

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

**Crystal structure,  
nucleic-acid binding properties,  
and dimerization model  
of Pur- $\alpha$**



**Almut Graebisch**

**aus Berlin**

**2010**

### **Erklärung**

Diese Dissertation wurde im Sinne von § 13 Abs. 3 bzw. 4 der Promotionsordnung vom 29. Januar 1998 von Herrn Dr. Dierk Niessing betreut und von Herrn Prof. Karl-Peter Hopfner vor der Fakultät für Chemie und Pharmazie vertreten.

### **Ehrenwörtliche Versicherung**

Diese Dissertation wurde selbstständig, ohne unerlaubte Hilfe erarbeitet.

München, am 31. März 2010

Almut Graebisch

Dissertation eingereicht am: 31. März 2010

1. Gutachter: Karl-Peter Hopfner

2. Gutachter: Klaus Förstemann

Mündliche Prüfung am: 17. Mai 2010

<b>Chapter 1 Introduction</b> .....	<b>1</b>
<b>1.1 Abstract</b> .....	<b>1</b>
<b>1.2 Scientific background</b> .....	<b>2</b>
1.2.1 Learning and memory .....	2
1.2.2 Active transport.....	2
1.2.3 Dendritic localization of mRNAs .....	3
Composition of the mRNA transport granule .....	5
Localization elements .....	7
Translocation .....	7
1.2.5 Activity-dependent translation.....	8
<b>1.3 Pur-<math>\alpha</math></b> .....	<b>9</b>
1.3.1 Cellular functions of Pur- $\alpha$ .....	11
Transcription.....	11
Replication.....	12
Cell cycle control.....	13
Tumor suppression .....	13
RNA binding .....	13
1.3.2 Fragile X Tremor/Ataxia syndrome.....	14
1.3.3 Pur- $\alpha$ knockout mouse.....	15
<b>1.4 Objectives</b> .....	<b>17</b>
<b>Chapter 2 Results</b> .....	<b>18</b>
<b>2.1 Structural studies</b> .....	<b>18</b>
2.1.1 Crystal structure of Pur- $\alpha$ .....	18
2.1.1.1 Human Pur- $\alpha$ .....	18
Crystallization trials of human Pur- $\alpha$ .....	18
Identification of PUR repeats in Pur- $\alpha$ .....	22
2.1.1.2 <i>Borrelia burgdorferi</i> Pur- $\alpha$ .....	23
Crystallization of <i>B. burgdorferi</i> Pur- $\alpha$ .....	25
Structure determination and refinement .....	28
Crystal structure of <i>B. burgdorferi</i> Pur- $\alpha$ .....	30
Surface assessment .....	32
Systematic structural comparison.....	33
2.1.1.3 <i>Drosophila melanogaster</i> Pur- $\alpha$ .....	35
Crystallization of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II .....	35
Structure determination and refinement .....	38
Crystal structure of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II .....	41
Surface assessment .....	42
Comparison with the crystal structure of <i>B. burgdorferi</i> Pur- $\alpha$ .....	44
Systematic structural comparison.....	45
Crystallization trials of <i>D. melanogaster</i> Pur- $\alpha$ repeat III .....	47
Co-crystallization trials of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II with nucleic acids .....	48
2.1.2 Solution structure of Pur- $\alpha$ .....	51
2.1.2.1 Analytical size-exclusion chromatography .....	51
2.1.2.2 Small angle X-ray scattering .....	52
SAXS of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II .....	55
SAXS of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-III .....	57

<b>2.2 Functional studies.....</b>	<b>60</b>
2.2.1 Nucleic-acid binding of Pur- $\alpha$ .....	60
2.2.1.1 Human Pur- $\alpha$ .....	60
Filter binding assays .....	60
2.2.1.2 <i>B. burgdorferi</i> Pur- $\alpha$ .....	62
Filter binding assays .....	62
Filter binding assays with mutant Pur- $\alpha$ .....	64
2.2.1.3 <i>D. melanogaster</i> Pur- $\alpha$ .....	66
Filter binding assays with Pur- $\alpha$ repeats I-III .....	67
Electrophoretic mobility shift assays with full length Pur- $\alpha$ .....	68
EMSA with Pur- $\alpha$ repeats I-II .....	69
EMSA with mutant Pur- $\alpha$ repeats I-II .....	70
Circular dichroism spectra of mutant Pur- $\alpha$ repeats I-II .....	74
EMSA with Pur- $\alpha$ repeat III .....	75
EMSA with quadruplex RNA and full length Pur- $\alpha$ .....	76
2.2.2 Protein binding of Pur- $\alpha$ .....	78
2.2.2.1 Yeast-two-hybrid screen .....	78
Interactions with known proteins .....	79
Arrestin1 .....	79
Btb VII .....	80
CG5758.....	80
Cka.....	80
Eye .....	81
Lamin C .....	81
Pur- $\alpha$ .....	82
Interactions with non-existing proteins .....	82
<b>Chapter 3 Discussion .....</b>	<b>84</b>
<b>3.1 Structure of Pur-<math>\alpha</math> .....</b>	<b>84</b>
3.1.1 PUR repeats are highly conserved structural units .....	84
3.1.2 <i>B. burgdorferi</i> Pur- $\alpha$ is a functional PUR protein .....	84
3.1.3 Two PUR repeats form a PUR domain.....	85
3.1.4 The PUR domain is similar to bacteriophage coat and Whirly proteins .....	86
3.1.5 Dimerization of Pur- $\alpha$ is mediated by PUR repeat III.....	88
3.1.6 PUR repeat III is predicted to form a PUR domain .....	90
3.1.7 Interaction of the PUR repeats is likely specific.....	91
3.1.8 The PUR domain is likely conserved in different orthologs .....	92
<b>3.2 Interaction with nucleic acids .....</b>	<b>93</b>
3.2.1 The PUR domain binds nucleic acids .....	93
3.2.2 PUR repeat III only weakly contributes to nucleic-acid binding.....	94
3.2.3 Pur- $\alpha$ binds DNA as well as RNA.....	95
3.2.4 The Pur- $\alpha$ dimer bears two nucleic-acid binding domains.....	97
3.2.5 The nucleic-acid binding surface of a PUR domain is reminiscent of an RRM domain .....	99



<b>3.3 Interaction with proteins .....</b>	<b>101</b>
3.3.1 Protein interactions are mediated by PUR repeats.....	101
3.3.2 Yeast-two hybrid screen .....	102
3.3.2.1 Potential artifacts .....	103
3.3.2.2 Arr1, LamC, Eye and Cka are putative binding partners of Pur- $\alpha$ .....	104
Arrestin1 .....	104
LaminC .....	105
Eye .....	106
Cka.....	106
Pur- $\alpha$ .....	106
<b>Chapter 4 Conclusion .....</b>	<b>107</b>
<b>Chapter 5 Materials and Methods.....</b>	<b>108</b>
<b>5.1 Consumables and chemicals .....</b>	<b>108</b>
<b>5.2 Oligonucleotides .....</b>	<b>108</b>
5.2.1 DNA oligonucleotides for molecular biology .....	108
4.2.2 DNA and RNA oligonucleotides for binding assays and crystallization.....	110
<b>5.3 Plasmids .....</b>	<b>111</b>
5.3.1 Commercial plasmids .....	111
5.3.2 Plasmids for recombinant protein expression in <i>E. coli</i> .....	111
5.3.3 Plasmids for baculovirus-based recombinant protein expression in insect cells .....	113
5.3.4 Plasmids for <i>in vitro</i> RNA transcription .....	114
<b>5.4 Bacterial strains.....</b>	<b>114</b>
<b>5.5 Media for cell culture .....</b>	<b>114</b>
<b>5.6 Molecular biology .....</b>	<b>115</b>
5.6.1 Standard cloning and site-directed mutagenesis .....	115
5.6.2 Transformation of <i>E. coli</i> and isolation of plasmid DNA.....	115
<b>5.7 Protein expression .....</b>	<b>115</b>
5.7.1 Recombinant protein expression in <i>E. coli</i> .....	115
Expression of selenomethionine-substituted protein.....	116
5.7.2 Baculovirus-based recombinant protein expression in insect cells .....	116
<b>5.8 Protein purification .....</b>	<b>117</b>
5.8.1 Purification of GST-tagged proteins .....	117
5.8.2 Purification of His-tagged proteins .....	118
4.8.3 Purification of MBP-tagged proteins .....	119
<b>5.9 Protein analysis.....</b>	<b>119</b>
5.9.1 Analytical size-exclusion chromatography .....	119
5.9.2 Circular dichroism spectroscopy .....	120
<b>5.10 Crystallization .....</b>	<b>120</b>
5.10.1 Crystallization and structure determination of <i>B. burgdorferi</i> Pur- $\alpha$ .....	120
5.10.2 Crystallization and structure determination of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II .....	121
5.10.3 Crystallization of <i>D. melanogaster</i> Pur- $\alpha$ repeat III.....	122
5.10.4 Co-crystallization with nucleic acids .....	122
<b>5.11 Limited proteolysis.....</b>	<b>123</b>
<b>5.12 Lysine methylation .....</b>	<b>123</b>
<b>5.13 Heavy atom derivatization by soaking .....</b>	<b>124</b>
<b>5.14 DNA soaking experiments .....</b>	<b>124</b>
<b>5.15 Small angle X-ray scattering .....</b>	<b>124</b>

<b>5.16 Nucleic-acid binding assays .....</b>	<b>125</b>
5.16.1 <i>In vitro</i> transcription of RNA.....	125
5.16.2 Isotopic labeling of oligonucleotides .....	126
5.16.3 Filter binding assays .....	126
5.16.4 Electrophoretic mobility shift assays .....	127
<b>5.17 Bioinformatics .....</b>	<b>128</b>
5.17.1 Protein parameters .....	128
5.17.2 Sequence alignment and secondary structure prediction .....	128
5.17.3 Structure visualization and analysis.....	129
<b>Chapter 6 References.....</b>	<b>130</b>
<b>Chapter 7 Appendix.....</b>	<b>144</b>
<b>7.1 Computational model of <i>D. melanogaster</i> Pur-<math>\alpha</math>.....</b>	<b>144</b>
<b>7.2 Sequence alignment and secondary structure prediction.....</b>	<b>145</b>
<b>7.3 Index of figures and tables .....</b>	<b>146</b>
<b>7.4 Abbreviations.....</b>	<b>148</b>
<b>Curriculum vitae .....</b>	<b>i</b>
Publications.....	ii
<b>Acknowledgements.....</b>	<b>iii</b>

## Chapter 1 Introduction

### 1.1 Abstract

This study characterizes Pur- $\alpha$  structurally and functionally. Pur- $\alpha$  is a highly conserved RNA- and DNA-binding protein involved in a multitude of cellular processes such as transcription, replication, cell cycle control, and mRNA transport. No homologous proteins with known structures are available.

X-ray crystallography is often hampered by the lack of diffraction-quality protein crystals. This study demonstrates how this bottleneck was overcome by the combination of iterative use of sensitive bioinformatics tools and structure determination of a bacterial homolog. The identification of three repeat regions (PUR repeats) in eukaryotic Pur- $\alpha$  enabled the detection of a bacterial homolog, which corresponds to one PUR repeat. The crystal structure of *Borrelia burgdorferi* Pur- $\alpha$  was solved at 1.9 Å and was employed for precise domain boundary prediction for the *Drosophila melanogaster* ortholog. Therewith it became possible to obtain diffraction-quality crystals of eukaryotic Pur- $\alpha$ . The crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II was solved at 2.1 Å and shares a highly conserved fold with *B. burgdorferi* Pur- $\alpha$ . One PUR repeat has an overall  $\beta\beta\beta\alpha$ - topology, and two PUR repeats interact with each other to form a globular PUR domain.

Small angle X-ray scattering (SAXS) analysis together with analytical size-exclusion chromatography provided evidence that dimerization of full length Pur- $\alpha$  requires PUR repeat III. PUR repeat III is proposed to form a PUR domain with a PUR repeat III from another Pur- $\alpha$  molecule. Surface envelopes calculated from SAXS data comply with this dimerization model.

DNA- as well as RNA-binding properties of Pur- $\alpha$  were examined by filter binding assays and electrophoretic mobility shift assays. Structure-guided mutagenesis identified the  $\beta$ -sheets of the PUR domain as the nucleic-acid binding surface.

To assess the protein-binding properties of *D. melanogaster* Pur- $\alpha$ , a yeast-two-hybrid screen was commissioned and evaluated. It confirmed the self-interaction of Pur- $\alpha$  and yielded Arrestin1, LaminC, Eye and Cka as putative previously unknown interaction partners.

## **1.2 Scientific background**

### **1.2.1 Learning and memory**

The biochemical basis for learning and memory is the ability of the nervous system to transfer neural activity into persistent changes in synaptic connectivity. Activity-dependent synaptic plasticity is established by long-term potentiation or long-term depression of the synapse (Steward and Schuman, 2003). This requires regional and temporal regulation of protein levels, substantially, but not exclusively, attained by localized protein synthesis in the dendrite (Schuman et al., 2006). The observation that translation does not only take place in the cell soma of a neuron, but also in dendrites, was groundbreaking in the early 1980s (Steward and Levy, 1982). The process allows a fine-tuned, spatially restricted gene expression in quick response to specific stimuli. Local expression of proteins requires the translocation of specific mRNAs from the site of transcription in the nucleus to the sites of translation in the distal dendrites. The delivery of mRNA is achieved by active transport with molecular motors (Kindler et al., 2005).

### **1.2.2 Active transport**

In contrast to passive transport (diffusion), active transport requires ATP-consumption and is performed by motor proteins that move bidirectionally through the cytoplasm, exploiting the polarity of the cytoskeleton (Mallik and Gross, 2004).

The cytoskeleton provides mechanical support to the cell and fulfills several functions: it spatially organizes the cell contents, connects the cell to the extra-cellular environment, segregates the chromosomes during mitosis, and establishes cellular forces and shape change. The cytoskeleton is constituted by three classes of highly dynamic polymers, that are microtubules, actin, and intermediate filaments (Fletcher and Mullins, 2010).

Microtubules are longitudinal polymers composed of  $\alpha$ - and  $\beta$ - tubulin, having a slow growing “minus end” and a fast growing “plus end”. In axons and distal dendrites, microtubule orientation is highly polarized with the minus ends directed to the cell body and the plus ends pointing away from the soma (Hirokawa and Takemura, 2005).

