studies. Here, I present the crystal structure of Pur-α/ssDNA co-complex from *Drosophila melanogaster* at 2.0 Å resolution. The structure explains Pur-α's dsDNA-binding and –unwinding, and its ssDNA stabilizing activity. The protein disrupts the base stacking of DNA by intercalation of a highly conserved phenylalanine. The importance of this structural feature was confirmed by *in vitro* unwinding assays. NMR titration experiments and EMSAs suggest that short RNA and DNA oligomers interact with Pur-α in identical ways. Filter-binding assays confirmed that the main nucleic acid binding domain of Pur-α binds two molecules of nucleic acid, as suggested by the crystal structure.

The second aim of this study was to investigate Pur-α's role in neurodegenerative diseases. For this, I generated inducible, mammalian expression vectors coding for the *fmr1* 5'UTR with normal and disease-related CGG-repeats. These vectors have been tested in COS7 and HeLa cells and can now be used for establishment of a stable cellular FXTAS model.

1. Introduction

1.1. Purine-rich element binding protein family

Pur (purine-rich element binding) proteins are nucleic acid-binding proteins that can be found from bacteria to mammals (Johnson, 2003). They bind to purine-rich elements conserved in origins of replication and gene flanking regions. The Pur family consists of 4 members, encoded by genes at three different loci (Bergemann and Johnson, 1992). These four members are Pur-α at chromosome 5q31, Pur-β at 7p13 and two isoforms of Pur-γ at 8p11 (Lezon-Geyda et al., 2001; Liu and Johnson, 2002; Ma et al., 1995). Different transcription termination sites generate the two isoforms of Pur-γ: Pur-γ A and B (Liu and Johnson, 2002). Except for Pur-γ B, all Pur proteins are expressed as a single, intronless coding sequence (Liu and Johnson, 2002). Transcription of Pur-γ B runs through the Pur-γ A termination signal, resulting in a very long transcript of which a 30 kb intron becomes spliced out. This splicing event results in a loss of the stop codon and a different C-terminus for the protein isoform B (Liu and Johnson, 2002).

Human Pur-α, Pur-β and Pur-γ (both isoforms) possess an N-terminal glycine-rich domain and, except for Pur-γ, a C-terminal gluatmine/glutamate-rich region. A so-called "Psycho" motif at the C-terminus describing the consensus motif of proline, serine, tyrosine and cysteine can be found in all Pur proteins, except for the isoform B of for Pur-γ.

All vertebrate Pur proteins contain three strongly conserved repeats of approximately 80 amino acids and are expressed at different time points during development (Graebsch et al., 2009). While Pur-γ is highly expressed at early stages of mouse embryo development (embryonic age 14), Pur-α expression is nearly undetectable at these early stages. Later Pur-γ protein levels decrease drastically whereas Pur-α expression reaches a peak at 18-25 days after birth, together with Pur-β. These observations implied that Pur-γ is an important factor for embryonic or fetal development that becomes replaced by Pur-α and Pur-β at a later developmental stage (Itoh et al., 1998; Johnson et al., 2006).

Pur-α plays multiple roles in cellular regulation including replication, transcription, mRNA transport and translation. Pur-α's various functions are further described in section 1.1.2.

Pur-β has been implicated in transcriptional repression of genes encoding for muscle-specific isoforms of actin and myosin in heart, skeletal muscle and vascular smooth muscle (Knapp et al., 2007; Rumora et al., 2013; Zhang et al., 2008). Both Pur-α and Pur-β have been shown to be present in the same mRNPs (messenger ribonucleoprotein particle) that is transported in dendrites along microtubules by a kinesin motor (Kanai et al., 2004). Interaction of Pur-α and Pur-β with nucleic acids results in the formation of multimeric complexes (reviewed by

Johnson, 2003; Kim et al., 2008). Still, direct interaction between Pur-α and Pur-β has not been shown so far.

Abberations in all three Pur genes have been implicated in multiple tumor types and cell proliferation disorders, including myelodysplastic syndrome, myelogenous leukemia and 5q31.3 microdeletion syndrome (Pur-α), brain tumors and glioblastoma (Pur-β), myeloproliferative syndrome (Pur-γ) (Brizard et al., 1988; Jackson et al., 2010; Lalani et al., 2014; Lopez-Gines et al., 2006).

1.2. Pur-α protein

Pur-α is the best-studied protein amongst the Pur family members. It was originally purified from mouse brain and identified as a ubiquitously expressed nucleic acid-binding protein (Haas et al., 1993; Haas et al., 1995). Human Pur-α consists of 322, *Drosophila melanogaster* Pur-α of 274 and mouse Pur-α of 321 amino acids. Whereas mouse Pur-α only differs from the human homolog by two amino acids, the *Drosophila* homolog shares 52% protein sequence identity with it. The Pur repeats in *Drosophila* Pur-α locate to residues 40-107 (repeat I), 117-185 (repeat II) and 193-256 (repeat III) (Graebsch et al., 2010).

Graebsch et al. solved the x-ray structure of residues 40-185 from *Drosophila* Pur-α, which constitutes the major part of the DNA/RNA-binding region. This region contains two almost identical Pur-repeats, so-called Pur repeat I and II (Pur-α I-II) (Graebsch et al., 2009). They interact with each other and form a nucleic acid binding PUR domain (Figure 1.1 A). Each repeat consists of a four-stranded anti-parallel β-sheet followed by a C-terminal α-helix. The two Pur-repeats are intertwined and thereby form a hydrophobic buried surface, which stabilizes the interaction between the two repeats (Figure 1.1 A). This so-called PUR domain, consisting of two Pur-repeats, is homologous to the Whirly class of nucleic acid binding folds and shows significant structural similarities to the mitochondrial RNA-binding protein-1 and 2 (MRP-2, MRP-2) as well as to the plant transcription regulator PBF-2 (P24) (Graebsch et al., 2009). Surface charge analysis revealed negative charges at the α-helices, whereas the β-sheets contain numerous positively charged residues. These findings implied that the β-sheets mediate DNA/RNA-binding whereas the amphiphatic helices might rather contribute to protein-protein interactions. Further evidence for nucleic acid binding to the β-sheets was given by electrophoretic mobility shift experiments (EMSA) with mutations in the fourth β-strand in both Pur-repeats, which resulted in reduction of nucleic acid binding (Graebsch et al., 2009).

Analysis of the oligomeric state of *Drosophila* Pur-α revealed that Pur-α I-II is monomeric in solution, whereas Pur-α comprising all three Pur-repeats forms a dimer (Figure 1.1 B) arranging in an unusual Z-like shape (Figure 1.1 C) (Graebsch et al., 2009). This dimerization is most likely mediated by the intermolecular interaction of two Pur III repeats (Graebsch et al., 2009).

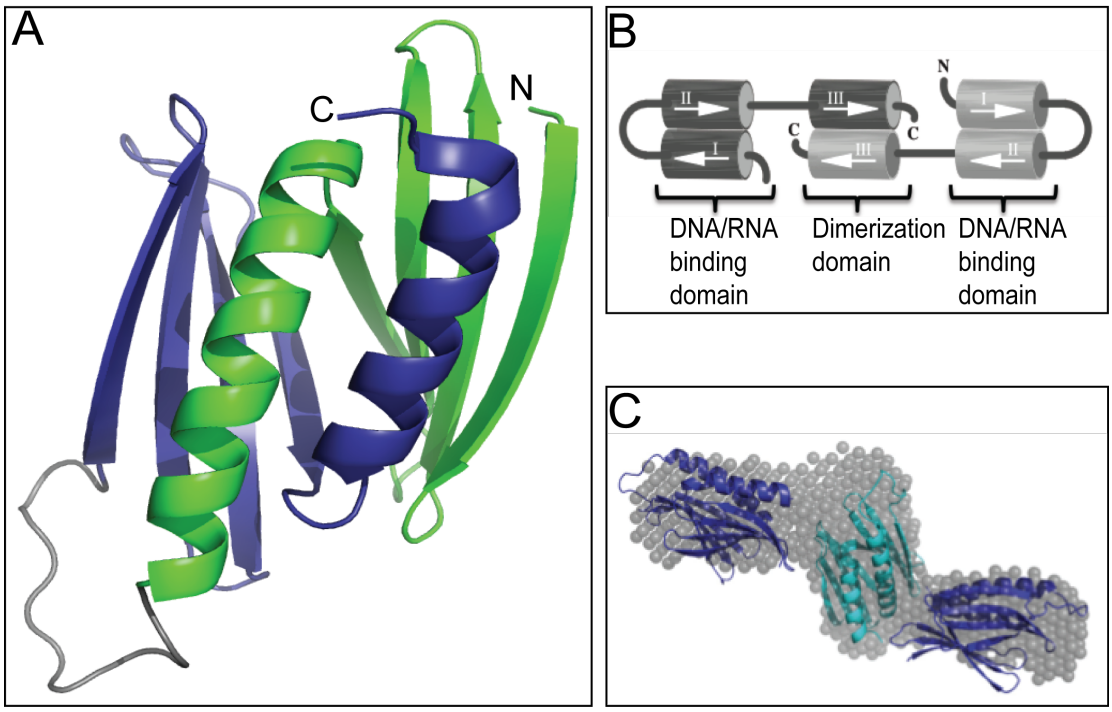


Figure 1.1 ***D. melanogaster*** **Pur-α protein. A:** Crystal structure of Pur-α I-II, showing the Whirly-like fold (ββββα) and the intercalation of the two repeats. Repeat I and II are depicted in green and blue, respectively. **B:** Schematic representation of two Pur-α molecules forming a dimer. Pur-repeat I and II constitute the DNA/RNA binding domain, two Pur-repeat III form the intermolecular dimerization domain. **C:** A fit of three Pur-α I-II molecules into the Z-like envelope calculated from the data obtain from Pur-α I-III measurements of small angle X-ray scattering (SAXS). (Figures adapted from (Graebsch et al., 2009).

1.3. Pur-α's cellular functions

Pur-α is a ubiquitous multifunctional protein that binds to both DNA and RNA and is known to regulate replication, transcription and translation. It has been shown that Pur-α can bind to single- and double-stranded nucleic acids that contain a GGN consensus motif. Pur-α also plays an important role in the transport of specific mRNAs to the dendrites in the developing brain. Besides being an essential neuronal factor, Pur-α is also involved in oncogenic transformation. In the following sections Pur-α's cellular functions will be summarized.

1.3.1. Replication

Initiation of replication requires unwinding of duplex DNA. It has been shown that Pur-α can bind to ssDNA and dsDNA and possesses DNA-helix-destabilizing activity (Darbinian et al., 2001). Furthermore it associates with DNA sequences that are close to viral and cellular origins of replication (Darbinian et al., 2001; Gallia et al., 2000). There is also evidence that Pur-α might be involved in replication-dependent repair of DNA lesions, as Pur-α deficient cells show enhanced sensitivity to the DNA replication inhibitor hydroxyurea and to the DNA-crosslinking anti-tumor drug cis-platin (Kaminski et al., 2008; Wang et al., 2007).

1.3.2. Transcription

Pur-α has been found to bind to the purine-rich region upstream of the human *c-myc* promoter (Bergemann and Johnson, 1992; Bergemann et al., 1992) and to regulate the transcription of more than 20 genes (reviewed in White et al., 2009). Important neuronal genes regulated by Pur-α are for example the myelin *proteolipid protein 1* (*Plp1*) gene (Dobretsova et al., 2008) and the *myelin basic protein* (*Mbp*) gene (Darbinian et al., 2001; Haas et al., 1995) both responsible for myelination of nerve cells in the central nervous system (CNS).

Pur-α-DNA binding results in formation of a multimeric complex accompanied by interaction with other transcription factors like SP1 (Tretiakova et al., 1999), YB1 (Kim et al., 2008), SP3 and Pur-β (Ji et al., 2007). SP1, for example, enhances the interaction of Pur-α with the *Mbp* promoter, thereby stimulating *Mbp* gene expression in glial cells (Haas et al., 1995; Tretiakova et al., 1999). Overexpression of Pur-α and SP1 in the CNS results in synergistic stimulation of MBP expression.

Another target gene of Pur-α is mouse *vascular smooth muscle* (VSM) *α-actin* (Kelm et al., 1997). Pur-α has been shown to bind to the purine-rich strand of the MCAT enhancer and interacts together with Pur-β and the mouse Y-box protein MSY1 (Kelm et al., 1997).

Pur-α is also involved in regulation of several human viruses. It activates the early and late promoters of the JC (John Cunningham) polyomavirus and the HIV-1 (human immunodeficiency virus 1) LTR (long terminal repeat) (reviewed in White et al., 2009).

Conversely, Pur-α was also reported to negatively regulate gene transcription including α-actin, amyloid-β protein, CD43, fas, gata2 and somatostatin (White et al., 2009) and even its own gene by binding to its own promoter and preventing its transcription (Muralidharan et al., 2001).

1.3.3. RNA transport and translational control

Pur-α plays an important role in the transport of specific mRNAs to the dendrites in the developing brain. For example, Pur-α binds to mouse non-coding *BC1* RNA and to its human counterpart *BC200* RNA (Johnson et al., 2006; Kobayashi et al., 2000), which are expressed almost exclusively in the nervous system and are distributed in neuronal dendrites as RNA:protein complexes. Pur-α has been suggested to link the *BC1* RNA complex to microtubules (Ohashi et al., 2000). The Pur-α binding site of the *BC1* RNA lies within its 5' proximal region and contains G/U- rich residues.

Consistently, PURA $^{-/-}$ mice show disturbed localization of the protein Staufen and FMRP (fragile X mental retardation protein). Both are involved in specific mRNA transport in dendrites (Johnson et al., 2006). RNA immunoprecipitation experiments indicated that FMRP and Pur-α interact simultaneously with *Map2* RNA, a dendrite-specific protein.

In addition, RNA interference experiments revealed that Pur-α is essential for mRNP transport (Kanai et al., 2004).

Aumiller et al. analyzed the localization of Pur-α in *Drosophila* oocytes and showed that Pur-α appears to associate with the oocyte mRNA transport system, and to shuttle between nucleus and cytoplasm. Actually, dimerization of Pur-α is necessary to achieve optimal transport into the oocyte (Aumiller et al., 2012).

Data from mouse brain support a role for Pur-α in dendritic protein translation and dendrite maturation, by regulating the neuronal levels of RhoA, a GTPase, critical for mRNA translation and dendritic maturation including subcellular compartmentalization and turn over.

1.3.4. Pur-α knock out mice

The crucial role of Pur-α for postnatal brain development is highlighted by its genetic inactivation in the mouse model. PURA $^{-/-}$ mice appear normal at birth, but after 2 weeks they develop neurological problems and die after 4 weeks (Khalili et al., 2003). These mice show fewer cells in the brain cortex, hippocampus and cerebellum as a consequence of decreased proliferation of the precursor cells. Depletion of Pur-α also causes pathological development of the astrocytes and oligodendrocytes. In addition, the number of synapses is significantly reduced in the hippocampus. Hokkanen et al. generated a Pur-α-deficient mouse model that allowed studying Pur-α's impact on brain growth and hippocampal and cerebellar development from newborn to the age of 6 months (Hokkanen et al., 2012). In contrast to Khalili et al. they observed enhanced proliferation of neuronal precursor cells and concluded that this may reflect a protraction of the brain development. Moreover, their PURA $^{-/-}$ mice

showed reduced expression of the neuron-specific cytoskeletal protein Map2 (microtubule-associated protein 2) and pathological hyperphosphorylation of axonal proteins. The adult Pur-α deficient mice suffered from continuous tremor, ataxia and showed enlarged brain size, which are the phenotypical aspects of the fragile X-associated tremor/ataxia syndrome (FXTAS). Further studies indicate that the PURA $^{-/-}$ mice showed altered dendritic localization of Staufen and FMRP, both being essential for specific mRNA transport to the dendrites (Johnson et al., 2006). Additionally, in Pur-α knockout mice RhoA levels are reduced and dendrite maturation is impaired (Mishra et al., 2013).

1.3.5. Pur-α and cancer

Several lines of evidence suggest that Pur-α is a major player in cell cycle control and oncogenic transformation. Pur-α binds to regulatory proteins such as retinoblastoma protein, E2F-1, cyclin A/Cdk2, etc. (Darbinian et al., 2004; Itoh et al., 1998; Johnson et al., 1995). Intracellular levels of Pur-α vary during the cell cycle, declining at the onset of S-phase and peaking during mitosis. Pur-α causes cell cycle arrest at either G1/S or G2/M phase (Itoh et al., 1998). Experiments showed that ectopic overexpression of Pur-α, e.g. in glioblastomas, suppresses their growth, whereas depletion of Pur-α have been reported in myelodysplastic syndrome and prostate cancer (Inoue et al., 2008; Johnson et al., 1995; Lezon-Geyda et al., 2001). Taken together, these results indicate that Pur-α might act as a tumor suppressor protein.

Thus, Pur-α is a multifunctional protein binding to ss/dsDNA and RNA and is involved in many cellular processes from replication and transcription to mRNA transport and translation in neurons. Hence, a better understanding of Pur-α binding mode to nucleic acids would help to comprehend the role of Pur-α in its various cellular functions and its role in cancer development.

1.4. Nucleotide repeat expansion diseases

Trinucleotide repeat disorders form a significant group of at least 16 neurologically inherited diseases such as Huntington disease and inherited ataxias (Orr and Zoghbi, 2007). Several cases have been reported where expansions of CAG trinucleotide repeats within coding regions of genes result in proteins with long repeating stretches of glutamine. Polyglutamine

peptides show enhanced propensity to form aggregates and large inclusions, resulting in neuronal dysfunction, neurodegeneration and neuronal loss (Ross and Poirier, 2004).

Also abnormal trinucleotide expansions in non-coding regions of genes can lead to severe clinical symptoms. For instance, the fragile X syndrome (FXS) (Oostra and Willemsen) is caused by abnormal trinucleotide expansions in the 5' untranslated region (5'UTR) of the *fmr1* gene. Myotonic dystrophy type 1 (DM1) is a dominant disorder caused by CTG repeat expansion within the 3' untranslated region (3'UTR) of the *dystrophia myotonica-protein kinase* (*dmpk*) gene. Investigations on DM1 have demonstrated that repeat expansions can be pathogenic at the RNA level (Day and Ranum, 2005; Liquori et al., 2001). Similar evidence now supports that an RNA gain-of-function mechanism may also be involved in the pathogenesis of the non-coding expansion disorder FXTAS (Oostra and Willemsen, 2009).

1.4.1. Fragile X syndrome

Fragile X syndrome (FXS) is the most common monogenetic form of autism and intellectual disability and was the first example of a trinucleotide repeat expansion disease (Oostra and Willemsen, 2009). In 1991 the responsible gene was identified and named *fragile X mental retardation 1* gene (*fmr1*) (Verkerk et al., 1991). The gene is located on the X-chromosome (Xp27.3), is highly expressed in the brain and testis (Devys et al., 1993) and encodes for the FMR1 protein (FMRP). Although ubiquitously expressed, the FMRP is an essential neuronal protein that regulates translation of hundreds of mRNAs, mostly involved in synaptic plasticity (Darnell and Richter, 2012). FXS is less severe in affected females, presumably because of X inactivation (Kaufmann et al., 1999; Tassone et al., 1999). Many affected females have some form of learning difficulties and behavioral problems (Hull and Hagerman, 1993). For males the prevalence for FXS is estimated at 1/6,000 to 1/4,000 in the general population (Turner et al., 1996; Youings et al., 2000).

FXS is caused by an expansion of a CGG repeat in the 5'UTR of *fmr1*. In the normal population the repeat length ranges from 5-54 CGGs. Patients with FXS show an expansion of more than 200 CGG repeats (termed full mutation range) leading to a transcriptional shut down of the gene and a lack of FMRP (Figure 1.2). Suppression of FMRP expression can be due to DNA- or/and RNA-mediated silencing. One scenario is that CGG repeat DNA can form hairpin structures, which trigger their own hypermethylation at the cytosine bases by DNA methyltransferases, as shown by *in vitro* experiments of Smith et al. (Smith et al., 1994). Another model suggests that repeat-binding proteins, for example transcription factors, form protein aggregates and thereby prevent transcription (Bulut-Karslioglu et al., 2012). An RNA-mediated mechanism of gene silencing has been suggested by Colak et al. (Colak et al., 2014).

In this model, the CGG-repeat *fmr1* mRNA hybridizes with the complementary DNA strand and thereby silences the *fmr1* promoter. Furthermore, unusual hairpins within the CGG-repeat mRNA might form and become a substrate for the enzyme Dicer, facilitating RNA-induced silencing (Handa et al., 2003; Usdin et al., 2014).

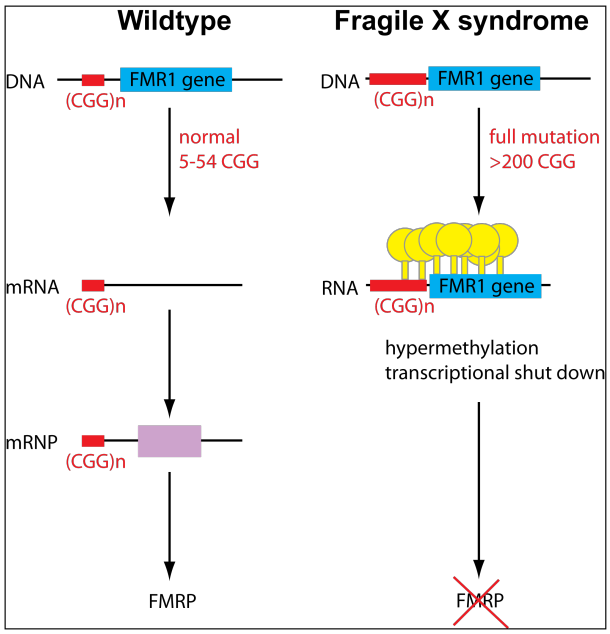


Figure 1.2 Fragile X syndrome. (Left) In the normal population (wild type) the CGG-repeat length in the 5'UTR of the *fmr1* gene ranges from 5-54 repeats. The gene becomes normally transcribed, incorporated into mRNPs and translated into the FMRP. (Right) When CGG repeats expand up to more than 200 repeats, the *fmr1* gene becomes hypermethylated and transcriptionally inactivated. FMRP is not expressed in the fragile X syndrome. (Figure adapted from Willemsen et al., 2011).

1.4.2. FXTAS

The fragile X-associated tremor/ataxia syndrome (FXTAS) is a late-onset (over 50 years) neurodegenerative disease, which is uncoupled from the FXS. Both disorders involve repeat expansion in the *fmr1* gene, but the clinical features and the molecular mechanism behind each disease is different (Hagerman and Hagerman, 2007; Hagerman et al., 2001; Jacquemont et al., 2003). However an overlap of these two diseases occurs when the full mutation (>200 CGG repeats) is partially or completely unmethylated or there is a high mosaicism in FXS. In this case FXS patients might also develop FXTAS.

FXTAS has been described in 2001 and is associated with a wide range of clinical features, such as tremor, ataxia, parkinsonism, dementia, neuropathy, executive function defects and disautonomia (reviewed in Hagerman and Hagerman, 2007). Life expectancy is between 5-25 years after onset of symptoms (Seritan et al., 2008), whereby CGG repeat size correlates with age of onset and age of death (Leehey et al., 2007; Ludwig et al., 2014).

FXTAS is one of the most common single-gene disorders leading to neurodegeneration in males. It is caused by moderate expansions (55-200 repeats) of a CGG trinucleotide in the 5'UTR of the *fmr1* gene (premutation range). 200 or more CGG repeats (full mutation) in the same gene cause FXS. About 1 in 800 males and 1 in 260 females of the general population

are carriers of *fmr1* premutations (Dombrowski et al., 2002). Nearly 1 in 3000 men have a lifetime risk of developing FXTAS (Jacquemont et al., 2003). The penetrance in females is much lower and female carriers only infrequently develop FXTAS. Approximately 4% of all female carriers develop milder FXTAS-related symptoms (Coffey et al., 2008) and 20% suffer from premature ovarian failure (POF) (Cronister et al., 1991).

The major neuropathological hallmark for FXTAS is eosinophilic, ubiquitin-positive, intranuclear, neuronal and astrocytic inclusions located in broad distribution throughout the brain (Greco et al., 2006; Iwahashi et al., 2006) and other organs such as adrenals, thyroid, heart, Leyding cells and pancreas (reviewed in Muzar and Lozano, 2014). Numbers of inclusions strongly correlate with the size of the CGG repeats. Analyses on the inclusions revealed association with *fmr1* mRNA and more than 20 proteins, including histones, intermediate filaments, microtubule components, myelin associated proteins. Among these proteins are the stress related protein αB-crystallin, the RNA-binding protein heterogeneous nuclear ribonucleoprotein A2/B1 (hnRNP A2/B1) (Iwahashi et al., 2006; Sofola et al., 2007), the splicing factor Sam68 (Sellier et al., 2010) and the DNA/RNA binding protein Pur-α (Jin et al., 2007). In addition it was revealed that no single protein is predominant or accounts for more than 10% of the protein mass (Iwahashi et al., 2006).

Sellier and colleagues proposed a mechanism, in which Sam68-responsive splicing is altered in FXTAS patients (Sellier et al., 2010). Its depletion in a mouse model leads to motor coordination defects (Lukong and Richard, 2008). It was found that Sam68 is sequestered by mRNAs containing expanded CGG repeats and thereby loses its splicing-regulatory function (Sellier et al., 2010). The same study showed that recruitment of Sam68 is a rather late event during FXTAS particle formation. Therefore its interaction with CGG repeats is thought to be indirect.

Another working hypothesis was that inclusions arise in part of impeded proteasomal degradation and subsequently accumulation of multiple polyubiquitinated protein species. However, it was shown that only a small number of proteins appear to be ubiquitinated and that polyubiquitination is unlikely (Iwahashi et al., 2006). These observations argue against the aggregation models that occur in many expanded CAG (polyglutamine) repeat disorders, in which the misfolded and/or polyubiquitinated proteins fail proteasomal degradation.

1.4.2.1. RNA toxicity – sequestration model

Yrigollen et al. observed that within the CGG repeat track of the *fmr1* gene usually 1-3 AGG interruptions are found in the normal population. These interruptions occur after 9 or 10

uninterrupted CGG repeats. However, none or only one AGG interruption is found in premutation carriers, and the length of pure CGG repeats strongly correlates with *fmr1* mRNA levels (Yrigollen et al., 2011). Furthermore, Kiliszek et al. could show that CGG-RNA repeats form hairpins with non-canonical GG pairs that are thermodynamically more stable than CAG and CUG repeats (Kiliszek et al., 2011).

Although the molecular pathogenesis of FXTAS remains unclear, an RNA gain-of-function mechanism is generally assumed to be the mechanistic trigger of this disease. This is in part based on the observation that levels of premutation *fmr1* mRNA are elevated 5-8-fold, while FMRP levels are decreased due to inefficient translation (Ludwig et al., 2014; Pretto et al., 2014). Also FXTAS has not been observed among older adults with full mutation alleles in whom the *fmr1* gene is silenced (Iwahashi et al., 2006). However recent evidence supports that both FMRP deficits and elevated *fmr1* mRNA levels might contribute to the pathomechanism of FXTAS.

Elevated mRNA levels can be caused by different events. For instance, histones at the *fmr1* promoter with FXTAS premutation show increased acetylation, which in turn leads to increased gene transcription (Todd et al., 2010). Another reason for elevated mRNA levels was suggested by Wang et al. (Wang et al., 1996). They showed that CGG-repeats exclude nucleosomes *in vitro*, which may facilitate the binding of transcription factors *in vivo*. Additionally, CGG-repeats form so-called R-loops, which lead to chromatin decondensation and increased transcription (Loomis et al., 2014; Powell et al., 2013).

Already in 2003 Jin et al. have proposed a mechanism by which elevated levels of CGG-repeat RNA can cause neurodegeneration. Here, CGG-repeat binding proteins become sequestered from their normal functions by the abundant premutation RNA (Jin et al., 2003) (Figure 1.3). Several of such RNA-binding proteins have been uncovered, including Pur-α, Rm62, CUGBP1, hnRNP A2/B1, Sam 68 and DROSHA-DGCR8 (reviewed by Lozano et al., 2014) that are also present in FXTAS inclusions. Sequestration of these proteins might lead to altered transcription, RNA processing, RNA trafficking causing altered protein conformation and function.

Antisense transcripts of the *fmr1* gene might also contribute to RNA toxicity. Similar to *fmr1*, the *antisense fmr1* (*asfmr1*) transcript is elevated in permutation carriers and is not expressed in FXS, suggesting that the antisense transcription may contribute to the pathogenesis of FXTAS and FXS. In addition, the *asfmr1* transcript exhibits premutation-specific alternative splicing that might contribute to disease-related abnormalities (Ladd et al., 2007).

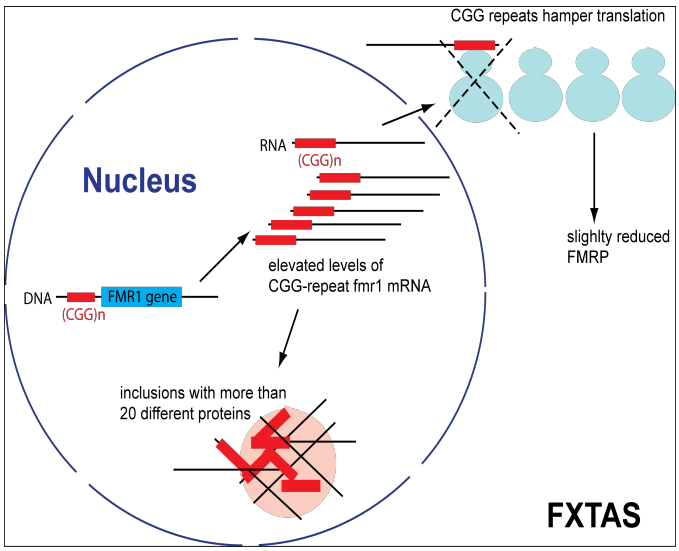


Figure 1.3 Schematic representation of the RNA gain-of-function mechanism proposed for FXTAS. The premutated *fmr1* gene causes enhanced transcription and therefore elevated levels of the *fmr1* mRNA. Though, the expanded CGG repeats hamper the translation at the ribosomes leading to slightly lower levels of FMR1 protein. Amongst others, CGG-binding proteins like Pur-α or hnRNP A2/B1 become recruited to the *fmr1* mRNA, which leads to the formation of intranuclear inclusions and likely results in the pathology of FXTAS. (Figure adapted from Oostra and Willemsen, 2009).

1.4.2.2. RAN translation – homopolymeric proteins

RAN (repeat-associated non-ATG) translation occurs across long, hairpin forming repeats and was first discovered in SCA8 (spinocerebellar ataxia 8), a neurodegenerative trinucleotide repeat-expansion disease (Daughters et al., 2009). The *sca8* gene carries a CTG/CAG repeat expansion in its 3'UTR, which is translated in an ATG-independent manner in all three reading frames, encoding homopolymeric (poly-glutamine, poly-alanine and poly-serine) proteins (Zu et al., 2011). RAN translation has also been shown for the *myotonic dystrophy type 1* gene, which also contains a trinucleotide repeat expansion (CTG) in the 3'UTR and results in poly-glutamine proteins (Zu et al., 2011).

RAN translation in at least two out of three reading frames has also been found for FXTAS. Todd et al. demonstrated that non-canonical translation of CGG-repeat expansions in FXTAS results in expression of polyglycine and polyalanine-containing products. Theses products are toxic in *Drosophila* and in human cell lines and induce intranuclear inclusion formation. Furthermore, the polyG and polyA-proteins are also present in FXTAS patient brains (Reddy and Pearson, 2013; Todd et al., 2013).

1.4.2.3. Therapy

There are no effective targeted therapies for FXTAS treatment, only medications to alleviate some of the symptoms. Memantine (FDA approved drug for Alzheimer's disease) is the only targeted trial for FXTAS. It reduces symptoms by selectively blocking the excitotoxic effects caused by abnormal transmission of glutamate in neurons. Patients treated with memantine showed improvements in cued-recall (Yang et al., 2014). Beta-blocker medications, such as propranolol and primidone have reduced tremor in some FXTAS patients (Hagerman et al.,

2012; Hagerman et al., 2008; Leehey, 2009). Current research focuses on preventing the neurobiological abnormalities in FXTAS with pharmaceutical compounds. The neurosteroid allopregnanolone was shown to ameliorate clustered burst firing in hippocampal permutation-neurons (Cao et al., 2012) and is a potential treatment for FXTAS. Other approaches for treatment focus on the molecular mechanisms of the disease. To understand the molecular causes of FXTAS, animal models (mouse, *Drosophila*) have been developed (Berman et al., 2014; Sofola et al., 2007).

1.4.3. ALS/FTLD

ALS (amyotrophic lateral sclerosis) is a late-onset neurodegenerative disease with symptoms occurring mostly at the age of 40-60 years and affecting men more frequently than women (reviewed by McCombe and Henderson, 2010). Juvenile ALS is rare. However, the disease may start early in life and show clinical features at a much later stage. Most patients diagnosed with ALS die within 3-5 years after disease onset. ALS is the most common motor neuron disease. The key features of the disease are muscle atrophy and weakness, fasciculation and spasticity due to degeneration of lower motor neurons in the brain stem and spinal cord, and loss of the upper motor neurons in the motor cortex (Horton et al., 1976; Robberecht and Philips, 2013). Neurons of the prefrontal and temporal cortex are also affected in ALS, leading to frontotemporal dementia (FTD) in about 15 % of patients (also known as frontotemporal lobar degeneration [FTLD]) (Ringholz et al., 2005). FTLD is the second most common dementia after Alzheimer's disease (Harvey et al., 2003; Van Langenhove et al., 2012) and is clinically, pathologically and mechanistically linked to ALS (Ringholz et al., 2005; Wheaton et al., 2007). Post-mortem examinations of ALS and FTLD patients revealed nuclear and cytoplasmic inclusions of aggregated proteins in neurons as the major pathological hallmark (Neumann et al., 2006). FTLD is classified based on the main protein component (Tau, TDP, FUS and UPS) of inclusions found in cortical neurons (reviewed in Ling et al., 2013). TAR DNA-binding protein 43 (TDP43) is the major ubiquitinated protein present in FTLD inclusions, which is similar to what is found in motor neurons of ALS patients (Arai et al., 2006; Neumann et al., 2006).

With the identification of a disease-causing hexanucleotide (G4C2) repeat expansion mutation in the *C9orf72* (chromosome 9, open reading frame 72) gene, another molecular link between ALS and FTLD was established (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Moreover, Pur-α was found to bind G4C2-repeat RNA. Expression of mutant G4C2-repeat RNA in mouse motoneuronal-like NSC34 and HeLa cells affected the distribution of Pur-α

and FMRP (Rossi et al., 2015). Repeat expansions of the *C9orf72* resembles CGG repeats of the *fmr1* gene. Both repeat RNAs cause sequestration of RNA-binding proteins, inclusion formation and in particular mislocalization of Pur-α (Jin et al., 2007; Sofola et al., 2007).

ALS is classified into two categories: familial ALS (FALS) and sporadic ALS (SALS). 90-95 % of ALS cases are sporadic and do not have an obvious family history. FALS is mostly inherited in an autosomal dominant way and only rarely X-linked or recessive (Andersen and Al-Chalabi, 2011; Pramatarova et al., 1995).

1.4.3.1. Protein toxicity/aggregation

Protein aggregates are the hallmark of many neurodegenerative diseases. The initiating event, which may induce ALS is aggregation of mutant proteins like SOD1 (superoxide dismutase 1), UBQLN2 (ubiquillin-2), VCP (vasolin-containing protein), TDP43 (Tar-DNA binding protein) or FUS/TLS (fused in sarcoma/translocated in liposarcoma) disrupting the normal protein homeostasis, intracellular transport, cytoskeletal architecture and mitochondrial function (reviewed in Robberecht and Philips, 2013). These defects induce cellular stress and in turn lead to axonal retraction and death of motor neurons (Bendotti et al., 2012; Saxena and Caroni, 2011).