There are three classes of molecular motor proteins: myosins, dyneins and kinesins. Myosin-mediated cargo transport is performed by the class of non-conventional myosins,

that are neither muscle nor filamentous myosins. Most non-conventional myosins advance towards the plus end of actin filaments. Cytoplasmic dyneins perform minus-end directed movement on microtubules (Mallik and Gross, 2004). The kinesin superfamily encompasses 45 members in mice and human, based on 14 evolutionary related families (Lawrence et al., 2004). All kinesins are highly homologous in their globular motor domain, which contains the ATP-binding site and a microtubule-binding site. In contrast, the C-terminal cargo-binding domain shows a high degree of diversity, enabling various cargoes to travel with the help of kinesins (Hirokawa and Takemura, 2005). Most kinesins (KIFs) support anterograde transport by moving along microtubules in plus-end direction (Lawrence et al., 2004), as it is required for mRNA transport from the cell soma to the distal dendrites.

Active transport of mRNA is a key mechanism in multiple cellular processes, as reviewed by Martin and Ephrussi (Martin and Ephrussi, 2009). For instance, the formation of the body axis in the developing *Drosophila* embryo is established by the localization of mRNAs such as *bicoid*, *nanos* and *oskar* to the anterior respective posterior pole of the oocyte (Becalska and Gavis, 2009; Johnstone and Lasko, 2001). Another well-studied example is the regulation of mating-type switch in budding yeast, which requires the localization of *ASH1* mRNA to the daughter cell (Muller et al., 2007). Active transport does not only apply to mRNA, but to a diversity of cargoes including proteins, vesicles, organelles, and cytoskeletal filaments (Vale, 2003).

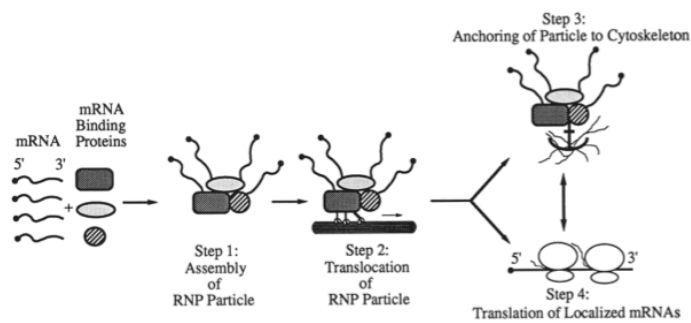
### **1.2.3 Dendritic localization of mRNAs**

The mRNAs locally translated in dendrites encode for a multitude of proteins with diverse functions such as receptors, protein kinases, cytoskeletal proteins, and scaffolding proteins (Table 1-1) (Eberwine et al., 2001; Kosik and Krichevsky, 2002; Steward and Schuman, 2003). Recent evidence supports there are hundreds of localized mRNAs as opposed to a handful, as previously thought (Eberwine et al., 2001; Martin and Ephrussi, 2009) (Table 1-1).

<b>mRNA</b>	<b>Abbreviation</b>	<b>Classification</b>
Activity-regulated cytoskeleton-associated protein	Arc	cytoskeletal protein
Brain-derived neurotrophic factor	BDNF	growth factor
Calcium calmoduline dependent kinase II- $\alpha$	CaMKII- $\alpha$	protein kinase
Cyclic AMP response element binding factor	CREB	transcription factor
Glutamate receptor subunit 1	GluR1	receptor
Glutamate receptor subunit 2	GluR2	receptor
Microtubule-associated protein 2	MAP2	cytoskeletal protein
N-methyl- <i>D</i> -aspartate receptor	NMDAR1	receptor
Neurotropic tyrosine kinase receptor type 2	TrkB	receptor
Shank	-	scaffolding protein
vasopressin	-	neuropeptide
$\beta$ -Actin	-	cytoskeletal protein

**Table 1-1. Examples of mRNAs transported in dendrites (Kosik and Krichevsky, 2002; Martin and Zukin, 2006).**

An ordered multistep process is suggested for the transport of dendritic mRNAs (Figure 1-1) (Wilhelm and Vale, 1993). The starting step constitutes the assembly of the transport granule, a large ribonucleoprotein particle (RNP). It contains the mRNA together with a variety of protein factors (Figure 1-1, Step 1) (Kanai et al., 2004). Then the transport granule is equipped with a motor protein, which moves along the cytoskeleton (Figure 1-1, Step 2). At the site of destination, the transport granule is anchored at the local cytoskeleton, presumably to the actin network of distal dendrites (Figure 1-1, Step 3). In a regulated process in response to stimulation, the granule is rearranged and the translation of the mRNA starts (Figure 1-1, Step 4) (Krichevsky and Kosik, 2001) (Wilhelm and Vale, 1993).



**Figure 1-1. Schematic drawing of the mRNA transport mechanism in neurons. Step 1: mRNA and RNA-binding proteins assemble to form the transport granule (RNP). Step 2: Motor proteins actively transport the granule along the cytoskeleton. Step 3: At the site of destination, the RNP is anchored to the cytoskeleton. Step 4: Rearrangement of the granule and initiation of translation. Picture taken from (Wilhelm and Vale, 1993).**

### Composition of the mRNA transport granule

The mRNA transport granules are highly complex macromolecular structures (Kanai et al., 2004) and contain mRNAs together with a multitude of protein factors (Barbarese et al., 1995; Kanai et al., 2004; Knowles et al., 1996). One of the factors essential for the transport is Pur- $\alpha$ , on which this study focuses.

Kanai and colleagues purified large RNA-transporting granules from mouse brain lysate by exploiting their affinity to the cargo-binding domain of conventional kinesin (KIF5) (Kanai et al., 2004). The granules had a size of about 1000 S, contained *CamKII- $\alpha$*  as well as *Arc* mRNA and were resistant against detergent, but RNase sensitive. Characterization of the protein content of the granule revealed at least 42 different factors, including RNA-binding proteins, RNA transport proteins, RNA helicases, heterogenous ribonucleoproteins (hnRNPs), and proteins involved in translation (Kanai et al., 2004). Following the hypothesis that some factors might constitute a functional core complex while additional factors might have regulatory function, Kanai and colleagues washed the complex with increasing salt concentrations. They identified four proteins that were more strongly bound to KIF5 than the other factors, namely Pur- $\alpha$ , Pur- $\beta$ , hnRNP-U and Staufen. Their importance is underlined by RNAi knock down experiments, which revealed that those four (among others) are essential for mRNA transport in dendrites (Kanai et al., 2004). However, a potential diversity in granule composition could not be resolved by this approach.

Barbee and colleagues examined the composition of single granules using GFP-labeled Staufen and antibody staining of RNPs in *Drosophila* neurons (Barbee et al., 2006). They assessed the co-localization of known granule components and revealed heterogeneity among the RNPs. For example, only 45 % of Staufen-positive granules also contained Fmr protein (Barbee et al., 2006). Similarly, only 50 % of  $\beta$ -actin mRNA containing granules co-localized with zipcode-binding protein1 (ZBP1) (Tiruchinapalli et al., 2003). Barbee and colleagues propose that different classes of mRNAs are repressed by different mechanisms. This opens up the possibility to specifically release mRNAs for translation (Barbee et al., 2006).

The transport of the mRNAs and the regulation of localized translation are thought to be functionally interdependent, and there are various trans-action factors that bind localization elements in the mRNA and are responsible for transport as well as for translational control (Kindler et al., 2005).

Besides its crucial role for transport (Kanai et al., 2004), Pur- $\alpha$  seems to be also involved in translational regulation. Pur- $\alpha$  is associated with polysomes in mouse brain, together with Staufen, Fmr protein, and Myosin Va (Ohashi et al., 2002). Together with additional factors, these four proteins form a large RNP, which is RNase sensitive, indicating that RNA is required for complex stability. Pur- $\alpha$  has also been suggested to play a role in the RNA-dependent assembly of the granule, a process not yet understood (Ohashi et al., 2002).

Further RNA-binding proteins required for mRNA transport are Staufen, zipcode-binding protein1 and 2 (ZBP-1 and ZBP-2), and cytoplasmic polyadenylation element-binding protein (CPEB) (Kindler et al., 2005; Martin and Ephrussi, 2009). Staufen is a double-stranded RNA-binding protein, which associates with microtubules and is crucial for somatodendritic localization of mRNAs, possibly by linking the microtubule-associated transport machinery to the mRNAs (Martin and Zukin, 2006). Staufen is also a key factor of mRNA localization in the *Drosophila* oocyte and embryo (Kiebler et al., 1999; St Johnston, 2005). ZBP1 is a nucleo-cytoplasmic shuttling protein that binds  $\beta$ -actin mRNA in the nucleus and promotes its translational silencing as well as its incorporation in RNPs. This indicates that mRNA targeting is already initiated in the nucleus (Bramham and Wells, 2007; Kindler et al., 2005). The cis-acting recognition element of CPEB, cytoplasmic polyadenylation element (CPE) is contained in a subset of dendritic



mRNAs. Binding of CPEB mediates both translational control and transport to dendrites (Huang et al., 2003).

### **Localization elements**

It is assumed that many different mRNAs travel in one granule (Gao et al., 2008; Kosik and Krichevsky, 2002). The destination of a transcript is encoded by cis-acting elements in the mRNA sequence, which are referred to as “zipcodes” (Martin and Ephrussi, 2009; Martin and Zukin, 2006; Singer, 1993).

Some general principles for cis-acting localization elements emerged from the investigation of related processes such as transport of *bicoid* mRNA in *Drosophila* oocytes or *ASH1* mRNA localization in budding yeast. Most of the localization elements identified so far lie in the 3'-untranslated region (UTR) of the transcript. One mRNA often carries several repetitive and redundant localization elements, but there are also examples where different cis-acting elements are responsible for distinct steps in localization (Martin and Ephrussi, 2009). A common feature is the formation of secondary structures that are required for localization, e. g. stem loops (Martin and Ephrussi, 2009). However, the search for dendritic targeting elements (DTEs) in mRNA is just emerging and reveals great complexity. The so far described DTEs differ vastly in length, sequence and number per mRNA (Kindler et al., 2005). The interplay between distinct DTEs in one mRNA, e. g. in *CaMKII- $\alpha$*  mRNA, is not yet understood and might direct the mRNA to different target sites within dendritic branches (Kindler et al., 2005).

### **Translocation**

The current model of mRNA transport in dendrites proposes kinesin-based translocation along microtubules for long distances from the soma to the dendrites (Ohashi et al., 2002). Myosin-driven motility is proposed for the peripheral regions of the dendrites, which are rich in actin fibers (Ohashi et al., 2002). Not all granules translocate at all times, some appear docked to microtubules, while others move with a speed of 4-6  $\mu\text{m}/\text{min}$ , consistent with fast transport motors such as kinesins or dyneins (Kosik and Krichevsky, 2002). Kanai and colleagues characterized the movement of RNA transporting granules in mouse dendrites and found that they move bidirectional, driven by a kinesin motor and an opposing motor, presumably dynein (Kanai et al., 2004).

### 1.2.5 Activity-dependent translation

Krichevsky and Kosik propose a model which links RNA localization to activity-dependent translation and synaptic plasticity (Figure 1-2) (Krichevsky and Kosik, 2001). The research group sedimented RNA fractions from cultured neurons on a sucrose gradient and assigned a distinct fraction heavier than polysomes to the transport granules (Krichevsky and Kosik, 2001). They suggest that during transport, the mRNAs are translationally silent, as shown by the failure of the granule fraction to incorporate  $^{35}\text{S}$ -labelled amino acids. Furthermore, the transport granules lack tRNAs and the rate-limiting initiation factors eIF4E and 4G. Upon depolarization by incubation with potassium chloride, the granules reorganized and specific mRNAs (*CaMKII-a*, *TrkB*, *NMDAR1*) were shifted to the polysome fraction (Krichevsky and Kosik, 2001).

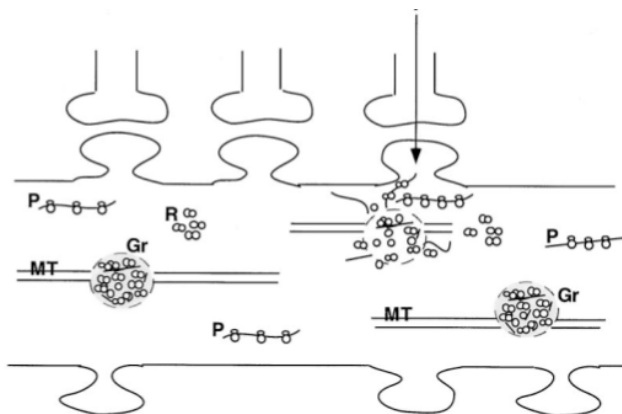


Figure 1-2. Schematic drawing of localized translation in neurons. An RNA granule (GR) containing mRNA, the translation machinery and regulatory factors is transported along a microtubule (MT). During transport, the granule is translationally silent. Upon depolarization (indicated by an arrow), the granule reorganizes, releasing mRNAs, ribosomes (R) and polysomes (P). Global protein synthesis is reduced, but specific mRNAs are shifted to the polysome fraction and translated. Figure taken from (Krichevsky and Kosik, 2001).

Several lines of evidence support the idea that neuronal activity triggers translational initiation. Ouyang and colleagues found that induction of long term potentiation in hippocampal neurons enhances the expression of CaMKII- $\alpha$  (Ouyang et al., 1999). Furthermore, stimulation with a glutamate receptor agonist increased reporter protein translation in living dendrites (Job and Eberwine, 2001). Consistently, rapid initiation of translation in response to NMDA receptor activation was shown for *Arc* (Steward et al., 1998; Steward and Worley, 2001).

Not only translation, but also localization of specific mRNAs is triggered by neuronal activity, as demonstrated for *Arc*, *TrkB*,  *$\beta$ -actin*, *GluR1*, *GluR2*, and *BDNF* (Martin and Zukin, 2006). This requires a pathway for synapse-to-nucleus signaling, whose mechanism is still unclear but seems to involve the active nuclear import pathway (Thompson et al., 2004).