20% of FALS is caused by a mutation in SOD1. Most mutations are missense mutations, which lead to misfolding of the protein. Mutant SOD1 protein usually becomes ubiquitinated and subsequently degraded by the proteasome, however, in ALS mutant SOD1 fails to be degraded and even impairs the proteasomal pathway as well as autophagy (Basso et al., 2006). Mutant SOD1 accumulates as oligomers and later as aggregates thereby causing stress response followed by loss of motor neurons.

UBQLN2 belongs to the ubiquitin-like protein family and delivers ubiquitinated proteins to the proteasome. Mutations in UBQLN2 occur in X-linked FALS, FALS, FTLD and SALS. UBQLN2 has been found in skein-like inclusions within motor neurons and the spinal cord together with TDP43 and FUS/TLS (Deng et al., 2011; Williams et al., 2012). Mutations in its conserved proline-rich region (P497H, P497S, P506T, P509S and P525S) have been shown to be associated with ALS and lead to overall impairment of protein degradation (Deng et al., 2011; Gellera et al., 2013; Williams et al., 2012).

Another key factor for proteasomal protein degradation and autophagy is VCP. Being an ubiquitin-sensitive chaperone, VCP unfolds and disassembles protein complexes and thereby enables degradation or recycling. Depletion of VCP leads to protein accumulation and immature autophagosomes (Ju et al., 2009; Tresse et al., 2010). Johnson et al. identified

mutations in the N-terminal region of VCP from ALS patients, which impair the fusion of lysosomes with autophagosomes and thereby disrupt autophagy (Johnson et al., 2010; Ju et al., 2009; Watts et al., 2004).

FUS/TLS and TDP43 are the major protein components of pathogenic inclusions observed in 90 % of ALS and >50 % of FTLD (Arai et al., 2006; Neumann et al., 2006). Both, TDP-43 and FUS/TLS are RNA-binding proteins that are involved in mRNA splicing, transport and translation regulation. Defects in RNA processing were therefore thought to be the major cause for ALS and FTD. Normally TDP43 shuttles from the nucleus to the cytoplasm and associates with RNA transport complexes (Ayala et al., 2008; Kiebler and Bassell, 2006; Zinszner et al., 1997). Mutations in TDP43 cause a shift of nuclear to cytoplasmic localization of TDP43 and an increase of aggregation propensity (Johnson et al., 2009). As a result TDP43 is no longer abundant in the nucleus (loss of function), which induces abnormal RNA processing. Elevated levels of mutant TDP43 in the cytoplasm tend to aggregate and might in turn sequester other proteins and RNAs (gain-of-function) and therefore contribute to the pathomechanism of ALS.

FUS/TLS also shuttles from the nucleus to the cytoplasm. FUS mutations have been found in ALS patients, who also showed FUS-positive inclusions (Kwiatkowski et al., 2009; Sun et al., 2011). All mutations in FUS/TLS found in ALS patients are missense mutations that affect the C-terminal domain containing the NLS (nuclear localization signal) or the glycine-rich region (reviewed in Lagier-Tourenne et al., 2010). These mutations often interrupt the interaction with transportin, which mediates nuclear import of FUS and lead to accumulation of FUS into stress granules that may form inclusions (Dormann et al., 2012; Dormann et al., 2010). Similar to TDP43, it is still unknown if mutant FUS/TLS contributes to ALS by a toxic gain-of-function mechanism (aggregation in the cytoplasm), a loss-of-function mechanism (depletion of FUS in the nucleus) or both.

1.4.3.2. RNA toxicity

Another disease-causing mechanism found in ALS and FTLD are the hexanucleotide (G4C2) repeat expansion mutations located in the first intron of the *C9orf72* gene. The *C9orf72* gene is mostly expressed in the central nervous system (Renton et al., 2011). In the normal population the *C9orf72* gene bears two to five G4C2-repeats and never exceeds 30 repeats. ALS patients show an expansion of ~600-2000 G4C2 repeats (DeJesus-Hernandez et al., 2011; Gijselinck et al., 2012; Renton et al., 2011). ALS caused by *C9orf72* mutations is dominantly inherited. Abnormal repeat expansions are found in 80 % of familial ALS-FTLD, 20-50 % of FALS, 5-10 % of SALS, 10-30 % of FTLD cases and are therefore the most

common cause of ALS and FTLD (Boeve and Graff-Radford, 2012; Chio et al., 2012; Mahoney et al., 2012). The *C9orf72* gene comprises 12 exons, of which 2 are non-coding. By normal alternative splicing, 2 protein-coding mRNAs are transcribed (Renton et al., 2011). The function of the protein isoforms is still unknown, although they are highly conserved in plants, fungi and animals. However, *Drosophila melanogaster* lacks the *C9orf72* gene, but expresses a structural homologue called (differentially expressed in normal and neoplastic cells) DENN-like protein, which plays a role in synapse formation and function (Levine et al., 2013; Zhang et al., 2012).

There are three pathogenic mechanisms that could contribute to the pathogenesis of ALS/FTLD linked to *C9orf72* mutations: Reduced expression of the C9orf72 protein (loss of function), accumulation of repeat-RNA trapping RNA-binding proteins and thereby disrupting RNA processing (gain of RNA toxicity) and accumulation of toxic poly-dipeptides by RAN translation (gain of protein toxicity). The leading pathological mechanism is still unknown.

ALS patients with a *C9orf72* expansion showed a reduction of *C9orf72* mRNA levels by 50 % due to partial or complete silencing of the expanded allele. However reduction of protein level has not been demonstrated (DeJesus-Hernandez et al., 2011; Gijselinck et al., 2012).

Neuronal, intranuclear RNA foci containing *C9orf72* hexanucleotide repeat RNA have been detected by fluorescence *in situ* hybridization experiments using a G4C2 probe (DeJesus-Hernandez et al., 2011; Rossi et al., 2015). These RNA foci might trap RNA-binding proteins, which in turn cannot fulfill their normal functions leading to disruption in RNA processing. Furthermore, protein binding might be enforced by G-quadruplexes formed by G4C2-repeats (Haeusler et al., 2014).

1.4.3.3. RAN translation - dipeptide repeat proteins

Recent studies have demonstrated that RAN translation also occurs at the *C9orf72* transcript. Other than FXTAS, the G4C2-repeat RNA generates dipeptide repeat (DPR) proteins (glycine-proline, glycine-arginine and glycine-alanine). These polydipeptides can be detected with antibodies in neuronal aggregates of *C9orf72*-linked ALS/FTLD patients (Mori et al., 2013). Glycine-alanine dipeptides were the most abundant proteins detected in inclusions of the cerebellum, hippocampus and other brain regions of postmortem patients. The inclusions were similar in shape and abundance to typical ALS/FTD inclusions. This data suggests that dipeptide-repeat proteins generated by RAN translation contribute to ALS/FTLD pathogenesis.

1.4.3.4. Therapy

It has been shown that chronic glutamate excitotoxicity may accumulate to toxic levels and contribute to neuronal death in ALS. Riluzole, an inhibitor of presynaptic glutamate release, is the only currently FDA-approved drug for treatment of ALS. However, riluzole can only prolong the patient's life for a few months (Rothstein, 1996).

One target for therapeutic treatment is the SOD1 protein. Anti-mutant SOD1 antibody and anti-sense oligonucleotides targeting *Sod1* mRNA were shown to reduce the synthesis of the pathogenic mutant SOD1 protein in transgenic mice and rats expressing human SOD1 (Ralph et al., 2005; Raoul et al., 2005; Smith et al., 2006). Reduction of mutant SOD1 protein slowed down the disease progression and is therefore an appealing approach for disease treatment (Gros-Louis et al., 2010).

In general, decreasing aggregation of misfolded proteins, like mutant SOD1, could be one way to delay disease progression. Heat shock proteins play a crucial role in folding/unfolding of proteins and thereby protect cells against stress and apoptosis (reviewed in Li and Srivastava, 2004). Kieran et al. treated mice expressing human mutant SOD1, which is prone to aggregate, with arimoclomol. Arimoclomol induces the heat shock proteins Hsp70 and Hsp90 and thereby slows down protein aggregation and the progressive loss of motor neurons (Kieran et al., 2004).

Another promising treatment for ALS represents transplantation of induced pluripotent stem cells (Robberecht and Philips, 2013) generated from fibroblasts of the ALS patient himself. These reprogrammed pluripotent cells can differentiate again into all cell types and therefore replace diseased motor neurons (Dimos et al., 2008; Glass et al., 2012).

1.5. Pur-α in FXTAS and ALS/FTLD

Pur-α is a DNA/RNA binding protein preferentially binding to purine-rich sequences (GGN motif), which can be found in both disease-related genes *fmr1* (FXTAS) and *C9orf72* (ALS-FTLD). Electrophoretic mobility shift assays (EMSA) with recombinant protein showed that Pur-α binds directly to FXTAS-linked CGG-repeat DNA and RNA with reasonable affinity (Graebsch et al., 2009). Additionally, Pur-α was pulled down from different tissues (mouse and human neurons, COS7 and HeLa cells) using biotinylated CGG- or G4C2-repeat RNA (Jin et al., 2007; Rossi et al., 2015; Sofola et al., 2007). Furthermore, Pur-α has been found to co-localize in FXTAS- and ALS/FTLD-related nuclear and cytoplasmic inclusions of repeat-

RNA expressing cell lines (COS7, HeLa), animal models (mouse, *Drosophila*) and patients (Jin et al., 2007; Sofola et al., 2007; Xu et al., 2013).

Being an important neuronal factor, sequestration of Pur-α into FXTAS and ALS/FTLD inclusions and thereby loss of function might cause disturbance of replication, transcription, mRNA transport and translation leading to neuronal cell death. Overexpression of Pur-α has been shown to rescue the CGG- and G4C2-mediated eye neurodegeneration phenotype of FXTAS and ALS in the fly model (Jin et al., 2007; Sofola et al., 2007). Pur-α also rescued cell viability when co-expressed in mammalian neuronal cells together with G4C2-repeat RNA (Xu et al., 2013).

Qurashi et al. showed that many Pur-α interaction partners are involved in CGG-repeat RNA-mediated neuronal toxicity. One interaction partner is Rm62, the *Drosophila* ortholog of p68 RNA helicase that is implicated in transcriptional regulation, pre-mRNA splicing, RNA interference and nucleo-cytoplasmic shuttling. It was shown that CGG-repeat RNA expression in *Drosophila* decreased the expression of Rm62, leading to nuclear accumulation of mRNAs involved in stress response (Qurashi et al., 2011). Notably, this process might be mediated by the interaction with Pur-α, which is affected in its localization by *C9orf72* expression, and might therefore also play a role in *C9orf72*-mediated neurodegeneration in *Drosophila* (Xu et al., 2013).

Taken together, there are multiple evidences that Pur-α plays a crucial role in the pathogenesis of several neurodegenerative diseases. Therefore, insights into the molecular interactions of Pur-α with disease-related repeat-RNA would significantly help to elucidate its role in ALS/FTLD and FXTAS.

1.6. Objectives

This study consists of two parts. The first part focuses on the molecular principles of Pur-α's binding to nucleic acids and its normal cellular function. The second part deals with Pur-α's role in neurodegenerative diseases.

The first aim of this study was to obtain the crystal structure of a Pur-α/nucleic acid co-complex to understand Pur-α's binding mode, complex stoichiometry and potential differences between DNA- and RNA-binding. Since Pur-α is a multifunctional protein, results could then be used to further clarify its cellular functions and furthermore its role in

neurodegenerative diseases. Additional structural methods, like NMR and SAXS, as well as biochemical assays have been used to accomplish this task.

Pur-α is an important neuronal factor that is involved in replication and transcription of several neuronal genes. Mice with disruption in both alleles of the Pur-α gene die after birth within 4 weeks, suffering from severe neurological defects. By establishing a stable cell line expressing GFP-tagged Pur-α, I wanted to perform Chip-Seq (Chromatin Immunoprecipitation DNA Sequencing) experiments that could reveal which genes become activated by Pur-α, whether they share a consensus sequence and if they can be classified into sub-categories. Findings would also contribute to understand why Pur-α deficient mice die after birth.

Pur-α has been shown to be involved in the pathomechanism of neurodegenerative diseases like ALS/FTLD and FXTAS. In order to understand Pur-α's role in these diseases, the aim of this last part was to establish a stable cell line expressing normal and FXTAS-related CGG-repeat RNA to identify proteins that directly bind to the CGG-repeat RNA and to validate if Pur-α is amongst these proteins. This way, findings for cells expressing normal and disease-related repeat RNA could be compared. Furthermore, it would be possible to identify potential protein interaction partners of Pur-α in the disease context.

2. Results

2.1. Structural and functional studies

Pur-α is a ubiquitous multifunctional protein that binds to both DNA and RNA and is known to regulate replication, transcription and translation (reviewed in White et al., 2009; Gallia et al., 2000). Pur-α also plays an important role in the transport of specific mRNAs to the dendrites in the developing brain and is involved in transcription of several neuronal genes (Darbinian et al., 2001; Dobretsova et al., 2008; Haas et al., 1995). Besides being an essential neuronal factor, Pur-α has been implicated in the pathomechanism of the fragile X-associated tremor/ataxia syndrome (FXTAS), a heritable neurodegenerative disease caused by a moderate expansion of a CGG trinucleotide in the 5'UTR of the *fragile X mental retardation 1* (*fmr1*) gene (Jacquemont et al., 2003). Consistent with its role in FXTAS, Pur-α binds to single- and double-stranded nucleic acids that contain a GGN consensus motif (White et al., 2009). To date Pur-α's binding mode to DNA and RNA is unknown. Hence, different biochemical and biophysical analyses of Pur-α's interaction with nucleic acids have been performed to comprehend the role of Pur-α in its various cellular functions.

2.1.1. Expression and purification of Drosophila and human Pur-α derivatives

For all structural and functional studies Pur-α derivatives (*Drosophila* Pur-α repeat I-II (17 kDa), repeat I-III (26 kDa), repeat III (10 kDa) and human Pur-α repeat I-III (27 kDa)) were expressed and purified as described in chapter 4.13.1 and 4.13.2. After the final purification step by size exclusion chromatography with the corresponding buffer, pooling and concentrating of the peak fractions resulted in protein purities of ≥ 95 % for all constructs. This was monitored on chromatograms and SDS PAGE (Figure 2.1 A-D).

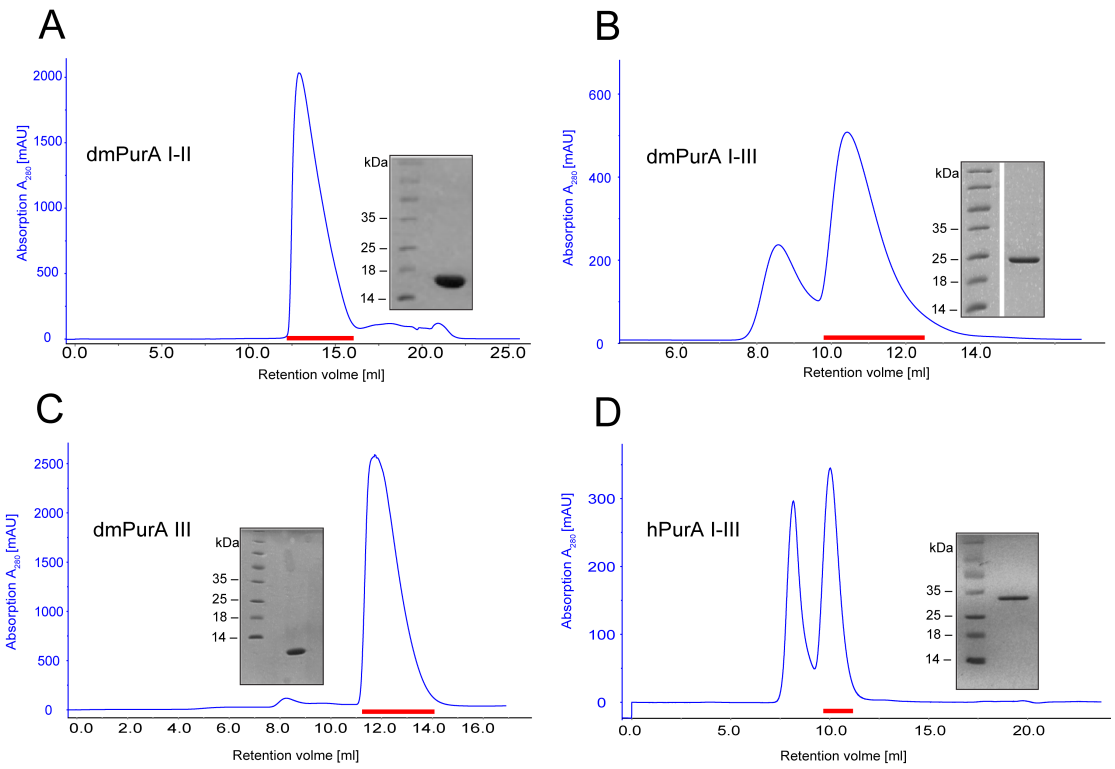


Figure 2.1 Purification of Pur-α protein derivatives used in this study. A-D: Size exclusion chromatogram (blue) of the final purification step with the Superdex 75 10/300 GL column. Peak fractions (red dash) were pooled, concentrated and analyzed on SDS PAGE. **A:** *D. melanogaster* Pur-α I-II (17 kDa). **B:** *D. melanogaster* Pur-α I-III (26 kDa). **C:** *D. melanogaster* Pur-α III (10 kDa). **D:** human Pur-α I-III (27 kDa).

2.1.2. EMSA with Pur-α and DNA/RNA

In order to assess if Pur-α has binding preference for ssDNA or RNA EMSA experiments were performed with the nucleic-acid binding domain of *Drosophila* Pur-α (repeat I-II) and radioactively labeled DNA or RNA oligonucleotides (24 nt) of identical sequence. The MF0677 sequence was chosen as a physiological Pur-α target found upstream of the human *c-myc* gene (Haas et al., 1993; Haas et al., 1995)(Figure 2.2 A). The CGG-repeat sequence was chosen because Pur-α has been found in FXTAS inclusions (Jin et al., 2007; Sofola et al., 2007) and

binds to the CGG repeats at the 5'UTR of the *fmr1* mRNA. When comparing the binding affinity for ssDNA and RNA of the same sequence, only a slight difference is visible (Figure 2.2 A left and right; B left and right). For both nucleic-acid sequences Pur-α has a slightly stronger affinity to ssDNA than to RNA. However, the affinity for the physiological Pur-α target MF0677 is much higher (K_D ~200 nM) than for the disease-related CGG-repeat sequence (K_D ~2 µM) (Figure 2.2 compare A and B; K_D estimated from EMSAs). In all EMSA experiments aggregation and therefore additional shifts were detectable at higher protein concentrations.

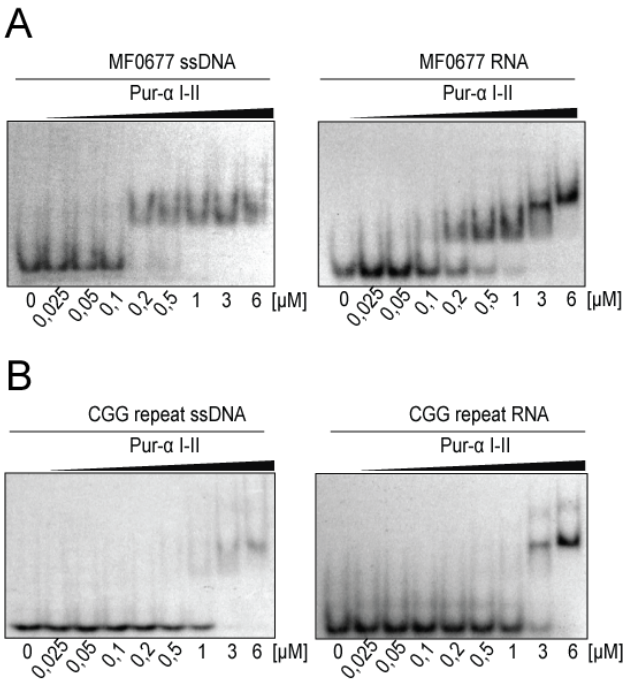


Figure 2.2 Pur-α I-II binds with similar affinities to DNA and RNA. Radioactive EMSA experiments with *D. melanogaster* Pur-α I-II. A: Pur-α I-II binds to MF0677 ssDNA (left) and ssRNA (right) with similar affinity. **B:** Pur-α I-II binds to CGG-repeat ssDNA (left) and RNA (right) also with similar affinity, but less strong than to the MF0677 sequence.

2.1.3. NMR with Pur-α and DNA/RNA

To further examine Pur-α's binding mode to nucleic acids, NMR titration experiments were performed with *Drosophila* Pur-α I-II and DNA/RNA oligonucleotides. ^{15}N-labeled Pur-α I-II was expressed and purified as described in chapter 4.13.1. and 4.13.2. The size exclusion chromatogram and SDS PAGE showed that the protein was pure and stable in NMR buffer (Figure 2.3 A). The aim was to compare DNA with RNA binding and to identify which residues are involved in interaction with nucleic acids and if RNA differs from DNA binding. Additionally, the stoichiometry of the protein-nucleic acid-complex should to be determined. However, since most chemical shift perturbations experienced an intermediate exchange regime throughout the titration steps, saturation and stoichiometric ratio could not be determined. Other problems occurred with the high protein concentration (50 µM) and buffer requirements (low pH, low salt) needed for NMR measurements. Under these conditions the

protein/DNA complex tended to aggregate and precipitate while measuring. Therefore, NMR experiments could not be used for stoichiometry determination.

The $^{1}H^{15}N$ HSQC spectrum of Pur-α alone was nicely dispersed. However, initial RNA- and DNA-titration experiments with the MF0677 and CGG oligomers (24 nt) used in EMSA experiments (chapter 2.1.2) at the high concentrations required for NMR measurements, resulted in protein precipitation. This issue was resolved by performing titrations with short 5'-GCGGA-3' 5mer DNA and RNA fragments. Such short GGN motif RNA and DNA oligonucleotides still bound to Pur-α with reasonable affinity (Figure 2.3 B), but did not induce precipitation. Addition of both oligonucleotides resulted in similar chemical shift perturbations, regardless of whether it was ssDNA or RNA (Figure 2.3 C and D). Also, the spectra revealed that binding most likely occurs at the glutamine/asparagine/arginine side chains, as well as at some backbone residues. In summary, the NMR titration experiments suggest similar binding modes of Pur-α for DNA and for RNA involving identical residues in both cases.

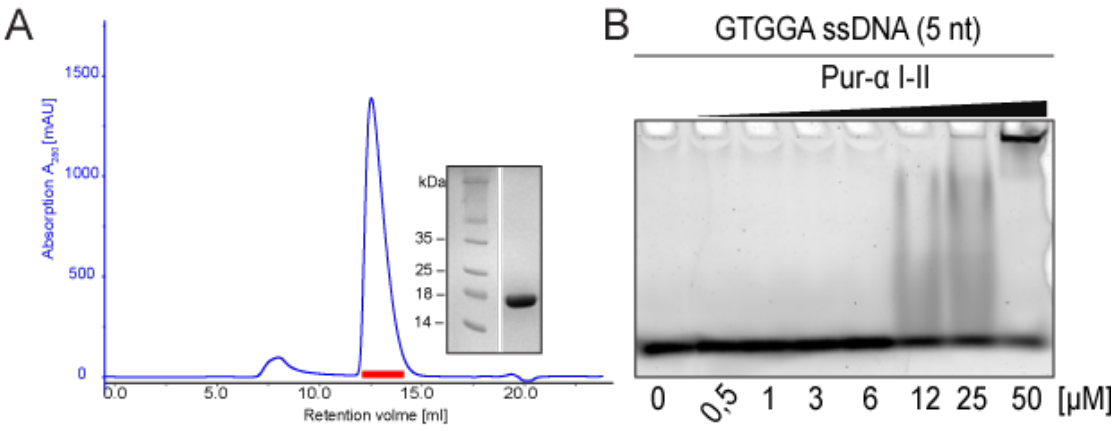


Figure 2.3 Pur-¤ I-II is stable in NMR buffer and binds to short GGN (5mer) oligonucleotides. A: Size exclusion chromatogram (blue) of the final purification step of ^{15}N-labeled Pur-¤ I-II with the Superdex 75 10/300 GL column in NMR buffer. The protein purity was verified by SDS PAGE. **B:** EMSA with unlabeled *Drosophila* Pur-¤ I-II and fluorescent-labeled 5'-GTGGA-3' ssDNA (5 nt).

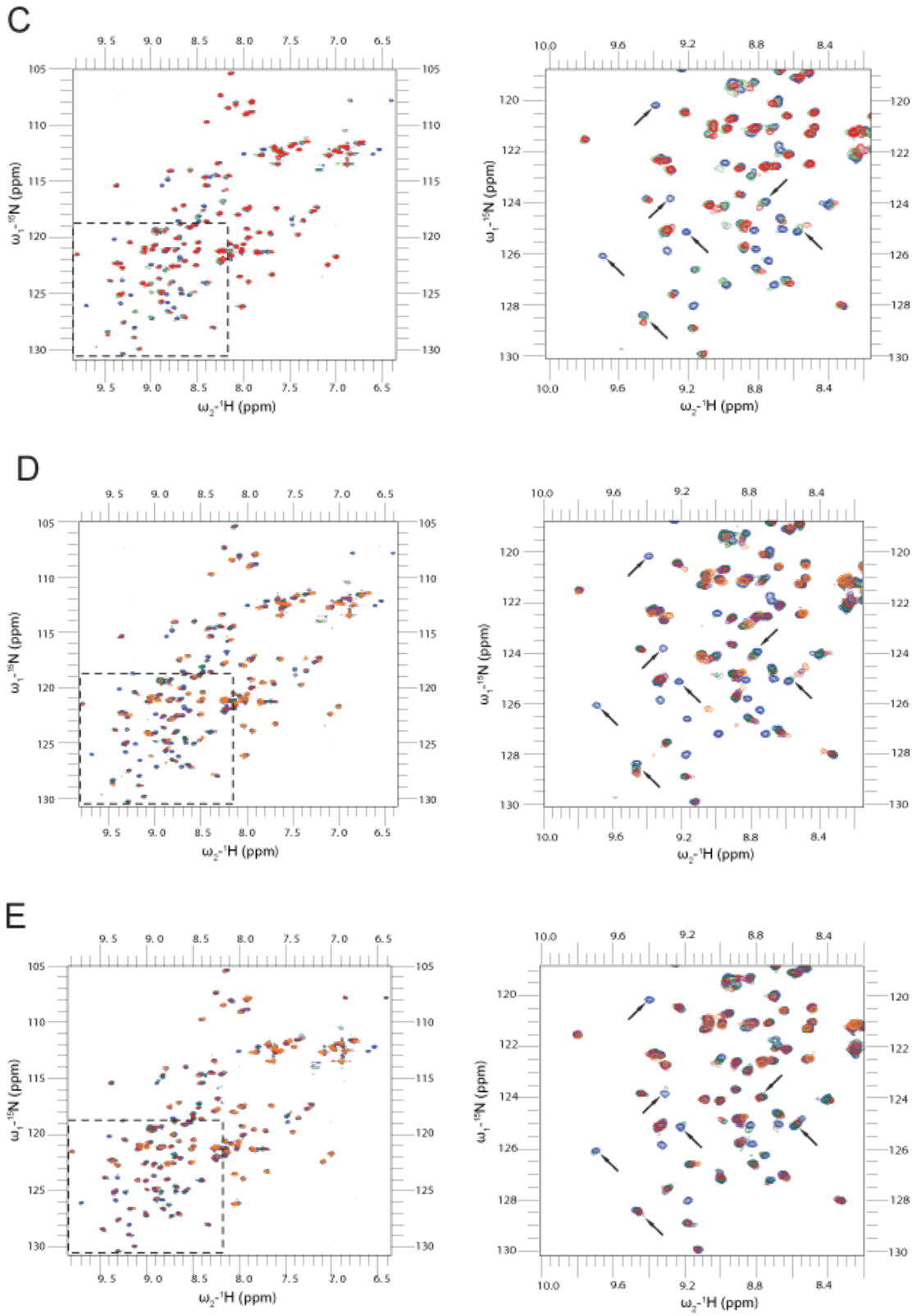


Figure 2.3 NMR titration experiments show that Pur-α's binding mode to DNA and RNA is similar. C-E: ^{1}H-^{15}N HSQC spectra representing NMR titration of the ^{15}N-labeled Pur-alpha I-II with increasing amounts of unlabeled 5'-GCGGA-3' ssDNA and RNA, respectively. Arrows indicate chemical shift pertubations seen upon nucleic acid binding. **C:** (Left) Overlayed spectra of titration with DNA and RNA. The peaks corresponding to the free, DNA-bound (1:1 ratio) and RNA-bound (1:1 ratio) protein states are blue, red and green, respectively. (Right) Close-up on the dashed area with the same color code. **D:** Titration with DNA. (Left) Peaks corresponding to the free and DNA-bound (protein:DNA 1:0.5, 1:1, 1:1.5 and 1:5 ratio) protein states are represented in blue, green, red, mauve and orange, respectively. (Right) Close-up on the dashed area with the same color code. **E:** Titration with RNA. (Left) Peaks corresponding to the free

and RNA-bound (protein:RNA 1:0.5, 1:1, 1:1.5 and 1:5 ratio) protein states are represented. Color code as in **D**. (Right) Close-up on the dashed area with the same color code.

2.1.4. Crystallization of Pur-α/DNA co-complex

Selenomethionine (SeMet)-labeled *Drosophila* Pur-α, comprising Pur-repeat I and II (aa 40-185), was expressed in *E.coli* and purified as described in chapter 4.13.2. Crystallization trials with Pur-α-DNA co-complexes were set up as described in chapter 4.15.1.

Rod-shaped crystals of 160 x 20 µm size (Figure 2.4 A) appeared within 4 days at 21°C in 0.05 M MES; pH 5,6; 8 % PEG400; 0,4 M ammonium sulfate and 1 mM TCEP.

Crystallization trials with native Pur-α protein and DNA in the same conditions did not result in crystals. Alike no protein-RNA crystals grew.

2.1.4.1. Structure determination and refinement

Synchrotron radiation diffraction data were collected at the ESRF (Grenoble, France). SeMet-Pur-α:DNA crystals diffracted up to 2.0 Å resolution (Figure 2.4 B) and belonged to spacegroup P 2_12_12. Cell constants are a = 81.94, b = 40.19, c = 48.81 Å and angles α = β = γ = 90º. A native dataset was recorded at beamline ID 23-2 with a wavelength of 0.8726 Å. A Mar/Rayonix 3x3 Mosaic 225 detector was used to collect 180 frames with an oscillation range of 2°. The detector distance was set to 265.357 mm (Table 2.1).

An anomalous dataset was not recorded.

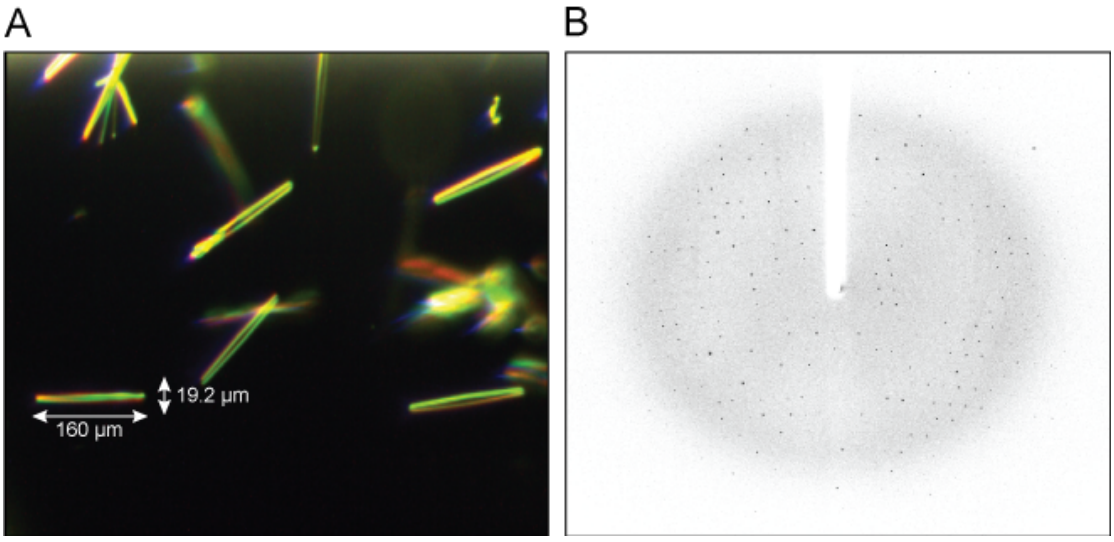


Figure 2.4 Pur-α:DNA crystals diffracted up to 2.0 Å. A: Rod-shaped crystals of SeMet-substituted *Drosophila* Pur-α I-II in complex with CGG DNA (7nt). **B:** Diffraction image of the crystal in **A**.

Data collection		Refinement	
X-ray source	ID23-2 (ESRF)	Resolution (Å)	41.9-2.0
Space group	$P2_12_12$	No. Reflections	11349
		R_{work}/R_{free}	16.3/21.5
Cell dimension		Completeness (%)	99.4
a, b, c (Å)	81.9, 40.2, 48.8		
α, β, γ (°)	90, 90, 90	**No. Atoms**	
		Total	1484
Wavelength (Å)	0.8726	Protein	1207
Resolution (Å)	50-2.0	DNA	145
R_{meas}	12.5 (79.3)	Water	126
I/σ|	18.85 (2.61)		
Completeness (%)	99.4 (94.3)	**B-factors**	
Redundancy	13.1 (7.6)	Protein	24.8
		DNA	30.4
		Water	35.2
		R.m.s deviations	
		Bond lengths (Å)	0.008
		Bond angles (Å)	1.25
		Ramachandran plot (%)	
		Favored	96.03
		Allowed	3.31
		Outliers	0.66
		Rotamer outliers (%)	0.76

Table 2.1 Data collection for the crystal structure of *Drosophila* Pur-α I-II in complex with DNA

2.1.4.2. Crystal structure of Pur-α/DNA co-complex

The structure was solved and refined as described in chapter 4.15.1. and Table 2.1. The asymmetric unit consists of one Pur-α and one ssDNA molecule. The DNA molecule is clamped between the interacting residues of Pur-repeat I and II (Figure 2.5 A).

Plotting the structure with the programme Nucplot (Figure 2.5 B) revealed the following interaction sites: K138 (β-sheet), N140 and R142 (loop region, Pur-repeat II) (Figure 2.5 C), Q52, S53 and K54 (loop region, Pur-repeat I) (Figure 2.5 D). Pur-α rather binds to guanine bases, but also to one of the cytosines (C5) and the sugar phosphate backbone (Figure 2.5 B). Within the crystal lattice the first two DNA bases (G1 and C2) of the 5'-end are base pairing with the 5'-end of the symmetry related DNA molecule (Figure 2.5 F and H). The cytosine C5 in the middle of the DNA strand is twisted and does not stack with the neighboring guanines (Figure 2.5 E). Instead F145 (β-sheet) on Pur-repeat II blocks the space for the cytosine C5

and undertakes the stacking with the upstream guanine G4 (Figure 2.5 E). This might contribute to the stabilization of ssDNA, a known property of Pur-α (Darbinian et al., 2001).

Pur-α consists of three Pur-repeats that share a moderate sequence identity (~30 %) and adopt the same fold (Graebsch et al., 2009 & 2010). Pur-repeat I and II together form the DNA/RNA binding domain whereas repeat III functions as a dimerization domain (Graebsch et al., 2009). Hence, the DNA binding motifs found on Pur-repeat II can also be found on Pur-repeat I. The interacting residues K138, N140, R142 and F145 on Pur-repeat II have their counterpart on Pur-repeat I in position K61 (β-sheet), N63, R65 (loop region) (KNR I) and F68 (β-sheet) (F I). Except for the K61 none of the other residues on Pur-repeat I have been found to interact with the DNA in the crystal structure. Additional binding sites on Pur-repeat I on the symmetry related protein molecule interact with the 3'-end of the DNA (Figure 2.5 F and G): Y57, D59, K61 and K70 and R80 (β-sheet). This interface might therefore represent a second potential DNA/RNA binding site, which is asymmetric to the binding site on Pur-repeat II.

Otherwise, crystallographic packing forces might have prevented DNA binding to the identical motif (KNR) on Pur-repeat I and therefore shifted the binding to a similar motif further downstream composed of the same residues (lysine, arginine) and a tyrosine for guanine base stacking.

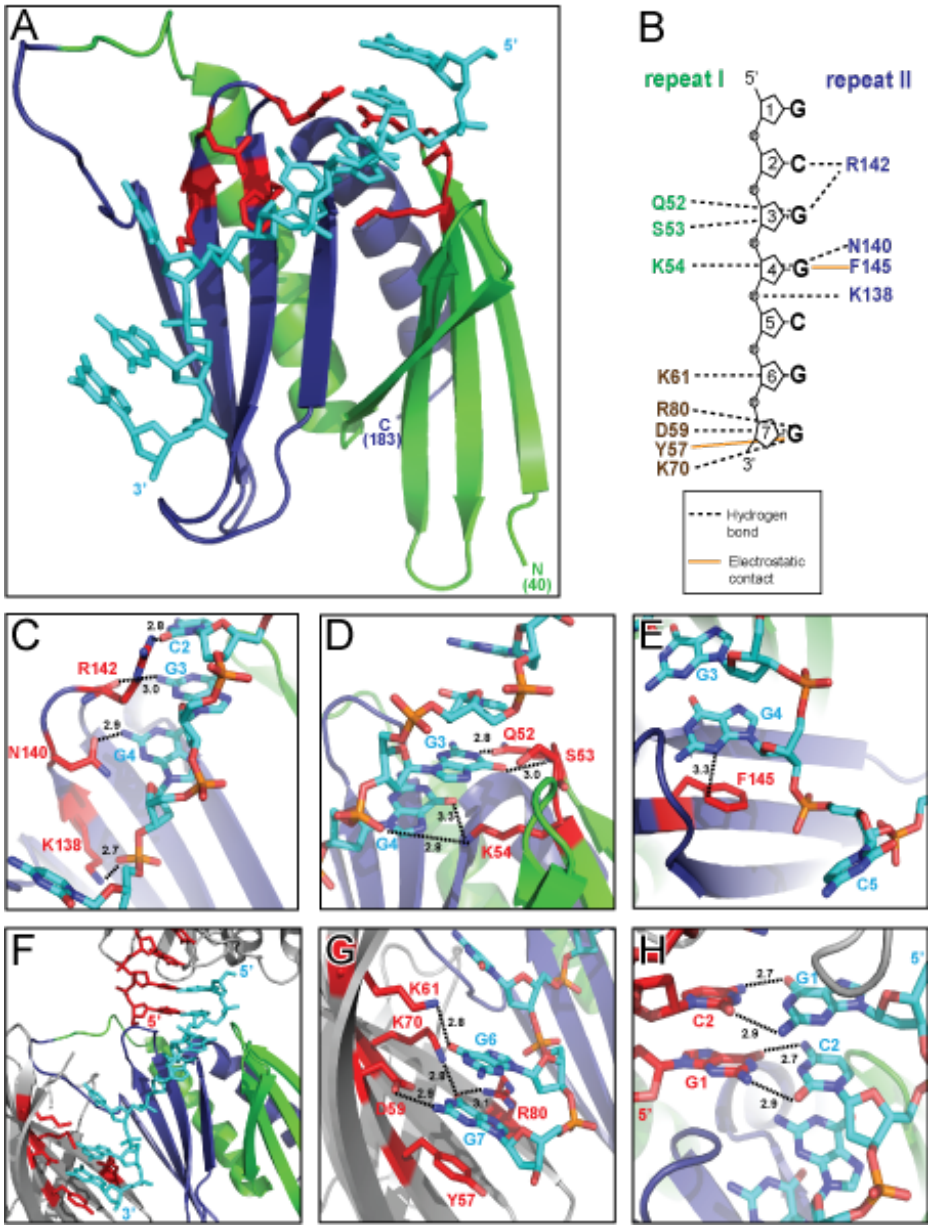


Figure 2.5 Crystal structure of Pur-α I-II in complex with the 5'-GCGGCGG-3' ssDNA. A: Ribbon backbone model of the DNA/RNA-binding domain formed by PUR repeat I (green) and II (blue) in complex with DNA (cyan), represented in sticks. Important protein residues involved in DNA interactions are depicted in red sticks. **B:** Schematic representation of Pur-α interaction with DNA. Both PUR repeats are involved in DNA binding. Pur-α rather binds to guanine bases, but also to one of the cytosines and the sugar phosphate backbone. Residues of the symmetry related protein molecule that interact with DNA are depicted in brown**. C-E:** Detail of the protein-DNA interaction sites. **E:** Phenylalanine 145 stacks with the guanine base G4 and blocks the space for the following cytosine C5. **F:** Within the crystal lattice the 5'-end of the DNA (cyan) anneals with the 5'-end of the symmetry related DNA molecule (red). The 3'-end of the DNA (cyan) is bound by several residues (red) of repeat I of the symmetry related protein molecule (gray). **G:** Detail of the DNA interaction sites with residues (red) of repeat I of the symmetry related protein (grey). **H:** Detail of the G1 and C2 (cyan) base pairing with the symmetry related DNA molecule (red).

2.1.4.3. Structural comparison of Pur-α co-complex and apo-structure

In order to verify if Pur-α undergoes conformational changes upon DNA binding a structural comparison of the Pur-α I-II apo-structure (PDB ID 3K44) and the structure of the protein-DNA co-complex was performed using the SSM (secondary-structure matching) superpose algorithm (Krissinel and Henrick, 2004) of the program COOT. Alignment of these two structures (Figure 2.6 A and B) did not show major conformational differences of free and DNA-bound protein. The root-mean-square deviation (RMSD) of atomic positions obtained a score of 1.14 Å. Excluding the flexible loop region (aa 107-120) from the calculations, the RMSD even improved to a score of 0.83 Å and confirmed that DNA binding does not induce a conformational change of Pur-α.

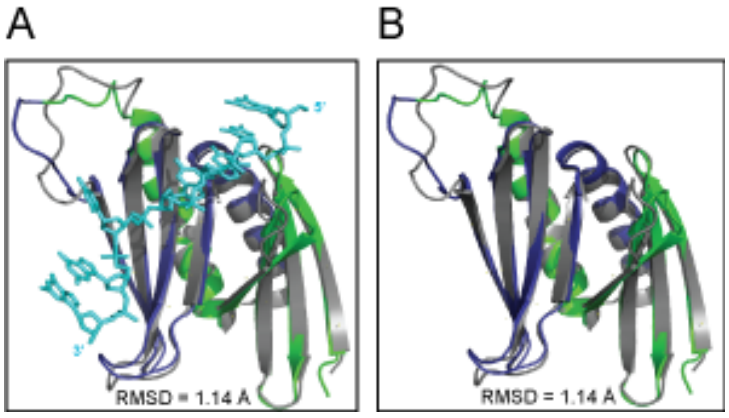


Figure 2.6 Pur-α does not undergo conformational changes upon DNA binding. A and B: Structure alignment of *D. melanogaster* Pur-α I-II apo-structure and in complex with ssDNA (**A**, DNA displayed; **B**, DNA not diplayed). Root-mean-square deviation (RMSD) is indicated in the figures.

2.1.4.4. Sequence alignment of Pur-α from different species

Pur-α is a highly conserved protein from bacteria to mammals. An amino acid sequence alignment of Pur-α from different species (*Drosophila melanogaster*, *Homo sapiens, Danio reiro and Caenorhabditis elegans*) was done to see if the interaction sites seen in the crystal structure are conserved residues (Figure 2.7). Indeed, Q52, S53 and K54, which form one of the two main interaction sites on Pur-repeat I (QSK I) are highly conserved. The opposing binding site on Pur-repeat II composed of K138, N140 and R142 (KNR II) is also highly conserved. The DNA twisting seen in the crystal structure is a special feature of Pur-α's interaction with DNA. Flipping out of the cytosine C5 is caused by stacking of F145 on Pur-repeat II (F II) with the guanine G4 base and thereby blocking the space for the cytosine. The sequence alignment shows that also the F145 (F II) is a highly conserved residue and might therefore be of functional importance. The high conservation of the interacting residues presents the basis for further DNA/RNA interaction studies with Pur-α mutants. The binding motif KNR and F on Pur-repeat II can also be found on Pur-repeat I (KNR I and F I) and

share a high conservation score. Hence, mutations of these residues were also considered for functional assays with Pur-α.

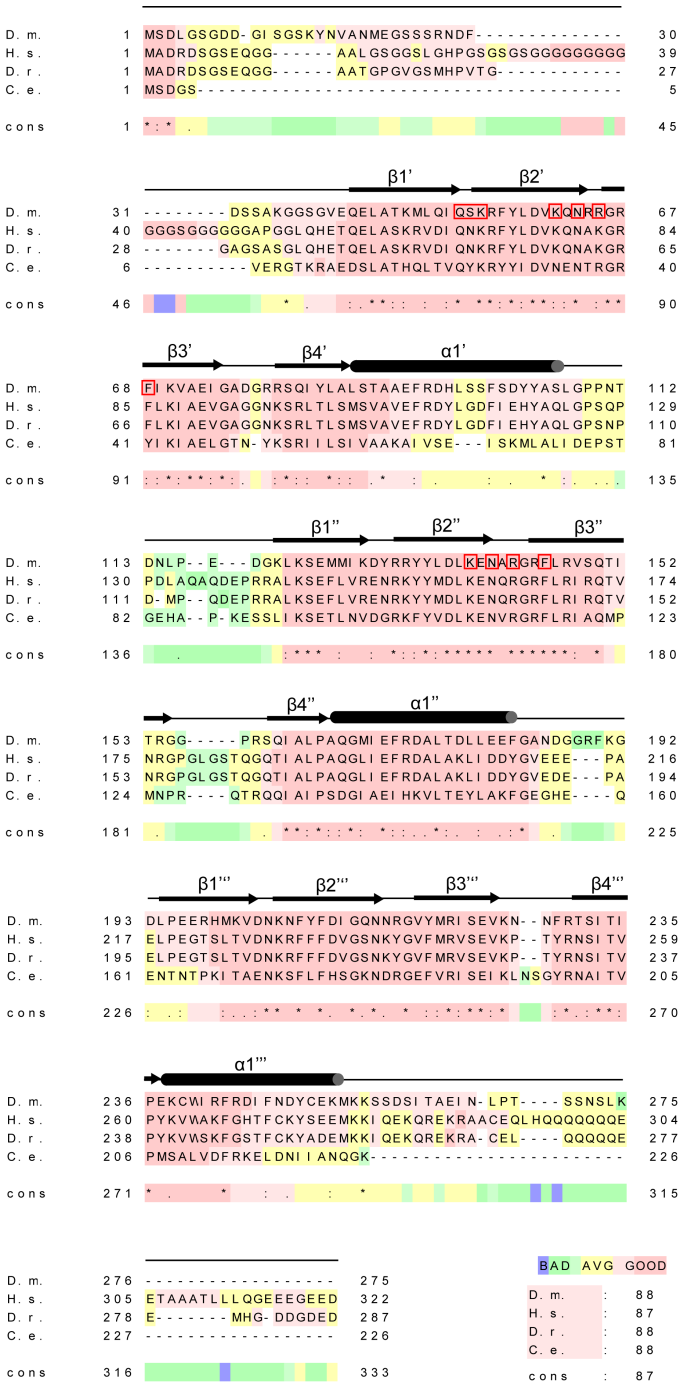


Figure 2.7 Amino acid sequence alignment of Pur-α from different species. D.m., *Drosophila melanogaster*; H.s., *Homo sapiens*; D.r., *Danio rerio*; C.e., *Caenorhabditis elegans*. Color-coding from blue to red reflects the range of sequence conservation from 0-100 %. Asterisk indicates positions, which have a single, fully conserved residue. Colon indicates conservation between groups of strongly similar properties. Period indicates conservation between groups of weakly similar properties. Secondary structure assignment is based on the crystal structure of Pur-α I-II. Secondary structure prediction of Pur-repeat III is based on sequence alignment with Pur-repeats I and II. Red boxes indicate mutation sites for DNA/RNA interaction studies.

2.1.5. EMSA with mutant Pur-α and DNA/RNA

To verify the interaction sites seen in the crystal structure and to test if there is a second binding site on Pur-repeat I, several point mutations have been introduced and their effect on DNA/RNA binding has been analyzed in radioactive EMSA experiments. Residues of the DNA binding site on Pur-repeat I and II and the potential second binding site on Pur-repeat I have been chosen for site-directed mutagenesis. The respective residues have been replaced by alanine: K61, N63, R65 (KNR I, triple mutant), K138, N140, R142 (KNR II, triple mutant), F68 (F I), F145 (F II). For the triple mutant Q52, S53, K54 (QSK I) residues were replaced by glycine, since the alanine-mutant versions of the protein tended to aggregate during purification. Likewise, the residues of the mutant Q52, S53, K54, K138, N140, R142 (QSK I - KNR II) were replaced by glycine (Q52, S53, K54) and by alanine (K138, N140, R142), respectively. The mutant proteins were expressed and purified as described in chapter 4.13.1. and 4.13.2. EMSA experiments were performed with CGG-repeat and MF0677 DNA/RNA oligomers (24 nt). All Pur-α I-II mutants show decreased binding to both DNA/RNA motifs (Figure 2.8 A-H) and, as in chapter 2.1.2, showed similar reductions in affinity for DNA and RNA oligomers with identical sequences. However, for the F II mutant binding to RNA seems to be only slightly decreased (Figure 2.8 E and F). Circular dichroism (CD) spectroscopy showed that all generated Pur-α I-II mutants were correctly folded (Figure 2.9) and verified the composition of α-helix and β-sheet structures.

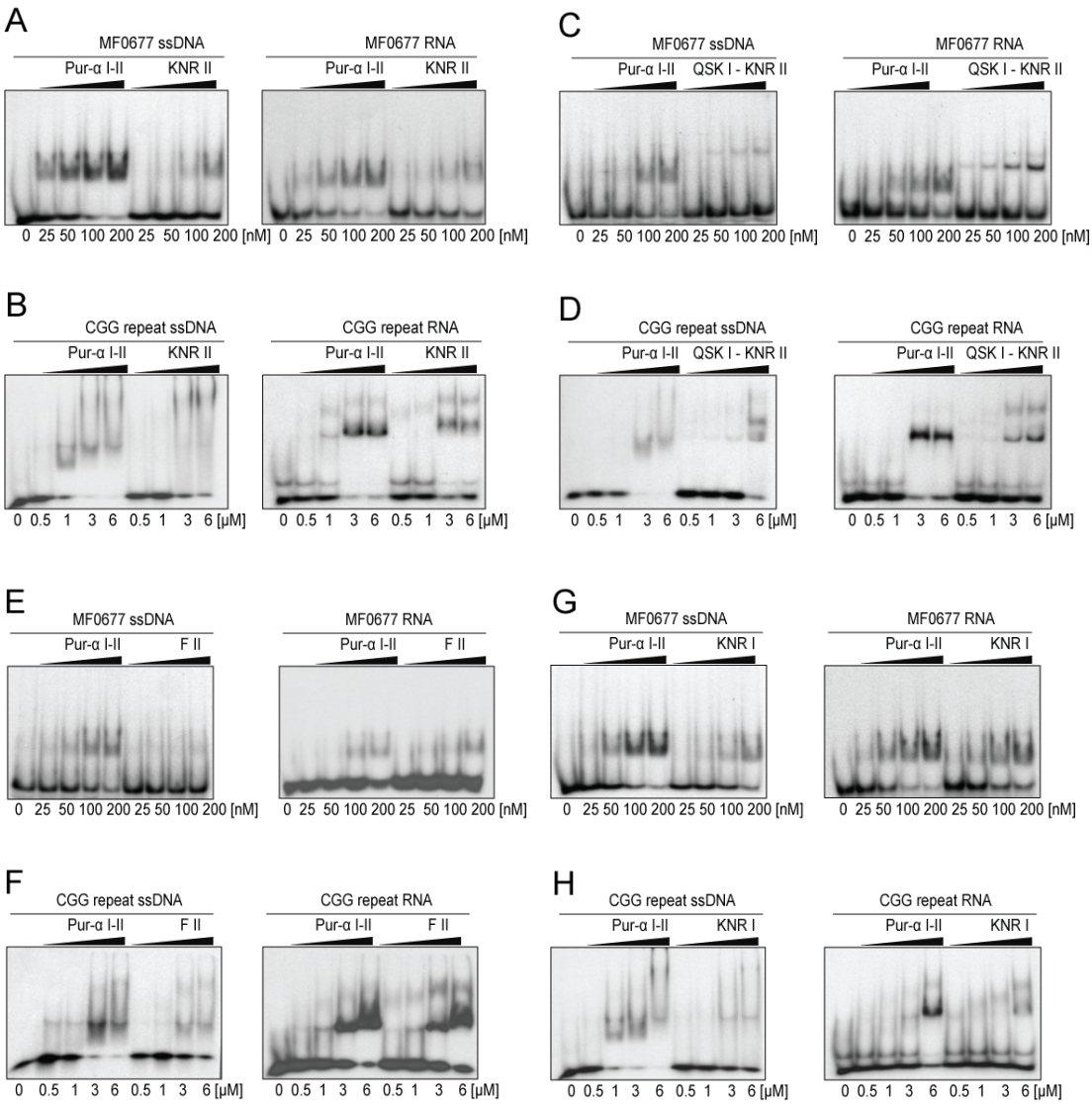


Figure 2.8 *Drosophila* Pur-α I-II mutants show decreased binding affinity to DNA and RNA. A, C, E, G: EMSA with wild type and mutant Pur-α I-II and MF067 ssDNA (left) and RNA (right), respectively. **B, D, F, H:** EMSA with wild type and mutant Pur-α I-II and CGG-repeat ssDNA (left) and RNA (right).

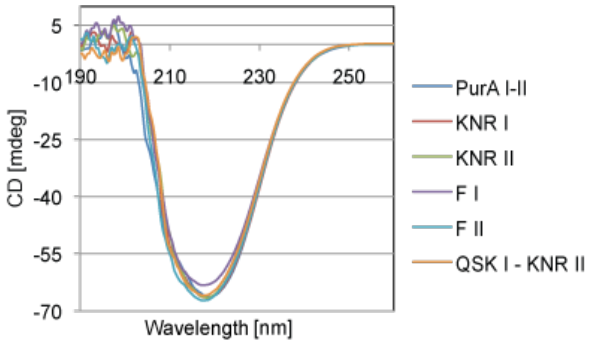


Figure 2.9 Folding of the wild type Pur-α I-II and the mutants are equal. Overlay of CD spectra of *Drosophila* Pur-α I-II (depicted in blue) and mutants (each depicted in a different color).

2.1.6. EMSA with Pur-α III and DNA/RNA

Previously it has been shown that the third Pur-repeat hardly contributes to DNA/RNA binding, but that it is necessary for Pur-α dimerization (Graebsch et al., 2009). Here DNA/RNA binding of Pur-α III alone was analyzed in radioactive EMSA experiments with CGG-repeat and MF0677 DNA/RNA oligomers (24 nt). Neither the CGG-repeat sequence nor the physiological Pur-α target sequence MF0677 bound to Pur-α III as strong as to Pur-α I-II (Figure 2.10 A and B). The main DNA/RNA interaction might therefore occur via the first two Pur-repeats.

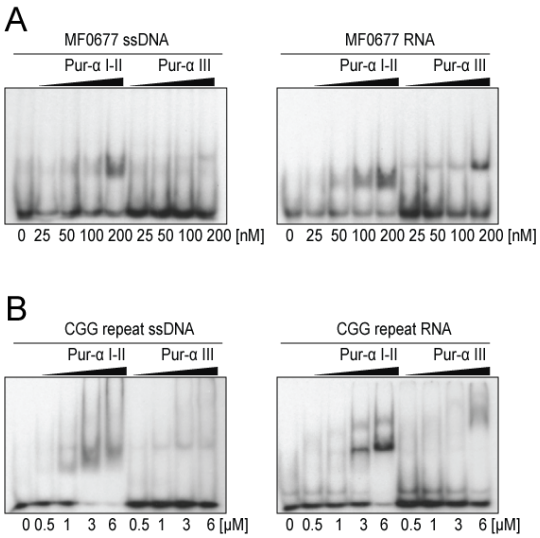


Figure 2.10 *Drosophila* Pur-α III shows only weak binding affinity to DNA and RNA. A: EMSA with wild type Pur-α I-II and Pur-α III and MF067 ssDNA (left) and RNA (right), respectively. **B:** EMSA with wild type Pur-α I-II and Pur-α III and CGG-repeat ssDNA (left) and RNA (right), respectively.

2.1.7. EMSA with Pur-α I-II and FTD/ALS-related repeat RNA

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease, which causes death of motor neurons (reviewed in Robberecht and Philips, 2013). 15% of ALS patients also suffer from frontotemporal lobar degeneration (FTLD) (Ringholz et al., 2005), which is the second most common dementia after Alzheimer's disease (Harvey et al., 2003; Van Langenhove et al., 2012). Similar to the Fragile X-associated tremor/ataxia syndrome (FXTAS) the major neuropathological feature of ALS/FTLD are neuronal, proteinaceous inclusions. Inclusions in ALS/FTLD are caused by hexanucleotide (G4C2) expansion mutations in the non-coding region of *C9orf72* locus, which accounts for up to 25% of FTD and 40 % of familiar ALS (Stepto et al., 2014). The pathogenic function of the repeat-expanded RNA arises from the formation of RNA foci containing hairpin structures and G-quadruplexes. These RNA structures cause sequestration of RNA binding proteins and, alike in FXTAS, Pur-α has been found to be part of FTD/ALS inclusions. In order to validate if Pur-α can bind to the ALS-related RNA radioactive EMSA experiments with (G4C2)4 RNA oligomers have been performed. Pur-α shows a similar affinity to G4C2-repeat RNA like to CGG-repeat RNA

(estimated K_D ~2 µM). Mutation in the DNA/RNA binding site (KNR II) decreases the affinity and leads to the assumption that both RNA sequences are bound in a similar way (Figure 2.11).

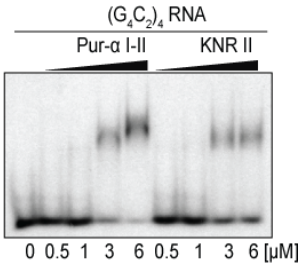
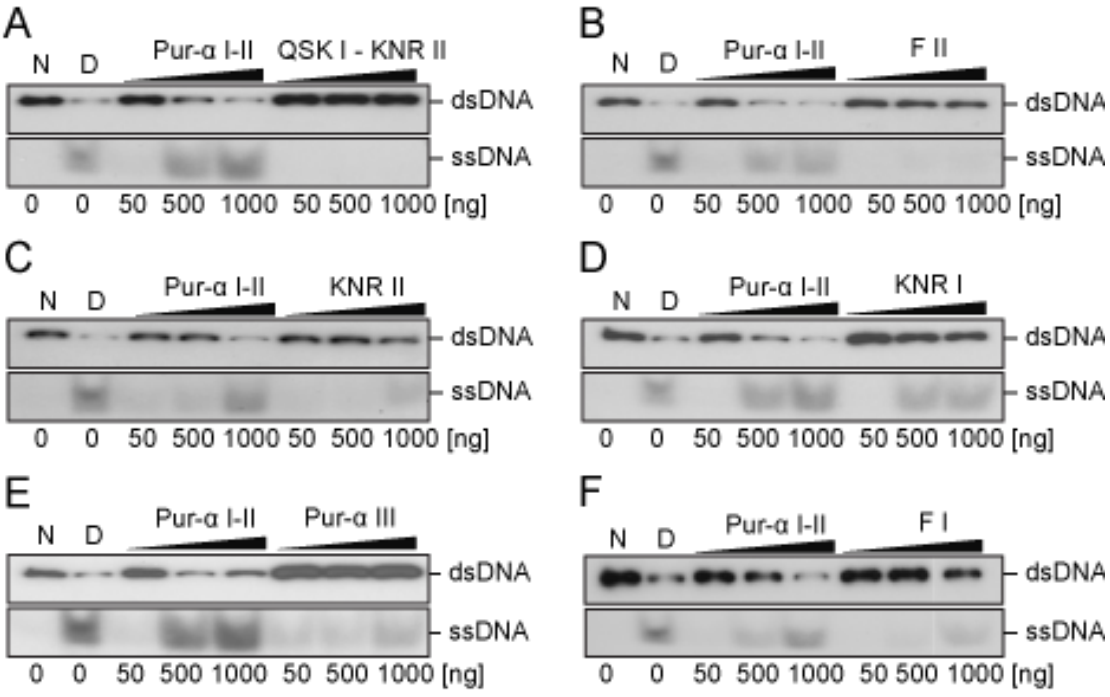


Figure 2.11 *Drosophila* Pur-α I-II KNR II mutant shows decreased binding affinity to ALS-related repeat-RNA. EMSA with wild type Pur-α I-II and KNR II mutant and (G4C2)4 RNA.

2.1.8. Unwinding assay

Pur-α is able to unwind dsDNA containing a GGN motif in an ATP-independent manner (Darbinian et al., 2001; Wortman et al., 2005). For this the protein contacts the purine-rich strand and displaces the pyrimidine-rich strand (Wortman et al., 2005). It has been postulated that Pur-α, being a transcription activator, binds to promoter regions, displaces the two DNA strands so other proteins can bind and activate transcription (Darbinian et al., 2001).

The crystal structure of the Pur-α-DNA co-complex shows that the normal base stacking of the ssDNA is disrupted and that the cytosine C5 is flipped out because the F145 takes over the stacking with the upstream guanine G4 (chapter 2.1.4.2). Additionally, the 5'-end and the 3'-end of the ssDNA are bound by two Pur-α molecules within the crystal lattice, which enforces the unusual twisting. This particular feature of DNA binding might explain how DNA double-strand separation could occur. To test if the binding mode seen in the crystal structure is essential for the strand displacement of dsDNA unwinding assays with mutant Pur-α and dsDNA substrates were performed. Radiolabeled ssDNA oligomer containing a GGN motif was annealed to an ssDNA plasmid. The resulting partly double-stranded substrate was incubated with Pur-α wild type and mutants and strand separation has been monitored on polyacrylamide gels. When both binding sites on Pur-repeat I and II were mutated (QSK I – KNR II) unwinding was not possible anymore (Figure 2.12 A), most likely because DNA binding was impaired as seen in EMSA experiments (chapter 2.1.5, Figure 2.8 C). An unwinding assay with the F145 (F II) mutant confirmed that flipping out of one base is essential for unwinding and that F145 plays an important role in this context. Mutating only the F145 already abolished the unwinding activity (Figure 2.12 B), whereas all other mutations showed only decreased unwinding (Figure 2.12 C-F). Results of all binding and unwinding experiments for all Pur-α derivatives are summarized in Figure 2.12 G.



G

+++ strong (wt-like) ++ moderate + weak – abolished		biniding affinity				unwinding ability
		CGG repeats (µM)		MF0677 (nM)		GGN
		ssDNA	ssRNA	ssDNA	ssRNA	dsDNA
WT	Pur-α I-II	+++	+++	+++	+++	+++
WT	Pur-α III	+	+	+	+	+
Mutants	KNR I	+	+	++	++	++
Mutants	KNR II	+	++	+	+	+
Mutants	F I	+++	+++	+++	+++	++
Mutants	F II	+	++	+	++	–
Mutants	QSK I - KNR II	+	+	+	+	–

Figure 2.12 Mutations in Pur-α I-II decrease dsDNA unwinding ability. Unwinding assay with different *D. melanogaster* Pur-α I-II mutants and Pur-α III. Protein was titrated to a partly dsDNA substrate containing a GGN motif. Pur-alpha I-II is able to separate the DNA strands, whereas mutations in both repeats **(A)** and the mutation of the phenylalanine 145 in repeat II **(B)** abolish the unwinding ability. Mutations in repeat II only **(C)** or in the identical motif of repeat I **(D)** decrease the unwinding ability. **(E)** Pur-alpha III shows only weak unwinding activity. Decreased unwinding also occurs upon mutation of phenylalanine 68 in repeat I **(F)**. The table **(G)** summarizes the results of all EMSA and unwinding experiments of all Pur-α derivatives and mutants.

2.1.9. Small angle X-ray scattering (SAXS)

Pur-repeat I and II share a high sequence and structure similarity and together form the DNA/RNA binding domain. The crystal structure of Pur-α I-II in complex with DNA shows that there are two binding sites and demonstrates that one DNA/RNA binding domain binds

two molecules of nucleic acid. Due to crystal packing only one molecule of DNA might be able to bind to each DNA/RNA binding domain. Since stoichiometry of the Pur-α/DNA co-complex could not be determined in NMR titration experiments, small angle x-ray scattering (SAXS) was applied. Therefore, several measurements were done with the MF0677 (24 nt) and CGG (7 nt) ssDNA and either Pur-α I-II or Pur-α I-III, a dimer in solution. SAXS measurements with the MF0677 oligonucleotide failed due to precipitation and addition of CGG ssDNA to Pur-α I-II or Pur-α I-III at a 1:1 ratio resulted in formation of polydisperse oligomers, therefore these data were not further evaluated. Addition of a two-fold excess of DNA caused precipitation and therefore samples could not be measured. Precipitation was probably caused by the relatively high protein concentration (50 µM) needed for signal detection. Calculation of a stoichiometric ratio for a Pur-α/DNA co-complex was therefore impossible. To prevent oligomerisation the ionic strength of the buffer was increased from 250 mM to 500 mM and 1 M NaCl. This prevented on the one hand uncontrolled oligomerisation, but on the other hand hindered dimerization of Pur-α I-III so that the results obtained from SAXS measurements were not reliable anymore.

2.1.10. Dot blot assay

Since the stoichiometry of the Pur-α/DNA co-complex could neither be determined in NMR titration experiments nor in SAXS measurements a different technique was used that does not need a high protein concentration and therefore avoids oligomerisation and aggregation. Though, to determine the complex stoichiometry the nucleic acid concentration needs to be above the K_D to ensure that all protein is bound. The estimated K_D for Pur-α I-II and MF0677 (24 nt) ssDNA was approximately 150 nM (Figure 2.2 A). In a so-called dot blot assay Pur-α I-II was titrated to a constant amount (1 µM) of radiolabeled MF0677 (24 nt) ssDNA and blotted onto a nitrocellulose membrane (Figure 2.13 A). Only protein or protein/DNA complexes can bind to the nitrocellulose membrane whereby unbound oligomers are washed away. The radioactive signal on the membrane was measured with a phosphor imager system as an indicator for radiolabeled DNA bound to Pur-α. By plotting the signal intensity against the protein concentration saturation curves were obtained. In the case of a 1:1 ratio (protein:DNA), saturation would be reached at a protein concentration of 1 µM. In the case of a 1:2 ratio (protein:DNA) saturation would already occur at 0.5 µM. As shown in Figure 2.13 A (from left to right), in three independent experiments saturation was already reached at protein concentrations of 0.45, 0.64 and 0.64 µM, respectively. Taken together, saturation

occurs at a mean value of 0.58 ± 0.1 µM of Pur-α (Figure 2.13 B), which indicates a stoichiometric ratio of 1:2 (protein:DNA). Thus, in solution Pur-α I-II can bind two molecules of ssDNA. This result is consistent with the two binding sites observed in the crystal structure (Figure 2.5).

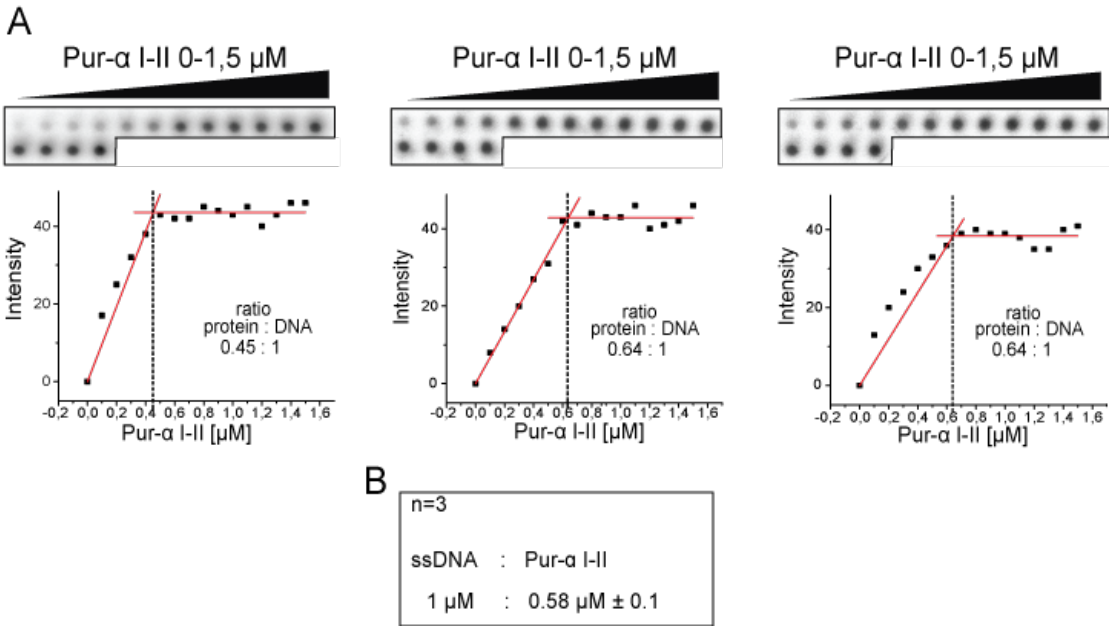


Figure 2.13 One Pur-α I-II molecule can bind two ssDNA molecules. A: (From left to right) Three independent filter binding assays with *D. melanogaster* Pur-α I-II and MF0677 ssDNA. Nitrocellulose filters (top) show the titration of Pur-α I-II (0-1.5 µM) to a constant amount of ssDNA (1 µM). The measured intensities from the filters were quantitatively analyzed. The graphs (bottom) show when saturation is reached. **B:** The mean protein concentration at which saturation is reached was calculated from all three assays. Results indicate that one Pur-α I-II molecule binds two ssDNA molecules.

2.1.11. DNase I footprint

Pur-α is a DNA/RNA binding protein that binds to purine-rich elements, which can be found at promoter regions (Haas et al., 1995) and origins of replication (Darbinian et al., 2001, Gallia et al., 2000). One known promoter region where Pur-α binds to is the MF0677 sequence upstream of the *c-myc* gene (Haas et al., 1995). To investigate Pur-α's binding mode to its physiological target, DNaseI footprint assays were performed using the DNA sequence upstream of the *c-myc* gene (myc-MF0677).

First, radioactive EMSA experiments with Pur-α and the physiological target sequence were performed in order to estimate the required protein concentration for footprint assays. Estimated from the remaining band of free DNA seen in the gel, *Drosophila* Pur-α I-II bound to myc-MF0677 ssDNA (150 nt) with a K_D of approximately 5 µM (Figure 2.14A). Human Pur-α I-III bound to the myc-MF0677 dsDNA (160 bp) substrate with a similar affinity

(Figure 2.14 B). The range of protein concentration chosen for footprint assays was therefore 2-10 µM.

DNase I footprints were carried out as described in chapter 4.16.8. Pur-α was incubated with radiolabeled myc-MF0677 DNA previous to DNase I digestion. Sites where Pur-α binds to should be protected from DNase I cleavage and therefore appear in the gel as "missing bands" when compared to the digested DNA not incubated with Pur-α (control). Different amounts of Pur-α and DNase I were applied. The sequencing reaction loaded on the same gel allows for the determination of the nucleotide sequence where Pur-α binds.