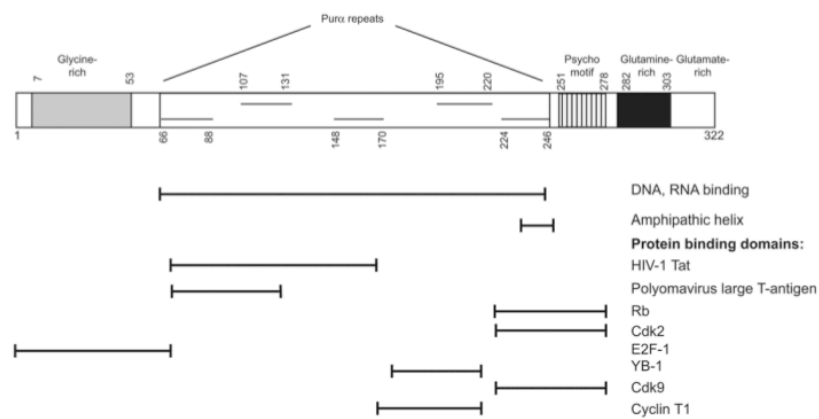
### 1.3 Pur- $\alpha$

Pur- $\alpha$  and its family members Pur- $\beta$  and Pur- $\gamma$  constitute the *purine-rich element (PUR)* binding protein family, which is highly conserved from bacteria to metazoan (White et al., 2009). PUR proteins are named after their specific binding to purine-rich nucleic acids and show sequence-specific affinity to DNA as well as RNA (Bergemann et al., 1992; White et al., 2009). They preferably bind repeat regions of their consensus sequence (GGN)<sub>n</sub>, with N unlike guanine. The PUR family members are highly homologous to each other, especially in their central nucleic-acid binding region. The overall sequence identity of human Pur- $\alpha$  and Pur- $\beta$  is 61 %, whereas it is 74 % in their central region. PUR proteins are also highly conserved between species, as illustrated by the fact that human and mouse Pur- $\alpha$  differ only in 2 amino acids (Liu and Johnson, 2002).

Pur- $\alpha$  is a ubiquitous protein, which is expressed in nearly every metazoan tissue. To perform its various cellular functions, it is present in the cytoplasm as well as in the nucleus (White et al., 2009). Pur- $\alpha$  binds sequence-specifically to dsDNA, ssDNA and RNA and is involved in a variety of cellular processes such as transcription, replication, cell cycle control, and, as already stated, transport of mRNA in dendrites (Gallia et al., 2000; White et al., 2009).

Previous work described the modular structure of human Pur- $\alpha$  (GeneID 5813), which is composed of 322 amino acids (34.9 kDa) (Figure 1-3). Pur- $\alpha$  has a glycine-rich N-terminus, a highly conserved central core region, and a C-terminus bearing a Psycho motif as well as a region rich in glutamines and glutamates (Gallia et al., 2000). The glycine-rich N-terminus (amino acids 7 to 53) is predicted to be unfolded and contains a stretch of 18 glycines interrupted by only one single serine. At the C-terminus, the Psycho motif (amino acids 251 to 278) shows limited homology to the large T-antigen of several polyomaviruses and is named after its content of proline (P), serine (S), tyrosine

(Y), and cysteine (C). In the glutamate/glutamine rich region (amino acids 276 to 321), half of the residues are either glutamates or glutamines (Gallia et al., 2000). The core region (amino acids 54 to 250) was found to be the RNA- and DNA-binding region as well as the interaction domain for many protein partners (White et al., 2009). It was previously proposed that the core region contains repetitive sequence elements that can be classified into three class I repeats (amino acids 66-88, 148-170, and 224-246) and two class II repeats (amino acids 107-131 and 195-220) (Bergemann et al., 1992; Gallia et al., 2000; White et al., 2009).



**Figure 1-3. Schematic representation of human Pur- $\alpha$ .** The modular protein is composed of a glycine-rich domain at the N-terminus, a central core region bearing the repetitive elements, and a C-terminus with a Psycho motif as well as a region rich in glutamines and glutamates. The central core region is reported to be the nucleic acid binding region. It is also the interaction domain for various proteins, as indicated by the black bars. Picture taken from (White et al., 2009).

### 1.3.1 Cellular functions of Pur- $\alpha$

#### Transcription

Appropriate to its role in replication and transcription, Pur- $\alpha$  is reported to have DNA double strand destabilizing properties (Darbinian et al., 2001a).

Pur- $\alpha$  interacts with TATA-less promoters and functions as transcriptional activator as well repressor. There are at least 19 target genes known whose transcription is regulated by Pur- $\alpha$  (White et al., 2009). These findings are mainly based on the use of a fusion plasmid bearing the promoter of the gene of interest and a reporter gene, monitoring the transcriptional activity in dependence of Pur- $\alpha$  co-expression in cell culture.

Interestingly, among the gene targets of Pur- $\alpha$ , several gene products are involved in apoptosis or cell proliferation, e.g. Bax (Kim et al., 2008), tumor necrosis factor  $\alpha$  (Darbinian et al., 2001c), transforming growth factor  $\beta$  (Thatikunta et al., 1997), platelet-derived growth factor subunit A (Zhang et al., 2005), GATA binding protein2 (Penberthy et al., 2004), and somatostatin (Sadakata et al., 2000). This underlines the involvement of Pur- $\alpha$  in cell cycle regulation and tumor suppression, as described below.

Furthermore, several target gene products have a specific importance for neuronal cells like myelin basic protein (Dobretsova et al., 2008), myelin proteolipid protein1 (Dobretsova et al., 2008), FE65 (Zambrano et al., 1997), CD11c $\beta$ -2 integrin (Shelley et al., 2002), and somatostatin (Sadakata et al., 2000). The importance of Pur- $\alpha$  for the nervous system is described below (section 1.3.2 and section 1.3.3).

Pur- $\alpha$  often exerts its regulatory function in multimeric complexes in concert with other factors. For example, the repression of *smooth muscle  $\alpha$ -actin* gene transcription is mediated by both Pur- $\alpha$  and Pur- $\beta$  via binding to the *MCAT* enhancer element (Kelm et al., 1999a; Knapp et al., 2006).

Pur- $\alpha$  auto-regulates its own expression by binding to its own promoter and preventing its transcription (Muralidharan et al., 2001). E2F-1, which is a crucial transcription factor for cell cycle control and an interaction partner for tumor suppressor proteins, acts as an antagonist and enhances transcription of the *PURA* gene (Darbinian et al., 2006).

Pur- $\alpha$  is also hijacked by viral genes, and activates transcription of human immunodeficiency virus (HIV-1) and human polyomavirus (JCV). Pur- $\alpha$  activates HIV-1 transcription by binding to the HIV-1 *TAR* element and interacting with the HIV-1 Tat

protein (Chepenik et al., 1998; Krachmarov et al., 1996). Further, Pur- $\alpha$  interacts with HIV-1 Rev protein and the *RRE* element in HIV-1 genomic RNA and thereby enhances the export of unspliced viral RNA from the nucleus (Kaminski et al., 2008a).

JC virus is a ubiquitous virus, which is latent in approximately 80 % of the population and usually subclinical. In severely immunosuppressed individuals, like AIDS patients, it causes progressive multifocal leucoencephalopathy (PML) (Cinque et al., 2009). PML leads to demyelination in regions of the central nervous system, severe brain lesions and death. HIV-1 protein Tat acts synergistically with Pur- $\alpha$  to activate the major late promoter of JCV via the Tat-responsive element, *upTAR*. Pur- $\alpha$  was shown to bind the *upTAR* element, which contains the Pur- $\alpha$  consensus sequence (GGN)<sub>n</sub> (Krachmarov et al., 1996).

Pur- $\alpha$  also controls gene expression at the translational level. It was shown to bind the mRNA transcript of the *vascular smooth muscle  $\alpha$ -actin* gene to suppress translation (Kelm et al., 1999b).

## **Replication**

Pur- $\alpha$  associates with zones of initiation of DNA replication. Pur- $\alpha$  was first discovered by its specific affinity to the *purine-rich element (PUR)* near the center of chromosomal replication upstream the human *c-myc* gene (Bergemann and Johnson, 1992; Bergemann et al., 1992). In eukaryotes, the PUR element is conserved in origins of replication and in gene flanking regions, such as near hamster *dhfr* and *rhodopsin* genes, the human  *$\beta$ -globin* gene, and the mouse *adenosine deaminase* gene (Itoh et al., 1998). The function of Pur- $\alpha$  is not known for most of these interactions.

Pur- $\alpha$  is also involved in the replication of several viruses. It associates with the origin of DNA replication and the plasmid maintenance region of bovine papilloma virus (BPV) (Habiger et al., 1992; Jurk et al., 1996). The replication of the human neurotrophic polyomavirus (JCV) is negatively regulated by Pur- $\alpha$ , possibly by interfering with JCV T-antigen (Chang et al., 1996).

There is also recent evidence that Pur- $\alpha$  might be involved in replication-dependent repair of DNA lesions, as Pur- $\alpha$  deficient cells show an enhanced sensitivity to the DNA-

replication inhibitor hydroxyurea and to the DNA-crosslinking anti-tumor drug cis-platin (Kaminski et al., 2008b; Wang et al., 2007).

### **Cell cycle control**

Various protein interaction partners have been identified that suggest a role for Pur- $\alpha$  in cell cycle regulation. These are the retinoblastoma tumor suppressor protein (Johnson et al., 1995), E2F-1 (Darbinian et al., 1999), Sp1 (Tretiakova et al., 1999), the Y-box binding protein1 (Safak et al., 1999), CyclinT1/Cdk9 (Darbinian et al., 2001c), and CyclinA /Cdk2 (Liu et al., 2005). The binding of Pur- $\alpha$  to most of these cell cycle regulating proteins was demonstrated with purified protein to be direct and specific. Although the functional consequence for most of these interactions is not explored, binding to this conglomerate of cell cycle regulating proteins suggests a role for Pur- $\alpha$  in cell cycle control. This interpretation is supported by the cell-cycle dependent variation of Pur- $\alpha$  expression levels. Pur- $\alpha$  levels in monkey CV-1 cells peak during mitosis and drop at the onset of S-phase (Itoh et al., 1998). Consistently, cell cycle arrest at G<sub>1</sub>/S and G<sub>2</sub>/M checkpoints is caused by microinjection of Pur- $\alpha$  in NIH-3T3 cells (Stacey et al., 1999).

### **Tumor suppression**

Cell cycle control is highly connected to oncogenic transformation. Pur- $\alpha$  acts as a tumor suppressor in human brain tumor (glioblastoma cells) (Darbinian et al., 2001b) and prevents growth of Ras-transformed NIH-3T3 cells in soft agar (Barr and Johnson, 2001). Furthermore, *PURA* gene deletions are frequently associated with myelogenous leukemia (Lezon-Geyda et al., 2001). For prostate cancer, loss of Pur- $\alpha$  is reported to promote cancer progression by triggering androgen repressor over-expression (Inoue et al., 2008; Liu et al., 2010). Taken together, these results indicate that Pur- $\alpha$  is a tumor suppressor protein.

### **RNA binding**

In mouse brain, Pur- $\alpha$  binds to an abundant non-translated RNA homologous to *7SL RNA*, named *PU-RNA*. The 290 nt long *PU-RNA* contains Alu core sequences with the

repeat motif GGAGGC and is predicted to form a stem-loop structure (Tretiakova et al., 1998). The binding of Pur- $\alpha$  to the *MBP* promoter sequence is specifically reduced by *PU-RNA* and is restored by the addition of RNase (Gallia et al., 1999; Tretiakova et al., 1998). Tretiakova and colleagues propose that *PU-RNA* is a cofactor that negatively regulates the expression of the *MPB* gene.

Pur- $\alpha$  was also shown to bind to rat *BCI* RNA and is proposed to enhance its transcription, which relies on RNA polymerase III (Kobayashi et al., 2000). The non-translated *BCI* RNA is preferentially expressed in neurons, and is specifically transported to the dendrites in RNPs (Tiedge et al., 1991). *BCI* RNA is suggested to control activity-dependent neuronal plasticity by directing the mRNA transport granule to its destination and by regulating translational silencing during transport (Lin et al., 2008; Wang et al., 2002). Ohashi and colleagues propose that the binding of Pur- $\alpha$  and Pur- $\beta$  to the two dendritic targeting elements in *BCI* RNA cross-links the *BCI* RNA to microtubules (Ohashi et al., 2000). Wang and colleagues showed that *BCI* RNA represses translation by directly interacting with the translation initiation apparatus (Wang et al., 2005; Wang et al., 2002).

In addition, Pur- $\alpha$  binds the functional homolog of *BCI* RNA, human *BC200* RNA (Johnson et al., 2006) and the HIV-1 *TAR* element (Chepenik et al., 1998). Further, Pur- $\alpha$  is proposed to bind r(CGG)-expansions in the 5'-untranslated region of the *Fmr1* mRNA in the neurodegenerative disease Fragile X Tremor/Ataxia syndrome (section 1.3.2) (Jin et al., 2007).

### **1.3.2 Fragile X Tremor/Ataxia syndrome**

Recent evidence establishes a crucial role for Pur- $\alpha$  in the pathogenesis of the neurodegenerative disorder Fragile X Tremor/Ataxia syndrome (FXTAS) (Oostra and Willemsen, 2009). The predominantly male FXTAS patients suffer from progressive action tremor with ataxia and general brain atrophy leading to cognitive decline and dementia (Berry-Kravis et al., 2007; Hagerman et al., 2001). FXTAS is caused by (CGG)<sub>n</sub>-repeats in the 5'-untranslated region of the *fragile X mental retardation1 (FMRI)* gene. Normal alleles contain <55 repeats, while FXTAS is associated with the so-called premutation between 55 and 200 repeats. The prevalence of the premutation is estimated



to be 1 out of 1000 males with an onset of the disease in the 5<sup>th</sup> decade (O'Donnell and Warren, 2002). The so-called full mutation with >200 repeats induces transcriptional shut down of the *FMRI* gene, resulting in Fragile X syndrome. This disease is a common form of inherited mental retardation with an estimated prevalence of 1 out of 4500 males (O'Donnell and Warren, 2002). A pathological hallmark of FXTAS are ubiquitin-positive nuclear inclusions in neurons and astrocytes, which contain *Fmr1* mRNA and more than 20 proteins, including Pur- $\alpha$  as well as LaminA/C (Greco et al., 2002; Iwahashi et al., 2006; Jin et al., 2007). In FXTAS carriers, Fmr1 protein levels are reduced, while *Fmr1* mRNA levels are elevated (Kenneson et al., 2001; Tassone et al., 2000). A recent model suggests that the toxic r(CGG)<sub>n</sub>-repeat RNA is the underlying cause of the neurodegenerative disease. Jin and colleagues showed that the expression of r(CGG)<sub>90</sub>-repeats is sufficient to cause neurodegeneration in a *Drosophila* model of FXTAS (Jin et al., 2007). They propose a mechanism by which r(CGG)<sub>n</sub>-binding proteins are sequestered from their normal functions by the abundant premutation RNA (Jin et al., 2003). The group further found that Pur- $\alpha$  binds directly and specifically to r(CGG)<sub>n</sub>-repeat RNA. Moreover, over-expression of Pur- $\alpha$  rescues the r(CGG)<sub>n</sub>-mediated eye neurodegeneration phenotype of the FXTAS fly model. Jin and colleagues propose that loss of functional Pur- $\alpha$  could impair mRNA transport in neurons, leading to neuronal cell death (Jin et al., 2007).