The footprint assay with *Drosophila* Pur-α I-II and myc-MF0677 ssDNA (150 nt) showed a protected site that is located at the 3'-end of the MF0677 subfragment used in EMSA experiments (chapter 2.1.2) (Figure 2.15 A). Surprisingly the protected region did not include a GGN motif. To test if Pur-α really binds to the identified sequence, EMSA experiments with DNA oligomers of the protected sequence and the unprotected 5'-end of the MF0677 sequence were performed. The EMSA assays showed that Pur-α binds with high affinity to the 5'-end of the MF0677 oligomer that contains a GGN motif and was not protected in the DNase I footprint (Figure 2.16 B), but only with a very low affinity to the 3'-end of the MF0677 oligomer that does not contain a GGN motif, but was protected in the footprint (Figure 2.16 A). Results of this footprint assay were therefore unreliable and could not be reproduced either.

Also footprint assays with dsDNA (myc-MF0677, 160 bp) and human Pur-α containing all three Pur-repeats did not give any conclusion about Pur-α's binding mode to its physiological target, since no protected sites were detectable (Figure 2.15 B).

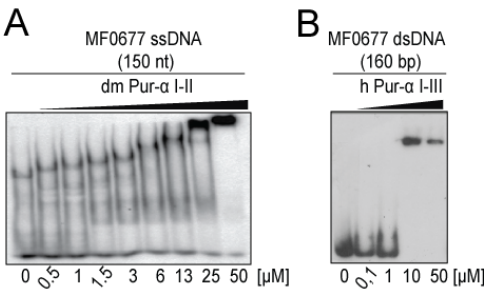


Figure 2.14 Pur-α binds to the MF0677 ss/dsDNA used in DNaseI footprints.
A: EMSA with *Drosophila* Pur-α I-II and radiolabeled MF0677 ssDNA 8150 nt).
B: EMSA with human Pur-α I-III and radiolabeled MF0677 dsDNA (160 bp).

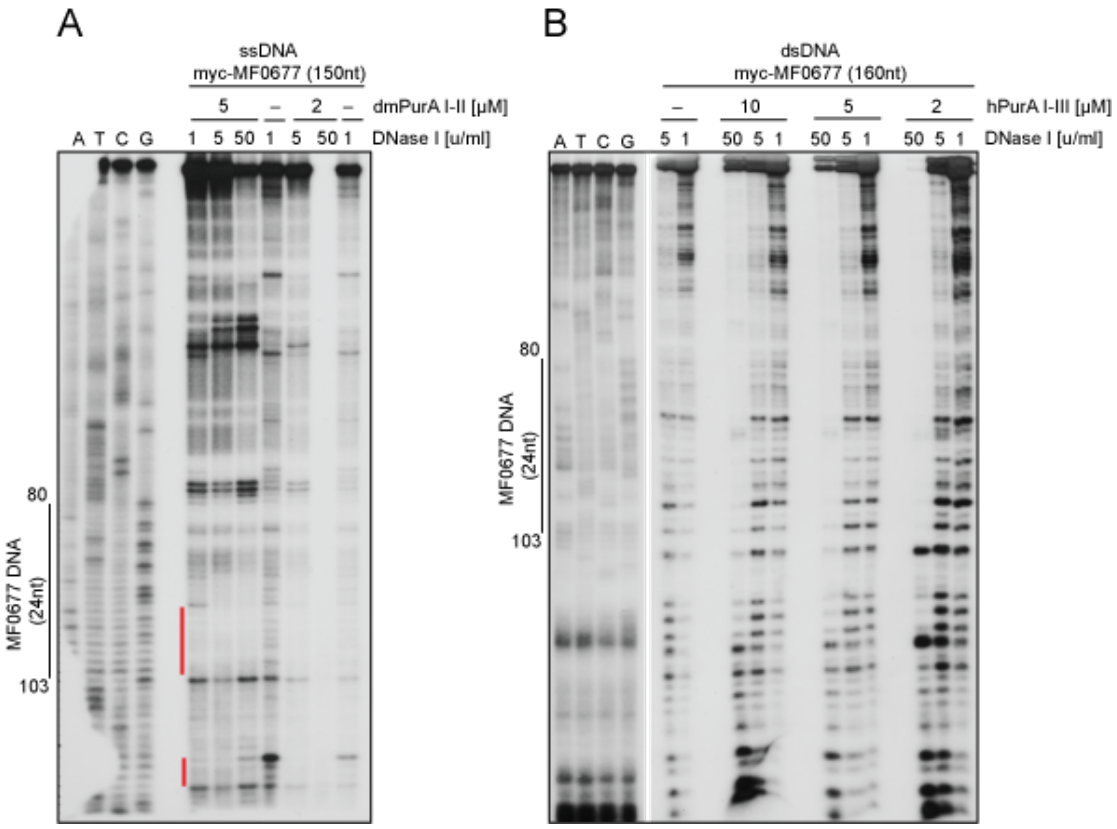


Figure 2.15 DNase I footprint assays. A: ssDNA myc-MF0677 (150 nt) incubated with *D. melanogaster* Pur-α I-II was digested with DNase I**. B:** DsDNA myc-Mf0677 (160 nt) incubated with human Pur-α I-III was digested with DNase I. Digestion of ssDNA and dsDNA, respectively, without protein was used as a control. Protein concentrations and DNase I concentrations were varied as depicted in the figure. Probes were run on a polyacrylamide gel. For sequence determination a sequencing reaction was also loaded on the gel. The region of the MF0677 DNA (24 nt) sequence used in EMSA experiments and filter-binding assays is indicated. Protected sites are depicted with a red dash. For the dsDNA footprint no protected sites were detectable.

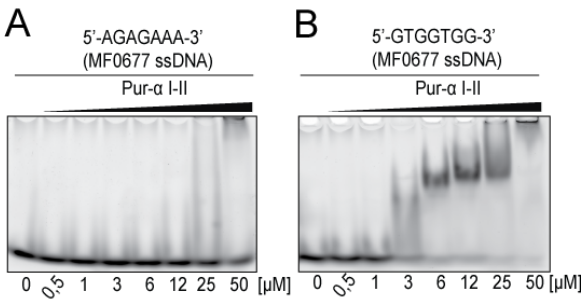


Figure 2.16 Pur-α does not bind to the sequence, which is protected in DNase I footprint assays, but to the GGN motif of the MF0677 sequence, which is not protected in the footprint assay. EMSA with Drosophila Pur-α I-II and fluorescently labeled ssDNA. **A:** EMSA with the "protected" DNA sequecence. **B:** EMSA with "unprotected" GGN motif of the MF0677 sequence.

2.2. Cell culture studies

2.2.1. Cellular function of Pur-α

Many reports suggest a role of Pur-α in activation of transcription and replication. It has been shown that Pur-α possesses dsDNA destabilizing activity (this study, Darbinian et al., 2001; Wortman et al., 2005) and is involved in transcription of several neuronal genes including the

myeline basic protein gene that is important for myelination of nerves in the central nervous system (CNS) (Darbinian et al., 2001; Haas et al., 1995). Mice with disruption in the *Pur-α* gene in both alleles (PURA -/-) appear normal at birth, but after 2 weeks they develop neurological problems and die after 4 weeks (Khalili et al., 2003). To gain information about Pur-α's role in replication and transcription and to better understand why Pur-α knockout mice die after birth, Chip-Seq experiments with human Pur-α were planned.

2.2.1.1. CHIP-Seq

CHIP-Seq is an *in vivo* method that combines chromatin immunoprecipitation (CHIP) with DNA sequencing (Seq). This method uses antibodies to capture proteins that are either directly bound to DNA or indirectly through protein-protein interaction. Thus, binding sites of DNA-associated proteins can be identified. Chip-Seq assays could thereby answer the question, which genes become activated by Pur-α binding, if they share a consensus sequence, and if they can be classified into functional groups.

2.2.1.2. Antibody test against human Pur-α

For CHIP an antibody is needed that specifically targets the protein of interest. Therefore, one commercial and two in-house-produced antibodies against human Pur-α were tested in western blot assays for Pur-α specificity. 20 µg of total protein from HeLa cells and 10 µg of recombinant *Drosophila* Pur-α I-II protein were loaded on SDS PAGE and blotted onto a nitrocellulose membrane. The blot was then incubated with the primary anti-Pur-α antibodies to detect human Pur-α (~ 37 kDa) and the primary anti-α-actin (~ 43 kDa) antibody to detect α-actin as a loading control. After incubation with the corresponding HRP-conjugated secondary antibody, protein signals were detected using the ECL substrate and light-sensitive films. None of the self-produced Pur-α antibodies showed specific binding to human Pur-α from HeLa cells (Figure 2.17 A, lane 2 and 3), but detected the recombinantly expressed *Drosophila* protein (Figure 2.17 A, lane 4). Likewise, the commercial antibody did not show specific binding to human Pur-α from HeLa cells (Figure 2. B, lane 2), but detected the recombinant *Drosophila* protein (Figure 2.17 B, lane 3). Specific detection of α-actin as a loading control worked nicely (Figure 2.17 A and B, lane 1). As a conclusion, the tested antibodies were not suitable for immunoprecipitation (IP) experiments because of their unspecific binding.

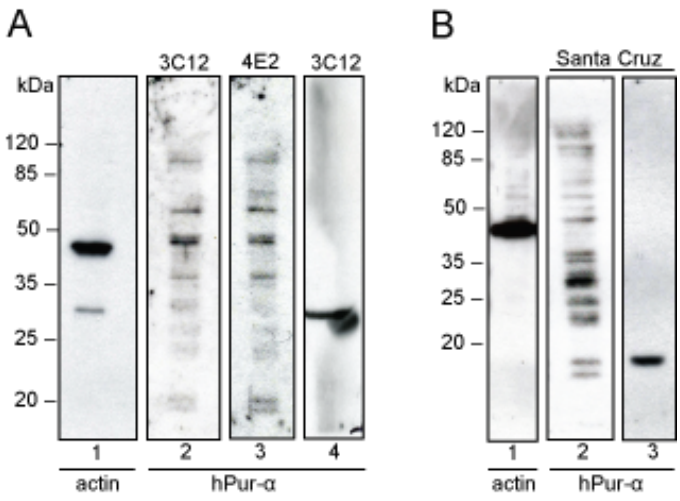


Figure 2.17 Unspecific binding of human Pur-α (hPur-α) antibody. A: Tissue culture supernatants (TCS) of 2 different cell clones were tested for human Pur-α (32 kDa) detection in western blot experiments with HeLa protein lysate. None of the TCS (lane 2 and 3) specifically detected human Pur-α. Lane 4 was loaded with recombinant *D. melanogaster* Pur-α I-II as an indicator for unspecific protein binding. Detection of α-actin (43 kDa) served as a loading control (lane 1). **B:** Similar to **A**, using a commercial human Pur-α antibody (lane 2 and 3). No specific detection of human Pur-α. Lane 3 was loaded with recombinant *D. melanogaster* Pur-α I-II as an indicator for unspecific protein binding. Detection of α-actin (43 kDa) served as a loading control (lane 1).

2.2.1.3. Generation of vectors for expression tagged human Pur-α

Since the tested antibodies against human Pur-α did not show specificity and commercial antibodies might additionally cross-react with other proteins of the Pur-family, tagging Pur-α would enable IPs with a more specific antibody against the tag. Therefore, I used a BAC (bacterial artificial chromosome) clone that contains the genomic human Pur-α sequence, including upstream and downstream regulatory regions (Figure 2.18 A). Pur-α was either N-terminally or C-terminally tagged (Figure 2.18 B and C, top) via homologous recombination using an eGFP-tagging cassette containing a geneticin-neomycin resistance for clone selection. Successful tagging was verified by amplification of eGFP or eGFP-tagged human Pur-α via PCR (Figure 2.18 B and C, bottom). Because the BAC constructs contain the genomic promoter and regulatory sequences, tagged-Pur-α protein expression underlies the same regulation as the protein expression of the endogenous *Pur-α* gene and does not have to be induced artificially. Eventually IPs can be carried out with an antibody against eGFP.

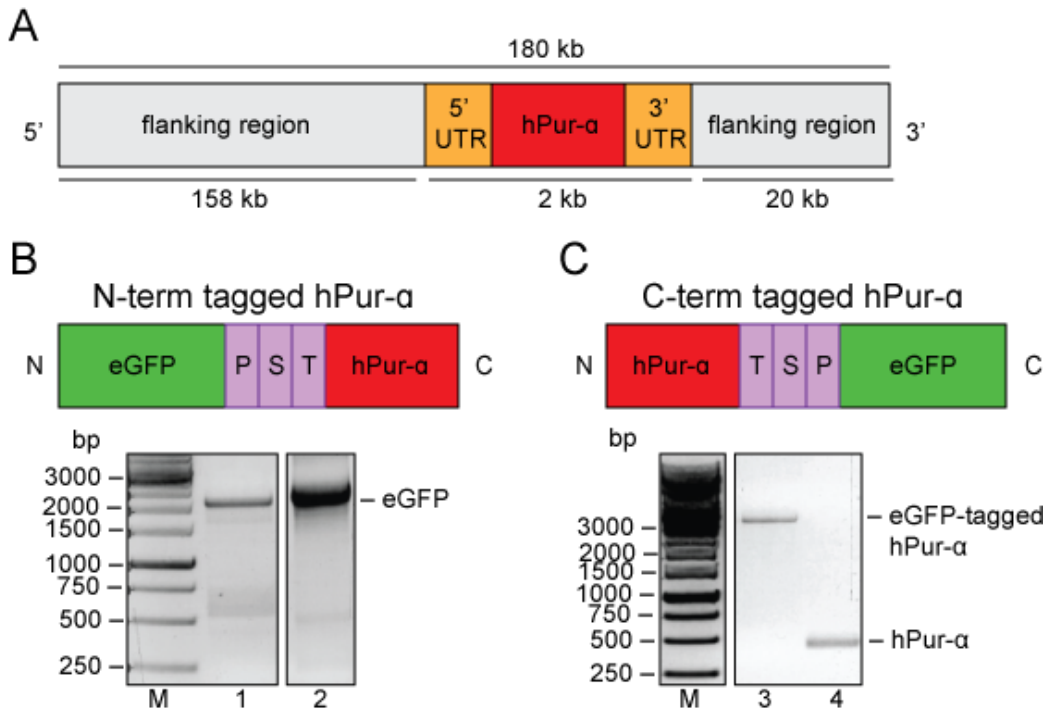


Figure 2.18 BAC constructs of human Pur-α. A: Linear scheme of the unmodified BAC containing the human Pur-α DNA sequence and the regulatory regions up- and downstream of the gene. **B:** (Top) Scheme of N- (left) and C-terminally (right) tagged human Pur-α. The tag consists of eGFP, Prescission protease cleavage site (P), S-peptide (S) and TEV cleavage site (T). Positive BAC clones were verified by amplification of the eGFP sequence (**B**, bottom, lane 1) with primers that bind within the eGFP sequence or eGFP tagged human Pur-α (**C**, bottom, lane 3) with primes that bind within the human Pur-α gene and the BAC backbone. As a control for N-terminally tagged human Pur-α, eGFP was amplified from the N-terminal tagging cassette (**B**, bottom, lane 2). As a control for C-terminally tagged human Pur-α PCR was performed with the unmodified BAC as template (**C**, bottom, lane 4). M, Molecular weight marker.

2.2.1.4. BAC transfection and cell selection

The BAC constructs carrying either N- (Figure 2.18 B, top) or C-terminally (Figure 2.18 C, top) tagged human Pur-α were transfected into HeLa cells. Clones carrying tagged human Pur-α constructs were selected for antibiotic resistance with increasing geneticin (G418) concentrations in order to establish a stable cell line. After 4 weeks of G418 selection all cells transfected with the untagged BAC, lacking the antibiotic resistance, were dead (Figure 2.19 C). However, three clones containing the N-terminally and three clones containing the C-terminally tagged human Pur-α BAC construct survived and grew to bigger colonies within a few weeks (Figure 2.19 A and B). Eventually, from the C-terminally tagged BAC colonies only one persisted.

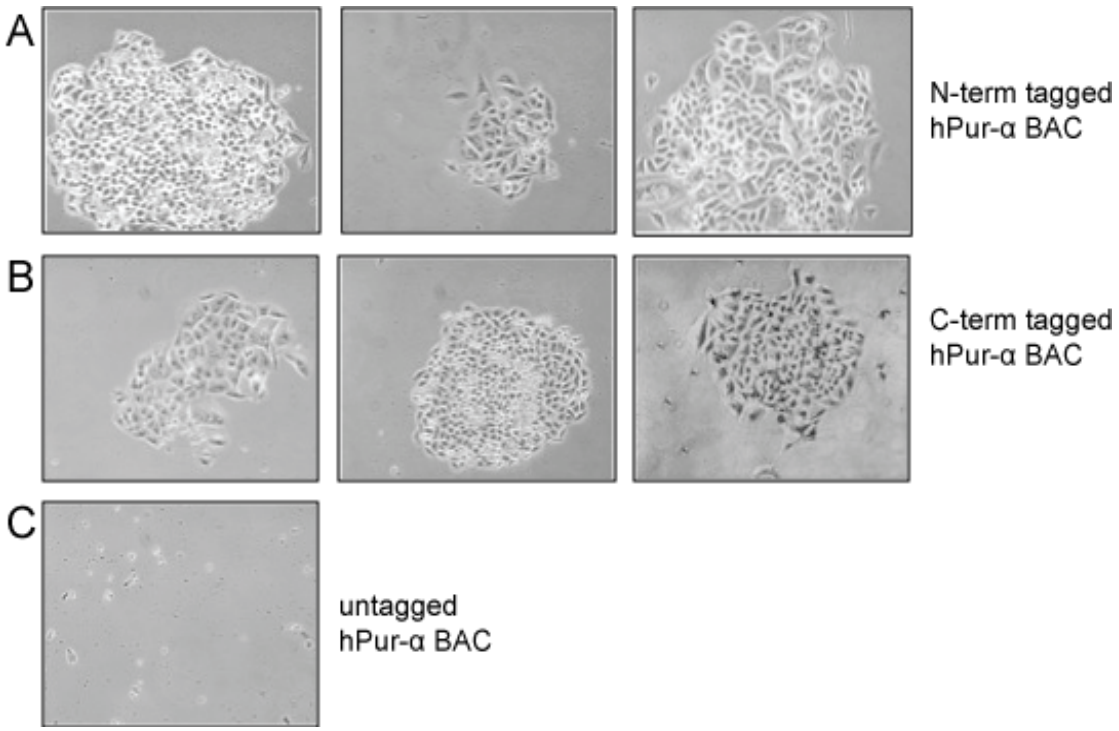


Figure 2.19 HeLa cell colonies carrying BAC constructs with eGFP tagged human Pur-α after 4 weeks of antibiotic selection. A: (From left to right) three different cell colonies holding the BAC construct with N-terminally tagged human Pur-α. **B:** (From left to right) three different cell colonies holding the BAC construct with C-terminally tagged human Pur-α. **C:** HeLa cells transfected with unmodified BAC died during antibiotic selection.

2.2.1.5. Validation of GFP-Pur-α expression

Geneticin resistant colonies carrying either N- or C-terminally eGFP-tagged human Pur-α BAC constructs were further amplified and tested for eGFP expression in western blot assays as described in chapter 4.18. Cells of one clone that contains C-terminally tagged Pur-α, cells of three different clones with N-terminally tagged Pur-α and untransfected HeLa cells were

harvested. Of each cell line 20 µg of total protein per lane were run on SDS PAGE and blotted onto a nitrocellulose membrane. The blot was then incubated with primary anti-GFP antibody to detect eGFP-tagged Pur-α (~ 62 kDa) and primary anti-α-actin (~ 43 kDa) antibody to detect α-actin as a loading control. After incubation with the corresponding HRP-conjugated secondary antibody, protein signals were detected using the ECL substrate and light-sensitive films. The blot did not show any specific protein band for eGFP-tagged Pur-α (~ 62 kDa), neither for the N-terminally, nor for the C-terminally tagged constructs (Figure 2.20). The lane loaded with untransfected HeLa protein lysate showed the same unspecific band pattern like the other cell lines. Only α-actin gave a strong and specific protein signal and confirmed that the same amount of total protein was loaded for each cell line (Figure 2.20).

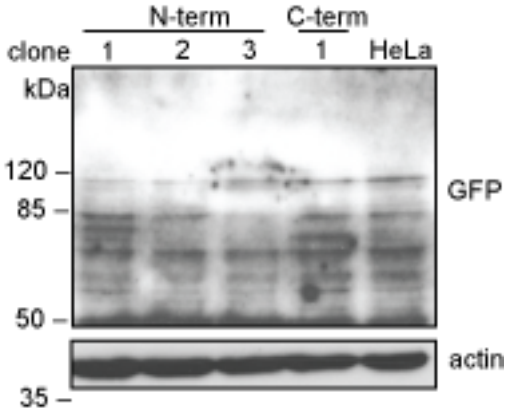


Figure 2.20 Expression test of GFP-tagged hPur-α from HeLa protein lysate. Western blot of HeLa cells stably transfected with BAC constructs containing N- or C-terminally eGFP-tagged human Pur-α. No GFP signal was detected for the N-terminally tagged (N-term, lane 1-3) or the C-terminally tagged human Pur-α (C-term, lane 1). Protein lysate of untransfected HeLa cells were loaded as a control (HeLa).

As the transfected HeLa cells are resistant to geneticin, but do not show expression of eGFP-tagged Pur-α in western blot assays, fluorescence microscopy was used as an alternative technique to detect GFP signal. However none of the stable cell lines gave a fluorescence signal. This can be explained by either silenced eGFP-Pur-α expression or exclusion of the *eGFP-Pur-α* gene from the genome during cell selection, while maintaining the antibiotic resistance.

2.2.2. Pur-α and FXTAS

It has been implied that Pur-α is involved in the pathomechanism of the fragile X-associated tremor/ataxia syndrome (FXTAS) (Jin et al., 2007). FXTAS is caused by premutation expansions (55-200 CGG repeats) in the 5'UTR in the *fragile X mental retardation 1* (*fmr1*) gene (Oostra Willemsen 2009), which lead to intranuclear and astrocytic inclusions throughout the brain (Greco et al., 2006; Iwahashi et al., 2006). Amongst other proteins, Pur-α has been found to be part of these inclusions together with the *fmr1* mRNA (Iwahashi et al., 2006; Jin et al., 2007). Since CGG repeats are known binding sites for Pur-α (Jin et al., 2003),

it has been suggested that Pur-α binds to the CGG-repeat RNA leading to sequestration of Pur-α and other CGG-binding proteins in the pathogenesis of this disease. CGG-repeat induced intranuclear inclusions have also been found in cell and *Drosophila* models (Jin et al., 2007; Sofola et al., 2007).

Therefore, it was planned to further investigate Pur-α's role in FXTAS by performing cell-culture studies with a mammalian cell line stably expressing CGG-repeat RNA.

2.2.2.1. CGG-repeat RNA expression in COS7 cells

Expression of CGG-repeat RNA (with 60 or more repeats) in COS7 cells has been shown to result in formation of nuclear inclusions (Sellier et al., 2010). These inclusions can be purified from nuclear extracts. I planned to perform *in vivo* cross-linking of the *fmr1* 5'UTR, followed by either 30 (normal repeat length) or 95 (FXTAS-related repeat length) CGG repeats, with its bound proteins. After isolation of the cross-linked RNA-protein complexes and subsequent mass spectrometry analysis, I wanted to determine the proteins that directly bind to the CGG-repeat RNA, identify potential interaction partners of Pur-α and compare the results found for the normal and disease-related CGG-repeat RNA.

2.2.2.2. Generation of vectors for CGG-repeat RNA expression

The *fmr1* 5'UTR containing either 30 or 95 CGG repeats was cloned into a mammalian expression vector (pRTS-1) (Figure 2.21 A, left), which contains a GFP reporter and hygromycin resistance. The CGG-repeat RNA and the GFP reporter are both under the control of a bi-directional CMV (cytomegalovirus) promoter. Expression is inducible by doxycycline (TET-ON advanced system).

Furthermore, the CGG-repeat RNA has been cloned into another mammalian expression vector (pRTR) that additionally contains a *CD2* (*cluster of differentiation 2*) gene expressed from the SV40 (simian virus 40) promoter (Figure 2.21 B, left). The *CD2* gene encodes for a cell adhesion protein on the cell surface that enables cell selection via magnetic beads shortly after transfection.

For RNA-isolation, the CGG-repeat RNA was tagged using a two-component system consisting of a bacteriophage coat protein PP7 (serine/threonine protein phosphatase 7, PP7-CP) and 24 RNA stem-loop structures (24x PP7-SL), which are recognized by the PP7-CP. The PP7-CP was cloned into the pRTS-1 and the pRTR vector as a fusion protein at the N-terminus of eGFP. The 24x PP7-SL were cloned upstream of the CGG-repeat RNA (Figure 2.21 A, right). Induction with doxycycline leads to simultaneous expression of PP7-CP fused

to eGFP and CGG-repeat RNA fused to PP7-SL. The PP7-CP can recognize and bind to the PP7-SL and via immunoprecipitation (IP) with an antibody against eGFP the repeat RNA can be isolated.

These four vector constructs (pRTS-1: 30 and 95 CGG repeats, pRTR: 30 and 95 CGG repeats) were generated to obtain a stable cell line with strong eGFP and CGG-repeat RNA expression, respectively.

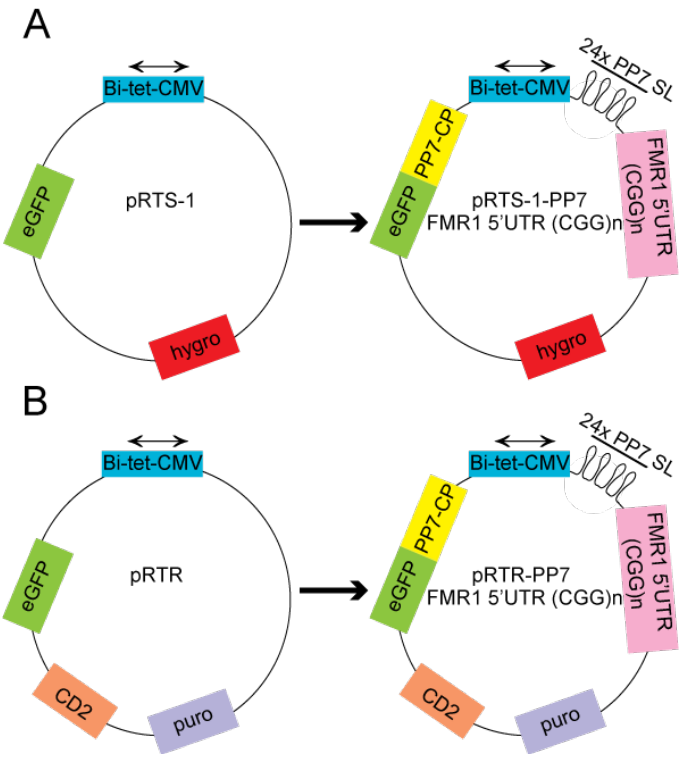


Figure 2.21 Vector design for CGG-repeat RNA expression in COS7 and HeLa cells. A: Scheme of the initial pRTS-1 vector (left) and the self-constructed pRTS-1-PP7 FMR1 5'UTR (CGG)n vector (right). Both vectors contain a bidirectional, inducible CMV promoter (blue), the gene for eGFP (green) and hygromycin resistance (red). The FMR1 5' UTR followed by either 30 or 95 CGG repeats was introduced into the pRTS-1 vector. For tagging the CGG-repeat RNA 24 PP7 stem loops were fused upstream of the FMR1 5'UTR. The eGFP gene was N-terminally fused to the PP7 coat protein, which recognizes and binds to the PP7 stem loop structures of the CGG-repeat RNA. **B:** Scheme of the initial pRTR vector (left) and the self-constructed pRTR-PP7 FMR1 5'UTR (CGG)n vector (right). Both vectors contain a bidirectional, inducible CMV promoter (blue), the gene for eGFP (green) and puromycin resistance (purple). As in **A** the FMR1 5' UTR followed by either 30 or 95 CGG repeats was introduced into the pRTR vector, likewise the PP7 tagging elements. Additionally, the pRTR vector contains a *CD2* gene (orange) that encodes for a cell adhesion protein on the cell surface to enable cell selection via magnetic beads shortly after transfection.

2.2.2.3. Test transfection of initial vectors

As a preliminary test, the initial vectors pRTS-1 and pRTR (Figure 2.21 A and B, left) were transiently transfected into COS7 cells. Twenty-four hours after doxycycline induction, expression of the GFP reporter was visualized by fluorescence microscopy. Cells containing the pRTS-1 plasmid gave a strong GFP signal compared to cells that were only treated with transfection reagent (control) (Figure 2.22 A). Hence, transfection efficiency was reasonable and induction with doxycycline worked well.

Cells containing the pRTR plasmid also gave a strong GFP signal compared to the control, but less strong than the pRTS-1-transfected cells. Transfection was less efficient than for the pRTS-1 construct (Figure 2.22 B).

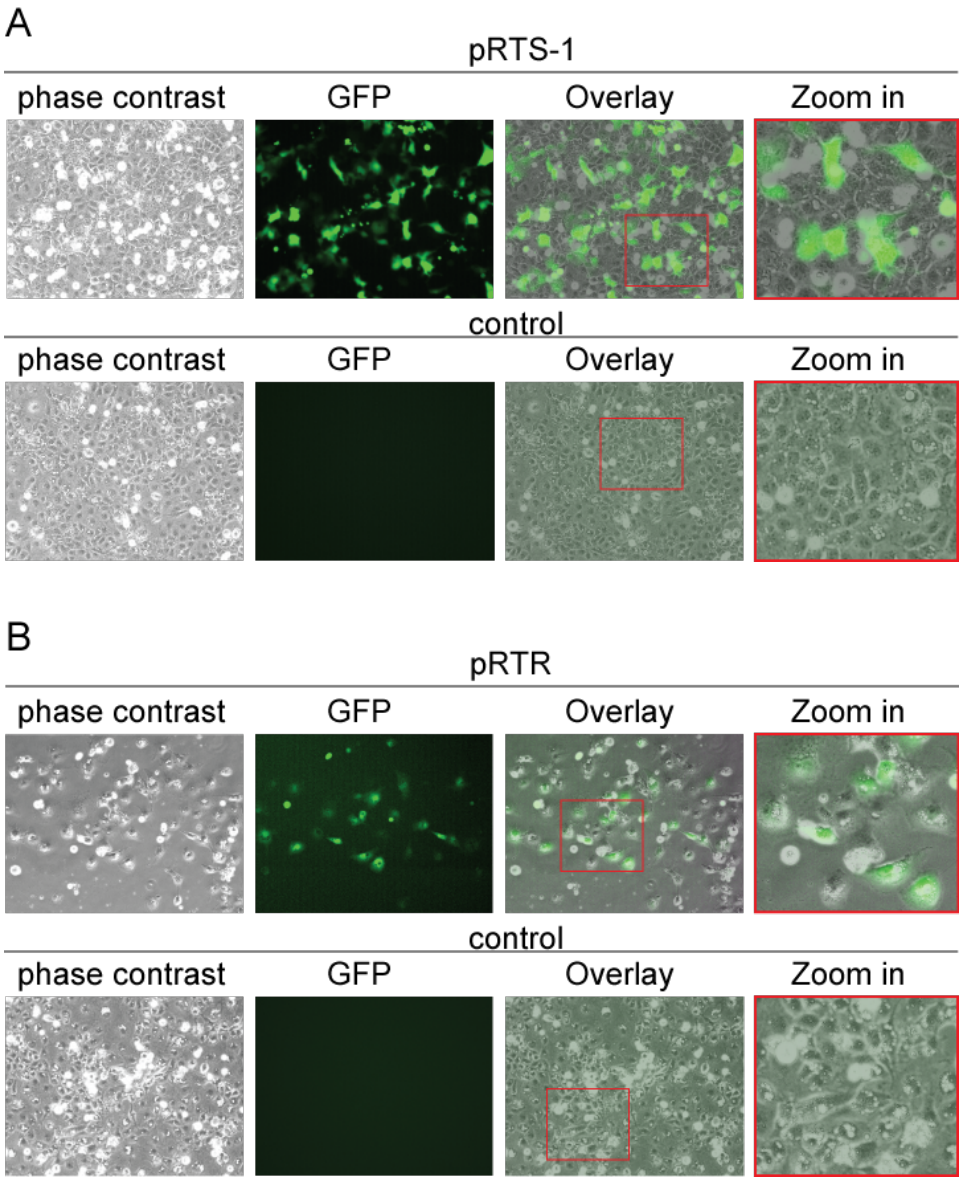


Figure 2.22 Test expression of initial vectors in COS7 cells. The initial vectors pRTS-1 **(A)** and pRTR **(B)** were transiently transfected into COS7 cells (top). After induction with doxycycline eGFP expression was analyzed with a fluorescence microscope. The images of phase contrast and GFP channel were overlayed. The red box indicates the region, which is shown as zoom in. Untransfected cells, which have been treated only with transfection reagent, were used as control (bottom).

2.2.2.4. Transient transfection and induction of CGG-repeat RNA

Next, COS7 cells were transiently transfected with the pRTS-1 vector containing only the untagged CGG-repeat RNA carrying either 30 or 95 CGG repeats and the GFP reporter, both under the control of the CMV promoter. After doxycycline induction, GFP fluorescence was detected by fluorescence microscopy. A strong GFP signal was detectable for both RNA constructs (30 and 95 CGG repeats), compared to the control (Figure 2.23). However, a very small amount of cells showed GFP expression compared to the cells transfected with the initial pRTS-1 vector (compare Figure 2.22 A). Transfection efficiency of the CGG-repeat

RNA constructs was much lower. Additionally, cells transfected with an RNA of 95 CGG repeats were more prone to die upon transfection than cells transfected with 30 CGG-repeat RNA.

Cells were then selected for hygromycin resistance for several weeks in order to obtain a cell line stably expressing the CGG-repeat RNA and the eGFP reporter. However, GFP fluorescent cells could not be accumulated during antibiotic selection. Although all cells were hygromycin resistant, only a small portion gave a fluorescence signal. Since antibiotic selection seems to be insufficient for generating a stable cell line, selection via magnetic beads coupled to an antibody against the cell-surface protein CD2 could increase the number of cells expressing the transfected construct. Hence, for further CGG-repeat RNA constructs, the pRTR vector, containing the *CD2* gene for magnetic bead cell selection, was used.

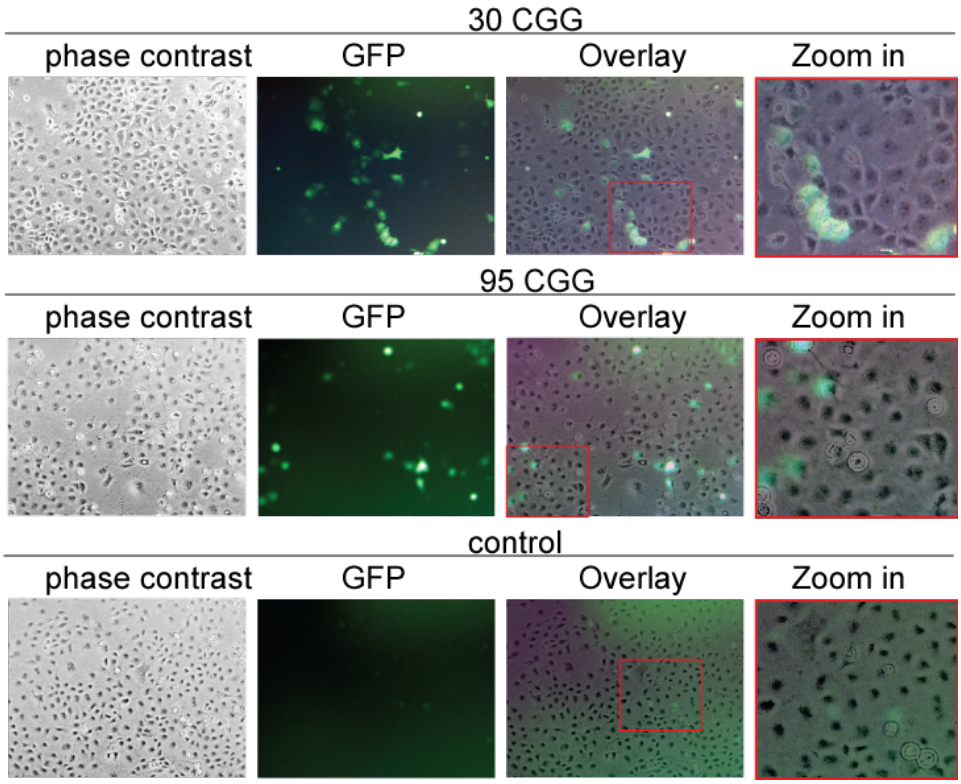


Figure 2.23 Co-expression of eGFP and CGG-repeat RNA in COS7 cells. The pRTS-1 vector containing the CGG-repeat RNA with either 30 (top) or 95 CGG (middle) repeats were transiently transfected into COS7 cells. After induction with doxycycline eGFP signal was analyzed with a fluorescence microscope as indicator for successful transfection and CGG-repeat RNA expression. The images of phase contrast and GFP channel were overlayed. The red box indicates the region, which is shown as zoom in. Untransfected cells, which have been treated only with transfection reagent, were used as control (bottom). Transfection efficiency was decreased compared to the transfection with the initial pRTS-1 vector without CGG-repeat RNA.

2.2.2.5. Transient transfection and induction of PP7-tagged CGG-repeat RNA

The pRTR vector containing tagged CGG-repeat RNA and PP7-GFP fusion protein was transiently transfected into COS7 cells. Upon transfection many cells died before protein expression could be induced by doxycycline.

A fluorescent microscope monitored expression of the GFP reporter twenty-four hours after induction. Again, many cells have died upon doxycycline induction. Of the surviving cells, expressing either 30 or 95 CGG repeat-RNA, only a small amount showed a GFP signal (Figure 2.24 A). However, when compared with the control, it was impossible to distinguish if the GFP signal comes from GFP expression within the cells or if it is only auto-fluorescence of dying cells. Therefore, antibiotic selection against puromycin for establishing a stable cell line was not done.

To test if the low transfection efficiency and the high death rate of the cells depend on the cell line used, the same constructs were transfected into HeLa cells (Figure 2.24 B). But similarly to the transfected COS7 cells, also many HeLa cells died upon transfection and induction. The GFP signal was very weak compared to the control, so it could not be distinguished between dying cells that auto-fluoresce or living HeLa cells that express the GFP reporter.

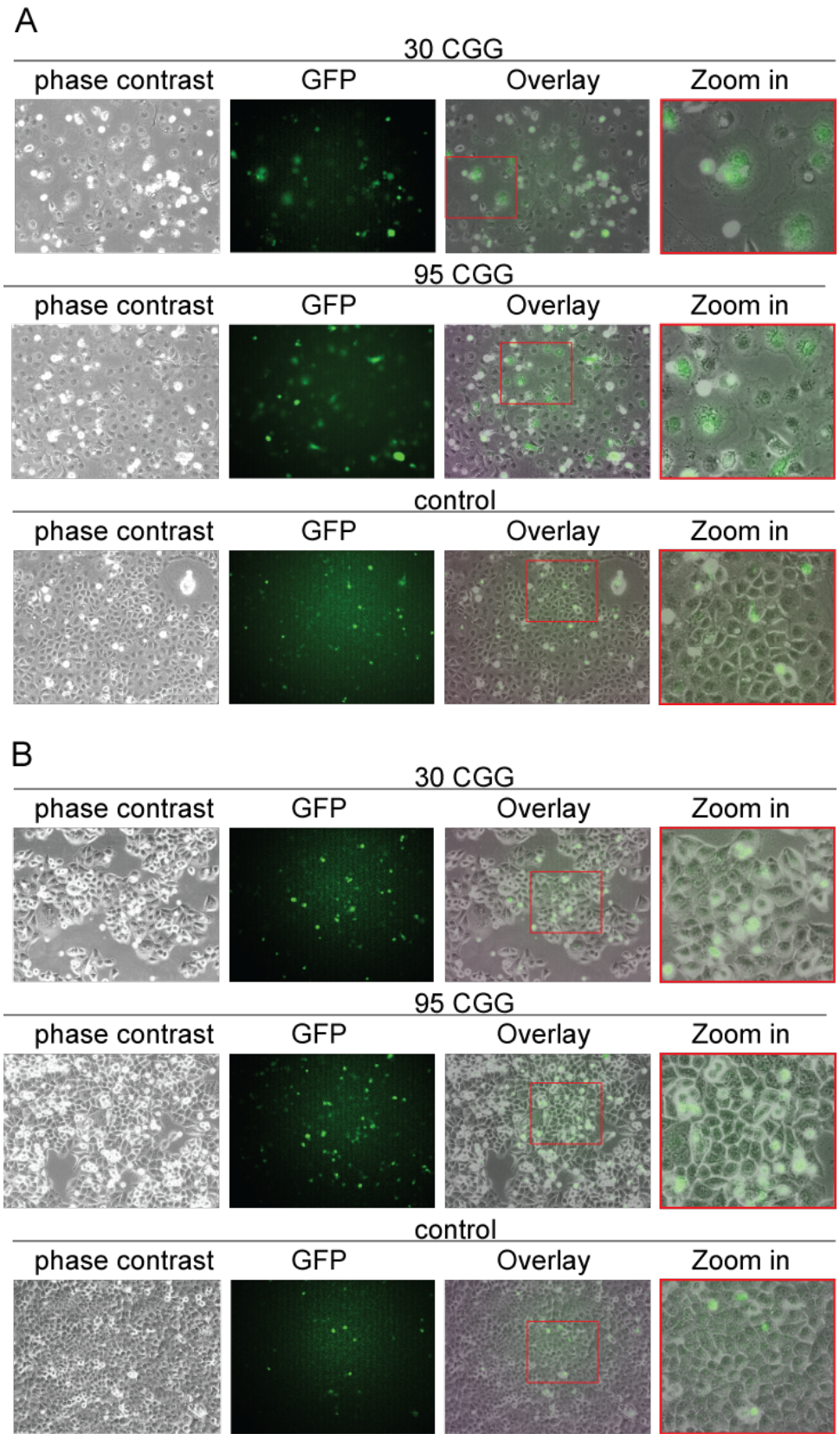


Figure 2.24 Co-expression of PP7CP-tagged eGFP and PP7SL-tagged CGG-repeat RNA in COS7 and HeLa cells. The pRTR-PP7 FMR1 5'UTR (CGG)n vector containing the CGG-repeat RNA with either 30 (top) or 95 CGG (middle) repeats was transiently transfected into COS7 **(A)** and HeLa **(B)** cells. After induction with doxycycline eGFP signal was analyzed with a fluorescence microscope as indicator for successful transfection and CGG-repeat RNA expression. The images of phase contrast and GFP channel were overlayed. The red box indicates the region, which is shown as zoom in (on right side). Untransfected cells, which have been treated only with transfection reagent, were used as control (bottom). Transfection efficiency was very low compared to the transfection with the initial pRTS-1 vector without CGG-repeat RNA.

2.2.2.6. Antibody test against αB-crystallin

In a recent study more than 20 proteins have been found within the inclusions from a postmortem FXTAS brain (Iwahashi et al., 2006; Sofola et al., 2007). Proteins identified include αB-crystallin, which could therefore be used as an inclusion marker for future imaging assays. Fifteen in-house-made antibodies against the C-terminus of αB-crystallin were tested on COS7 protein lysate in western blot assays (Figure 2.25). None of the tested antibodies did specifically detect αB-crystallin.

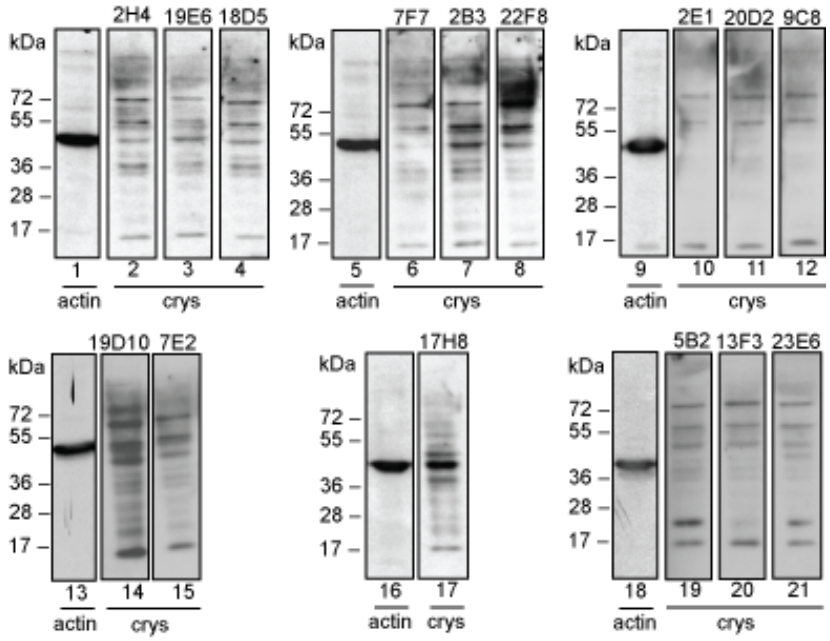


Figure 2.25 Unspecific binding of αB-crystallin (crys) antibody. Tissue culture supernatants (TCS) of 15 different cell clones were tested for αB-crystallin (23 kDa) detection in western blot experiments with COS7 protein lysate. None of the TCS (lane 2-4, 6-8, 10-12, 14, 15, 16, 19-21) specifically detected αB-crystallin. Detection of α-actin (43 kDa) served as a loading control (lane 1, 5, 9, 13, 16, 18).

3. Discussion

3.1. Interaction of Pur-α with nucleic acids

3.1.1. Affinity for DNA and RNA

Pur-α I-II shows strong and specific binding to its physiological target MF0677 DNA. Located upstream of the *c-myc* gene (Haas et al., 1995), but much weaker binding to the CGG-repeat RNA sequence (Graebsch et al., 2009). For this reason it has been suggested that the binding of Pur-α to DNA is stronger than to RNA and as a consequence that there might be differences in the binding modes to both nucleic acid targets. However, this study shows that Pur-α does not generally bind less effective to RNA than to DNA. Instead, EMSA experiments demonstrate that Pur-α has similar binding affinities for DNA and RNA of the same sequence. Thus, the higher affinity is observed for MF0677 DNA and RNA sequence (K_D ~200 nM) (Figure 2.2 A) when compared to CGG-repeat DNA and RNA (K_D ~2 µM) (Figure 2.2 B).

Complementing results were obtained from NMR titration assays with ^{15}N-labeled *Drosophila* Pur-α I-II and oligonucleotides. Initial RNA- and DNA-titration trials with the MF0677 and CGG-repeat oligomers (24 nt) used in EMSA experiments failed due to precipitation. Most likely this was caused by the high protein concentrations and buffer requirements (low pH, low salt) needed for NMR measurements. In fact, aggregation of Pur-α/DNA complexes was also detectable as higher band shifts in the EMSA experiments (Figure 2.2 A and B). NMR titrations with short (5mer) GGN-motif DNA and RNA oligomers that still bound to Pur-α with reasonable affinity (Figure 2.3 B) did not induce precipitation. The spectra showed similar chemical shift perturbations, regardless of whether it was ssDNA or RNA, indicating that both nucleic acids have the same affinity for Pur-α and that they are likely bound in the same way (Figure 2.3 C-E).

Also the crystal structure of Pur-α/DNA co-complex did not provide any evidence why Pur-α should have a higher binding affinity for DNA than for RNA. Since a hydroxyl-group on the 2' position of the pentose ring of the RNA sugar backbone would not cause steric clashes (Figure 2.5) the crystal structure did also not suggest any differences in binding modes between DNA and RNA.

To further understand why Pur-α has a higher affinity for the MF0677 sequence than for pure CGG-repeat oligomers and to define its exact binding motif, DNaseI footprint assays were performed using the promoter region upstream of the *c-myc* gene. However, binding to the GGN motif within the MF0677 subfragment was not detectable. Several attempts using

ssDNA or dsDNA substrates, different protein derivatives comprising only Pur-repeat I-II (nucleic acid binding domain) or all three Pur-repeats, or using protein from different species (human, *Drosophila*) did not provide any meaningful results. No Pur-α binding sites were detected (Figure 2.15). Therefore, the questions why Pur-α shows such high affinity for the MF0677 sequence and how it binds to the *c-myc* promoter region could not be answered.

3.1.2. Stoichiometry of the Pur-α/nucleic acid co-complex

Stoichiometry of protein/nucleic acid co-complexes can be assessed by several techniques, such as NMR (nuclear magnetic resonance) titrations, SAXS (small angle X-ray scattering) and ITC (isothermal titration calorimetry). The first two methods have been applied for determination of the stoichiometric ratio of a Pur-α/nucleic acid co-complex.

NMR titration steps resulted in nicely dispersed spectra (Figure 2.3 C-E). However, most chemical shift perturbations experienced an intermediate exchange regime. Even measurements at higher or lower temperatures (203 K and 293 K) did not alter the exchange rate from intermediate to fast or slow. Additionally, by providing an increasing amount of nucleic acid the samples tended to aggregate and became more prone to precipitation while measuring. Therefore, saturation could not be reached and the stoichiometric ratio of the complex could not be determined.

Since NMR experiments could not be used for stoichiometry determination SAXS measurements were performed. Several measurements were done with the MF0677 (24 nt) ssDNA and either Pur-α I-II or Pur-α I-III, a dimer in solution. Alike in NMR titration assays, polydispers oligomers and aggregates of the Pur-α/nucleic acid co-complex occurred during SAXS measurements. Probably again due to the relatively high protein concentration (50-100 µM) needed for signal detection. Calculation of a stoichiometric ratio was therefore impossible. Even when the ionic strength in the buffer was increased from 250 mM to 500 mM NaCl oligomerisation still occurred. Only 1 M of NaCl prevented aggregation of the complex, but unfortunately, also impeded dimerization of Pur-α I-III. Results obtained from SAXS measurements were therefore not reliable and could not be used for stoichiometry determination.

ITC represents another possibility for complex analysis. However, also ITC measurements require relatively high protein (10-100 µM) and ligand (50-500 µM) concentration for signal detection, depending on the K_D of the reactants (Duff et al., 2011; Pierce et al., 1999). Considering the problems of aggregation and oligomerisation in NMR and SAXS

measurements, ITC did not show great promise for successful stoichiometry determination and was therefore not explored.

To overcome the high aggregation and oligomerisation problem, a method was chosen that only needs low protein concentrations and is less sensitive to salt concentration or pH of the buffer. Though, to determine the complex stoichiometry the nucleic acid concentration needs to be above the K_D to ensure that all protein is bound. The K_D for Pur-α I-II and MF0677 (24 nt) ssDNA was estimated to be approximately 150 nM (Figure 2.2 A). In a so-called dot blot assay Pur-α I-II was titrated to a constant amount (1 μM) of radiolabeled MF0677 (24 nt) ssDNA and blotted onto a nitrocellulose membrane. The signal intensities of protein/DNA complexes bound to the nitrocellulose membrane were plotted against the protein concentration. The saturation curves of three independent experiments clearly showed a stoichiometric ratio of 1:2 (Pur-α:DNA) (Figure 2.13). Thus, in solution Pur-α I-II can bind two molecules of ssDNA. This result is consistent with the two binding sites observed in the crystal structure of Pur-α I-II in complex with ssDNA (Figure 2.5), where the ssDNA oligomers become clamped between Pur-repeat I and II. The interacting residues are Q52, S53, K54 (QSK I) on Pur-repeat I and K138, N140, R142 (KNR II) on Pur-repeat II. A particularly interesting feature of this co-structure is the flipping-out of cytosine C5 and the resulting twist of the 3'-end of the DNA strand, which then interacts with a Pur-repeat I of another symmetry related protein molecule (K61, R80, D59, Y57 and K70) (Figure 2.5). Although both Pur-repeats share a moderate sequence identity (~30 %), share the identical binding motif (KNR), and adopt the same fold (Graebsch et al., 2009 & 2010), the second binding event on Pur-repeat I appears at overlapping but non-identical residues. Crystallographic packing forces might have prevented DNA binding to the identical motif (KNR) on Pur-repeat I. One could also speculate that symmetric and asymmetric binding depends on the GGN-motif of the nucleic acids. For instance, different motifs (GGA, GGG, GGC, GGT) might bind to slightly shifted interaction sites of Pur-α. Moreover, this might explain why CGG-repeats bind less strongly to Pur-α than the MF0677 sequence that mostly consists of GGA and GGT motifs.

3.1.3. Interaction sites of Pur-α

The Pur-α/DNA crystal structure shows that both Pur-repeats interact with nucleic acids and that the DNA strands become clamped between the two repeats. Pur-α mostly interacts with the guanine bases. Only R142 and K138 interact with the base of cytosine C2 and the sugar

phosphate backbone of cytosine C5, respectively (Figure 2.5 B). Binding therefore occurs sequence specific and confirms the GGN binding motif that has already been postulated by Bergemann and Johnson, 1992.

Previous findings implied that the positively charged β-sheets mediate DNA/RNA-binding whereas the amphiphatic helices might rather contribute to protein-protein interactions (Graebsch et al., 2009). The crystal structure of the protein-DNA co-complex confirms that the β-sheets, including their short linkers, are involved in DNA binding, in contrast to the α-helices that show no interaction with the nucleic acid (Figure 2.5). The residues involved in DNA binding are highly conserved (Figure 2.7). Mutation of the interacting residues resulted in a decreased binding affinity and therefore confirmed the interaction sites seen in the crystal structure (Figure 2.8 A-F). Moreover, a second potential binding site was found on Pur-repeat I (KNR I), which upon mutation causes a decrease in binding affinity as seen in EMSA experiments (Figure 2.8 G-H).

Within the crystal lattice the first two DNA bases (G1 and C2) of the 5'-end are base pairing with the 5'-end of the symmetry related DNA molecule (Figure 2.5 H). This base pairing might be due to crystal packing forces. At least, there is no evidence that the base pairing has any physiological relevance.

Structural comparison of the Pur-α I-II apo-structure (PDB ID 3K44) and the structure of the protein-DNA co-complex revealed that Pur-α does not show significant conformational changes upon nucleic acid binding (Figure 2.6).

3.1.4. Unwinding of dsDNA

The most interesting feature of Pur-α's binding mode to ssDNA represents the non-canonical twist of the DNA strand. Pur-α is known to unwind dsDNA, but so far it has not been shown how this is done in an ATP-independent manner (Darbinian et al., 2001; Wortman et al., 2005). Here, the crystal structure of the protein/DNA co-complex provides an explanation how stabilization of ssDNA is enabled. The phenylalanine F145 (F II) on Pur-repeat II undertakes base stacking with the guanine G4 and thereby blocks the space for the neighboring cytosine C5 (Figure 2.5 E). Thereupon, the cytosine flips out and the 3'-end of the DNA strand becomes distorted. The interaction of the K138 with the phosphate backbone of the cytosine enforces this unusual turn (Figure 2.5 B and C), which additionally becomes stabilized by binding of the 3'-end to another protein molecule (Figure 2.5 F). F II is highly

conserved throughout different species (Figure 2.7). When it becomes mutated to an alanine unwinding of dsDNA becomes abolished (Figure 2.12 B).

In contrast, the general ability of DNA/RNA-binding is only decreased in this mutant (Figure 2.8. E and F), probably because the main interaction sites (QSK I, KNR II) are sufficient to facilitate binding. When the main interaction sites are mutated, nucleic acid binding (Figure 2.8 C and D) is strongly decreased and consequently unwinding of dsDNA is abolished (Figure 2.12 A). Base stacking with a guanine base also occurs on the potential second binding site of Pur-repeat I (Figure 2.5 B and G).

F II has its structural counterpart in position F68 (F I) on Pur-repeat I. Although F I is also highly conserved, the guanine base stacking is not mediated by F I, but by another conserved residue, the Y57 (Figure 2.5 B and G; Figure 2.7). As mentioned before, the two binding sites of Pur-α seen in the crystal structure are asymmetric and might account for sequence specific binding to nucleic acids with different GGN motifs. Base stacking of the phenylalanine and tyrosine with the guanine bases underlines once more the binding preference of Pur-α to GGN motifs (Bergemann and Johnson, 1992).

Pur-α binds to origins of replication and promoter regions (Bergemann and Johnson, 1992; Bergemann et al., 1992) and regulates the transcription of more than 20 genes (White et al., 2009). Pur-α's unwinding ability of dsDNA might therefore play an important role in the initiation of replication and transcription. In this study, the stoichiometry of Pur-α bound to short ssDNA molecules was determined to have a ratio of 1:2 (protein:DNA), but in this context it could also be possible that longer ssDNA fragments might wrap around the nucleic acid binding domain (Pur-repeat I-II) interacting with the two binding surfaces, seen in the crystal structure.

Moreover, it is still unknown how dsDNA binding to an intermolecular Pur-α dimer providing two nucleic acid binding domains occurs. Analysis of the electrostatic surface charges show that there are two opposing patches of positively charged residues, one on each Pur-repeat (Figure 3.1 A and B). These residues correspond to the interaction sites seen in the crystal structure, whereby DNA-binding to Pur-repeat I has only been observed for the symmetry-related Pur-α molecule within the crystal lattice. However, in solution each Pur-repeat might bind to one of the strands of a duplex DNA molecule (Figure 3.1 C). This might be enabled through spontaneous openings and re-closings of the dsDNA helix, called DNA breathing, by which DNA bases become accessible for DNA-binding proteins such as Pur-α (Peyrard et al., 2009; Jose et al., 2012). Further separation of the two DNA strands might be caused by the intercalating residues (phenylalanine, tyrosine) that undergo the base stacking with the guanines and thereby cause the non-canonical twist of the DNA strands. The partly melted

duplex DNA could then be further unwound by DNA helicases, which are required for initiation of transcription and replication.

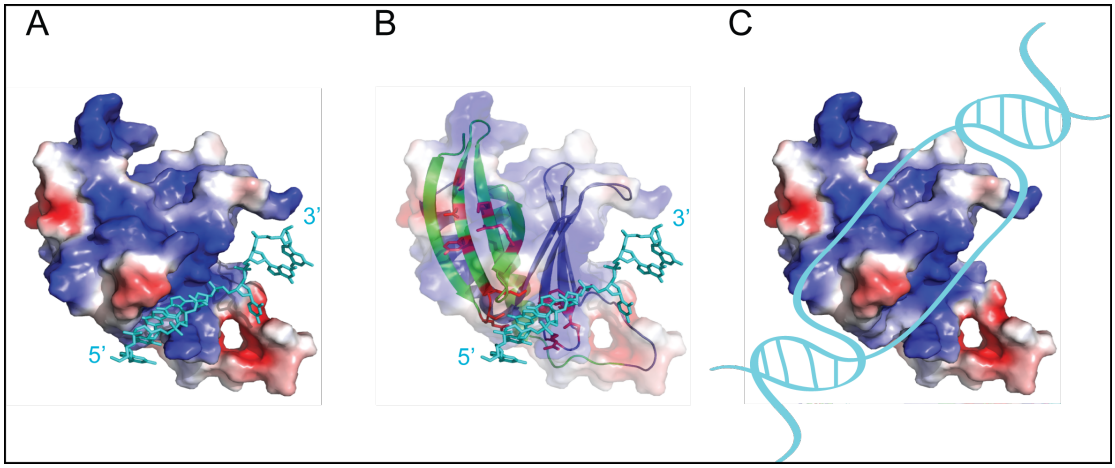


Figure 3.1 Unwinding model for Pur-α and dsDNA. A: Electrostatic surface model of Pur-α I-II in complex with ssDNA (cyan). Red and blue coloration indicate negative and positive electrostatic potentials, respectively. **B:** Representation as in **A**, additionally showing the ribbon backbone model of Pur-repeat I (green) and II (blue). DNA interaction sites, seen in the crystal structure, are shown as red sticks. **C:** Representation as in A. Model showing dsDNA (cyan) bound to Pur-α I-II. The double-strand is locally unwound and the two separated strands bind to the two opposing binding sites on the protein.

Pur-repeat III binds only weakly to DNA and RNA (Figure 2.10) and unwinds dsDNA only slightly (Figure 2.12 E) compared to Pur-repeat I-II. On the basis of the crystal structure it seems unlikely that Pur-repeat III would interfere with the dsDNA-binding of Pur-repeat I-II, since it arranges on the backside of their nucleic acid binding surface (Figure 3.2 A and B). Hence, Pur-repeat III might only facilitate dimerization, thereby guiding a second DNA/RNA-binding domain (Pur-repeat I-II) to another GGN motif further upstream or downstream where additional DNA-unwinding could take place (Figure 3.2 C).

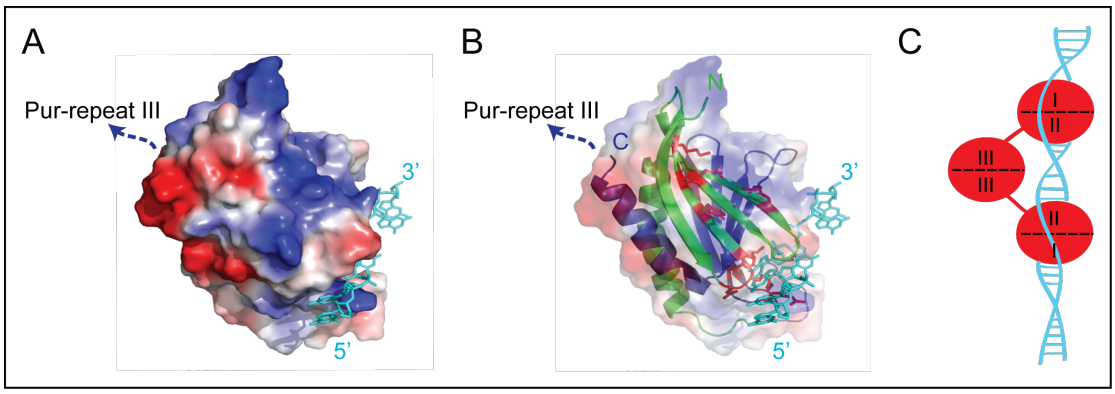


Figure 3.2 Pur-repeat III might guide two nucleic acid binding domains to dsDNA. A: Electrostatic surface model of Pur-α I-II in complex with ssDNA (cyan). Red and blue coloration indicate negative and positive electrostatic potentials, respectively. Pur-repeat III likely arranges at the opposing site of the nucleic acid binding region. **B:** Representation as in **A**, additionally showing the ribbon backbone model of Pur-repeat I (green) and II (blue) and its C-terminus connecting to Pur-repeat III. DNA interaction sites, seen in the crystal structure, are shown as red sticks. **C:** Schematic drawing of a intermolecular Pur-α dimer (red) bound to dsDNA (cyan). Pur-repeat III mediates dimerization, leading both nucleic acid binding domains (repeat I and II) to the duplex DNA where they unwind the strands.

Pur-α binds to CGG-repeats found in the 5'UTR of the *fmr1* mRNA (Jin et al., 2007; Sofola et al., 2007). CGG-RNA repeats are known to form thermodynamically stable structures by non-canonical GG pairing (Kiliszek et al., 2011). One interaction partner of Pur-α is the Rm62, the *Drosophila* ortholog of p68 RNA helicase that is implicated in transcriptional regulation, pre-mRNA splicing, RNA interference and nucleo-cytoplasmic shuttling (Qurashi et al., 2011). Hence, another function of Pur-α could be the initial unwinding of dsRNA to allow its interacting helicase to subsequently regulate RNA processing, translation and transport. In this study, several attempts were made to establish an dsRNA-unwinding assay. Due to technical limitiations no suitable dsRNA substrate could be generated. CGG-repeat RNA oligonucleotides did not anneal to a dsRNA substrate. Annealing of the CGG-repeat RNA to a complementary GCC-strand resulted in a dsRNA substrate, but could not be unwound by Pur-α, implying that Pur-α needs unpaired nucleotides that provide a contact point for unwinding. Hence, Pur-α's role in unwinding of dsRNA still needs to be analyzed.

3.1.5. Pur-α mutations cause 5q31.3 microdeletion syndrome

Recently, de novo mutations in Pur-α have been found to cause a so-called microdeletion syndrome, which is characterized by neonatal hypotonia, encephalopathy and severe developmental delay (Lalani et al., 2014). Of the identified mutations two missense mutations (K97E, A89P) are of particular interest. Sequence alignment of Pur-α from different species shows that the residues A89 and K97 of the human Pur-α protein correspond to the residues

A72 and R80 of the *Drosophila* protein, respectively. These residues are highly conserved (Figure 2.7).

In the crystal structure of the protein/DNA co-complex A72 does not directly interact with the DNA molecule but forms backbone hydrogen bonds between the β-strands of Pur-repeat I to stabilize the β-sheet (Figure 3.3 A top), which has been shown to be the nucleic acid binding surface (this study, Graebsch et al., 2009). Proline is very rigid compared to other amino acids and acts as a disruptor for α-helices and β-sheets. When A72 and its counterpart A98 in the human protein (Figure 3.3 A middle), are substituted by a proline the backbone interactions that stabilize the β-sheet become disrupted (Figure 3.3 A bottom). This likely leads to misfolding of the protein and furthermore to abolishment of nucleic acid binding.

R80 in the *Drosophila* protein directly binds to the guanine base G7 (Figure 2.5 B and G; Figure 3.3 B top). Mutation of R80 has been shown to result in reduced nucleic acid binding (Graebsch et al., 2009). The positively charged residue K97 in human Pur-α is analogous to *Drosophila* R80. Most likely also K97 is involved in DNA interaction as modeled in the crystal structure (Figure 3.3 B middle). This might explain why a mutation to glutamate, a negatively charged residue likely impairing the DNA interaction, causes dysfunction of Pur-α (Figure 3.3 B bottom).

As described in chapter 3.1.1. Pur-α binds to DNA and RNA in undistinguishable mode. In the previous section it has already been discussed that Pur-α regulates replication, transcription and moreover, has been implied to regulate mRNA transport and translation (Ohashi et al., 2000; Johnson et al., 2006; Kanai et al., 2004; Aumiller et al., 2012). Hence, impaired binding to nucleic acids would lead to misregulation of plenty of cellular functions.

Taken together, the crystal structure of the Pur-α/DNA co-complex reveals how these two missense mutations (K97E, A89P) found in the microdeletion syndrome might have an effect on the nucleic binding ability and therefore on the proper function of Pur-α leading to the disease phenotype.

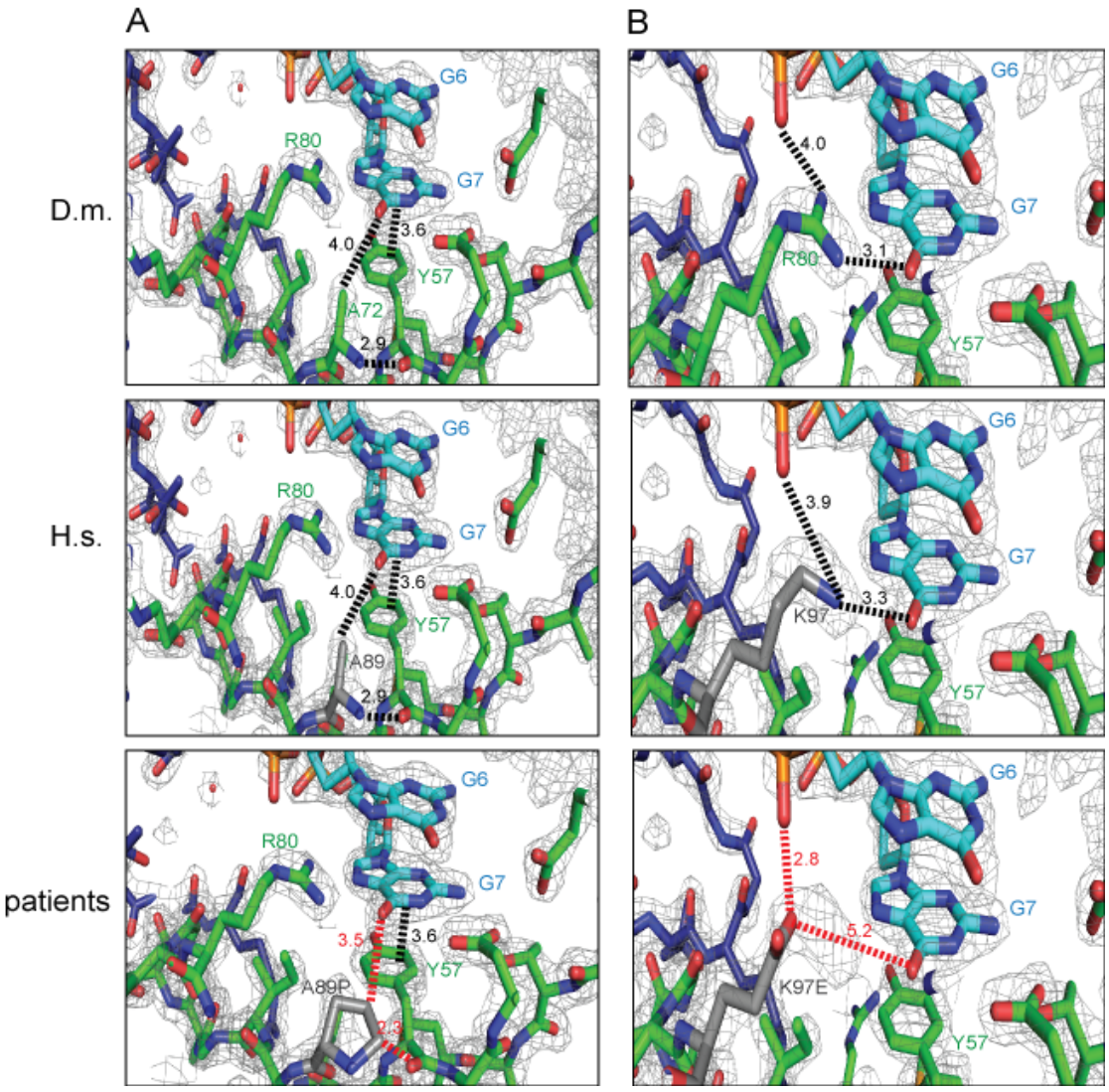


Figure 3.3 Pur-α mutations found in the 5q31.3 microdeletion syndrome can be modeled into the crystal structure of *Drosophila* Pur-α I-II (green) in complex with DNA (cyan). A: Residue A72 of the *Drosophila* protein (top) corresponds to the residues A89 (grey) of the human protein modeled into the *Drosophila* crystal structure (middle). Both alanines form backbone hydrogen bonds. When A89 is mutated to proline, backbone interactions are disrupted (bottom). **B:** Residue R80 of the *Drosophila* protein (top) corresponds to the residues K97 (grey) of the human protein, which was modeled into the crystal structure (middle). Both R80 and K97 are positively charged residues. R80 interacts with the guanine G7. The same interaction could also be mediated by K97. Mutation of K97 to a glutamate would impair DNA binding due to its negative charges (bottom). D.m.: *Drosophila melanogaster*; H.s.:*Homo sapiens*.

3.2. Outlook

3.2.1. Doxycycline-inducible CGG-repeat/reporter expression system - Cellular FXTAS model

FXTAS is caused by moderate expansions (55-200 repeats) of a CGG trinucleotide in the 5'UTR of the *fmr1* gene. The major neuropathological hallmark for FXTAS is intranuclear, neuronal and astrocytic inclusions throughout the brain (Greco at al., 2006; Iwahashi et al., 2006). Analyses on the inclusions from post-mortem brain tissues revealed association with *fmr1* mRNA and more than 20 proteins, including Pur-α (Iwahashi et al., 2006; Sofola et al., 2007 Jin et al., 2007). To date Pur-α's role in the disease pathomechanism is not understood.