### 1.3.3 Pur- $\alpha$ knockout mouse

The crucial role of Pur- $\alpha$  for brain development is highlighted by its genetic inactivation in the mouse. Homozygous animals lacking both alleles of the *PURA* gene (*PURA*<sup>-/-</sup>) appear normal at birth, but after two weeks, they develop neurological problems. Suffering from severe tremor and spontaneous seizures, they die four weeks after birth (Khalili et al., 2003). In *PURA*<sup>-/-</sup> mice, the number of neuronal cells in hippocampus, brain cortex and cerebellum is reduced. Lack of proliferation of precursor cells might reflect the role of Pur- $\alpha$  in DNA replication (Khalili et al., 2003). Possibly Pur- $\alpha$  is required for developmentally timed DNA replication, leading to cell differentiation. In *PURA*<sup>-/-</sup> mice, astrocytes, oligodendrocytes and Purkinje cells show a pathological development and the number of synapses is significantly decreased in the hippocampus (Khalili et al., 2003). Lamination is defective in the hippocampus and cerebellum of

*PURA*<sup>-/-</sup> mice. This is in accordance to a role of Pur- $\alpha$  in the transcription of *myelin basic protein (MBP)* and *myelin proteolipid protein1 (Plp1)* gene (Dobretsova et al., 2008; Tretiakova et al., 1999). However, Pur- $\alpha$  is not absolutely required for MBP expression, as the myelin sheath is present in the spinal cord and optic nerve of the *PURA*<sup>-/-</sup> animals (Khalili et al., 2003). A crucial role for Pur- $\alpha$  in dendritic mRNA transport is underlined by the dramatic mislocalization of Staufen and FMRP, both components of mRNA-transporting granules (Johnson et al., 2006; Kanai et al., 2004; Ohashi et al., 2002). It is speculated that despite the essential role of Pur- $\alpha$ , survival of the *PURA*<sup>-/-</sup> animals is prolonged by its homolog Pur- $\beta$ , which might rescue loss of Pur- $\alpha$  (Khalili et al., 2003).

In summary, Pur- $\alpha$  is multifunctional protein with diverse roles in cellular gene regulation. By interacting with dsDNA, ssDNA, RNA as well as proteins, Pur- $\alpha$  is involved in transcription, replication, cell cycle control, and mRNA transport in neurons. Multifunctional proteins such as Pur- $\alpha$  provoke the conventional “one protein – one function” paradigm, and transcription factors that interact with both DNA and RNA raise the intriguing question how specificity is achieved (Cassiday and Maher, 2002).

## 1.4 Objectives

In a reductionist approach, this study aims to contribute to the understanding of mRNA localization by the structural and functional characterization of one essential factor, Pur- $\alpha$ . Since PUR proteins show no sequence homology to other protein families with previously determined structures, their 3-dimensional fold is unknown. It was aimed at obtaining the crystal structure of Pur- $\alpha$  to gain an understanding of its interaction with various protein partners as well as with nucleic acids. It would be especially interesting to find out whether there are distinct or common binding modes for DNA and RNA targets and what determines specificity.

X-ray crystallography provides structural information at the atomic level and has led to a tremendous increase in biological understanding since the 1960s (Kendrew et al., 1958). In recent years, the combination of structural methods has proven very successful when a classic approach was aggravated by technical limitations (e. g. hampered crystallization) or the problem's complexity (e. g. multi-protein complexes) (Cowieson et al., 2008). Small angle X-ray scattering (SAXS) is a technique complimentary to X-ray crystallography, as it allows the investigation of macromolecules in solution, providing additional information about oligomeric state, shape and structural dynamics (Putnam et al., 2007). Moreover, SAXS independently assesses whether fold and oligomeric state of the crystallized protein are predicated on the crystal lattice or if they are also found in solution and thus have biological relevance.

Structural information was related with functionality by the identification of the RNA- and DNA- binding surface of Pur- $\alpha$ . Therefore, structure-guided mutagenesis was performed and nucleic-acid binding properties of wild type and mutant Pur- $\alpha$  were assessed.

Since Pur- $\alpha$  is a multi-functional protein, this study also aimed at identifying new protein-binding partners, which might act together with Pur- $\alpha$  to exert its diverse cellular functions. Therefore, a yeast-2-hybrid screen was commissioned and evaluated.

## **Chapter 2 Results**

The crystal structures of two orthologs, *B. burgdorferi* Pur- $\alpha$  and *D. melanogaster* Pur- $\alpha$  repeats I-II, were solved (section 2.1.1.1 and section 2.1.1.2). The solution structure of *D. melanogaster* Pur- $\alpha$  repeats I-II was obtained by SAXS and complemented the crystals structure (section 2.1.2.2). SAXS analysis further showed that *D. melanogaster* Pur- $\alpha$  repeats I-III is dimeric and lead to a dimerization model (section 2.1.2.2). Nucleic-acid binding properties of *B. burgdorferi*, *D. melanogaster*, and human Pur- $\alpha$  were characterized by electrophoretic mobility shift assays and filter binding assays (section 2.2.1). Binding affinities were also determined for truncated and mutated protein fragments of *B. burgdorferi* and *D. melanogaster* Pur- $\alpha$ . This allowed for the mapping of the nucleic-acid binding surface on the respective crystals structure (section 2.2.1).

### **2.1 Structural studies**

#### **2.1.1 Crystal structure of Pur- $\alpha$**

##### **2.1.1.1 Human Pur- $\alpha$**

Human Pur- $\alpha$  is comprised of 322 amino acids with a molecular weight of 34.9 kDa (GeneID 5813). In humans, the PUR family is also represented by Pur- $\beta$  and two isoforms of Pur- $\gamma$ . The latter appears in two isoforms that arise from different polyadenylation sites (Liu and Johnson, 2002).

#### **Crystallization trials of human Pur- $\alpha$**

Initially, crystallization trials of human Pur- $\alpha$  were performed. Because of its functional relevance for nucleic-acid binding, crystallization trials focused on protein fragments of the previously described core region (Figure 1-3 and section 1.3) (Gallia et al., 2000). The glycine-rich stretch at the N-terminus is predicted to be unstructured and thus likely preventing crystallization. Accordingly, expression constructs were designed that lacked the N-terminal glycine-rich region and also had different parts of the C-terminus deleted (Table 2-1).

The design of expression constructs also relied on previous work by D. Niessing and limited proteolysis (section 5.11). This method identifies stable protein fragments via the enzymatic digestion of flexible regions with sequence-specific proteases. Protein

fragments that yield diffraction quality crystals are typically resistant to proteolytic digestion (Gao et al., 2005).

Fragments of human Pur- $\alpha$  were expressed in different expression systems (*E. coli* or a baculovirus-based insect cell expression system) and purified using different affinity tags (Histidin, MBP, or GST) (section 5.7 and section 5.8). They were purified to homogeneity using affinity chromatography, anion-exchange chromatography and size-exclusion chromatography (section 5.8 and section 5.9). Numerous constructs tested were unstable, i.e. degraded or precipitated (Table 2-1).

Construct	Tag	Expression	Purification/ Stability	Initial Crystals	Refined Crystals	Comment
hPurA1-322	GST	+	+			a)
hPurA140-268	GST	+/-	-			
hPurA191-259	GST	+/-	-			
hPurA42-186	GST	+/-	-			
hPurA56-245	MBP	+	+	-		b)
hPurA56-245	His	+	+	+/-	-	c)
hPurA56-245	GST	+	-			
hPurA56-274	His	+/-	-			
hPurA56-274	GST	+	-			
hPurA56-287 C272S	MBP	+	+	+/-	-	d)
hPurA56-287 C272S	GST	+	+	+	-	e)
hPurA60-105	GST	-				
hPurA60-202	GST	+/-	-			
hPurA60-245	His	+	+	+/-	-	
hPurA96-287	GST	+/-	-			

**Table 2-1. Protein fragments of human Pur- $\alpha$  for crystallization trials. hPurA refers to human Pur- $\alpha$ . The numbers indicate the start and stop site of the amino acid sequence. a) Full length protein, b) crystallization trials as MBP-fusion protein, also after reductive methylation, c) crystallization trials with and without RNA or DNA, d) crystallization trials as MBP-fusion protein, e) crystallization trials with and without RNA or DNA.**

Fragments of human Pur- $\alpha$  comprised of amino acids 56-245 and 60 to 245, respectively, were expressed in *E. coli* and were purified employing a Histidin-Tag. The fragment Pur- $\alpha$  56-287 C272S was expressed in *E. coli* and purified with a GST-Tag. It had cysteine at position 272 replaced by serine because previous work of D. Niessing hinted at improved purification properties of this fragment (Table 2-1).

For these fragments, protein stability was highly depended on the ionic strength of the buffer. Initially, buffer containing 200 mM sodium chloride and 50 mM potassium

phosphate buffer pH 8.0 was used. The protein fragments could be concentrated only to 0.5 mg/mL, as precipitation occurred at higher concentrations. Initial crystallization trials (section 5.10) yielded crystals in various conditions, but when their diffraction was tested at Deutsches Elektronen Synchrotron (DESY, Hamburg, Germany), they were identified as salt crystals. Obviously, the phosphate buffer had to be replaced.

To improve the solubility of the protein fragments, several different buffer compositions were screened, including the variation of the buffer and the pH (HEPES, TRIS, pH 6-9), the variation of concentration and type of salt (100-500 mM sodium chloride or potassium chloride) and the use of solubility-enhancing additives (15 % glycerol or 50 mM arginine/glutamic acid). It was found that a salt content as high as 500 mM was required to prevent degradation and precipitation. Best stability was achieved in buffer containing 500 mM potassium chloride and 100 mM TRIS pH 8.4 at 4°C. The fragments could be concentrated up to approximately 2 mg/mL in this buffer. Unfortunately, no initial crystals were obtained with this buffer, presumably because the high salt content averted crystallization. It was then tried to replace a part of the potassium chloride by the volatile salt ammonium carbonate, but no crystals were obtained either (Table 2-1).

The maltose-binding protein (MBP) is used as an affinity-tag for protein purification. It was also reported to enhance the solubility of proteins and promote their crystallization (Smyth et al., 2003). Therefore, human Pur- $\alpha$  56-287 C272S and Pur- $\alpha$  56-245 were expressed and purified as MBP fusion proteins. Solubility of the fragments was dramatically increased and initial crystal trials were set up with concentrations varying from 5 mg/mL to 65 mg/mL. For the Pur- $\alpha$  56-287 C272S MBP fusion protein, crystalline material was obtained in 5 out of 384 conditions of the initial screen. Unfortunately, the crystals could not be reproduced in the refinement screen (Table 2-1).

Chemical modifications of surface residues change the surface properties of the protein of interest and can therefore constitute a rescue strategy for proteins that resist crystallization. For instance, methylation of lysine residues can reduce solubility and promote crystallization (Kobayashi et al., 1999). Reductive methylation of a fusion protein of human Pur- $\alpha$  56-245 and MBP and was carried out as described (Walter et al., 2006). Unfortunately, the methylated protein yielded no initial crystals, either (section 5.12 and Table 2-1).

The expression of Pur- $\alpha$  56-245 and Pur- $\alpha$  56-287 in a baculovirus-based insect cell expression system (section 5.7.2) did not improve the solubility of the fragments and no initial crystals were obtained (Table 2-1).

Since also the crystal structure of Pur- $\alpha$  in complex with nucleic acids was desired, filter binding assays were performed (section 2.2.1.1). The oligomers JCVupTar ssDNA, TAR14AU ssRNA and TAR14GC ssRNA (Table 5-2) showed good affinity to human Pur- $\alpha$  56-287 C272S (Table 2-12). The addition of the nucleic acid improved the solubility of the protein fragments and initial crystallization screens were performed at concentrations of 2-5 mg/mL. Pur- $\alpha$  56-245 mixed with JCVupTar ssDNA (2 mg/mL, 1:1.25) yielded crystals that turned out to be salt crystals when tested at a synchrotron beamline. No initial crystals were obtained for Pur- $\alpha$  56-287 C272S and TAR14AU (5 mg/mL, 1:1.1). Pur- $\alpha$  56-287 C272S and TAR14GC (2 mg/mL, 1:1.1) yielded initial crystals in 7 out of 380 conditions. In the refinement, crystals were reproducible only in one condition (Figure 2-1 and Table 2-1). The crystals were very small with the smallest dimension below 5  $\mu$ m. For improvement, the composition of the crystallization solution was varied (pH, concentration of salt and concentration of MPD). Furthermore, an additive screen with 48 conditions was performed (custom screen by D. Niessing). Unfortunately, no crystal improvement was achieved. Due to their small size, the crystals were not fished and tested for diffraction. It is thus well possible they were salt crystals.



**Figure 2-1. Crystals of human Pur- $\alpha$ . Pur- $\alpha$  56-287 C272S and TAR14GC RNA (1:1.1) were mixed with crystallization solution containing 100 mM sodium cacodylate pH 6.5, 100 mM sodium acetate, and 30 % MPD.**

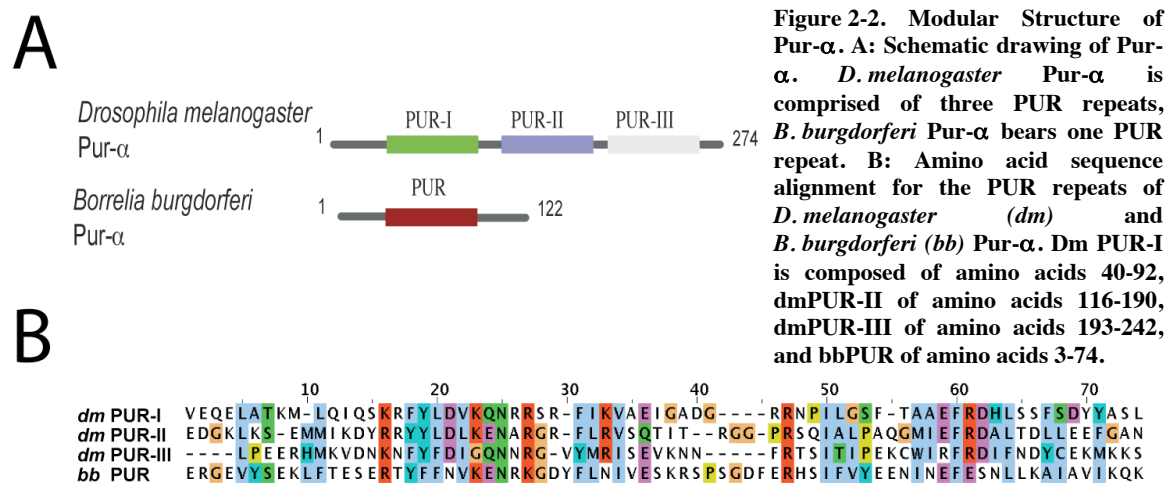
### **Identification of PUR repeats in Pur- $\alpha$**

Protein expression constructs based on the previously reported boundaries of the class I and class II repeats of Pur- $\alpha$  (Figure 1-3) either failed to yield soluble proteins or resulted in degradation of the protein fragments (Table 2-1). This indicated that these repeats might not represent folded domains, or at least no independent structural entities. It was thus questioned if the class I and class II repeats are accurately defined. To further investigate this issue, a cooperation with J. Söding (Gene Center, Munich, Germany) was initiated. J. Söding performed a computational search for repetitive elements in Pur- $\alpha$  using the HHrepID server (Biegert and Söding, 2008; Söding et al., 2005). HHrepID is very sensitive to distant homology, especially for repeat-proteins. It was found that human Pur- $\alpha$  contains three independent and consecutive repeat elements in its core region (Appendix 7-2). These repeats were termed PUR repeats. They are not identical with the previously described class I and class II repeats (Figure 1-3 and Appendix 7-2). PUR repeats are conserved throughout species and various metazoan Pur- $\alpha$  orthologs have three PUR repeats (Appendix 7-2). The sequence identity of the PUR repeats compared to each other is about 30 % (Table 2-5).