In pull down experiments from different tissues (mouse and human neurons, COS7 and HeLa cells) with biotinylated RNA Pur-α has been shown to bind to CGG-repeat RNA (Jin et al., 2007; Sofola et al., 2007). However, in these experiments RNA-binding might occur via protein-protein rather than via direct protein-RNA interaction or might be induced by unnaturally high concentrations of the RNA. To date no protein-interaction partner of Pur-α has been identified that could mediate Pur-α's sequestration into the inclusions.

Expression of CGG-repeat RNA (with 60 or more repeats) in COS7 cells has been shown to result in formation of nuclear inclusions (Sellier et al., 2010), which can be purified from nuclear extracts. Based on this, stable cell lines expressing normal and FXTAS-related CGG repeat RNA would help to identify proteins that directly bind to the CGG-repeat RNA and to validate that Pur-α is amongst these proteins. By performing *in vivo* cross-linking of the CGG-repeat RNA with its bound proteins, the cross-linked RNA-protein complexes could be isolated and subsequently analyzed by mass spectrometry. Furthermore, findings for cells expressing normal and disease-related repeat RNA could be compared. This way, it would also be possible to identify potential protein interaction partners of Pur-α in the disease context.

Towards this goal, I generated mammalian vector constructs (pRTS-1 and pRTR), expressing the *fmr1* 5'UTR, followed by either 30 (normal repeat length) or 95 (FXTAS-related repeat length) CGG repeats, which can be applied for the establishment of a stable cell line (Figure 2.21). Special features of these vectors are the inducible bi-directional promoter, the eGFP reporter, the two-component system (PP7 coat protein binding to PP7 stem loops) for RNA-isolation and, only present in the pRTR vector, a cell surface protein (CD2) that enables cell selection via magnetic beads that are coupled to the respective antibody (see chapter 2.2.2.2.).

Test transfections of the empty vectors (pRTS-1 and pRTR) in COS7 cells were successful, meaning the expression of the eGFP reporter could be induced and the eGFP signal could be

monitored by fluorescence microscopy (Figure 2.22). However, first attempts of generating a stable cell line expressing only the CGG-repeat RNA (pRTS-1) in COS7 cells were not satisfactory (Figure 2.23). After several weeks of selection, cells were resistant to the antibiotic, but only few cells expressed the reporter GFP upon induction. Either the cells have integrated only the antibiotic resistance gene into their genome or the GFP reporter together with the CGG-repeats became silenced. A faster and maybe more reliable alternative to antibiotic cell selection is FACS (Fluorescence-activated cell sorting). FACS is a laser-based, biophysical technology, which can be employed in cell sorting by suspending cells in a stream of fluid and passing them by an electronic detection apparatus. Based on the specific light scattering and fluorescent characteristics cells of interest can be sorted from a heterogeneous mixture. Here, using the eGFP signal to perform cell selection via FACS could not be applied because induction of the GFP reporter also induces the expression of the CGG-repeat RNA and cells containing disease-related repeat length (95 CGG) show decreased cell viability and die 72-96 hours post induction (Arocena et al., 2005; Hoem et al., 2011;Selier et al., 2010). To avoid this problem the second vector construct (pRTR) was used enabling cell selection via magnetic beads coupled to an antibody against the cell-surface protein CD2. However, upon transfection of the pRTR vector, containing CGG-repeat RNA, GFP reporter and the PP7 two-component system for RNA-isolation, many COS7 cells died before protein expression could be induced. Also, the eGFP signal was not distinguishable from auto-fluorescence of dying cells (Figure 2.24 A). Overall, the more components the vectors contained and the larger the size of the vector, the more cells were prone of dying during transfection. This effect was not cell specific, since transfection of HeLa cells with the same construct showed the same result (Figure 2.24 B). For time limitations, this problem has not been resolved and therefore, cell selection for establishing a stable cell line has not been done, yet.

Taken together, the stable cell lines expressing CGG-repeat RNA could not be established, but the necessary vector constructs for expression of normal and disease-related CGG-repeat RNA have been generated (Figure 2.21). Only transfection and selection methods have to be further improved to obtain a stable cell line with strong eGFP and CGG-repeat RNA expression, respectively.

3.2.2. Pur-α's role in transcription

Pur-α is involved in transcription of several neuronal genes including the *myeline basic protein* (*Mbp*) gene (Darbinian et al., 2001; Haas et al., 1995) and the *myelin proteolipid protein 1* (*Plp1*) gene (Dobretsova et al., 2008), both responsible for myelination of nerve cells in the central nervous system (CNS). Pur-α deficient (PURA -/-) mice develop severe neurological problems and die within a few weeks after birth (Khalili et al., 2003). Still, Pur-α's role in transcription and why a lack of Pur-α is lethal in knockout mice has to be elucidated. Chip-Seq (chapter 2.2.1.1.) experiments with Pur-α could reveal which genes become activated by Pur-α, whether they share a consensus sequence and if they can be classified into sub-categories. Moreover, such findings would also contribute to understand why Pur-α deficient mice suffer from severe neurological defects and die shortly after birth.

In this study, BAC (bacterial artificial chromosome) constructs were generated to establish a stable cell line expressing either N-terminally or C-terminally eGFP-tagged human Pur-α (Figure 2.18). Tagging of Pur-α was necessary, since tested anti-Pur-α antibodies did not show specificity and would cross-react with other members of the Pur-family (Figure 2.17).

A Pur-α BAC clone was chosen because it contains the genomic Pur-α promoter and regulatory sequences. This way, tagged Pur-α protein expression underlies the same regulation as the protein expression of the endogenous *Pur-α* gene and does not have to be induced artificially. The BAC constructs were transfected into HeLa cells and subsequently selected for antibiotic resistance to establish a stable cell line. Again, after several weeks of selection, antibiotic resistant colonies were obtained (Figure 2.19). However, expression of eGFP-tagged Pur-α could neither be detected in Western Blot assays (Figure 2.20) nor by fluorescence microscopy. This can be explained by either silenced eGFP-Pur-α expression or by exclusion of the *eGFP-Pur-α* gene from the genome during cell selection, while maintaining the antibiotic resistance. Since antibiotic selection seems to be insufficient for generating the stable cell line, future clone selections will be done by performing FACS. The eGFP signal will facilitate to sort the cells expressing eGFP-tagged Pur-α and to accumulate positive clones for further amplification and establishment of a stable cell line.

3.2.3. Expression of Pur-α mutants in the FXTAS Drosophila model

Jin et al have generated transgenic *Drosophila* flies that selectively express FXTAS-related CGG-repeat RNA in the eyes (Jin et al., 2003). These flies show a CGG-mediated eye neurodegeneration phenotype, which can be rescued by overexpression of Pur-α. Furthermore,

Pur-α co-localizes in nuclear inclusions induced by CGG-RNA expression in the fly model (Ji et al., 2003; Jin et al., 2007; Sofola et al., 2007).

In this study, the crystal structure of the Pur-α/DNA co-complex revealed the residues that are involved in nucleic acid binding. Binding assays with mutant Pur-α further confirmed the interaction sites seen in the crystal structure. To test whether these Pur-α mutants also have an effect *in vivo*, I am collaborating with the Jin lab to use the FXTAS fly model for mutant Pur-α expression. For this purpose, I generated vectors containing the full length *Drosophila Pur-α* gene harboring the respective mutations: OSK I - KNR II, F II, KNR II (Chapter 2.1.4.2.) and R80A/R158A (Graebsch et al., 2009). These vectors are currently used in Jin's lab to generate transgenic flies expressing mutant Pur-α together with CGG-repeat RNA in the fly's eyes. Once these flies are established, they will be first studied by light-microscopic analysis of the facette eyes and then further investigated by using several techniques, such as confocal imaging and transmission electron microscopy of brain sections as well as immunohistochemistry and SEM (scanning electron microscopy) of fly eyes.

The generated transgenic flies could answer the following questions: Do the mutations of Pur-α weaken the nucleic acid binding ability also *in vivo* and thereby impair Pur-α's normal function? Does mutant Pur-α have an effect on nuclear inclusion formation? Does mutant Pur-α become sequestered into the nuclear inclusions? Can mutant Pur-α still rescue the eye neurodegeneration phenotype? Since one mutant was primarily defective in its unwinding activity, it might also be possible to distinguish between different functions.

In summary, my structure-to-functional analysis should yield a comprehensive understanding of Pur-α's binding mode to nucleic acids and therefore its function in cellular processes, such as transcription, where unwinding of duplex DNA might be essential for transcription regulation. Furthermore, my results might help to understand Pur-α's role in neurodegenerative diseases, such as FXTAS. Here, RNA-binding might lead to sequestration into the inclusions and loss of unwinding might even contribute to pathogenesis of this disease.

4. Materials and Methods

4.1. Consumables and chemicals

All common chemicals used were purchased from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany), Roche (Mannheim, Germany) and Sigma-Aldrich (Hamburg and Seezle, Germany), unless stated otherwise. Enzymes and nucleotides for molecular biology as well as molecular weight markers and loading dyes for gel electrophoresis were ordered from Fermentas (St. Leon-Rot, Germany), New England Biolabs (Frankfurt, Germany) and Invitrogen (Karlsruhe, Germany). Reagents for bacterial cell cultures were obtained from Becton, Dickinson & Co (Heidelberg, Germnay) and Sigma Aldrich. Media for mammalian cell culture was purchased from Invitrogen. Chromatography was performed with materials and columns of GE Healthcare (Munich, Germany). Radioactive nucleotides (γ-^{32}P-ATP) for labeling of nucleic acids were obtained from Hartmann Analytic (Braunschweig, Germany). DNA and RNA oligonucleotides were ordered from Eurofins MWG (Ebersberg, Germany) and Thermo Fisher (Ulm, Germany). Crystallization screens, tools and reagents were purchased from Hampton Research (Aliso Viejo, USA) and Qiagen (Hilden, Germany).

4.2. Oligonucleotides

4.2.1. DNA oligonucleotides for cloning

No.	Name	Sequence 5'-3'
1	PurA-KNRII-for	GTATTACTTGGACTTAGCAGAAGCTGCGGCTGGCCGATTTTTACG
2	PurA-KNRII-rev	CGTAAAAATCGGCCAGCCGCAGCTTCTGCTAAGTCCAAGTAATAC
3	PurA-F145A for	GAAAATGCGCGTGGCCGAGCTTTACGGGTATCGCAAAC
4	PurA-F145A rev	GTTTGCGATACCCGTAAAGCTCGGCCACGCGCATTTTC
5	PurA-R80A for	GATTGGCGCTGATGGTAGAGCAAGTCAAATTTACTTGGC
6	PurA R80A rev	GCCAAGTAAATTTGACTTGCTCTACCATCAGCGCCAATC
7	PurA-R158A for	CAATAACAAGAGGGGGGCCTGCATCTCAAATCGCTTTACCG
8	PurA-R158 rev	CGGTAAAGCGATTTGAGATGCAGGCCCCCCTCTTGTTATTG
9	PurA-QSK for1	GCGTGCAGAGATTTCAATGCTTGAAG
10	PurA-QSK-G-revMut	CGTCCACCTCCTATTTGCAACATTTTC
11	PurA-QSK-G-forMut	AATAGGAGGTGGACGATTTTATTTGGATG
12	PurA-QSK rev1	GAGGCAGATCGTCAGTCAGTCAC
72	PurA-KNR I for mut	GTAGCACAAGCTAGAGCAGGCCGTTTTAT
73	PurA-KNR I rev mut	GCCTGCTCTAGCTTGTGCTACATCCAAAT
74	PurA-KNR I rev1	GAGGCAGATCGTCAGTCAGTCACG
13	PurA-F68A-for	GAAGAGGCCGTGCTATAAAGGTTGC
14	PurA-F68A-rev	GCAACCTTTATAGCACGGCCTCTTC
15	PP7CP-SwaIfor	AAAATTTAAATATGTCCAAAACCATCGTTCTTTCGGTCGG
16	PP7CP-SwaIrev	TTTATTTAAATTACGGCCCAGCGGCACAAG

Table 4. 1 DNA oligonucleotides for cloning

4.2.2. DNA oligonucleotides for interaction studies and crystallization

No.	Name	Sequence 5'-3'
53	GGN	TCAGAGCCGCCACCCTCA
54	Fluo-SL DNA	FAM-CCAGGGCACTTAAAAAAATTCGCCTGG-DAB
57	CGG DNA (24nt)	CGGCGGCGGCGGCGGCGGCGGCGG
55	CGG DNA (7nt)	GCGGCGG
88	Fluo-MF0766-5'end (7nt)	GTGGTGG-fluorescein
89	Fluo-MF0677-3'end (7nt)	AGAGAAA-fluorescein
58	MF0677 DNA (24nt)	GGAGGTGGTGGAGGGAGAGAAAAG
62	myc-MF0677 (150nt)	AACTCAACGGGTAATAACCCATCTTGAACAGCGTACATGCTATAC ACGCACCCCTTTCCCCCGAATTGTTTTCTCTTTTGGAGGTGGTGGA GGGAGAGAAAAGTTTACTTAAAATGCCTTTGGGTGAGGGACCAAG GATGAGAAGAATGT
63	myc-MF0677-revC (150nt)	ACATTCTTCTCATCCTTGGTCCCTCACCCAAAGGCATTTTAAGTAA ACTTTTCTCTCCCTCCACCACCTCCAAAAGAGAAAACAATTCGGGG GAAAGGGGTGCGTGTATAGCATGTACGCTGTTCAAGATGGGTTATT ACCCGTTGAGTT
71	NMR DNA (5nt)	GCGGA

Table 4. 2 DNA oligonucleotides for interaction studies and crystallization

4.2.3. RNA oligonucleotides for interaction studies and crystallization

No.	Name	Sequence 5'-3'
51	CGG (25nt)	GCGGCGGCGGCGGCGGCGG
52	Fluo-SL RNA	FAM-CCAGGGCACUUAAAAAAAUUCGCCUGG-DAB
56	CGG RNA (7nt)	GCGGCGG
59	CGG RNA (24nt)	CGGCGGCGGCGGCGGCGGCGGCGG
60	MF0677 RNA (24nt)	GGAGGUGGUGGAGGGAGAGAAAAG
61	(G4C2)4 (24nt)	GGGGCCGGGGCCGGGGCCGGGGCC

Table 4. 3 RNA oligonucleotides for interaction studies and crystallization

4.3. Plasmids

4.3.1. Commercial plasmids

Name	Application	Tag	Antibiotic	Source
pGEX-6P-1	protein expression in bacteria	GST	Amp	GE Healthcare
M13mp18 ssDNA	DNA unwinding assay	-	-	New England Biolabs
pEX-A-MF0677	DNase I footrint	-	Amp	Eurofins MWG

Table 4. 4 Commercial plasmids

4.3.2. Plasmids for recombinant protein expression in *E. coli*

No.	Name/Insert*	Vector	Primer	Restriction sites	Template	Note
1	dmPurA 40-185	pGEX-6P-1	-	BamHI/XhoI	-	a)
2	dmPurA 40-255	pGEX-6P-1	-	BamHI/XhoI	-	a)
4	dmPurA 40-185 I69M	pGEX-6P-1	39/40	BamHI/XhoI	1	
5	dmPurA 40-185 S88M	pGEX-6P-1	43/44	BamHI/XhoI	1	
6	dmPurA 40-185 L135M	pGEX-6P-1	47/48	BamHI/XhoI	1	
7	dmPurA 40-185 F145A	pGEX-6P-1	3/4	BamHI/XhoI	1	
8	dmPurA 40-185 F68A	pGEX-6P-1	13/14	BamHI/XhoI	1	
9	dmPurA 40-185 QSK I – KNR II	pGEX-6P-1	9/10/11/12	BamHI/XhoI	11	
10	dmPurA 40-180 KNR I	pGEX-6P-1	9/72/73/74	BamHI/XhoI	1	
11	dmPurA 40-185 KNR II	pGEX-6P-1	1/2	BamHI/XhoI	1	
12	dmPurA 40-185 R80A	pGEX-6P-1	5/6	BamHI/XhoI	1	
13	dmPurA 40-185 R80A/R158A	pGEX-6P-1	7/8	BamHI/XhoI	12	
14	dmPurA 185-260	pGEX-6P-1	-	BamHI/XhoI	-	a)
15	hPurA 56-287 C272S	pGEX-6P-1	-	BamHI/XhoI	-	b)

Table 4. 5 Plasmids for recombinant protein expression in *E. coli*

*dmPurA refers to *D.melanogaster* Pur-α, hPurA refers to human Pur-α. The numbers in the name indicate the start and stop site of the amino sequence.
a) Created by Dr. Almut Graebsch, Gene Center Munich (Germany)
b) Obtained from Prof. Dierk Niessing, Helmholtz Zentrum München (Germany)

4.3.3. Plasmids for RNA/protein expression in mammalian cell lines

No.	Name	Vector	Insert*	Restriction sites	Template	Note
17	pRTS-1	pRTS-1	-	-	-	a)
18	pRTS-1_30CGG	pRTS-1	*fmr1* 5'UTR-$(CGG)_{30}$	XhoI/BlpI	46	
19	pRTS-1_95CGG	pRTS-1	*fmr1* 5'UTR-$(CGG)_{95}$	XhoI/BlpI	47	
20	pRTS-1_30CGG_PP7	pRTS-1	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{30}$, eGFP-PP7 CP	SfiI (PP7 SL-fmr1 5'UTR-(CGG)n) AscI/SwaI (eGFP-PP7 CP)	46/43/44	
21	pRTS-1_95CGG_PP7	pRTS-1	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{95}$, eGFP-PP7 CP	SfiI (PP7 SL-fmr1 5'UTR-(CGG)n) AscI/SwaI (eGFP-PP7 CP)	47/43/44	
22	pRTS-1_30CGG_PP7SL	pRTS-1	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{30}$	SfiI	46/43	
23	pRTS-1_95CGG_PP7SL	pRTS-1	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{95}$	SfiI	47/43	
25	pUC19-SfiI	pUC19	-	-	-	a)
26	pUC-30CGG	pUC19	*fmr1* 5'UTR-$(CGG)_{30}$	EcoRV	46	
27	pUC-95CGG	pUC19	*fmr1* 5'UTR-$(CGG)_{95}$	EcoRV	47	
28	pUC-30CGG_PP7SL	pUC19	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{30}$	BamHI (PP7 SL) EcoRV (*fmr1* 5'UTR-$(CGG)_{30}$	46/43	
29	pUC-95CGG_PP7SL	pUC19	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{95}$	BamHI (PP7 SL) EcoRV (*fmr1* 5'UTR-$(CGG)_{95}$	47/43	
31	pRTR-CD2-NIG	pRTR	-	-	-	a)
50	pRTR-1_30CGG_PP7	pRTR	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{30}$, eGFP-PP7 CP	SfiI (PP7 SL-fmr1 5'UTR-(CGG)n) AscI/SwaI (eGFP-PP7 CP)	20	
51	pRTR-1_95CGG_PP7	pRTR	27xPP7 SL, *fmr1* 5'UTR-$(CGG)_{95}$, eGFP-PP7 CP	SfiI (PP7 SL-fmr1 5'UTR-(CGG)n) AscI/SwaI (eGFP-PP7 CP)	21	
43	pCR4-24xPP7-SL	pCR4	24xPP7 SL	BamHI/?	-	b)
44	pET22HT-PP7delFG	pET22	PP7 CP		-	b)
46	pCEP4-TRE-T-30-EGFP2	pCEP4	*fmr1* 5'UTR-$(CGG)_{30}$	XhoI/BlpI	-	c)
47	pCEP4-TRE-T-95-EGFP2	pCEP4	*fmr1* 5'UTR-$(CGG)_{95}$	XhoI/BlpI	-	c)

Table 4. 6 Plasmids for RNA/protein expression in mammalian cell lines

*SL refers to stem loop, CP refers to coat protein.
a) provided by Prof. Georg Bornkamm, Helmholtz Zentrum München (Germany)
b) provided by Prof. Ralf Jansen, MPI Tübingen (Germany)
c) provided by Prof. Flora Tassone, UC Davis (USA)

4.4. BAC (Bacterial artificial chromosome) clone, tagging cassettes, tagging and verification oligonucleotides

BAC name	Vector	Insert	length (bp)	Gene start	Gene end	Source
RP11-1106D15	pBAC 3.6	Chromosome 5 genomic sequence of human Pur-alpha flanked by regulatory regions	180122	131561	140520	Empire Genomics (Buffalo, USA)

Table 4. 7 BAC clone

Tagging cassette	Tag	Source
N-term: R6K-NFLAP	eGFP/PreScission site/S-peptide/TEV site/	Dr. Ina Poser, MPI Dresden (Germany)
C-term: R6K-LAP	Tev site/S-peptide/PreScission site/eGFP	Dr. Ina Poser, MPI Dresden (Germany)

Table 4. 8 Tagging cassettes

Tagging oligonucleotide	Sequence 5'-3'
N-term GFP tag for	GCGGCGGGCGGAGCGGCAGGCGGCGGCGGCGCGGCAGCGGAGCGCAGCATCATG GTGTCCAAGGGCGAGGAACTG
N-term GFP tag rev	GCCCGAACCCAGCGCCGCACCACCCTGCTCGCTGCCGCTGTCTCGGTCCGCGGCCC TGGGCAGGTCGTCGGTCAG
C-term GFP tag for	CCGCCGCTGCCACCCTGCTACTGCAGGGTGAGGAAGAAGGGGAAGAAGATGATTA TGATATTCCAACTACTG
C-term GFP tag rev	TGTGTGTGTGTATGCATGTGTGTGTGTGGGGGTTTCATTCAGTTTGATTCAGAAG AACTCGTCAAGAAG

Table 4. 9 Tagging oligonucleotides

Verification oligonucleotide	Sequence 5'-3'
N-term verify for	CCTGAAGTTCATCTGCACCA
N-term verify rev	CCTTCAGCTCGATCCTGTTC
C-term verify for	GTGTTTATGCGAGTGAGCGA
C-term verify rev	AGTTGGTGGAGGTTCTGTGG

Table 4. 10 Verification oligonucleotides

4.5. *E. coli* strains

Name	Genotype	Source
XL-1 Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lac1^q ZΔM15Tn10 (Tet^r)]	Stratagene (La Jolla, USA)
BL21 Star (DE3)	B F$^-$ ompT hsdS($r_B^-m_B^-$) dcm+ Tet^r galλ (DE3) EndA Hte [argU ile Y leu W Cam^r]	Stratagene (La Jolla, USA)
B834 (DE3)	F$^-$ ompT gal met r_Bm_B	Novagen (Schwalbach, Germany)

Table 4. 11 *E. coli* strains

4.6. Mammalian cell lines

Name	Description	Source
HeLa (**He**nrietta **La**cks)	Human epithelial cells from cervical carcinoma	Obtained from Dr. Marta Pabis, Helmholtz Zentrum München (Germany)
COS7	Fibroblast-like cells from african green monkey kidney tissue	Provided by Prof. Daniel Krappmann, Helmholtz Zentrum München (Germany)

Table 4. 12 Mammalian cell lines

4.7. Media and supplements for bacterial cell culture

Medium	Composition
LB (Luria-Bertani medium)	1 % (w/v) bacto tryptone, 0.5 % (w/v) bacto yeast extract, 0.5 % (w/v) NaCl, add 1.5 % (w/v) agar for plates
SeMet Medium	7.5 mM $(NH_4)_2SO_4$, 8.5 mM NaCl, 55 mM KH_2PO_4, 100 mM K_2HPO_4, 1 mM $MgSO_4$, 20 mM glucose, 1 mg/l $CaCl_2$, 1 mg/l $FeCl_2$, 1 µg/l of the following trace element: Cu^{2+}; Mn^{2+}; Zn^{2+}; MoO_4^{2+}, 1 mg/l Thiamine, 1 mg/l Biotin, 100 mg/l of the following amino acids (L-alanine, L-arginine, L-aspartic acid, L-cysteine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tyrosine, L-valine, L-selenomethionine
NMR Medium	33.7 mM Na_2HPO_4-$2H_2O$, 22 mMl KH_2PO_4, 8.55 mM NaCl, 9,35 mM $^{15}NH_4Cl$, 0.4 % (w/v) glucose, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$, 1 µg biotin, 1 µg thiamin and trace elements as follows: 50 mg/l EDTA, 8.3 mg/l $FeCl_3$-$6H_2O$, 0.84 mg/l $ZnCl_2$, 0.13 mg/l $CuCl_2$-$6H_2O$, 0.1 mg/l H_3BO_3, 0.016 mg/l $MnCl_2$-$4H_2O$

Table 4. 13 Media for bacterial cell culture

Supplements	Application	Final concentration
Ampicillin	*E. coli* selection	100 µg/ml
Kanamycin	*E. coli* selection	50 µg/ml
Chloramphenicol	*E .coli* selection	34 µg/ml
IPTG (Isopropyl-β-D-thiogalactopyranosid)	Induction of protein expression in *E. coli*	0.25 mM

Table 4. 14 Supplements for bacterial cell culture

4.8. Medium, supplements and reagents for mammalian cell culture

Medium	Composition	Application
Complete Medium	(v/v) 90 % DMEM, (v/v) 10 % FBS, 100 U/ml penicillin/streptomycin	Normal cell culture
Reduced serum medium	(v/v) 100 % OptiMEM	Transfection
Freezing medium	(v/v) 70 % DMEM, (v/v) 20 % FBS, 100 U/ml Pen/Strep, (v/v) 10 % DMSO	Cell freezing

Table 4. 15 Media for mammalian cell culture

Supplements	Application	Final concentration	Source
Gibco® DMEM (4.5 g/l glucose, L-glutamine, phenol red)	Cell culture	(v/v) 70 or 90 %	Invitrogen (Karlsruhe, Germany)
Gibco® OptiMEM (L-glutamine, phenol red)	Transfection	(v/v) 100 %	Invitrogen (Karlsruhe, Germany)
FBS (fetal bovine serum)	Cell culture	(v/v) 10 or 20 %	Invitrogen (Karlsruhe, Germany)
Penicillin / Streptomycin (Pen/Strep)	Antibiotic / cell culture	100 U/ml	Invitrogen (Karlsruhe, Germany)
DMSO (Dimethylsulfoxid)	Cell freezing	10 % (v/v)	Sigma-Aldrich (Hamburg and Seezle, Germany)
0.5 % Trypsin-EDTA	Cell splitting	(v/v) 0.05 %	Invitrogen (Karlsruhe, Germany)
Doxycycline hyclate	Induction of gene expression	1 µg/ml	Sigma-Aldrich (Hamburg and Seezle, Germany)
Geneticin (G418)	Cell selection	50-500 µg/ml	Invitrogen (Karlsruhe, Germany)
Hygromycin B	Cell selection	50-500 µg/ml	Invitrogen (Karlsruhe, Germnay)
1x PBS (phosphate-buffered saline)	Cell wash	1x	Invitrogen (Karlsruhe, Germany)

Table 4. 16 Supplements for mammalian cell culture

Reagents	Application	Source
FuGENE HD	Transfection	Promega (Mannheim, Germany)
Lipofectamine® 2000	Transfection	Invitrogen (Karlsruhe, Germany)

Table 4. 17 Reagents for mammalian cell culture

4.9. Antibodies

Primary antibody	Clonality / Source	Dilution / Application	Source
Anti-α-actin	Monoclonal / mouse	1:1000 / western blot	Millipore (Schwalbach, Germany)
Anti-GFP	Monoclonal / mouse	1:1000 / western blot	Covance/Hiss Diagnostics (Freiburg, Germany)
Anti-human-Pur-alpha	Monoclonal / mouse	1:200 /western blot	Santa Cruz (Heidelberg, Germany)
Anti-human-Pur-alpha (1D7, 3C12, 4E2)	Monoclonal / rat	1:50 / western blot	Dr. Elisabeth Kremmer, Helmholtz Zentrum München (Germany)
αB-crystallin (17H8, 19D10, 7E2, 5B2, 13F3, 23E6, 2H4, 19E6, 18D5, 7F7, 2B3, 22F8, 2E1, 20D2, 9C8)	Monoclonal / rat	1:50 / western blot	Dr. Elisabeth Kremmer, Helmholtz Zentrum München (Germany)

Table 4. 18 Primary antibodies

Secondary antibody	Clonality / Source	Dilution /Application	Source
Anti-mouse-HRP	Polyclonal / goat	1:3000 / western blot	BioRad (Munich, Germany)
Anti-rat-HRP	Monoclonal / mouse	1:2000 / western blot	Dr. Elisabeth Kremmer Helmholtz Zentrum München (Germany)

Table 4. 19 Secondary antibodies

4.10. General buffers and stock solutions

Name	Composition	Application
4x stacking gel buffer	0.5 M Tris, (w/v) 0.4 % SDS, pH 6.8	SDS PAGE
4x separating gel buffer	3 M Tris, (w/v) 0.4 % SDS, pH 8.5	SDS PAGE
10x TGS	0.25 M Tris, (w/v) 1 % SDS, 1.9 M glycine	SDS PAGE
4x loading dye	110 mM Tris/HCl pH 6.8, (v/v) 40 % glycerol, (w/v) 0.5 % bromophenol blue, (w/v) 4 % SDS, 40 mM DTT	SDS PAGE
Coomassie staining solution	(v/v) 50 % ethanol, (v/v) 7 % acetic acid, (w/v) 0.2 % Coomassie Brilliant blue R-250	SDS PAGE
1x TBE	8.9 mM Tris, 8.9 mM boric acid, 2 mM EDTA (pH 8.0)	Native PAGE
Formamide loading dye	(v/v) 80 % deionized formamide, 1 mg/ml xylene cyanol, 1 mg/ml bromphenol blue	Denaturing polyacrylamide/TBE gel electrophoresis
50x TAE	2 M Tris-base, (v/v) 5.71 % acetic acid, 50 mM EDTA (pH 8.0)	Agarose gel electrophoresis
10x PBS	80 g NaCl, 2 g KCl, 14,2 g Na_2HPO_4, 2,4 g KH_2PO_4 in 1 l H_2O; pH 7,4	Western blotting
PBS-T	1x PBS, 1:1000 Tween®20	Western blotting
10x Semi-dry transfer buffer	30 g Tris, 113 g Glycine in 1 l H_2O	Western blotting
1x Semi-dry transfer buffer	(v/v) 10 % semi-dry transfer buffer (10x), (v/v) 20 % methanol	Western blotting
Blocking buffer	PBS-T, (w/v) 5 % milk powder	Western blotting
2x magic mix	(w/v) 48 % Urea, 15 mM Tris-HCl pH 7.5, (v/v) 8.7 % glycerol, (w/v) 1 % SDS, (w/v) 0.004% bromophenolblue, 143 mM β-mercaptoethanol (add fresh)	Western blotting

Table 4. 20 General buffers and stock solutions

4.11. Molecular biology

4.11.1. Cloning

Cloning was done by standard methods as described in Sambrook and Russell 2001. Target genes were amplified by polymerase chain reaction (PCR). PCR products were separated by agarose gel electrophoresis in 1x TAE buffer, stained with GelRed (Biotium, Hayward, USA)

and subsequently purified with the NucleoSpin® PCR clean up kit (Macherey-Nagel, Düren, Germany). After digestion with restriction enzymes, DNA fragments were ligated into linearized and dephosphorylated vectors (FastAP, Fermentas, St. Leon-Roth, Germany). Restriction enzymes and buffers were used according to the manufacturer's protocols. Point mutations were introduced using either the QuickChange II Site-Directed Mutatgenesis Kit (Agilent Technologies, Waldbronn, Germany) or by PCR-based site-directed mutagenesis (Ho, Hunt et al. 1989). Cloned constructs and sequences of the corresponding oligonucleotides used are described in chapter 4.2. and 4.3, respectively. Nucleic acid concentrations were measured with NanoDrop Spectrophotometer (Thermo Fisher, Ulm, Germany) at 260 nm.

4.11.2. Transformation of E. coli and isolation plasmid DNA

Chemically competent *E. coli* cells were prepared according to Hanahan 1983. Plasmid DNA was transformed as described in Sambrook and Russell 2001. Transformants were selected by antibiotic resistence on LB agar plates. A single colony was used to inoculate 5 ml LB medium supplemented with respective antibiotics and grown over night at 37 °C. Plasmid DNA was isolated using the NucleoSpin® Plasmid kit (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol.

Cloned vector constructs were verified by PCR, restriction digestion or sequencing through Eurofins MWG (Ebersberg, Germany).

4.12. Bioinformatics

4.12.1. Protein parameters

Physical and chemical protein parameters were determined with the ExPASy Proteomics Server (http://web.expasy.org/protparam). The molar extinction coefficient, molecular weight and isoelectric point of protein fragments were calculated with the ProtParam tool (Gasteiger et al. 2005). Protein concentrations were calculated from the absorption at 280 nm measured with NanoDrop Spectrophotometer (Termo Fisher, Ulm, Germany) and the respective extinction coefficient according to the Lambert-Beer law ($E_\lambda = \varepsilon_\lambda * c * d$).

4.12.2. Nucleic acid parameters and primer design

The molar extinction coefficient and the molecular weight for oligonucleotides were either provided by the supplier (MWG, Thermo Fisher) or determined with the OligoCalc server (http://www.basic.northwestern.edu/biotools/oligocalc.html) (Kibbe, 2007) or the IDT Biophysics server (http://biophysics.idtdna.com/UVSpectrum.html). Concentrations were calculated from the absorption at 260 nm measured with NanoDrop Spectrophotometer and the respective extinction coefficient according to the Lambert-Beer law ($E_\lambda = \varepsilon_\lambda * c * d$). The ratio of absorbance at 260 and 280 nm was used to assess the purity of the nucleic acid. A ratio of ~1.8 was considered as pure DNA and a ratio of ~2.0 as pure RNA.

Primers for standard PCR were designed using Primer3web tool (http://primer3.ut.ee/) (Utergrasser et al. 2012). Mutagenic primers for site-directed mutagenesis were designed using the PrimerX tool (http://www.bioinformatics.org/primerx/).

4.12.3. Sequence alignment

DNA and protein sequences were obtained from the NCBI database (www.ncbi.nlm.gov). Multiple sequence alignments were performed with T-coffee (Notredame et al., 2000).

4.13. Protein expression and purification

4.13.1. Protein Expression

E. coli BL21 (DE3) cells transformed with pGEX-6P-1::Pur-alpha fragments were grown at 37 °C in 3 litres of LB medium supplemented with 100 µg/ml ampicillin.

For ^{15}N-labeling of protein cells were grown in NMR medium.

For selenomethionine-substituted protein the methionine auxotrophic *E. coli* strain B834 (DE3) was used and cells were grown in SeMet medium.

After reaching an OD_{600} of 0.6, cell cultures were cooled down to 18 °C and expression was induced at an OD_{600} of 0.8 by adding 0.25 mM IPTG. Cells were harvested after 12-16 hours of expression by centrifugation at 4 °C (4,000 rpm, 20 min). Pellets were resuspended in lysis buffer (500 mM NaCl, 50 mM Hepes, pH 7.5), flash frozen in liquid nitrogen and stored at -80 °C.

4.13.2. Protein Purification

All purification steps were carried out at 4 °C. After thawing, one tablet of EDTA (ethylenediaminetetraacetic acid)-free protease inhibitor cocktail (Roche, Germany) and 0.4 mM PMSF (phenylmethylsulfonyl fluoride) were added to the cell suspension. Cells were lysed by sonication with a Branson Sonifier 250 (Emerson, Danbury, USA) (4x 4 min, amplitude 40 %, output 4) and cell debris was removed by centrifugation (16,000 rpm, 20 min). GST-tagged proteins were purified by GST affinity chromatography (GE Healthcare). Therefore, the lysate was loaded onto a pre-equilibrated 5 ml GST-Trap FF column. After extensive washing with loading buffer, elution was done in 20 ml of elution buffer containing 500 mM NaCl, 20 mM Hepes pH 8.0 and 25 mM glutathione. The GST tag was cleaved off by adding 50 µg PreScission–protease (GE Healthcare) prior to dialysis against buffer composed of 250 mM NaCl, 20 mM Hepes pH 8.0. After protease cleavage, the GST tag was removed using a GST column and nucleic acids were removed using a HiTrapQ column (GE Healthcare). Pur-alpha was further purified by a Heparin column (GE Healthcare) and subsequently eluted from the Heparin column with buffer containing 2 M NaCl and 20 mM Hepes pH 8.0. The elution fraction was concentrated in a centrifugal filter device (Amicon Ultra, Millipore, Billerica, USA), then centrifuged (13,000 rpm, 20 min) to remove aggregates and finally purified by size exclusion chromatography with a Superdex 75 10/30 GL column (GE Healthcare) in buffer composed of 250 mM NaCl or 500 mM NaCl (F I and F II mutant) and 20 mM Hepes pH 8.0. Peak fractions were again concentrated, centrifuged and either used for crystallization or flash frozen in liquid nitrogen with 10 % glycerol added and stored at -80 °C. For cysteine-containing and for selenomethionine-substituted proteins 2 mM DTT was added to all buffers.

For NMR experiments size exclusion chromatography was performed in 50 mM potassium phosphate buffer pH 7.0 and 200 mM NaCl (NMR buffer).

Absence of nucleic acid contamination was confirmed with UV-spectroscopy by measuring the ratio of absorbance at 260 nm and 280 nm. Only samples with a ratio A260/A280 of ≤ 0.6 were used and considered as nucleic acid free.

4.14. Methods for protein analysis

4.14.1. SDS PAGE

For monitoring purification progress and protein purity, as well as for western blot analysis, proteins were separated by dodecylsulfate polyacrylamide gel electrophoresis (SDS PAGE),

as described in Laemmli, 1970. Depending on the protein size, 10-15 % polyacrylamide gels were run, followed by Coomassie blue staining.

4.14.2. Circular dichroism (CD) spectroscopy

To confirm proper protein folding of the Pur-alpha derivatives CD spectra (wavelength 190-260 nm) were recorded with a JASCO-715 spectropolarimeter at 5 °C in a 0.1-cm cuvette. Proteins were diluted in buffer containing 250 mM NaCl, 20 mM Hepes pH 8.0 and 2 mM DTT to a final protein concentration of 30 µM in 300 µl total volume. Five scans were taken with a speed of 50 nm/min.

4.15. Structural biology

4.15.1. Co-crystallization of Pur-α and nucleic acid and structure determination

Crystals of selenomethionine-substituted Pur-alpha repeat I-II in complex with CGG 7mer ssDNA (ratio 1:2.2) were grown at 21 °C by hanging-drop vapor-diffusion using a 2:1 mixture of protein-DNA-complex (final protein concentration 1.77 mg/ml) and crystallization solution containing 50 mM MES pH 5.2, 500 mM $(NH_4)_2SO_4$, 1 mM TCEP and 16% PEG400. The drop size was 3 µl and the total reservoir volume 500 µl using the 24-well EasyXtal Crystal Support (Qiagen, Hilden, Germany). Rod-shaped crystals of 160x19.2 µm size appeared within 4 days. Prior to data collection crystals were cryoprotected in mother liquor and flash frozen in liquid nitrogen. Native dataset was recorded at beamline ID23-2 [European Synchrotron Radiation Facility (ESRF) Grenoble, France]. Crystals diffracted up to 2.0 Å. Data were integrated and scaled with XDS (Kabsch 1993). Structure was solved by molecular replacement with PHASER (McCoy et al. 2007) using the apo-structure of *Drosophila* Pur-alpha 40-185 (PDB ID 3K44) and was completed using COOT (Emsley and Cowtan 2004). Refinement of the native data was performed with PHENIX (Adams et al. 2010) using NCS and TLS. The final model was analyzed with SFCHECK (Vaguine et al. 1999) PHENIX and REFMAC (Murshudov et al. 1997, Terwilliger 2002). Images of the crystal structure, superimpositions of the co-complex and apo-structure, as well as electrostatic surface potentials were prepared with PyMol (Schrodinger 2010) (DELano; http://www.pymol.org/).

Crystallization trials of native and selenomethionine-substituted Pur-alpha repeat I-II (1.4-2.0 mg/ml) in complex with RNA CGG 7mer (ratio 1:2.2) were performed in crystallization

solution containing 50 mM MES pH 5.0-5.8, 400-700 mM $(NH_4)_2SO_4$, 1 mM TCEP and 8-18 % PEG400, but yielded no crystals.

4.16. Protein-DNA/RNA interaction studies

4.16.1. RNase-free water

All RNA experiments were performed with RNase-free materials, reagents and buffers made from RNase free water. RNase-free water was prepared by adding (v/v) 0.05 % DEPC (diethylpyrocarbonate) (Roth, Karlsruhe, Germany) to Millipore-purified water. After incubation at 37 °C over night, the DEPC was inactivated by autoclaving.

4.16.2. NMR experiments

All NMR spectra were recorded in NMR buffer with 5 % D_2O at 298 K using Bruker Avance III spectrometer equipped with TCI cryogenic probe heads, at field strengths corresponding to 900 MHz proton Larmor frequency. To study DNA/RNA binding ^{1}H ^{15}N HSQC spectra were recorded of ^{15}N-labeled protein (50 µM) titrated with nucleic acids with different stoichiometric ratio of protein:nucleic acid (1:0.25, 1:0.5, 1:0.75, 1:1, 1:1.25, 1:1.5, 1:2.5 and 1:5). Spectra were acquired and processed with Topspin3.2 (Bruker, Karlsruhe, Germany) and analyzed with CCPNMR Analysis (Vranken et al., 2005).

4.16.3. Isotopic labeling of oligonucleotides and primer

For RNA-labeling RNase-free buffers, materials and reagents were used. 10 pmol of chemically synthesized DNA or RNA oligonucleotides were phosphorylated at the 5'-end with 10 pmol γ-^{32}P ATP by T4 polynucleotide kinase (New England Biolabs, Frankfurt, Germany) with buffer A in a final volume of 20 µl. Labeling reaction was carried out at 37 °C and stopped after 30 min by incubation at 70 °C for 10 min. Labeled oligonucleotides were purified by a NucAway™ Spin column (Ambion, Austin, USA) and stored at -20 °C.

For DNase I footprint experiments 10 pmol of MF0677 ssDNA (150nt) were labeled with 17 pmol γ-^{32}P ATP as described for oligonucleotides. The labeled DNA was then supplemented with 20 µl of formamide loading dye and purified by gel electrophoresis (400 V, 90 min) using a pre-warmed 6 % TBE polyacrylamide gel with 8 M urea. The DNA band was visualized by radiograph films, cut out and extracted from the gel by crush & soak technique. Thereby, the DNA gel was cut into little pieces, which were incubated with 300 µl of

extraction solution (0.5 M ammonium acetate, 0.1 mM EDTA, (v/v) 0.1 % SDS) at room temperature over night. Next day the labeled DNA had diffused into the extraction solution, which was then transferred into a fresh Eppendorf tube. After addition of 900 µl of (v/v) 100 % ethanol the labeled DNA was precipitated by incubation in a dry-ice/isopropanol bath for 30 min and subsequent centrifugation at 13,000 rpm for 30 min at 4 °C. After washing the DNA pellet with 1 ml of (v/v) 70 % ethanol and another centrifugation step (13,000 rpm, 10 min), the pellet was dried in a Concentrator plus (Eppendorf, Hamburg, Germany) centrifuge for 10 min. The dried, labeled DNA pellet was finally resuspended in 20 µl Millipore-purified water and stored at -20 °C.

4.16.4. Electrophoretic mobility shift assay (EMSA)

The protein-nucleic acid complexes were formed in RNase-free binding buffer containing 250 mM NaCl, 20 mM Hepes pH 8.0, 3 mM $MgCl_2$, 4 % glycerol, 2 mM DTT). Serial protein dilutions and a constant amount of radiolabeled nucleic acid (2.5 nM) were incubated for 20 min at 21 °C. DNA-binding experiments contained 25 µg/ml Salmon Sperm DNA, and RNA-binding experiments contained 100 µg/ml yeast tRNA competitor. Samples were loaded onto 6 % TBE polyacrylamide gels. After electrophoresis (45 min, 100 V) gels were incubated for 15 min in fixing solution ((v/v) 10% acetic acid, (v/v) 30% methanol), dried in a gel dryer (BioRad, Munich, Germany) and analyzed with radiograph films in a Protec Optimax developer (Hohmann, Hannover, Germany).

For fluorescence EMSA experiments 100 nM of fluorescein-conjugated nucleic acid was applied. Gels were analyzed directly after electrophoresis with a Typhoon 9200 scanner (Amersham Biosciences, Freiburg, Germany). Settings were as follows:

Emission filter: 526 SP Fluorescein/Alexa Fluor 488, PMT: 400-600, laser: Green (532), sensitivity: Normal, pixel size: 100 microns.

4.16.5. SAXS (small angle X-ray scattering)

All SAXS data were recorded on an in-house SAXS instrument (SAXSess mc2, Anton Paar, Graz, Austria) equipped with a Kratky camera, a sealed X-ray tube source and a two-dimensional Princeton Instruments PI•SCX:4300 CCD detector (Roper Scientific, Sarasota, USA). Scattering curves were measured in a final volume of 50 µl with different buffers and different concentrations of protein only or protein-DNA complexes (see table below) in line collimation mode. Data were evaluated and processed with different programs from the ATSAS 2.1 software package (Konarev, Petoukhov et al. 2006). Primary analysis of the data

was done with PRIMUS (Konarev, Volkov et al. 2003) and data processing was performed using SAXSQuant software (version 3.9). Desmearing of the data, the forward scattering, I(0), the radius of gyration, Rg, the maximum dimension, Dmax, and the inter-atomic distance distribution functions, P(r), were computed with the program GNOM (version 4.6).

Buffer	Protein concentration	Protein	DNA	Ratio protein:DNA
50 mM potassium phosphate buffer pH 7.0, 200 mM NaCl	50 µM	Pur-alpha repeat I-II	CGG DNA (7nt)	1:0, 1:1, 1:2
20 mM Hepes pH 8.0, 350 mM NaCl	50 µM	Pur-alpha repeat I-II	MF0677 DNA (24nt) CGG DNA (7nt)	1:0, 1:1, 1:2 1:0, 1:1, 1:2
20 mM Hepes pH 8.0, 500 mM NaCl, 2 mM DTT	50 µM	Pur-alpha repeat I-III	CGG DNA (7nt)	1:0, 1:1, 1:2
20 mM Hepes pH 8.0, 1 M NaCl, 2 mM DTT	50 µM	Pur-alpha repeat I-III	CGG DNA (7nt)	1:0, 1:1, 1:2

Table 4. 21 SAXS conditions: buffers, protein concentrations and ratios

4.16.6. Dot blot assay

Dot blot assays were performed as described in Wong and Lohman, 1993. Protein was titrated to a constant amount of 1 µM MF0677 ssDNA (thereof 2.5 nM radiolabeled) in a final volume of 80 µl and incubated for 20 min at 21 °C in binding buffer 150 mM NaCl, 20 mM Hepes pH 8.0. Nitrocellulose filter (Roth, Karlsruhe, Germany) was presoaked for 10 min in 0.4 M KOH followed by intensive washing with Milli-Q H_2O. Nitrocellulose and nylon filters (Roth, Karlsruhe, Germany) were then equilibrated in binding buffer for 15 min. Both filters (nitrocellulose, top; nylon filter, bottom) were placed into a dot-blot apparatus (BioRad, Munich, Germany). Vacuum was applied and the wells were washed once with 80 µl binding buffer before and after samples were loaded. The nitrocellulose filters were analyzed using a phosphor imager system to measure the retained radiolabeled oligonucleotides on the nitrocellulose filter. Quantification was done using the dot blot analyzer plug-in of the ImageJ 1.47v software (National Institute of Health, USA).

4.16.7. Unwinding assay

Unwinding assays were carried out according to Darbinian et al. 2001. A dsDNA substrate was prepared by annealing a complementary 18-mer oligonucleotide to a GGN motif of the M13mp18 ssDNA plasmid. The 18-mer (GGN) was labeled with γ-^{32}P ATP. Protein dilutions were added to a constant amount of dsDNA substrate (100 ng) in binding buffer composed of 150 mM NaCl, 20 mM Hepes pH 8.0. Samples were incubated at 37 °C for 1 h. The unwinding reaction was stopped by adding SDS to a final concentration of (v/v) 0.3 %. Samples were run on 9 % native polyacrylamide gels in 1x TBE buffer for 2 h 30 min at

200 V. Gels were incubated for 15 min in fixing solution ((v/v) 10 % acetic acid, (v/v) 30 % methanol), dried in a gel dryer and analyzed with radiograph films.

4.16.8. DNase I footprint

Radioactivity of labeled oligonucleotides was measured with the liquid scintillation analyzer TriCarb 2100TR (PerkinElmer, Massachusetts, USA) using the Cerenkov protocol (^{3}H channel). In a final volume of 25 µl, 100.000 counts per minute (cpm) of 5'end-radiolabeled ssDNA (150 nt) or dsDNA (160 nt), gained from restriction digest of pEX-A-MF0677 with HindIII and SacI, containing the myc-MF0677 sequence were applied for incubation with different protein concentrations for 20 min on ice. After addition of $MgCl_2$ and $CaCl_2$ to a final concentration of 7 mM and 3 mM, respectively, the protein-DNA-complex was digested at RT with DNase I (1-50 unit/ml). Cleavage was stopped after 1 min by adding an equal volume of DNase I stop solution (20 mM EDTA; 1 % SDS; 0.2 M NaCl, 100 µg/ml yeast tRNA). Subsequently, DNA fragments were extracted with phenol:chloroform followed by ethanol precipitation according to Sambrook and Russell, 2001. Dried DNA pellets were dissolved in formamide loading dye and boiled at 100 °C before loading on the gel.

For the sequencing reaction 100.000 cpm of radioactively labeled primer were annealed to 250 fmoles of unlabeled myc-MF0677 ssDNA (150 nt) in 5x reaction buffer (200 mM Tris-Cl pH 8,8 and 25 mM $MgCl_2$) for 2 min at 65°C and afterwards cooled to RT. Then 4 µl of each ddNTP Chain Extension/Termination mixture (15 µM of dNTP; ddNTP:dNTP ratios 30:1; 80:1, 80:1 and 30:1 for cytosine, thymine, adenine and guanine, respectively) was added together with 8 µl of Taq DNA-polymerase (1 unit/µl; diluted in 25 mM Tris pH 8,8; 0,01 mM EDTA pH 8,0; 0,15 % Tween-20 and 0,15 % Nonidet P-40). Probes were incubated at 72°C for 10 min. Sequencing reaction was stopped by adding formamide loading dye and boiling at 100°C.

The DNase I cleavage probes and the sequencing reactions were loaded with an equal amount of cpm onto 6 % TBE polyacrylamide gels with 8 M urea. After electrophoresis (90 min, 1800 V) gels were incubated for 15 min in fixing solution ((v/v) 10 % acetic acid, (v/v) 30 % methanol) and analyzed with radiograph films.

4.17. Cell culture

4.17.1. Cell cultivation

All reagents and materials used for cell culture were sterile. HeLa and COS7 cells were maintained in 75 cm^2 flasks (Greiner Bio-One, Frickenhausen, Germany) with Complete medium at 37°C in a humidified incubator of 5% CO_2. For splitting and seeding Complete medium was removed from the flasks and cells were washed with 10 ml 1x PBS. Trypsin-EDTA was diluted in 1x PBS to a final concentration of (v/v) 0.05 %. To detach the adherent cells from the flask 2 ml of diluted Trypsin-EDTA was added and flasks were incubated at 37 °C for 5 min. Trypsination was stopped with 10 ml Complete medium. Subsequently, the cell suspension was transferred to Falkon tubes (BD Biosciences, Heidelberg, Germany) and centrifuged at 1,200 rpm for 5 min. After discarding the supernatant, the cell pellet was resuspended in fresh Complete medium and the required amount of cells were transferred to new flasks (splitting) or 6-well plates (seeding). For seeding cells were counted with a Neubauer counting chamber and the required cell number depending on the experiment was transferred into 6-well plates.

4.17.2. Transfection

The previous day to transfection 0.5-2x10^5 cells were seeded into 6-well plates. When cells were 70-90 % confluent they were transfected using the Lipofectamine 2000 (Invitrogen, Karlsruhe, Germany) or the FuGENE HD (Promega, Mannheim, Germany) transfection reagent according to the manufacturer's protocol.

The amount of transfected DNA depended on the experiment:

Transfection of BAC (bacterial artificial chromosome) constructs: 100-250 µg DNA/well.

Transfection of CGG-repeat RNA constructs (pRTR, pRTS-1): 10 µg DNA/well.

4.17.3. Induction of protein/RNA expression

24 hours after transfection of the pRTR or pRTS-1 constructs, expression was induced by adding doxycycline to a final concentration of 1 µg/ml to the medium. Successful transfection and induction was verified by western blot or by monitoring the eGFP signal with a Biorevo BZ-9000 (Keyence, Neu-Isenburg, Germany) fluorescence microscope at 10x lens magnification and analysis with the BZ-II Analyzer software.

4.17.4. Establishment of a stable cell line expressing tagged human Pur-α

A BAC clone containing the genomic sequence of human Pur-alpha plus flanking regulatory regions was ordered from Empire Genomics (RP11-1106D15 in pBAC 3.6). For N- and C-terminal tagging of Pur-alpha two eGFP tagging cassettes (N-terminal tagging: R6K-NFLAP; C-terminal tagging: R6K-LAP) were kindly provided by Ina Poser (MPI Dresden). BAC and oligo sequences for BAC tagging were taken from the homepage of BACFinder-MitoCheck (www.mitocheck.org/cgi-bin/BACfinder). Tagging was performed using the Quick and Easy BAC Modification Kit (Gene Bridges) according to the manufacturer's protocol and Poser et al. 2008. Successful tagging was verified by amplification of eGFP via PCR.

The BAC constructs containing either N- or C-terminally eGFP tagged human Pur-alpha were then transfected into HeLa cells. Different amounts (6, 9, 12 or 15 µl) of Lipofectamine 2000 were used according to the manufacturer's protocol. Two days after transfection, cells were cultivated in media supplemented with 50 µg/ml G418 for selection. After additional 3 days the concentration of G418 was increased to 400 µg/ml and media was changed every 2-3 days to remove dead cells and cell debris. Within the next 3-4 weeks untransfected cells were dead. After 5 weeks G418 resistant colonies were further amplified into 75 cm^2 flasks. Cells were maintained in media containing 500 µg/ml G418. Expression of eGFP-tagged protein was analyzed with fluorescence microscopy and Western Blot. Cloning plasmids and oligonucleotides employed can be found in chapter 4.4.

4.17.5. Establishment of a stable cell line expressing CGG-repeat RNA

The two vectors, each containing the 5'UTR of the *fmr1* gene followed by either 30 or 95 CGG repeats (pCEP4-TRE-(CGG)n-eGFP) were a kind gift of Prof. Flora Tassone (University of California, USA). The inserts were cut out with BlpI/XhoI, sticky ends were filled up in a PCR (Sambrook and Russell 2001) with Vent®-polymerase (Hukema et al.) in order to ligate the construct into the cloning vector pUC19-SfiI (obtained from Prof. Georg Bornkamm, Helmholtz Zentrum München) linearized by enzymatic digestion with EcoRV. To tag the CGG repeat RNA, 27 PP7 stem loops from the pCR4-24xPP7-SL vector (created by Prof Ralf Jansen, MPI Tübingen) were cut out via BamHI and ligated upstream of the *fmr1* 5'UTR at the BamHI restriction site. The new insert, containing 27 PP7 stem loops fused to the *fmr1* 5'UTR followed by CGG repeats, was then cloned into the final mammalian expression vector pRTS-1 via "cut and paste" with SfiI restriction enzyme. Prior to that, the PP7 coat protein, which recognizes and binds to PP7 stem loop structures, was amplified by standard PCR (primer 15 and 16) from the pET22HT-PP7delFG plasmid (provided by Prof. Ralf Jansen, MPI Tübingen) and cloned into the pRTS-1 vector in frame at the N-term of the

eGFP gene at the SwaI restriction site. All cloning steps were verified either by sequencing or restriction digest.

The PP7 stem loop-fmr1 5'UTR-(30 or 95 CGG) insert and the PP7 coat protein fused to eGFP was also cloned into the pRTR vector, which additionally encodes for the CD2 gene and enables cell selection via Dynabeads® (Invitrogen, Karlsruhe, Germany).

Constructs were then transfected into HeLa or COS7 cells and protein/RNA expression was induced with doxycycline. Expression of eGFP-tagged PP7 protein was monitored by fluorescence microscopy.

4.18. Western Blot

4.18.1. Cell harvest and lysis

HeLa and COS7 cells were harvested from 6-well plates. The medium was removed and cells were washed once with 2 ml per well of 1x PBS. Depending on the density of the cells 75 µl (less dense) or 100 µl (dense) of 2x magic mix was added to each well. Cells were scraped off with a cell scraper (TPP, Trasdingen, Switzerland), transferred to a QIAshredder filter column (Qiagen, Hilden, Germany) and centrifuged at 13,000 rpm for 2 min. The flow through was flash frozen in liquid nitrogen and stored at -80 ˚C.

4.18.2. Determination of total protein concentration

Total Protein concentration was measured by using the Quant-iT™ Protein Assay Kit (Invitrogen, Karlsruhe, Germany) with the Qubit™ Fluorometer (Invitrogen, Karlsruhe, Germany) according to the manufacturer's protocol.

4.18.3. Blotting and immuno-staining of proteins

20 µg of total protein per lane was loaded onto a 10 % polyacrylamide gel (SDS PAGE) and separated for 1 h 15 min at 150 V. Subsequently the gel was equilibrated for 15 min in 1x Semi-dry transfer buffer. The nitrocellulose blotting membrane was equilibrated for 3 s in methanol, 2 min in bidestilled H_2O and 15 min in 1x Semi-dry transfer buffer. Proteins were then blotted onto the membrane in a Semi-dry-blotter (Peqlab (VWR), Erlangen, Germany) at 200 mA for 45 min. To saturate unspecific proteins, the blot was incubated in (w/v) 5 % milk powder/PBST over night at 4˚C. The next day, the blot was incubated with the primary antibody diluted in (w/v) 1 % BSA/PBST for 1h at RT or over night at 4˚C. Three washing

steps (5 min each) with PBST followed to remove unbound antibody. The corresponding HRP conjugated secondary antibody was diluted in PBST and was incubated with the blot for 45 min at room temperature, again followed by three washing steps with PBST. Protein signals were detected by incubating the blot with the Pierce ECL substrate (Thermo Fisher Scientific, Bonn, Germany) for 1 min and subsequent exposure to light-sensitive films (GE Healthcare, Munich, Germany).

4.18.4. Antibody testing (αB-crystallin, Pur-α)

15 in house produced tissue culture supernatants of the αB-crystallin antibody (raised against the C-terminal peptide R-E-E-K-P-A-V-T-A-A-P-K-K-Y-COOH (Bhat et al. 1991), three tissue culture supernatants of the human Pur-alpha antibody (Jurk et al. 1996) and a commercial human Pur-alpha antibody have been tested in western blot assays. Therefore $2x10^5$ COS7 or HeLa cells per well were seeded into 6-well plates. Next day cells were harvested and 20 µg of total protein per lane were separated on a 10 % SDS gel and blotted onto a nitrocellulose membrane. Each blot was first incubated with the primary antibody for 1 h at room temperature followed by incubation with the corresponding HRP conjugated secondary antibody for 45 min at room temperature. Protein signals were detected using the ECL substrate and blot exposition to light-sensitive films.

5. Appendix

5.1. Abbreviations

%	Per cent	m	Milli, meter
°C	Degree Celsius	M	mol per Liter, methionine
µ	Micro	MES	2-(*N*-morpholino)ethanesulfonic acid
A	Alanine (amino acid) or adenine (nucleic acid)	min	Minute
Å	Ångström	n	Nano
A280, A260, A254	Absorption at 83avelength 260 nm/ 280 nm / 254 nm	nt	Nucleotide
aa	Amino acid	nm	Nanometer
Amp	Ampicillin	NaCl	Sodium chloride
ATP	Adenosine triphosphate	n.d.	Not determined
bb,	*B. burgdorferi Borrelia burgdorferi*	OD600	Optical density at 600 nm
BAC	Bacterial artificial chromosome	N-term	Amino-terminus
BSA	Bovine serum albumin	P	Proline
C	Cysteine, cytosine	PAGE	Polyacrylamide gel electrophoresis
CD	Circular dichroism	PCR	Polymerase chain reaction
Ci	Curie	PEG	Polyethlyleneglycol
CNS	Central nervous system	pH	Potentia hydrogenii
C-term	Carboxy-terminus	PMSF	Phenylmethanesulphonylfluoride
D	Aspartic acid	PVFD	Polyvinylidene fluoride
Da	Dalton	R	Arginine
deg	Degree	rev	Reverse
DEPC	Diethylpyrocarbonate	RMSD	Root-mean-square deviation
dm, *D. melanogaster*	*Drosophila melanogaster*	(m)RNA	(messenger) Ribonucleic acid
DNA	Desoxyribonucleic acid	RNP	Ribonucleoprotein particle
ds	Double stranded	rpm	Rounds per minute
E	Glutamic acid	RT	Room temperature
E. coli	*Escherichia coli*	s	Second
EDTA	Ethylenediamine-tetraacetic acid	S	Serine
EMSA	Electrophoretic mobility shift assay	SAXS	Small angle X-ray scattering
F	Phenylalanine	SDS	Sodium docecyl sulfate
for	Forward	SeMet	Selenomethionine
g	Gram or standard accelaration	ss	Single stranded
G	Guanine	TCEP	Tris-(2-carboxyethyl)-phosphine hydrochlorid
GST	Glutatione S-transferase	TRIS	tris(hydroxymethyl)aminomethane
h	Hour or human	tRNA	Transfer RNA
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	UTR	Untranslated region
I	Isoleucine	UV	Ultraviolet

IPTG	Isopropyl-β-D-thiogalactopyranosid	v/v	Volume per volume
K	Lysine, Kelvin	WT	Wild type
k	Kilo	w/v	Weight per volume
Kan	Kanamycin	Y	Tyrosine
K_D	Equilibrium dissociation constant		
l	Liter		
L	Leucine		

5.2. Index of figures

Figure 1.1 *D. melanogaster* Pur-α protein. 4
Figure 1.2 Fragile X syndrome. 9
Figure 1.3 Schematic representation of the RNA gain-of-function mechanism proposed for FXTAS. 12
Figure 2.1 Purification of Pur-□ protein derivativess used in this study. 21
Figure 2.2 Pur-α I-II binds with similar affinities to DNA and RNA. 22
Figure 2.3 NMR titration experiments show that Pur-α's binding mode to DNA and RNA is similar. 23
Figure 2.4 Pur-α:DNA crystals diffracted up to 2.0 Å. 25
Figure 2.5 Crystal structure of Pur-α I-II in complex with the 5'-GCGGCGG-3' ssDNA. 28
Figure 2.6 Pur-α does not undergo conformational changes upon DNA binding. 29
Figure 2.7 Amino acid sequence alignment of Pur-α from different species. 30
Figure 2.8 *Drosophila* Pur-α I-II mutants show decreased binding affinity to DNA and RNA 32
Figure 2.9 Folding of the wild type Pur-α I-II and the mutants are equal. 32
Figure 2.10 *Drosophila* Pur-α III shows only weak binding affinity to DNA and RNA. 33
Figure 2.11 *Drosophila* Pur-α I-II KNR II mutant shows decreased binding affinity to ALS-related repeat-RNA. 34
Figure 2.12 Mutations in Pur-α I-II decrease dsDNA unwinding ability. 35
Figure 2.13 One Pur-α I-II molecule can bind two ssDNA molecules. 37
Figure 2.14 Pur-α binds to the MF0677 ss/dsDNA used in DNaseI footprints. 38
Figure 2.15 DNase I footprint assays. 39
Figure 2.16 Pur-α does not bind to the sequence, which is protected in DNase I footprint assays, but to the GGN motif of the MF0677 sequence, which is not protected in the footprint assay. 39
Figure 2.17 Unspecific binding of human Pur-α (hPur-α) antibody. 41
Figure 2.18 BAC constructs of human Pur-α.. 42
Figure 2.19 HeLa cell colonies carrying BAC constructs with eGFP tagged human Pur-α after 4 weeks of antibiotic selection. 43
Figure 2.20 Expression test of GFP-tagged hPur-α from HeLa protein lysate. 44
Figure 2.21 Vector design for CGG-repeat RNA expression in COS7 and HeLa cells.. 46
Figure 2.22 Test expression of initial vectors in COS7 cells. 47
Figure 2.23 Co-expression of eGFP and CGG-repeat RNA in COS7 cells. 48
Figure 2.24 Co-expression of PP7CP-tagged eGFP and PP7SL-tagged CGG-repeat RNA in COS7 and HeLa cells. 50
Figure 2.25 Unspecific binding of αB-crystallin (crys) antibody. 51
Figure 3.1 Unwinding model for Pur-α and dsDNA. 57
Figure 3.2 Pur-repeat III might guide two nucleic acid binding domains to dsDNA. 58
Figure 3.3 Pur-α mutations found in the 5q31.3 microdeletion syndrome can be modeled into the crystal structure of *Drosophila* Pur-α I-II in complex with DNA. 60

5.3. Index of tables

Table 2.1 Data collection for the crystal structure of *Drosophila* Pur-α I-II in complex with DNA. 26
Table 4.1 DNA oligonucleotides for cloning. 65
Table 4.2 DNA oligonucleotides for interaction studies and crystallization. 66
Table 4.3 DNA oligonucleotides for interaction studies and crystallization. 66
Table 4.4 Commercial plasmids. 66
Table 4.5 Plasmids for recombinant protein expression in *E. coli*. 67
Table 4.6 Plasmids for RNA/protein expression in mammalian cell lines. 67
Table 4.7 BAC clone. 68
Table 4.8 Tagging cassettes. 68
Table 4.9 Tagging oligonucleotides. 68
Table 4.10 Verification oligonucleotides. 68
Table 4.11 *E. coli* strains. 68
Table 4.12 Mammalian cell lines. 68
Table 4.13 Media for bacterial cell culture. 69
Table 4.14 Supplements for bacterial cell culture. 69
Table 4.15 Supplements for mammalian cell culture. 69
Table 4.16 Supplements for mammalian cell culture. 69
Table 4.17 Reagents for mammalian cell culture. 69
Table 4.18 Primary antibodies. 70
Table 4.19 Secondary antibodies. 70
Table 4.20 General buffers and stock solutions. 70
Table 4.21 SAXS conditions: buffers, protein concentrations and ratios. 77

6. References

Adams, P. D., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66(Pt 2): 213-221.

Andersen, P.M., and Al-Chalabi, A. (2011). Clinical genetics of amyotrophic lateral sclerosis: what do we really know? Nature reviews Neurology *7*, 603-615.

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y.*, et al.* (2006). TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochemical and biophysical research communications *351*, 602-611.

Aumiller, V., Graebsch, A., Kremmer, E., Niessing, D., and Forstemann, K. (2012). Drosophila Pur-alpha binds to trinucleotide-repeat containing cellular RNAs and translocates to the early oocyte. RNA biology *9*, 633-643.

Ayala, Y.M., Zago, P., D'Ambrogio, A., Xu, Y.F., Petrucelli, L., Buratti, E., and Baralle, F.E. (2008). Structural determinants of the cellular localization and shuttling of TDP-43. Journal of cell science *121*, 3778-3785.

Barmada, S. J. (2015). Linking RNA dysfunction and neurodegeneration in amyotrophic lateral sclerosis. Neurotherapeutics 12(2): 340-351.

Basso, M., Massignan, T., Samengo, G., Cheroni, C., De Biasi, S., Salmona, M., Bendotti, C., and Bonetto, V. (2006). Insoluble mutant SOD1 is partly oligoubiquitinated in amyotrophic lateral sclerosis mice. The Journal of biological chemistry *281*, 33325-33335.

Bhat SP, et al. (1991) Alpha B-crystallin exists as an independent protein in the heart and in the lens. Eur J Biochem. 202(3):775-781.

Bendotti, C., Marino, M., Cheroni, C., Fontana, E., Crippa, V., Poletti, A., and De Biasi, S. (2012). Dysfunction of constitutive and inducible ubiquitin-proteasome system in amyotrophic lateral sclerosis: implication for protein aggregation and immune response. Progress in neurobiology *97*, 101-126.

Bergemann, A.D., and Johnson, E.M. (1992). The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking regions and origins of DNA replication. Molecular and cellular biology *12*, 1257-1265.

Bergemann, A.D., Ma, Z.W., and Johnson, E.M. (1992). Sequence of cDNA comprising the human pur gene and sequence-specific single-stranded-DNA-binding properties of the encoded protein. Molecular and cellular biology *12*, 5673-5682.

Berman, R.F., Buijsen, R.A., Usdin, K., Pintado, E., Kooy, F., Pretto, D., Pessah, I.N., Nelson, D.L., Zalewski, Z., Charlet-Bergeurand, N.*, et al.* (2014). Mouse models of the fragile X premutation and fragile X-associated tremor/ataxia syndrome. Journal of neurodevelopmental disorders *6*, 25.

Boeve, B.F., and Graff-Radford, N.R. (2012). Cognitive and behavioral features of c9FTD/ALS. Alzheimer's research & therapy *4*, 29.

Brizard, A., Guilhot, F., Huret, J.L., Benz-Lemoine, E., and Tanzer, J. (1988). The 8p11 anomaly in "monoblastic" leukaemia. Leukemia research *12*, 693-697.

Bulut-Karslioglu, A., Perrera, V., Scaranaro, M., de la Rosa-Velazquez, I.A., van de Nobelen, S., Shukeir, N., Popow, J., Gerle, B., Opravil, S., Pagani, M.*, et al.* (2012). A transcription factor-based mechanism for mouse heterochromatin formation. Nature structural & molecular biology *19*, 1023-1030.

Cao, Z., Hulsizer, S., Tassone, F., Tang, H.T., Hagerman, R.J., Rogawski, M.A., Hagerman, P.J., and Pessah, I.N. (2012). Clustered burst firing in FMR1 premutation hippocampal neurons: amelioration with allopregnanolone. Human molecular genetics *21*, 2923-2935.

Chio, A., Calvo, A., Mazzini, L., Cantello, R., Mora, G., Moglia, C., Corrado, L., D'Alfonso, S., Majounie, E., Renton, A.*, et al.* (2012). Extensive genetics of ALS: a population-based study in Italy. Neurology *79*, 1983-1989.

Coffey, S.M., Cook, K., Tartaglia, N., Tassone, F., Nguyen, D.V., Pan, R., Bronsky, H.E., Yuhas, J., Borodyanskaya, M., Grigsby, J.*, et al.* (2008). Expanded clinical phenotype of women with the FMR1 premutation. American journal of medical genetics Part A *146A*, 1009-1016.

Colak, D., Zaninovic, N., Cohen, M.S., Rosenwaks, Z., Yang, W.Y., Gerhardt, J., Disney, M.D., and Jaffrey, S.R. (2014). Promoter-bound trinucleotide repeat mRNA drives epigenetic silencing in fragile X syndrome. Science *343*, 1002-1005.

Cronister, A., Schreiner, R., Wittenberger, M., Amiri, K., Harris, K., and Hagerman, R.J. (1991). Heterozygous fragile X female: historical, physical, cognitive, and cytogenetic features. American journal of medical genetics *38*, 269-274.

Darbinian, N., Gallia, G.L., and Khalili, K. (2001). Helix-destabilizing properties of the human single-stranded DNA- and RNA-binding protein Puralpha. Journal of cellular biochemistry *80*, 589-595.

Darbinian, N., White, M.K., Gallia, G.L., Amini, S., Rappaport, J., and Khalili, K. (2004). Interaction between the pura and E2F-1 transcription factors. Anticancer research *24*, 2585-2594.