### 2.1.1.2 *Borrelia burgdorferi* Pur- $\alpha$

The newly identified PUR repeats were employed in sequence homology searches to identify Pur- $\alpha$  from a prokaryotic species. Interest in a bacterial homolog arose when crystallization trials of metazoan Pur- $\alpha$  constantly failed (section 2.1.1), as the use of an ortholog has proven successful in many cases (Kundrot, 2004). Savchenko and colleagues performed a high throughput crystallization screen and found that the usage of one more ortholog doubles the success rate of structural studies (Savchenko et al., 2003). Since the ortholog approach utilizes sequence alteration between species and thus constitutes a trial and error game, it is not predictable which organism to favour. However, homologs with shorter loop regions show the tendency to crystallize more willingly, as do bacterial proteins when compared to their metazoan counterparts.



**Figure 2-2. Modular Structure of Pur- $\alpha$ .** A: Schematic drawing of Pur- $\alpha$ . *D. melanogaster* Pur- $\alpha$  is comprised of three PUR repeats, *B. burgdorferi* Pur- $\alpha$  bears one PUR repeat. B: Amino acid sequence alignment for the PUR repeats of *D. melanogaster* (*dm*) and *B. burgdorferi* (*bb*) Pur- $\alpha$ . Dm PUR-I is composed of amino acids 40-92, dmPUR-II of amino acids 116-190, dmPUR-III of amino acids 193-242, and bbPUR of amino acids 3-74.

J. Söding performed homology database searches with the consensus PUR motif using the HHblast server (Söding et al., 2005). Homology database searches with the PUR motif resulted in the identification of a bacterial (spirochete) homolog, that is *B. burgdorferi* B31 gene bank entry BB0047 hypothetical protein. Due to its homology to the PUR motif, it was named *B. burgdorferi* Pur- $\alpha$ . The protein is comprised of 122 amino acids and its core region (amino acid 3-74) corresponds to a single PUR repeat (Figure 2-2). The homology region is referred to as *B. burgdorferi* PUR in the following. Its sequence identity is compared to the three PUR repeats of *D. melanogaster* Pur- $\alpha$ , because the latter was investigated later on. The sequence identity (similarity) of

*B. burgdorferi* PUR is 19 % (47 %), 21 % (38 %) and 13 % (38 %) to PUR repeat I, PUR repeat II, respective PUR repeat III, of *D. melanogaster* Pur- $\alpha$  (Figure 2-2).

In the GeneBank entry, BB0047 is reported to contain 127 amino acids, but a slightly different open reading frame was found when the nucleotide sequence was translated using the Expasy Translate Tool (Gasteiger et al., 2003). Translation started 15 nucleotides later, resulting in a 122 amino acid protein (14.1 kDa). To assess if the later start of translation is correct, the sequences of homologous proteins from other *Borrelia* species were compared. The protein sequences of *B. garnii*, *B. afzelii*, *B. valisiana*, and *B. spielmanii* gene products were highly homologous (more than 95 % identity) and none of them included the first start of translation. Thus, the later start of translation chosen as the correct start for *B. burgdorferi* Pur- $\alpha$  and all numbering refers to this definition.

## Crystallization of *B. burgdorferi* Pur- $\alpha$

Various truncation mutants of *B. burgdorferi* Pur- $\alpha$  were designed, expressed in *E. coli* and purified using a GST-tag (section 5.7 and section 5.8). The protein fragments of *B. burgdorferi* Pur- $\alpha$  showed a much better stability and solubility than the fragments of human Pur- $\alpha$  (Table 2-2).

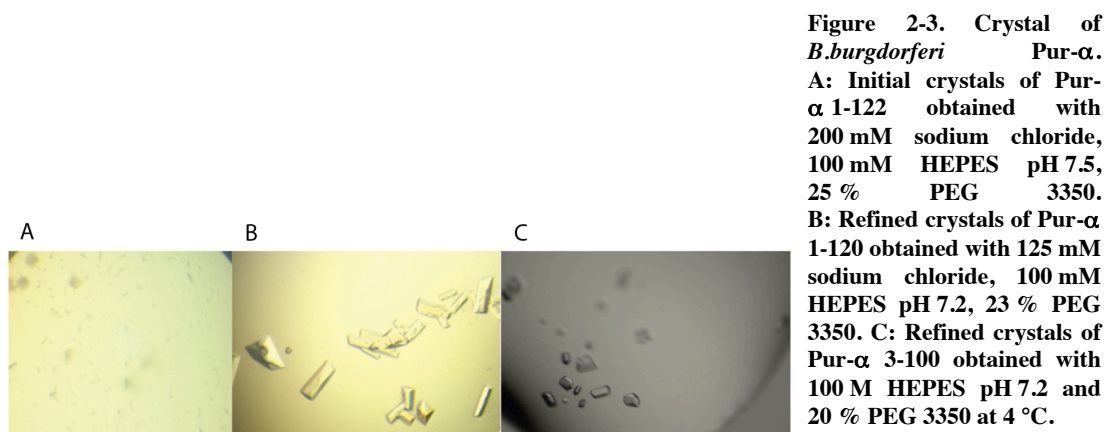
Construct	Tag	Expression	Purification/ Stability	Initial Crystals	Refined Crystals	Comment
bbPurA1-100	GST	++	++	+	+	
bbPurA1-120	GST	++	++	+	+	
bbPurA1-122	GST	++	++	+/-	-	a)
bbPurA3-100	GST	++	++	+	++	b)
bbPurA3-100 L12M,F22M,I59M	GST	+	+	+	++	c)
bbPurA3-105	GST	+	+	+	+	
bbPurA3-110	GST	+	+	+	+	
bbPurA3-120	GST	+	+	+	+	
bbPurA3-70	GST	+	-			
bbPurA3-74	GST	+	-			
bbPurA3-78	GST	+	-			
bbPurA3-82	GST	+	+	+		d)
bbPurA3-90	GST	+	+	+	+	e)
bbPurA3-95	GST	+	+	+	+	
bbPurA3-95 L12M,F22M,I59M	GST	+	+	+/-		
bbPurA5-100	GST	+	+	+	+	
bPurA5-95	GST	+	+	+	+	

**Table 2-2. Protein fragments of *B. burgdorferi* Pur- $\alpha$  for crystallization trials. bbPurA refers to *B. burgdorferi* Pur- $\alpha$ , the numbers indicate the start and stop site of the amino acid sequence. a) Full length protein, b) used for structure determination, also co-crystallization trials with DNA and heavy atom soaks, c) used for structure determination, also co-crystallization trials with DNA, d) also co-crystallization trials with DNA, e) also co-crystallization trials with DNA.**

The crystallization experiments were performed at room temperature, unless stated otherwise (section 5.10).

Full length *B. burgdorferi* Pur- $\alpha$  1-122 yielded initial crystals in 78 out of 480 crystallization conditions tested (Figure 2-3, A). The crystals could be refined to (40 x 40 x 20)  $\mu\text{m}$  size by varying the concentration of the protein and the crystallization solution composition. When the crystals were tested at DESY, no diffraction was

observed, possibly due to their small size. Therefore, it was aimed to increase their size. This was attempted by variation of the expression construct. *B. burgdorferi* Pur- $\alpha$  1-100 crystallized in 48 out of 384 conditions tested, Pur- $\alpha$  1-120 crystallized in 2 out of 192 conditions, and Pur- $\alpha$  3-120 crystallized in 25 out of 288 conditions. Refined crystals had a size of approximately (200 x 80 x 50)  $\mu\text{m}$  (Figure 2-3, B). The crystals were tested for

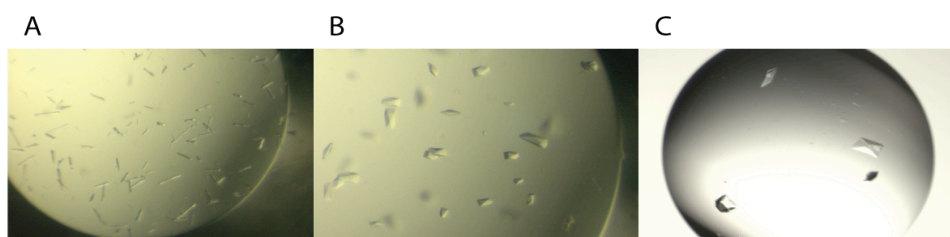


diffraction at a beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France) and diffracted up to 3-4 Å. Several cryo-protectants were tested, and PEG 400 and ethylene glycol were superior over MPD and glycerol (20 % in each case). Final refinement employed Pur- $\alpha$  3-100, which crystallized in 36 out of 193 conditions. Refinement of the crystallization conditions was also performed at 4 °C, which improved the diffraction quality of the crystals. Finally, *B. burgdorferi* Pur- $\alpha$  3-100 was crystallized at a concentration of 2.2 mg/ml at 4 °C. The crystallization solution contained 100 mM HEPES pH 7.2, and 20 % PEG 3350. After 2-4 days, cuboid crystals of approximately (200 x 100 x 30)  $\mu\text{m}$  size appeared, which diffracted up to 2.1 Å (section 5.10.1, Figure 2-3, C).

Selenomethionine-derivatized protein was produced as described (section 5.7.1). The fragment *B. burgdorferi* Pur- $\alpha$  3-100 does not contain any methionines, so they had to be introduced by site-directed mutagenesis. To reduce the probability of folding defects, the methionines were introduced in positions that bear methionines in Pur- $\alpha$  orthologs from other *Borrelia* species. Methionines were introduced in position leucine 12, phenylalanine 22, and isoleucine 59. Surprisingly, selenomethionine-derivatized

*B. burgdorferi* Pur- $\alpha$  3-100 L12M, F22M, I59M did not crystallize in the same conditions as the native protein. After refinement of the composition of the crystallization solution, only needle shaped crystals of approximately (50 x 5 x 5)  $\mu\text{m}$  size were obtained (Figure 2-4, A). In initial crystallization screens, the protein crystallized in 9 out of 193 conditions tested.

Because a co-structure with DNA was also desired, screening was also performed with the addition of DNA oligonucleotides. In filter binding assays, *B. burgdorferi* Pur- $\alpha$  showed affinity for the hTel12 ssDNA oligonucleotide (Table 5-2 and Table 2-13). Thus, initial crystallization trials were performed with Pur- $\alpha$  3-100 L12M, F22M, I59M and hTel12 ssDNA (1:1). Crystals were found in 36 out of 193 conditions (Figure 2-4, B). Finally, selenomethionine-derivatized *B. burgdorferi* Pur- $\alpha$  3-100 L12M, F22M, I59M was mixed with hTel12 DNA (1:1) and crystallized at a concentration of 1.2 mg/mL in 2.8 M sodium formate at 4 °C. Spindle-shaped and cuboid crystals of (100 x 50 x 50)  $\mu\text{m}$  appeared after 2-5 days and diffracted up to 1.9 Å (Figure 2-4, C and section 5.10.1).



**Figure 2-4. Crystals of selenomethionine-derivatized *B. burgdorferi* Pur- $\alpha$ .**  
**A:** Refined crystals of Pur- $\alpha$  3-100 L12M, F22M, I59M obtained with 200 mM sodium chloride, 100 mM HEPES pH 7.5, 25 % PEG 3350. **B:** Initial crystals of Pur- $\alpha$  3-100 L12M, F22M, I59M with hTel12 DNA (1:1) obtained with 3.5 M sodium formate pH 7.0. **C:** Refined crystals of Pur- $\alpha$  3-100 L12M, F22M, I59M with hTel12 DNA (1:1) obtained with 3.8 M sodium formate pH 7.0 at 4 °C.

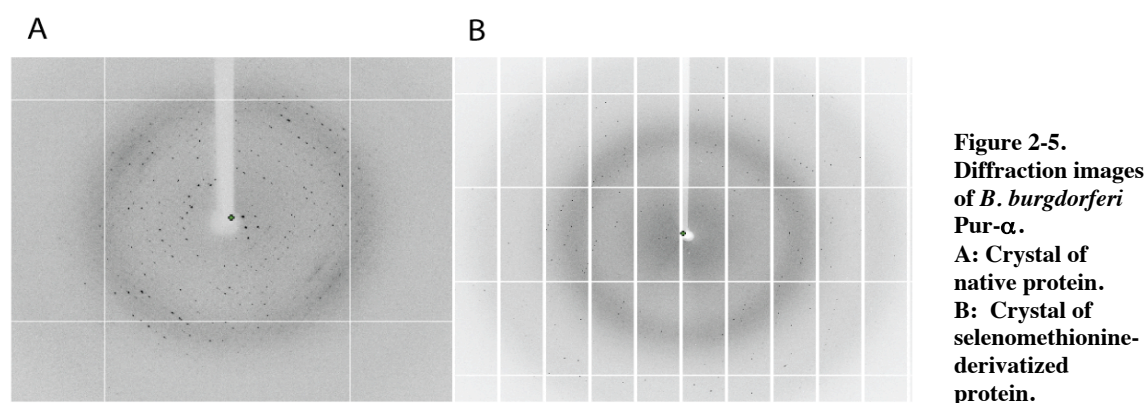
## Structure determination and refinement

Synchrotron radiation diffraction data were collected ESRF and Swiss Light Source (SLS, Villingen, Switzerland).

The crystals of native protein diffracted up to 2.2 Å and belong to spacegroup  $P2_12_12_1$  (Table 2-3 and Figure 2-5). Cell constants are  $a = 47.79$  Å,  $b = 57.75$  Å,  $c = 142.31$  Å and angles  $\alpha = \beta = \gamma = 90.00^\circ$ . A native dataset was recorded at beamline ID 23-1 (ESRF) with a wavelength of 0.9724 Å. A Q315r ADSC CCD detector was used to collect 360 frames with an oscillation angle of  $0.5^\circ$ .

For phasing, it was initially aimed at obtaining heavy-atom derivatized protein crystals by soaking. Different mercury and platinum derivates were added to the crystallization drop and incubated (section 5.13). The crystals were tested at the ESRF, and it was found that the diffraction quality was decreased. Most crystals showed total loss of diffraction, while some produced smeared spots with a diminished resolution (4.5 Å). Therefore, the selenomethione-approach was followed.

Selenomethionine-derivatized *B. burgdorferi* Pur- $\alpha$  crystallized in spacegroup  $I2_12_12_1$  with cell constants  $a = 48.17$  Å,  $b = 58.33$  Å,  $c = 141.84$  Å and angles  $\alpha = \beta = \gamma = 90.00^\circ$ . Crystals diffracted up to 1.9 Å (Table 2-3 and Figure 2-5).



An anomalous data set at the selenium K-edge was collected at the X06SA/PXI beamline (SLS). Using a PILATUS 6M detector, 180 frames with an oscillation angle of  $1^\circ$  per image were recorded. Prior to data collection, a fluorescence scan determined the optimal wavelengths. The peak wavelength was 0.9792 Å (12.6617 keV,  $f' = -7.276$ ,  $f'' = 6.838$ ) and the inflection point was 0.9797 Å (12.6556 keV,  $f' = -10.562$ ,  $f'' = 3.406$ ).

Dataset	Native	SeMet Peak
X-ray source	ID23-1 (ESRF)	X06SA/PXI (SLS)
Wavelength in Å	0.9724	0.9792
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	I2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Cell dimensions		
a, b, c in Å	47.79, 57.75 Å, 142.31	48.17, 58.33 Å, 141.84
α, β, γ	90.00°, 90.00°, 90.00°	90.00°, 90.00°, 90.00°
Datarange in Å	70.9-2.2	70.9-1.9
I/σI	14.99 (2.48)	12.19 (5.27)
Observations	78, 627	105, 665
Unique observations	20, 282	30, 605
Redundancy	3.9	3.5
Completeness in %	99.0 (97.6)	98.7 (95.5)
Rsym in %	7.4 (52.2)	8.2 (35.6)

**Table 2-3. Data collection for the crystal structure of *B. burgdorferi* Pur-α. SeMet refers to the selenomethionine-derivatized crystal, R<sub>sym</sub> refers to the unweighted R-value on I between symmetry mates. Numbers in parantheses indicate values for the highest resolution shell.**

Phase angles were determined by single wavelength anomalous dispersion (SAD) using the peak wavelength. The structural model was also build from this dataset. Refinement was performed with REFMAC (Murshudov et al., 1997) using non-crystallographic symmetry (NCS) with two molecules per asymmetric unit. Translation, Liberation, and Screw rotation (TLS) displacement tensors (Winn et al., 2001) were used and allowed to model anisotropic displacement of the atoms. Refinement resulted in final R<sub>work</sub> = 18.7 % and R<sub>free</sub> = 25.0 % (Table 2-4). The comparatively high difference of R<sub>work</sub> and R<sub>free</sub> can at least in part be explained with radiation damage of the selenium atoms. The structural model contains amino acids 3 to 79 as well as 161 water molecules. Although the DNA oligonucleotide hTel12 (Table 5-2) was present in the crystallization conditions (section 5.10.1), the DNA was not visible in the crystal structure. Bond lengths and angles are in good accordance with protein geometry as root mean square deviations (RMSD) from ideal values are very small (RMSD for bond length 0.012 Å, RMSD for bond angles 1.275 °). The Ramachandran plot visualizes possible conformations for the dihedral angle of a peptide bond in protein structures. For the structural model, 98 % of the amino acids lie in favored regions of the Ramachandran plot (Table 2-4). The B-

factor gives a measure for the effective diameter of an atom's electron density and reflects the fluctuation of the atom around its average position. The low B-factor (22.1 Å<sup>2</sup>) of the crystal structure corresponds to a high degree of order (Table 2-4).

Dataset	SeMet Peak
Data range in Å	70.9-1.9
Reflections	21, 157
R <sub>work</sub> in %	18.5
R <sub>free</sub> in %	23.0
RMSD bond length in Å	0.010
RMSD bond angles in deg	1.169
Ramachandran plot in %	
Favored	98
Allowed	2
Outlier	0
Average B-factor in Å <sup>2</sup>	23.5

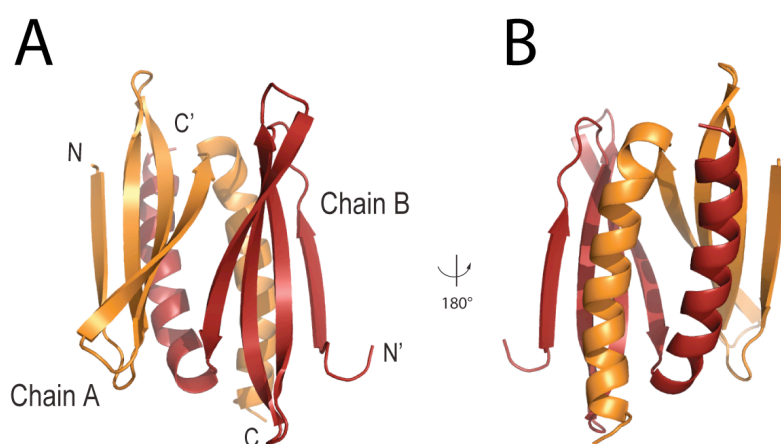
**Table 2-4. Refinement statistics for the crystal structure of *B. burgdorferi* Pur- $\alpha$ . RMSD refers to root mean square deviation.**

### Crystal structure of *B. burgdorferi* Pur- $\alpha$

The structural model of *B. burgdorferi* Pur- $\alpha$  revealed that it crystallized as a dimer. Each monomer is comprised of a 4-stranded anti-parallel  $\beta$ -sheet followed by an  $\alpha$ -helix, as shown in the ribbon backbone model (Figure 2-6). One peptide chain intertwines with the other chain, resulting in a buried surface interface of 2058 Å<sup>2</sup> per monomer (section 3.1.3).

The part of the crystallized protein that is visible in the structural model (amino acids 3 to 79) nicely matches the homology region that was characterized as *Borellia* PUR (amino acids 3 to 74) (Figure 2-2), thus indeed one PUR repeat corresponds to a structural entity. The interaction of two monomers results in a globular domain that is referred to as PUR domain. The PUR domain exhibits one side showing both  $\alpha$ -helices ( $\alpha$ -helical side) and one side showing both  $\beta$ -sheets ( $\beta$ -sheet side) (Figure 2-6).





**Figure 2-6. Crystal structure of *B. burgdorferi* Pur- $\alpha$ .**

**A:** Ribbon backbone model of the PUR domain formed by two monomers,  $\beta$ -sheet side. One monomer is depicted in red, the other in gold. **B:** View from (A), rotated by  $180^\circ$  around the vertical axis,  $\alpha$ -helical side.

The dimerization of *B. burgdorferi* Pur- $\alpha$  is mostly stabilized by hydrophobic interactions, as the assessment of the buried surface interface reveals a large number of aliphatic and aromatic residues. Hydrophobic amino acids on the inward oriented side of the  $\alpha$ -helices include phenylalanine, leucine, alanine, isoleucines and valine (F62, L66, A69, I70, I73, V72). They are complemented by hydrophobic residues on the inner side of the  $\beta$ -sheets, including valines, tyrosines, phenylalanine, leucine, and isoleucines (V7, V24, V54, Y8, Y20, F22, L34, I36, I52). Phenylalanine 62 and tyrosine 20 may additionally contribute to stability by  $\pi$ -stacking interaction. Putative electrostatic interactions between the two chains are rare, as charged residues are generally oriented towards the solvent. At the interface between the  $\alpha$ -helices, favourable electrostatic interaction may form between asparagine 60/lysine 74 and glutamate 63/lysine 74. At the edge between the  $\alpha$ -helix and the adjacent  $\beta$ -sheet, hydrogen bonds may form between glutamine 10/asparagine 65 and tyrosine 55/lysine 68.

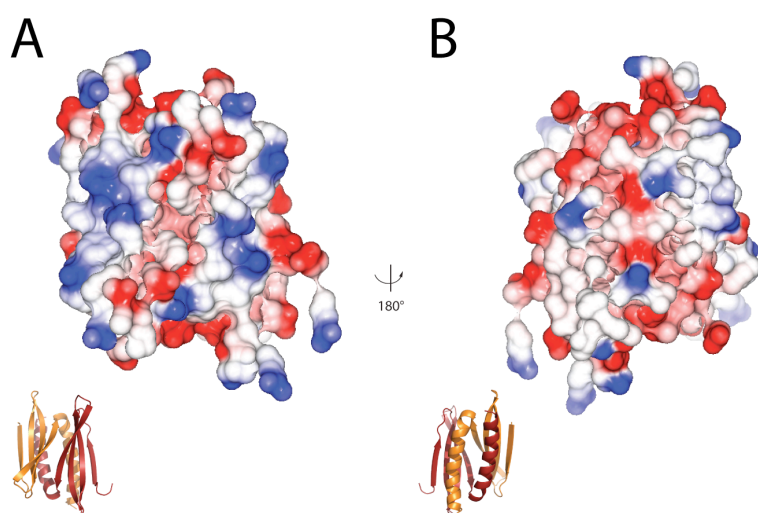
The submission of the *B. burgdorferi* Pur- $\alpha$  crystal structure to the RSCB protein data bank (PDB) is in preparation.

## Surface assessment

A plot of the electrostatic potential on the surface of the crystal structure of *B. burgdorferi* Pur- $\alpha$  reveals differences concerning both sides of the PUR domain (Figure 2-7).

The  $\alpha$ -helical side of the PUR domain is rich in negative charge mainly due to glutamic acid residues (E57, E56, E61, E63) and bears only a few positively charged residues, like lysines (K68, K74). In contrary, the  $\beta$ -sheets of the PUR domain have a more pronounced positive surface charge, mainly due to arginine and lysine residues (R4, K11, R18, K25, R28, R29, R41, R49) and exhibit fewer negatively charged residues like aspartic acid (D31, D46) (Figure 2-7).

This suggests the  $\beta$ -sheet side is a potential nucleic-acid binding surface, as nucleic-acid binding sites are typically enriched in positively charged residues and depleted in negatively charged residues (Bahadur et al., 2008).



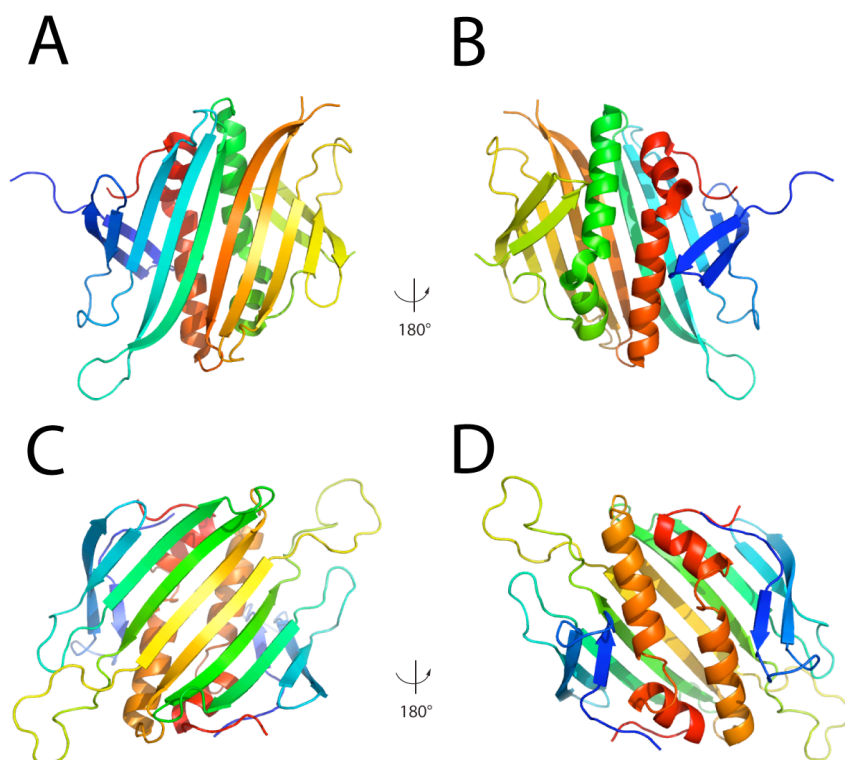
**Figure 2-7.** Electrostatic surface potential of *B. burgdorferi* Pur- $\alpha$ . **A:** Surface representation of the solvent-accessible surface of the PUR domain,  $\beta$ -sheet side. Dark blue coloration indicates positive charge, red indicates negative charge and white shows uncharged regions. **B:**  $\alpha$ -helical side, view from (A), rotated by 180° around the vertical axis. Orientations are as given in the ribbon backbone model on the bottom.

## Systematic structural comparison

A systematic structural comparison using DaliLite (Holm, 2008) unveiled moderate similarity to a number of crystal structures, two of them from bacteriophage coat proteins, which are dimeric RNA-binding proteins.

The highest score was obtained for bacteriophage coat protein from *Pseudomonas* phage pp7 (PDB-ID: 2QUD; Z-score: 5.2; RMSD: 4.8 for 68  $\alpha$ -carbon pairs) (Chao et al., 2008) (Figure 2-8). The other bacteriophage coat protein was from *Enterobacterio* phage ms2 (PDB-ID: 1MSC; Z-score: 4.6; RMSD: 4.8 for 67  $\alpha$ -carbon pairs) (Ni et al., 1995) (Figure 2-8).