Darnell, J.C., and Richter, J.D. (2012). Cytoplasmic RNA-binding proteins and the control of complex brain function. Cold Spring Harbor perspectives in biology *4*, a012344.
Daughters, R.S., Tuttle, D.L., Gao, W., Ikeda, Y., Moseley, M.L., Ebner, T.J., Swanson, M.S., and Ranum, L.P. (2009). RNA gain-of-function in spinocerebellar ataxia type 8. PLoS genetics *5*, e1000600.
Day, J.W., and Ranum, L.P. (2005). RNA pathogenesis of the myotonic dystrophies. Neuromuscular disorders : NMD *15*, 5-16.
DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J.*, et al.* (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72*, 245-256.
Deng, H.X., Chen, W., Hong, S.T., Boycott, K.M., Gorrie, G.H., Siddique, N., Yang, Y., Fecto, F., Shi, Y., Zhai, H.*, et al.* (2011). Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature *477*, 211-215.
Devys, D., Lutz, Y., Rouyer, N., Bellocq, J.P., and Mandel, J.L. (1993). The FMR-1 protein is cytoplasmic, most abundant in neurons and appears normal in carriers of a fragile X premutation. Nature genetics *4*, 335-340.
Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., Croft, G.F., Saphier, G., Leibel, R., Goland, R.*, et al.* (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science *321*, 1218-1221.
Dobretsova, A., Johnson, J.W., Jones, R.C., Edmondson, R.D., and Wight, P.A. (2008). Proteomic analysis of nuclear factors binding to an intronic enhancer in the myelin proteolipid protein gene. Journal of neurochemistry *105*, 1979-1995.
Dombrowski, C., Levesque, S., Morel, M.L., Rouillard, P., Morgan, K., and Rousseau, F. (2002). Premutation and intermediate-size FMR1 alleles in 10572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles. Human molecular genetics *11*, 371-378.
Dormann, D., Madl, T., Valori, C.F., Bentmann, E., Tahirovic, S., Abou-Ajram, C., Kremmer, E., Ansorge, O., Mackenzie, I.R., Neumann, M.*, et al.* (2012). Arginine methylation next to the PY-NLS modulates Transportin binding and nuclear import of FUS. The EMBO journal *31*, 4258-4275.
Dormann, D., Rodde, R., Edbauer, D., Bentmann, E., Fischer, I., Hruscha, A., Than, M.E., Mackenzie, I.R., Capell, A., Schmid, B.*, et al.* (2010). ALS-associated fused in sarcoma (FUS) mutations disrupt Transportin-mediated nuclear import. The EMBO journal *29*, 2841-2857.
Duff, M. R., Jr., et al. (2011). Isothermal titration calorimetry for measuring macromolecule-ligand affinity. J Vis Exp(55).
Emsley, P. and K. Cowtan (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60(Pt 12 Pt 1): 2126-2132.
Gallia, G.L., Johnson, E.M., and Khalili, K. (2000). Puralpha: a multifunctional single-stranded DNA- and RNA-binding protein. Nucleic acids research *28*, 3197-3205.
Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res 31, 3784-3788.
Gellera, C., Tiloca, C., Del Bo, R., Corrado, L., Pensato, V., Agostini, J., Cereda, C., Ratti, A., Castellotti, B., Corti, S.*, et al.* (2013). Ubiquilin 2 mutations in Italian patients with amyotrophic lateral sclerosis and frontotemporal dementia. Journal of neurology, neurosurgery, and psychiatry *84*, 183-187.
Gijselinck, I., Van Langenhove, T., van der Zee, J., Sleegers, K., Philtjens, S., Kleinberger, G., Janssens, J., Bettens, K., Van Cauwenberghe, C., Pereson, S.*, et al.* (2012). A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. The Lancet Neurology *11*, 54-65.
Glass, J.D., Boulis, N.M., Johe, K., Rutkove, S.B., Federici, T., Polak, M., Kelly, C., and Feldman, E.L. (2012). Lumbar intraspinal injection of neural stem cells in patients with amyotrophic lateral sclerosis: results of a phase I trial in 12 patients. Stem cells *30*, 1144-1151.
Graebsch, A., Roche, S., Kostrewa, D., Soding, J., and Niessing, D. (2010). Of bits and bugs--on the use of bioinformatics and a bacterial crystal structure to solve a eukaryotic repeat-protein structure. PloS one *5*, e13402.
Graebsch, A., Roche, S., and Niessing, D. (2009). X-ray structure of Pur-alpha reveals a Whirly-like fold and an unusual nucleic-acid binding surface. Proceedings of the National Academy of Sciences of the United States of America *106*, 18521-18526.
Greco, C.M., Berman, R.F., Martin, R.M., Tassone, F., Schwartz, P.H., Chang, A., Trapp, B.D., Iwahashi, C., Brunberg, J., Grigsby, J.*, et al.* (2006). Neuropathology of fragile X-associated tremor/ataxia syndrome (FXTAS). Brain : a journal of neurology *129*, 243-255.
Gros-Louis, F., Soucy, G., Lariviere, R., and Julien, J.P. (2010). Intracerebroventricular infusion of monoclonal antibody or its derived Fab fragment against misfolded forms of SOD1 mutant delays mortality in a mouse model of ALS. Journal of neurochemistry *113*, 1188-1199.
Haas, S., Gordon, J., and Khalili, K. (1993). A developmentally regulated DNA-binding protein from mouse brain stimulates myelin basic protein gene expression. Molecular and cellular biology *13*, 3103-3112.
Haas, S., Thatikunta, P., Steplewski, A., Johnson, E.M., Khalili, K., and Amini, S. (1995). A 39-kD DNA-binding protein from mouse brain stimulates transcription of myelin basic protein gene in oligodendrocytic cells. The Journal of cell biology *130*, 1171-1179.

Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R.*, et al.* (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. Nature *507*, 195-200.
Hagerman, P.J., and Hagerman, R.J. (2007). Fragile X-associated tremor/ataxia syndrome--an older face of the fragile X gene. Nature clinical practice Neurology *3*, 107-112.
Hagerman, R., Lauterborn, J., Au, J., and Berry-Kravis, E. (2012). Fragile X syndrome and targeted treatment trials. Results and problems in cell differentiation *54*, 297-335.
Hagerman, R.J., Hall, D.A., Coffey, S., Leehey, M., Bourgeois, J., Gould, J., Zhang, L., Seritan, A., Berry-Kravis, E., Olichney, J.*, et al.* (2008). Treatment of fragile X-associated tremor ataxia syndrome (FXTAS) and related neurological problems. Clinical interventions in aging *3*, 251-262.
Hagerman, R.J., Leehey, M., Heinrichs, W., Tassone, F., Wilson, R., Hills, J., Grigsby, J., Gage, B., and Hagerman, P.J. (2001). Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology *57*, 127-130.
Handa, V., Saha, T., and Usdin, K. (2003). The fragile X syndrome repeats form RNA hairpins that do not activate the interferon-inducible protein kinase, PKR, but are cut by Dicer. Nucleic acids research *31*, 6243-6248.
Harvey, R.J., Skelton-Robinson, M., and Rossor, M.N. (2003). The prevalence and causes of dementia in people under the age of 65 years. Journal of neurology, neurosurgery, and psychiatry *74*, 1206-1209.
Hoem, G., et al. (2011). CGG-repeat length threshold for FMR1 RNA pathogenesis in a cellular model for FXTAS. Hum Mol Genet 20(11): 2161-2170.
Hokkanen, S., Feldmann, H.M., Ding, H., Jung, C.K., Bojarski, L., Renner-Muller, I., Schuller, U., Kretzschmar, H., Wolf, E., and Herms, J. (2012). Lack of Pur-alpha alters postnatal brain development and causes megalencephaly. Human molecular genetics *21*, 473-484.
Horton, W.A., Eldridge, R., and Brody, J.A. (1976). Familial motor neuron disease. Evidence for at least three different types. Neurology *26*, 460-465.
Hull, C., and Hagerman, R.J. (1993). A study of the physical, behavioral, and medical phenotype, including anthropometric measures, of females with fragile X syndrome. American journal of diseases of children *147*, 1236-1241.
Inoue, T., Leman, E.S., Yeater, D.B., and Getzenberg, R.H. (2008). The potential role of purine-rich element binding protein (PUR) alpha as a novel treatment target for hormone-refractory prostate cancer. The Prostate *68*, 1048-1056.
Itoh, H., Wortman, M.J., Kanovsky, M., Uson, R.R., Gordon, R.E., Alfano, N., and Johnson, E.M. (1998). Alterations in Pur(alpha) levels and intracellular localization in the CV-1 cell cycle. Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research *9*, 651-665.
Iwahashi, C.K., Yasui, D.H., An, H.J., Greco, C.M., Tassone, F., Nannen, K., Babineau, B., Lebrilla, C.B., Hagerman, R.J., and Hagerman, P.J. (2006). Protein composition of the intranuclear inclusions of FXTAS. Brain : a journal of neurology *129*, 256-271.
Jackson, C.C., Medeiros, L.J., and Miranda, R.N. (2010). 8p11 myeloproliferative syndrome: a review. Human pathology *41*, 461-476.
Jacquemont, S., Hagerman, R.J., Leehey, M., Grigsby, J., Zhang, L., Brunberg, J.A., Greco, C., Des Portes, V., Jardini, T., Levine, R.*, et al.* (2003). Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. American journal of human genetics *72*, 869-878.
Ji, J., Tsika, G.L., Rindt, H., Schreiber, K.L., McCarthy, J.J., Kelm, R.J., Jr., and Tsika, R. (2007). Puralpha and Purbeta collaborate with Sp3 to negatively regulate beta-myosin heavy chain gene expression during skeletal muscle inactivity. Molecular and cellular biology *27*, 1531-1543.
Jin, P., Duan, R., Qurashi, A., Qin, Y., Tian, D., Rosser, T.C., Liu, H., Feng, Y., and Warren, S.T. (2007). Pur alpha binds to rCGG repeats and modulates repeat-mediated neurodegeneration in a Drosophila model of fragile X tremor/ataxia syndrome. Neuron *55*, 556-564.
Jin, P., Zarnescu, D.C., Zhang, F., Pearson, C.E., Lucchesi, J.C., Moses, K., and Warren, S.T. (2003). RNA-mediated neurodegeneration caused by the fragile X premutation rCGG repeats in Drosophila. Neuron *39*, 739-747.
Johnson, B.S., Snead, D., Lee, J.J., McCaffery, J.M., Shorter, J., and Gitler, A.D. (2009). TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. The Journal of biological chemistry *284*, 20329-20339.
Johnson, E.M. (2003). The Pur protein family: clues to function from recent studies on cancer and AIDS. Anticancer research *23*, 2093-2100.
Johnson, E.M., Chen, P.L., Krachmarov, C.P., Barr, S.M., Kanovsky, M., Ma, Z.W., and Lee, W.H. (1995). Association of human Pur alpha with the retinoblastoma protein, Rb, regulates binding to the single-stranded DNA Pur alpha recognition element. The Journal of biological chemistry *270*, 24352-24360.
Johnson, E.M., Kinoshita, Y., Weinreb, D.B., Wortman, M.J., Simon, R., Khalili, K., Winckler, B., and Gordon, J. (2006). Role of Pur alpha in targeting mRNA to sites of translation in hippocampal neuronal dendrites. Journal of neuroscience research *83*, 929-943.

Johnson, J.O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V.M., Trojanowski, J.Q., Gibbs, J.R., Brunetti, M., Gronka, S., Wuu, J., *et al.* (2010). Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron *68*, 857-864.
Jose, D., et al. (2012). Breathing fluctuations in position-specific DNA base pairs are involved in regulating helicase movement into the replication fork. Proc Natl Acad Sci U S A 109(36): 14428-14433.
Ju, J.S., Fuentealba, R.A., Miller, S.E., Jackson, E., Piwnica-Worms, D., Baloh, R.H., and Weihl, C.C. (2009). Valosin-containing protein (VCP) is required for autophagy and is disrupted in VCP disease. The Journal of cell biology *187*, 875-888.
Jurk, M., et al. (1996). Characterization of the single-strand-specific BPV-1 origin binding protein, SPSF I, as the HeLa Pur alpha factor. Nucleic Acids Res 24(14): 2799-2806.
Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr 66(Pt 2): 125-132.
Kaminski, R., Darbinyan, A., Merabova, N., Deshmane, S.L., White, M.K., and Khalili, K. (2008). Protective role of Puralpha to cisplatin. Cancer biology & therapy *7*, 1926-1935.
Kanai, Y., Dohmae, N., and Hirokawa, N. (2004). Kinesin transports RNA: isolation and characterization of an RNA-transporting granule. Neuron *43*, 513-525.
Kaufmann, W.E., Abrams, M.T., Chen, W., and Reiss, A.L. (1999). Genotype, molecular phenotype, and cognitive phenotype: correlations in fragile X syndrome. American journal of medical genetics *83*, 286-295.
Kelm, R.J., Jr., Elder, P.K., Strauch, A.R., and Getz, M.J. (1997). Sequence of cDNAs encoding components of vascular actin single-stranded DNA-binding factor 2 establish identity to Puralpha and Purbeta. The Journal of biological chemistry *272*, 26727-26733.
Khalili, K., Del Valle, L., Muralidharan, V., Gault, W.J., Darbinian, N., Otte, J., Meier, E., Johnson, E.M., Daniel, D.C., Kinoshita, Y., *et al.* (2003). Puralpha is essential for postnatal brain development and developmentally coupled cellular proliferation as revealed by genetic inactivation in the mouse. Molecular and cellular biology *23*, 6857-6875.
Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. Nucleic Acids Res 35 (Web Server issue): W43-46.
Kiebler, M.A., and Bassell, G.J. (2006). Neuronal RNA granules: movers and makers. Neuron 51, 685-690.
Kieran, D., Kalmar, B., Dick, J.R., Riddoch-Contreras, J., Burnstock, G., and Greensmith, L. (2004). Treatment with arimoclomol, a coinducer of heat shock proteins, delays disease progression in ALS mice. Nature medicine *10*, 402-405.
Kiliszek, A., Kierzek, R., Krzyzosiak, W.J., and Rypniewski, W. (2011). Crystal structures of CGG RNA repeats with implications for fragile X-associated tremor ataxia syndrome. Nucleic acids research *39*, 7308-7315.
Kim, K., Choi, J., Heo, K., Kim, H., Levens, D., Kohno, K., Johnson, E.M., Brock, H.W., and An, W. (2008). Isolation and characterization of a novel H1.2 complex that acts as a repressor of p53-mediated transcription. The Journal of biological chemistry *283*, 9113-9126.
Knapp, A.M., Ramsey, J.E., Wang, S.X., Strauch, A.R., and Kelm, R.J., Jr. (2007). Structure-function analysis of mouse Pur beta II. Conformation altering mutations disrupt single-stranded DNA and protein interactions crucial to smooth muscle alpha-actin gene repression. The Journal of biological chemistry *282*, 35899-35909.
Kobayashi, S., Agui, K., Kamo, S., Li, Y., and Anzai, K. (2000). Neural BC1 RNA associates with pur alpha, a single-stranded DNA and RNA binding protein, which is involved in the transcription of the BC1 RNA gene. Biochemical and biophysical research communications *277*, 341-347.
Konarev, P.V., Petoukhov, M.V., Volkov, V.V., and Svergun, D.I. (2006). ATSAS2.1, a program package for small-angle scattering data analysis. J Appl Cryst 39, 277-286.
Konarev, P.V., Vokov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). Primus: a Windows PC-based system for small angle-scattering data analysis. Journal of Applied Crystallography 36, 6.
Kwiatkowski, T.J., Jr., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., *et al.* (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science *323*, 1205-1208.
Ladd, P.D., Smith, L.E., Rabaia, N.A., Moore, J.M., Georges, S.A., Hansen, R.S., Hagerman, R.J., Tassone, F., Tapscott, S.J., and Filippova, G.N. (2007). An antisense transcript spanning the CGG repeat region of FMR1 is upregulated in premutation carriers but silenced in full mutation individuals. Human molecular genetics *16*, 3174-3187.
Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.
Lagier-Tourenne, C., Polymenidou, M., and Cleveland, D.W. (2010). TDP-43 and FUS/TLS: emerging roles in RNA processing and neurodegeneration. Human molecular genetics *19*, R46-64.
Lalani, S.R., Zhang, J., Schaaf, C.P., Brown, C.W., Magoulas, P., Tsai, A.C., El-Gharbawy, A., Wierenga, K.J., Bartholomew, D., Fong, C.T., *et al.* (2014). Mutations in PURA cause profound neonatal hypotonia, seizures, and encephalopathy in 5q31.3 microdeletion syndrome. American journal of human genetics *95*, 579-583.
Leehey, M.A. (2009). Fragile X-associated tremor/ataxia syndrome: clinical phenotype, diagnosis, and treatment. Journal of investigative medicine : the official publication of the American Federation for Clinical Research *57*, 830-836.

Leehey, M.A., Hagerman, R.J., and Hagerman, P.J. (2007). Fragile X syndrome vs fragile X-associated tremor/ataxia syndrome. Archives of neurology *64*, 289; author reply 289-290.
Levine, T.P., Daniels, R.D., Gatta, A.T., Wong, L.H., and Hayes, M.J. (2013). The product of C9orf72, a gene strongly implicated in neurodegeneration, is structurally related to DENN Rab-GEFs. Bioinformatics *29*, 499-503.
Lezon-Geyda, K., Najfeld, V., and Johnson, E.M. (2001). Deletions of PURA, at 5q31, and PURB, at 7p13, in myelodysplastic syndrome and progression to acute myelogenous leukemia. Leukemia *15*, 954-962.
Li, Z., and Srivastava, P. (2004). Heat-shock proteins. Current protocols in immunology / edited by John E Coligan [et al] *Appendix 1*, Appendix 1T.
Ling, S.C., Polymenidou, M., and Cleveland, D.W. (2013). Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis. Neuron *79*, 416-438.
Liquori, C.L., Ricker, K., Moseley, M.L., Jacobsen, J.F., Kress, W., Naylor, S.L., Day, J.W., and Ranum, L.P. (2001). Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science *293*, 864-867.
Liu, H., and Johnson, E.M. (2002). Distinct proteins encoded by alternative transcripts of the PURG gene, located contrapodal to WRN on chromosome 8, determined by differential termination/polyadenylation. Nucleic acids research *30*, 2417-2426.
Loomis, E.W., Sanz, L.A., Chedin, F., and Hagerman, P.J. (2014). Transcription-associated R-loop formation across the human FMR1 CGG-repeat region. PLoS genetics *10*, e1004294.
Lopez-Gines, C., Cerda-Nicolas, M., Gil-Benso, R., Pellin, A., Lopez-Guerrero, J.A., Benito, R., del Rey, J., Miro, R., Roldan, R., and Barbera, J. (2006). Primary glioblastoma with EGFR amplification and a ring chromosome 7 in a young patient. Clinical neuropathology *25*, 193-199.
Lozano, R., Rosero, C.A., and Hagerman, R.J. (2014). Fragile X spectrum disorders. Intractable & rare diseases research *3*, 134-146.
Ludwig, A.L., Espinal, G.M., Pretto, D.I., Jamal, A.L., Arque, G., Tassone, F., Berman, R.F., and Hagerman, P.J. (2014). CNS expression of murine fragile X protein (FMRP) as a function of CGG-repeat size. Human molecular genetics *23*, 3228-3238.
Lukong, K.E., and Richard, S. (2008). Motor coordination defects in mice deficient for the Sam68 RNA-binding protein. Behavioural brain research *189*, 357-363.
Ma, Z.W., Pejovic, T., Najfeld, V., Ward, D.C., and Johnson, E.M. (1995). Localization of PURA, the gene encoding the sequence-specific single-stranded-DNA-binding protein Pur alpha, to chromosome band 5q31. Cytogenetics and cell genetics *71*, 64-67.
Mahoney, C.J., Beck, J., Rohrer, J.D., Lashley, T., Mok, K., Shakespeare, T., Yeatman, T., Warrington, E.K., Schott, J.M., Fox, N.C.*, et al.* (2012). Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. Brain : a journal of neurology *135*, 736-750.
McCombe, P.A., and Henderson, R.D. (2010). Effects of gender in amyotrophic lateral sclerosis. Gender medicine *7*, 557-570.
McCoy, A. J., et al. (2007). Phaser crystallographic software. J Appl Crystallogr 40(Pt 4): 658-674.
Mishra, M., Del Valle, L., Otte, J., Darbinian, N., and Gordon, J. (2013). Pur-alpha regulates RhoA developmental expression and downstream signaling. Journal of cellular physiology *228*, 65-72.
Mori, K., Weng, S.M., Arzberger, T., May, S., Rentzsch, K., Kremmer, E., Schmid, B., Kretzschmar, H.A., Cruts, M., Van Broeckhoven, C.*, et al.* (2013). The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. Science *339*, 1335-1338.
Muralidharan, V., Sweet, T., Nadraga, Y., Amini, S., and Khalili, K. (2001). Regulation of Puralpha gene transcription: evidence for autoregulation of Puralpha promoter. Journal of cellular physiology *186*, 406-413.
Murshudov, G. N., et al. (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr 53(Pt 3): 240-255.
Muzar, Z., and Lozano, R. (2014). Current research, diagnosis, and treatment of fragile X-associated tremor/ataxia syndrome. Intractable & rare diseases research *3*, 101-109.
Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M.*, et al.* (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science *314*, 130-133.
Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. J Mol Biol 302, 205-217.
Ohashi, S., Kobayashi, S., Omori, A., Ohara, S., Omae, A., Muramatsu, T., Li, Y., and Anzai, K. (2000). The single-stranded DNA- and RNA-binding proteins pur alpha and pur beta link BC1 RNA to microtubules through binding to the dendrite-targeting RNA motifs. Journal of neurochemistry *75*, 1781-1790.
Oostra, B.A., and Willemsen, R. (2009). FMR1: a gene with three faces. Biochimica et biophysica acta *1790*, 467-477.
Orr, H.T., and Zoghbi, H.Y. (2007). Trinucleotide repeat disorders. Annual review of neuroscience *30*, 575-621.
Peyrard, M., et al. (2009). Nonlinear analysis of the dynamics of DNA breathing. J Biol Phys 35(1): 73-89. Poser, I., et al. (2008). BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. Nat Methods 5(5): 409-415.

Pierce, M. M., et al. (1999). Isothermal titration calorimetry of protein-protein interactions. Methods 19(2): 213-221.
Powell, W.T., Coulson, R.L., Gonzales, M.L., Crary, F.K., Wong, S.S., Adams, S., Ach, R.A., Tsang, P., Yamada, N.A., Yasui, D.H., et al. (2013). R-loop formation at Snord116 mediates topotecan inhibition of Ube3a-antisense and allele-specific chromatin decondensation. Proceedings of the National Academy of Sciences of the United States of America *110*, 13938-13943.
Pramatarova, A., Figlewicz, D.A., Krizus, A., Han, F.Y., Ceballos-Picot, I., Nicole, A., Dib, M., Meininger, V., Brown, R.H., and Rouleau, G.A. (1995). Identification of new mutations in the Cu/Zn superoxide dismutase gene of patients with familial amyotrophic lateral sclerosis. American journal of human genetics *56*, 592-596.
Pretto, D.I., Mendoza-Morales, G., Lo, J., Cao, R., Hadd, A., Latham, G.J., Durbin-Johnson, B., Hagerman, R., and Tassone, F. (2014). CGG allele size somatic mosaicism and methylation in FMR1 premutation alleles. Journal of medical genetics *51*, 309-318.
Putnam, C. D., et al. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40(3): 191-285.
PyMol (Schrodinger 2010) (DELano; http://www.pymol.org/).
Qurashi, A., Li, W., Zhou, J.Y., Peng, J., and Jin, P. (2011). Nuclear accumulation of stress response mRNAs contributes to the neurodegeneration caused by Fragile X premutation rCGG repeats. PLoS genetics 7, e1002102.
Ralph, G.S., Radcliffe, P.A., Day, D.M., Carthy, J.M., Leroux, M.A., Lee, D.C., Wong, L.F., Bilsland, L.G., Greensmith, L., Kingsman, S.M*., et al.* (2005). Silencing mutant SOD1 using RNAi protects against neurodegeneration and extends survival in an ALS model. Nature medicine *11*, 429-433.
Raoul, C., Abbas-Terki, T., Bensadoun, J.C., Guillot, S., Haase, G., Szulc, J., Henderson, C.E., and Aebischer, P. (2005). Lentiviral-mediated silencing of SOD1 through RNA interference retards disease onset and progression in a mouse model of ALS. Nature medicine *11*, 423-428.
Reddy, K., and Pearson, C.E. (2013). RAN translation: fragile X in the running. Neuron *78*, 405-408.
Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L*., et al.* (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72*, 257-268.
Ringholz, G.M., Appel, S.H., Bradshaw, M., Cooke, N.A., Mosnik, D.M., and Schulz, P.E. (2005). Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology *65*, 586-590.
Robberecht, W., and Philips, T. (2013). The changing scene of amyotrophic lateral sclerosis. Nature reviews Neuroscience *14*, 248-264.
Ross, C.A., and Poirier, M.A. (2004). Protein aggregation and neurodegenerative disease. Nature medicine *10 Suppl*, S10-17.
Rossi, S., Serrano, A., Gerbino, V., Giorgi, A., Di Francesco, L., Nencini, M., Bozzo, F., Schinina, M.E., Bagni, C., Cestra, G*., et al.* (2015). Nuclear accumulation of mRNAs underlies G4C2 repeat-induced translational repression in a cellular model of C9orf72 ALS. Journal of cell science.
Rothstein, J.D. (1996). Therapeutic horizons for amyotrophic lateral sclerosis. Current opinion in neurobiology *6*, 679-687.
Rumora, A.E., Wang, S.X., Ferris, L.A., Everse, S.J., and Kelm, R.J., Jr. (2013). Structural basis of multisite single-stranded DNA recognition and ACTA2 repression by purine-rich element binding protein B (Purbeta). Biochemistry *52*, 4439-4450.
Sambrook, J., and Russell, D.W. (2001). Molecular cloning: a laboratory manual, Third Edition edn (Cold Spring Harbor, Cold Spring Harbor Laboratory Press).Saxena, S., and Caroni, P. (2011). Selective neuronal vulnerability in neurodegenerative diseases: from stressor thresholds to degeneration. Neuron 71, 35-48.
Sellier, C., Rau, F., Liu, Y., Tassone, F., Hukema, R.K., Gattoni, R., Schneider, A., Richard, S., Willemsen, R., Elliott, D.J*., et al.* (2010). Sam68 sequestration and partial loss of function are associated with splicing alterations in FXTAS patients. The EMBO journal *29*, 1248-1261.
Seritan, A.L., Nguyen, D.V., Farias, S.T., Hinton, L., Grigsby, J., Bourgeois, J.A., and Hagerman, R.J. (2008). Dementia in fragile X-associated tremor/ataxia syndrome (FXTAS): comparison with Alzheimer's disease. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics *147B*, 1138-1144.
Smith, R.A., Miller, T.M., Yamanaka, K., Monia, B.P., Condon, T.P., Hung, G., Lobsiger, C.S., Ward, C.M., McAlonis-Downes, M., Wei, H*., et al.* (2006). Antisense oligonucleotide therapy for neurodegenerative disease. The Journal of clinical investigation *116*, 2290-2296.
Smith, S.S., Laayoun, A., Lingeman, R.G., Baker, D.J., and Riley, J. (1994). Hypermethylation of telomere-like foldbacks at codon 12 of the human c-Ha-ras gene and the trinucleotide repeat of the FMR-1 gene of fragile X. Journal of molecular biology *243*, 143-151.
Sofola, O.A., Jin, P., Qin, Y., Duan, R., Liu, H., de Haro, M., Nelson, D.L., and Botas, J. (2007). RNA-binding proteins hnRNP A2/B1 and CUGBP1 suppress fragile X CGG premutation repeat-induced neurodegeneration in a Drosophila model of FXTAS. Neuron *55*, 565-571.
Stepto, A., et al. (2014). Modelling C9ORF72 hexanucleotide repeat expansion in amyotrophic lateral sclerosis and frontotemporal dementia. Acta Neuropathol 127(3): 377-389.

Sun, Z., Diaz, Z., Fang, X., Hart, M.P., Chesi, A., Shorter, J., and Gitler, A.D. (2011). Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. PLoS biology *9*, e1000614.
Tassone, F., Hagerman, R.J., Ikle, D.N., Dyer, P.N., Lampe, M., Willemsen, R., Oostra, B.A., and Taylor, A.K. (1999). FMRP expression as a potential prognostic indicator in fragile X syndrome. American journal of medical genetics *84*, 250-261.
Todd, P.K., Oh, S.Y., Krans, A., He, F., Sellier, C., Frazer, M., Renoux, A.J., Chen, K.C., Scaglione, K.M., Basrur, V.*, et al.* (2013). CGG repeat-associated translation mediates neurodegeneration in fragile X tremor ataxia syndrome. Neuron *78*, 440-455.
Todd, P.K., Oh, S.Y., Krans, A., Pandey, U.B., Di Prospero, N.A., Min, K.T., Taylor, J.P., and Paulson, H.L. (2010). Histone deacetylases suppress CGG repeat-induced neurodegeneration via transcriptional silencing in models of fragile X tremor ataxia syndrome. PLoS genetics *6*, e1001240.
Tresse, E., Salomons, F.A., Vesa, J., Bott, L.C., Kimonis, V., Yao, T.P., Dantuma, N.P., and Taylor, J.P. (2010). VCP/p97 is essential for maturation of ubiquitin-containing autophagosomes and this function is impaired by mutations that cause IBMPFD. Autophagy *6*, 217-227.
Tretiakova, A., Steplewski, A., Johnson, E.M., Khalili, K., and Amini, S. (1999). Regulation of myelin basic protein gene transcription by Sp1 and Puralpha: evidence for association of Sp1 and Puralpha in brain. Journal of cellular physiology *181*, 160-168.
Turner, G., Webb, T., Wake, S., and Robinson, H. (1996). Prevalence of fragile X syndrome. American journal of medical genetics *64*, 196-197.
Untergasser, A., et al. (2012). "Primer3--new capabilities and interfaces." Nucleic Acids Res 40(15): e115.
Usdin, K., Hayward, B.E., Kumari, D., Lokanga, R.A., Sciascia, N., and Zhao, X.N. (2014). Repeat-mediated genetic and epigenetic changes at the FMR1 locus in the Fragile X-related disorders. Frontiers in genetics *5*, 226.
Vaguine, A. A., et al. (1999). "SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model." Acta Crystallogr D Biol Crystallogr 55(Pt 1): 191-205.
Van Langenhove, T., van der Zee, J., and Van Broeckhoven, C. (2012). The molecular basis of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum. Annals of medicine *44*, 817-828.
Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., *et al.* (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. Cell *65*, 905-914.
Vranken, W. F., et al. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59(4): 687-696.
Wang, H., Wang, M., Reiss, K., Darbinian-Sarkissian, N., Johnson, E.M., Iliakis, G., Amini, S., Khalili, K., and Rappaport, J. (2007). Evidence for the involvement of Puralpha in response to DNA replication stress. Cancer biology & therapy *6*, 596-602.
Wang, Y.H., Gellibolian, R., Shimizu, M., Wells, R.D., and Griffith, J. (1996). Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. Journal of molecular biology *263*, 511-516.
Watts, G.D., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P., and Kimonis, V.E. (2004). Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. Nature genetics *36*, 377-381.
Wheaton, M.W., Salamone, A.R., Mosnik, D.M., McDonald, R.O., Appel, S.H., Schmolck, H.I., Ringholz, G.M., and Schulz, P.E. (2007). Cognitive impairment in familial ALS. Neurology *69*, 1411-1417.
White, M.K., Johnson, E.M., and Khalili, K. (2009). Multiple roles for Puralpha in cellular and viral regulation. Cell cycle *8*, 1-7.
Willemsen, R., Levenga, J., and Oostra, B.A. (2011). CGG repeat in the FMR1 gene: size matters. Clinical genetics *80*, 214-225.
Williams, K.L., Warraich, S.T., Yang, S., Solski, J.A., Fernando, R., Rouleau, G.A., Nicholson, G.A., and Blair, I.P. (2012). UBQLN2/ubiquilin 2 mutation and pathology in familial amyotrophic lateral sclerosis. Neurobiology of aging *33*, 2527 e2523-2510.
Wortman, M. J., et al. (2005). Mechanism of DNA binding and localized strand separation by Pur alpha and comparison with Pur family member, Pur beta. Biochim Biophys Acta 1743(1-2): 64-78.
Xu, Z., Poidevin, M., Li, X., Li, Y., Shu, L., Nelson, D.L., Li, H., Hales, C.M., Gearing, M., Wingo, T.S.*, et al.* (2013). Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. Proceedings of the National Academy of Sciences of the United States of America *110*, 7778-7783.
Yang, J.C., Niu, Y.Q., Simon, C., Seritan, A.L., Chen, L., Schneider, A., Moghaddam, S.T., Hagerman, P.J., Hagerman, R.J., and Olichney, J.M. (2014). Memantine effects on verbal memory in fragile X-associated tremor/ataxia syndrome (FXTAS): a double-blind brain potential study. Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology *39*, 2760-2768.

Youings, S.A., Murray, A., Dennis, N., Ennis, S., Lewis, C., McKechnie, N., Pound, M., Sharrock, A., and Jacobs, P. (2000). FRAXA and FRAXE: the results of a five year survey. Journal of medical genetics *37*, 415-421.

Yrigollen, C.M., Tassone, F., Durbin-Johnson, B., and Tassone, F. (2011). The role of AGG interruptions in the transcription of FMR1 premutation alleles. PloS one *6*, e21728.

Zhang, A., David, J.J., Subramanian, S.V., Liu, X., Fuerst, M.D., Zhao, X., Leier, C.V., Orosz, C.G., Kelm, R.J., Jr., and Strauch, A.R. (2008). Serum response factor neutralizes Pur alpha- and Pur beta-mediated repression of the fetal vascular smooth muscle alpha-actin gene in stressed adult cardiomyocytes. American journal of physiology Cell physiology *294*, C702-714.

Zhang, D., Iyer, L.M., He, F., and Aravind, L. (2012). Discovery of Novel DENN Proteins: Implications for the Evolution of Eukaryotic Intracellular Membrane Structures and Human Disease. Frontiers in genetics *3*, 283.

Zinszner, H., Sok, J., Immanuel, D., Yin, Y., and Ron, D. (1997). TLS (FUS) binds RNA in vivo and engages in nucleo-cytoplasmic shuttling. Journal of cell science *110 (Pt 15)*, 1741-1750.

Zu, T., Gibbens, B., Doty, N.S., Gomes-Pereira, M., Huguet, A., Stone, M.D., Margolis, J., Peterson, M., Markowski, T.W., Ingram, M.A.*, et al.* (2011). Non-ATG-initiated translation directed by microsatellite expansions. Proceedings of the National Academy of Sciences of the United States of America *108*, 260-265.

Acknowledgements

I am very grateful to Dierk Niessing, who encouraged me to “jump into the cold water” and switch from cellular biology to biochemistry and structural biology. Thank you for giving me the opportunity to carry out my PhD in your lab, for believing in me, your guidance, your readiness for scientific discussions and your critical comments on this written work.

I would additionally like to thank Klaus Förstemann for officially supervising this thesis and being part of my thesis committee.

I thank the remaining members of my thesis committee Tobias Madl, Johannes Söding and Matthew Bennett for their scientific input and helpful suggestions. In particular I would like to thank Tobias Madl for agreeing on the collaboration and helping with NMR and SAXS data analysis.

Special thanks go to my collaboration partner Christoph Hartlmüller for his great NMR work.

I would also like to thank Christoph Göbl, who helped me with the SAXS measurements.

I extend my thanks to the people of the Madl lab for making me feel comfortable in Garching and giving me advice and support.

I am very grateful to Marta Pabis and Steffen Heuck for the nice atmosphere in the cell culture lab and for the mutual assistance.

My special thanks go to all the former and present members of the Niessing lab for the great atmosphere, scientific support, but also for everything beyond research – Teamstaffellauf, BBQ, skiing, coffee & cake sessions … and many nice chat.

I send big thanks to all my friends for their patience, support, and encouragement and for making my time in Munich memorable.

Zu guter Letzt möchte ich meinen Eltern, Lutz und Karin Weber, danken, die mich in allen Angelegenheiten immer unterstützt, mich motiviert und jederzeit an mich geglaubt haben.