The typical characteristics of bacteriophage coat proteins are a N-terminal  $\beta$ -hairpin, a five-stranded antiparallel  $\beta$ -sheet and two C-terminal  $\alpha$ -helices. To form anti-parallel dimers, the  $\alpha$ -helices of two protein chains intertwine into each other. On the opposing side, the dimer exhibits an extended  $\beta$ -sheet. The flat  $\beta$ -sheet is comprised of 10  $\beta$ -strands from both monomers and is reported to be the RNA binding surface for both bacteriophage coat proteins (Chao et al., 2008; Ni et al., 1995) (PDB-ID: 2QUX).



**Figure 2-8.** Structures from the PDB with similarity to *B. burgdorferi* Pur- $\alpha$ . A: Coat protein from *Pseudomonas* phage pp7 (PDB-ID: 2QUD). B: View from (A) rotated by 180° around the vertical axis. C: Coat protein from enterobacterio phage ms2 (PDB-ID: 1MSC). D: View from (C) rotated by 180° around the vertical axis.

The structural model of *B. burgdorferi* Pur- $\alpha$  lacks the N-terminal  $\beta$ -hairpin, has only a four-stranded  $\beta$ -sheet and only one  $\alpha$ -helix. However, the dimerization mode is similar: the  $\alpha$ -helices intertwine and the  $\beta$ -sheets locate at the opposing side. Another pronounced difference is that in *B. burgdorferi* Pur- $\alpha$ , the  $\beta$ -sheets are positioned in a steep angle with respect to each other, resulting in a roof-like topology (Figure 2-6). In the bacteriophage coat proteins, the  $\beta$ -sheets form flat surface (Figure 2-8).

### 2.1.1.3 *Drosophila melanogaster* Pur- $\alpha$

Pur- $\alpha$  is the only member of the Pur family found in *Drosophila* (GeneID 43797). The protein is comprised of 274 amino acids (31.0 kDa). Sequence identity to human Pur- $\alpha$  is 49 %, and sequence similarity is 66 %.

The self-alignment of *D. melanogaster* Pur- $\alpha$  yielded the positions of its three PUR repeats (section 2.1.1.1) (Figure 2-2). PUR repeat I is comprised of amino acids 40-92, PUR repeat II of amino acids 116-190, and PUR repeat III of amino acids 193-242. The sequence identity of repeat I and II is 33 %, whereas the sequence similarity is 61 % (Table 2-5). Repeat III is slightly more divergent from repeat I and repeat II, with an identity of 25 % to repeat I (similarity 53 %) and an identity of 32 % (similarity 56 %) to repeat II (Table 2-5).

Identity/Similarity in %	PUR Repeat I	PUR Repeat II	PUR Repeat III
PUR Repeat I	-	33 / 61	25 / 53
PUR Repeat II	33 / 61	-	32 / 56
PUR Repeat III	24 / 53	34 / 56	-

Table 2-5. Sequence conservation between the three PUR repeats of *D. melanogaster* Pur- $\alpha$ .

### Crystallization of *D. melanogaster* Pur- $\alpha$ repeats I-II

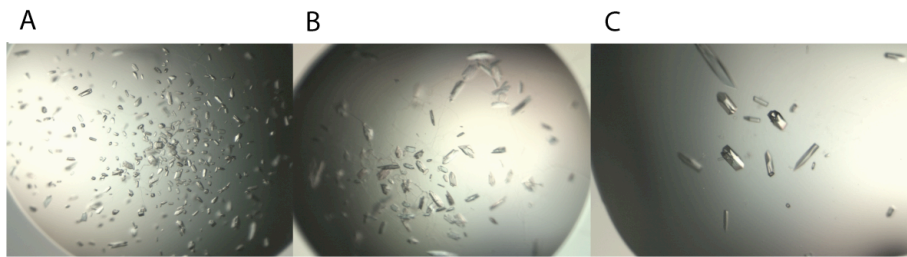
In order to guide the design of *D. melanogaster* Pur- $\alpha$  expression constructs, the *B. burgdorferi* crystal structure was used to build a homology model of *D. melanogaster* Pur- $\alpha$ . The homology model of *D. melanogaster* Pur- $\alpha$  was done by J. Söding using the server HHpred (Söding et al., 2005) (Appendix 7-1). The computational model successfully predicted folded domains, as the protein fragments derived from it were stable and crystallizable (Table 2-6). The protein fragments were expressed in *E. coli* and purified using a GST-tag (section 5.7.1 and section 5.8.1).

Construct	Tag	Expression	Purification/ Stability	Initial Crystals	Refined Crystals	Comment
dmPurA1-274	GST	+	+			a)
dmPurA115-185	GST	+	-			
dmPurA181-260	GST	+	+	+	+	
dmPurA185-260	GST	+	+	+	+	
dmPurA188-258	GST	+	+	+	+	
dmPurA36-185	GST	+	-			
dmPurA40-105	GST	-				
dmPurA40-185	GST	++	++	++	++	b)
dmPurA40-245	GST	+	-			
dmPurA40-255	GST	+	+	-		
dmPurA40-274	GST	+	+/-			
dmPurA46-185	GST	+	-			

**Table 2-6. Protein fragments of *D. melanogaster* Pur- $\alpha$  for crystallization trials. dmPurA refers to *D. melanogaster* Pur- $\alpha$ , the numbers indicate the start and stop site of the amino acid sequence. a) Full length protein, b) used for structure determination and trials of heavy atom soaking, DNA soaking and co-crystallization with DNA.**

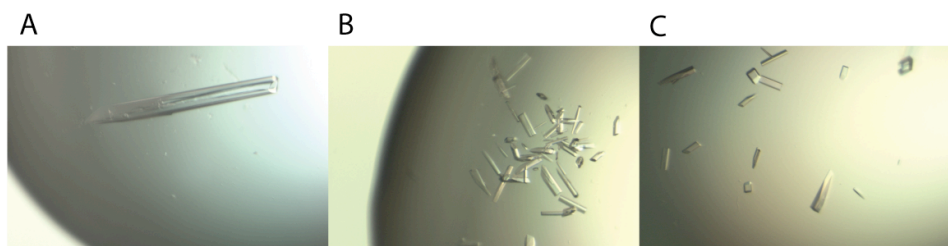
The crystallization experiments were performed at room temperature, unless stated otherwise (section 5.10). A fragment of *D. melanogaster* Pur- $\alpha$  40-255, which comprises all three PUR repeats, did not yield any initial crystals in 384 conditions tested.

Then crystallization of a fragment only containing PUR repeat I and PUR repeat II was tried. Pur- $\alpha$  40-185 crystallized in 17 out of 480 conditions at 21 °C and in 5 out of 480 conditions at 4 °C. Several conditions were refined and screened for single three-dimensional crystals. Best crystals were found in solutions containing MES buffer and PEG 3350. It was found that the crystal size was dependent on the PEG content, as bigger crystals were grown at lower PEG concentration (Figure 2-9). Finally, *D. melanogaster* Pur- $\alpha$  40-185 was crystallized at a concentration of 3-5 mg/ml at 21 °C. The crystallization solution contained 50 mM MES pH 5.9, 200 mM magnesium chloride, and 25 % PEG 3350. Cuboid crystals of approximately (100 x 50 x 50)  $\mu$ m appeared within 2-4 days and diffracted up to 2.1 Å.



**Figure 2-9.** Crystals of *D. melanogaster* Pur- $\alpha$  repeats I-II. Pur- $\alpha$  40-185 was crystallized in 50 mM MES pH 5.9, 200 mM magnesium chloride and 35 % (A), 30 % (B), respectively 25 % (C) PEG 3350.

Selenomethionine-derivatized protein was produced and purified as described (section 5.7.1 and 5.8.1). The crystallized fragment of *D. melanogaster* Pur- $\alpha$  40-185 contains three methionines, so the mutational introduction of methionines was not necessary. The selenomethionine-derivatized protein crystallized in the same conditions as the native protein, however, crystals grew very big with the largest dimension up to 600  $\mu\text{m}$ . The crystals broke or cracked during transfer and cryo-protection. Thus crystallization solution and protein concentration had to be refined. Finally, crystals of selenomethionine-derivatized *D. melanogaster* Pur- $\alpha$  40-185 were grown at a concentration of 2.7 mg/mL in buffer containing 200 mM magnesium chloride, 100 mM HEPES pH 7.8, and 22 % PEG 3350. They had a size of approximately (150 x 50 x 50)  $\mu\text{m}$  and diffracted up to 2.1  $\text{\AA}$ .



**Figure 2-10.** Crystal of selenomethionine-derivatized *D. melanogaster* Pur- $\alpha$  repeats I-II. Initial crystals of Pur- $\alpha$  40-185 obtained with 200 mM magnesium chloride, 100 mM MES pH 6.8, and 22 % PEG 3350. B: Refined crystals of Pur- $\alpha$  40-185 obtained with 200 mM magnesium chloride, 100 mM HEPES pH 7.2, and 22 % PEG 3350. C: Refined crystals of Pur- $\alpha$  40-185 obtained with 200 mM magnesium chloride, 100 mM HEPES pH 7.8, and 22 % PEG 3350.

## Structure determination and refinement

Synchrotron radiation diffraction data were collected at ESRF (Figure 2-11). The native crystals of *D. melanogaster* Pur- $\alpha$  repeats I-II belong to spacegroup P2 with cell constants  $a = 63.17 \text{ \AA}$ ,  $b = 62.54 \text{ \AA}$ ,  $c = 64.48 \text{ \AA}$  and angles  $\alpha = \gamma = 90.00^\circ$ ,  $\beta = 90.01^\circ$ . A native dataset was recorded at beamline ID14-1 (ESRF) with a wavelength of  $0.934 \text{ \AA}$  (Table 2-7). A Q210 CCD ADSC detector was used to collect 600 frames with an oscillation angle of  $0.5^\circ$  each.

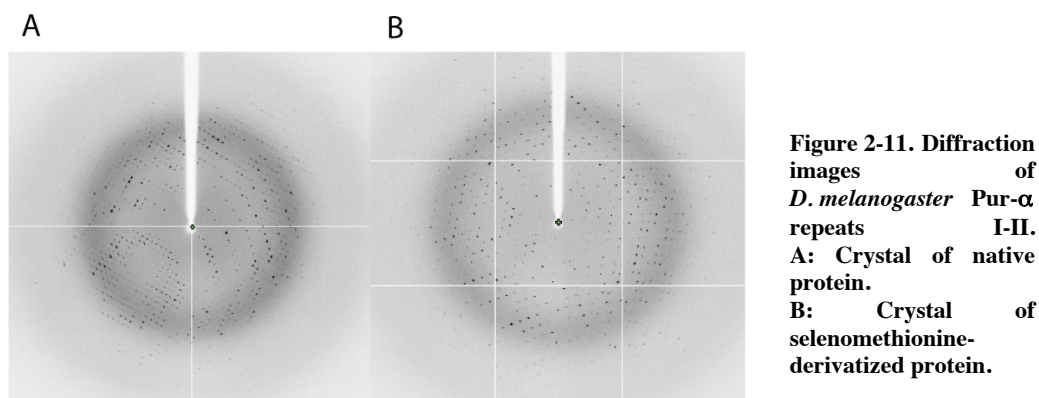
Initially, molecular replacement was tried using the program PHASER (McCoy et al., 2007) with the structural model of *B. burgdorferi* Pur- $\alpha$  as a search ensemble. Unfortunately, no convincing solution was found.

Thus, selenomethionine-derivatized protein crystals were produced. An anomalous data set was recorded at beamline ID14-4 (ESRF) (Table 2-7). The crystals diffracted up to  $2.1 \text{ \AA}$  and are isomorphous to the crystals obtained with native protein (Figure 2-11). The beamline was equipped with a Q315r ADSC CCD detector. From one crystal, datasets at three different wavelengths (peak, inflection, and remote) were recorded (Table 2-7). For the peak wavelength, 720 frames with an oscillation angle of  $0.5^\circ$  per image were collected. Since the crystal showed signs of radiation damage, for the inflection and remote wavelengths, the number was reduced to 360 frames (oscillation angle  $0.5^\circ$  per image). Prior to data collection, a fluorescence scan determined the optimal wavelengths. Peak wavelength was  $0.9789 \text{ \AA}$  ( $12.6656 \text{ keV}$ ,  $f' = -7.95$ ,  $f'' = 6.20$ ), inflection point was  $0.9791 \text{ \AA}$  ( $12.662679 \text{ keV}$ ,  $f' = -10.33$ ,  $f'' = 2.39$ ), and for high remote  $0.9567 \text{ \AA}$  ( $12.9602 \text{ keV}$ ) was chosen.



Dataset	Native	SeMet Peak	SeMet Inflection	SeMet Remote
X-ray source	ID14-1 (ESRF)	ID14-4 (ESRF)	ID14-4	ID14-4
Wavelength in Å	0.934	0.9789	0.9791	0.9567
Spacegroup	P2	P2		
Cell dimensions				
a, b, c in Å	63.17, 62.54, 64.48	63.17, 62.54, 64.48		
$\alpha, \beta, \gamma$	90.00°, 90.00°, 90.01°	90.00°, 90.00°, 90.04°		
Data range in Å	64-2.1 (2.23-2.1)	64-2.1 (2.23-2.1)	64-2.1 (2.23-2.1)	64-2.1 (2.23-2.1)
I/ $\sigma$ I	16.00 (4.35)	16.15 (5.28)	18.04 (6.43)	13.45 (3.25)
Observations	174, 369	212, 070	104, 025	112, 661
Unique observations	28, 601	56, 933	54, 477	58, 851
Redundancy	6.1	3.7	1.9	1.9
Completeness in %	98.9 (94.6)	96.0 (80.5)	91.9 (75.5)	92.1 (75.8)
R <sub>sym</sub> in %	7.0 (40.3)	5.4 (20.8)	2.9 (10.7)	3.7 (26.6)

**Table 2-7. Data collection for the crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II. SeMet refers to the selenomethionine-derivatized crystal, R<sub>sym</sub> refers to the unweighted R-value on I between symmetry mates. Numbers in parentheses indicate values for the highest resolution shell.**



**Figure 2-11. Diffraction images of *D. melanogaster* Pur- $\alpha$  repeats I-II. A: Crystal of native protein. B: Crystal of selenomethionine-derivatized protein.**

Phases were determined by multiple wavelength anomalous dispersion (MAD) and extended to 2.1 Å. The structural model was built from the native dataset. Refinement was done with REFMAC (Murshudov et al., 1997) using NCS and TLS. Final R-factors are R<sub>work</sub> = 22.3 % and R<sub>free</sub> = 24.0 %. Four copies of the protein fragment were found in the asymmetric unit. The structural model consists of amino acids 40 to 184, eight

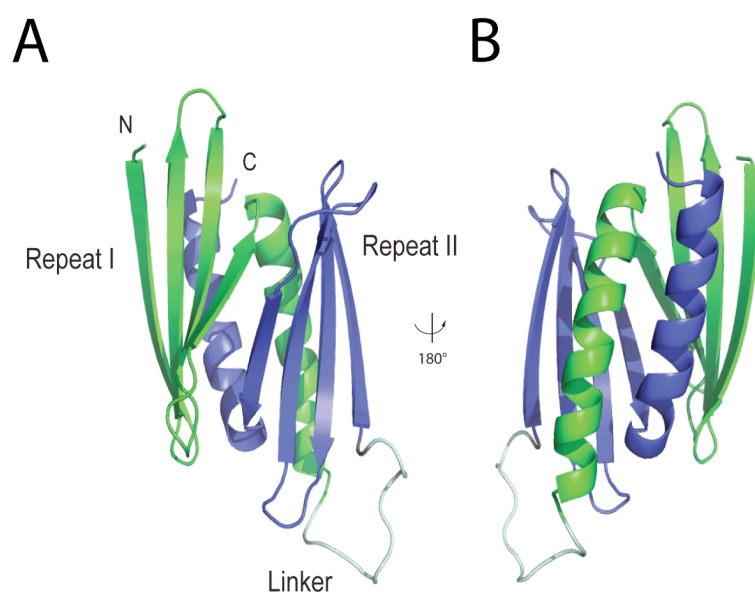
chloride ions, three ethylene glycol molecules (from the cryo-protectant), and 117 water molecules. In the Ramachandran plot, 95 % of the residues are in the allowed regions and bond lengths and bond angles comply with protein geometry (RMSD 0.016 Å, and 1.77°, respectively) (Table 2-8).

<b>Dataset</b>	<b>Native</b>
Data range in Å	64.0-2.1
Reflections	27, 900
Rwork	22.3
Rfree	24.0
RMSD bond length in Å	0.016
RMSD bond angles in deg	1.77
Ramachandran plot in %	
Favored	95
Allowed	4.5
Outlier	0.5
Average B factor in Å <sup>2</sup>	36.6

**Table 2-8. Refinement statistics for the crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II.**

### Crystal structure of *D. melanogaster* Pur- $\alpha$ repeats I-II

The structural model of *D. melanogaster* Pur- $\alpha$  repeats I-II shows that both PUR repeats adopt the same fold. Each repeat is comprised of a four-stranded anti-parallel  $\beta$ -sheet, followed by a C-terminal  $\alpha$ -helix (Figure 2-12). A short linker region connects both repeats, which intertwine into each other to form an intra-molecular dimer. Therein, the  $\alpha$ -helix from one repeat intercalates into the other repeat. This results in a large hydrophobic buried surface interface of approximately 33 % of each repeat's surface (1830  $\text{\AA}^2$  per repeat). The globular domain formed by the intra-molecular dimer corresponds to one PUR domain, which was already described in this study for *B. burgdorferi* Pur- $\alpha$  (section 2.1.1.2). The PUR domain exhibits one  $\beta$ -sheet side (Figure 2-12, A) and one  $\alpha$ -helical side (Figure 2-12, B).



**Figure 2-12.** Crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II, PDB-ID 3K44. **A:** Ribbon backbone model of the globular PUR domain formed by two PUR repeats. The first repeat is depicted in green, the second in blue. The linker is shown in grey. **B:** View from (A), rotated by 180° around the vertical axis.

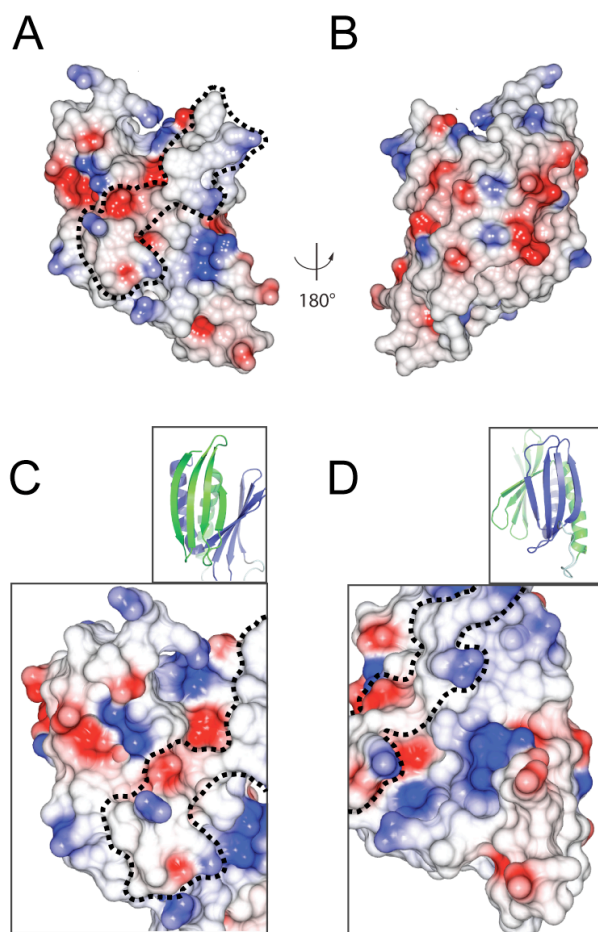
In the crystal structure, the PUR domain is formed by intra-molecular dimerization of PUR repeats I and II. Their attachment seems to be mainly stabilized by hydrophobic interactions. The inward oriented face of the  $\alpha$ -helices is rich in aromatic and aliphatic side chains such as phenylalanines, leucines, and tyrosin (F93, F100, F171, F182, L97, L175, L178, Y107). They pack against the hydrophobic inside of the  $\beta$ -sheets, which

includes leucines, phenylalanines, valines, isoleucines, and tyrosines (L44, L49, L58, L121, L135, L137, L146, L163, F56, V60, V71, V148, I69, I83, I161, Y133). A possible  $\pi$ -stacking interaction is seen between phenylalanines 97 and 171, which are located at the interface of the  $\alpha$ -helices. Charged residues are generally oriented towards the solvent and seem to play a minor role in stabilizing the PUR domain. However, some potential hydrogen bonds are located at the solvent exposed interfaces of the secondary structure elements. The interaction between both  $\alpha$ -helices is stabilized by hydrogen-bonds between serine 98 and arginine 172 and between arginine 94 and threonine 176. At the interface between the  $\alpha$ -helices and the  $\beta$ -sheets, hydrogen bond form between: lysine 47/asparagine 177, lysine 47/glutamate 181, glutamine 52/glutamine 166, histidin 96/glutamate 124, and serine 99/lysine 122.

The crystal structure of *D. melanogaster* Pur- $\alpha$  40-185 has been submitted to the RSCB protein data bank, PDB-ID 3K44.

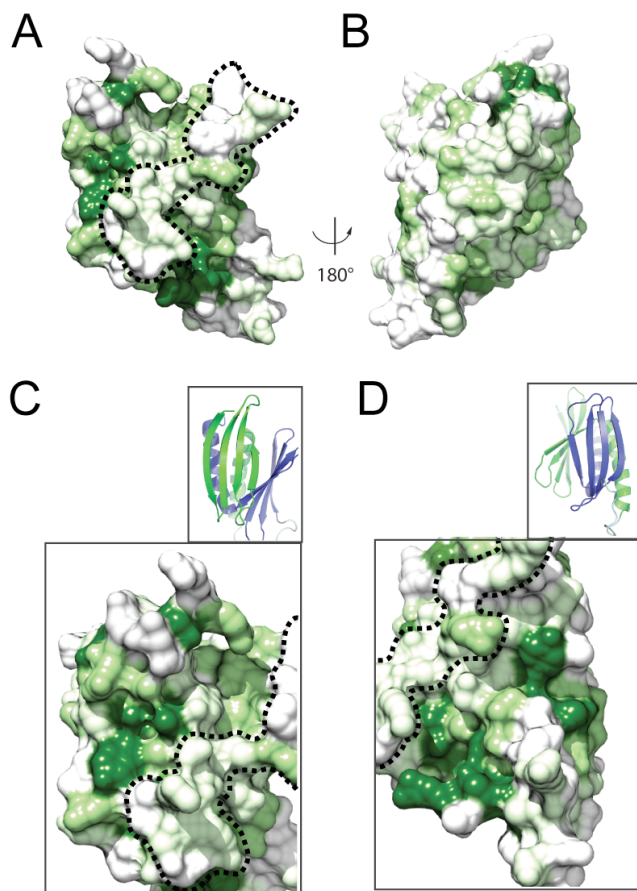
### **Surface assessment**

The electrostatic surface potential of the crystal structure of *D. melanogaster* Pur- $\alpha$  repeat I and II reveals a different charge distribution for the  $\alpha$ -helical and the  $\beta$ -sheet side of the PUR domain (Figure 2-13). The  $\alpha$ -helical side shows moderate negative charge. In contrast, a number of positively charged residues is located at the opposing  $\beta$ -sheet side. The  $\beta$ -sheet formed by PUR repeat II (Figure 2-13, D) appears to bear more pronounced positive charge (K121, R132, K138, R142, R147, R155, R158) than the  $\beta$ -sheet formed by PUR repeat I (R55, K61, R65, K70, R80) (Figure 2-13, C). The clustering of positive charge on the  $\beta$ -sheets suggests that they are potential nucleic-acid binding surfaces (Bahadur et al., 2008).



**Figure 2-13.** Electrostatic surface potential of *D. melanogaster* Pur- $\alpha$  repeats I-II. **A:** Surface representation of the solvent-accessible surface of the PUR domain,  $\beta$ -sheet side. Dark blue coloration indicates positive charge, red indicates negative charge and white shows uncharged regions. The dashed line indicates the unconserved rim (Figure 2-12). **B:** View from (A), rotated by 180° around the vertical axis,  $\alpha$ -helical side. **C:** Close-up of the  $\beta$ -sheet of PUR repeat I. **D:** Close-up of the  $\beta$ -sheet of PUR repeat II. Orientations are as given in the ribbon backbone model above.

Both sides of the PUR domain also reveal great differences in sequence conservation on the surface. The amino acid sequences of Pur- $\alpha$  from *D. melanogaster*, human, *Caenorhabditis elegans*, and *Arabidopsis thaliana* were aligned and plotted on the surface of the crystal structure (Appendix 7-2 and Figure 2-14). Whereas the  $\alpha$ -helical side is almost devoid of highly conserved residues, the opposing  $\beta$ -sheet side reveals considerable sequence conservation. Both conserved  $\beta$ -sheets are clearly separated by an un-conserved rim (dashed line in Figure 2-14). This suggests the two separated conserved regions might act independently from each other.

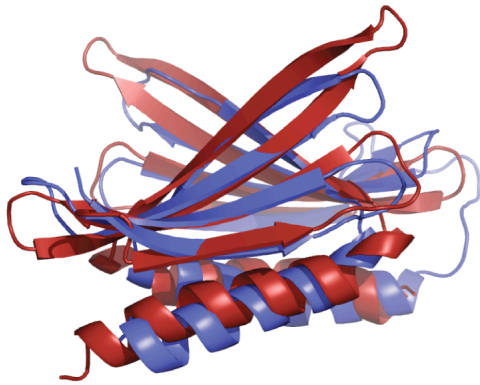


**Figure 2-14:** Surface conservation of *D. melanogaster* Pur- $\alpha$  repeats I-II. **A:** Surface representation of sequence conservation at the solvent-accessible surface,  $\beta$ -sheet side. Dark green coloration indicates complete conservation, light green indicates partial conservation, white shows unconserved regions. **B:** View from (A), rotated by  $180^\circ$  around the vertical axis,  $\alpha$ -helical side. **C:** Close-up of the  $\beta$ -sheet of PUR repeat I. **D:** Close-up of the  $\beta$ -sheet of PUR repeat II. Orientations are as given in the ribbon backbone model above. The sequence alignment is given in (Appendix 7-2).

Taken together, electrostatic potential and surface conservation hint at the  $\beta$ -sheet side for functional importance and nucleic-acid binding.

### Comparison with the crystal structure of *B. burgdorferi* Pur- $\alpha$

The crystal structures of *B. burgdorferi* Pur- $\alpha$  and *D. melanogaster* Pur- $\alpha$  repeats I-II show a pronounced structural conservation. Both proteins have the same  $\beta\beta\beta\alpha$ -topology and superpose nicely (Figure 2-15). Pairwise comparison with DaliLite (Holm and Park, 2000) delivers a Z-score of 6.1 and an RMSD of  $2.3 \text{ \AA}$  for 68  $\alpha$ -carbon pairs.



**Figure 2-15.** Superposition of the crystal structures of *B. burgdorferi* Pur- $\alpha$  (red) and *D. melanogaster* Pur- $\alpha$  repeats I-II (blue).

The most prominent difference is that the PUR domain is constituted by inter-molecular dimerization in the *B. burgdorferi* structure, whereas is built from an intra-molecular interaction in *D. melanogaster* Pur- $\alpha$ . In the *B. burgdorferi* structure,  $\beta$ -strand 3 and  $\beta$ -strand 4 are subtly distorted and slightly longer compared to their *D. melanogaster* counterparts. Also the  $\alpha$ -helices are slightly longer in the *B. burgdorferi* structure.

The observed RMSD of about 2 Å is in the expectation range for evolutionary related proteins with a sequence identity of about 20 % (section 3.1.2) (Koehl and Levitt, 2002).

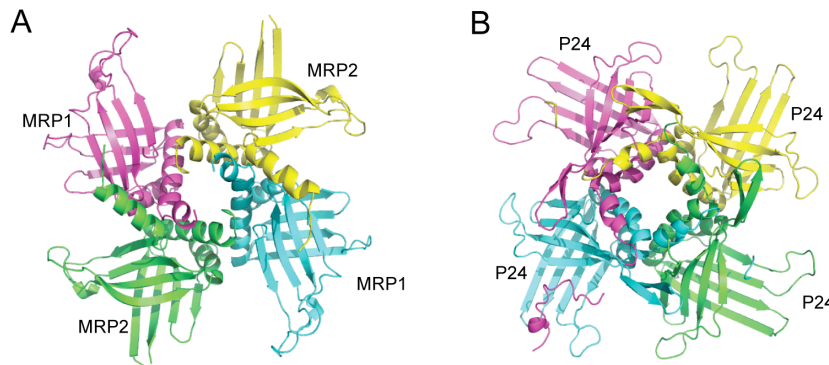
### **Systematic structural comparison**

A systematic structural comparison was performed using DaliLite (Holm et al., 2008). Significant structural similarities were found to a set of proteins, while none of the proteins had a sequence identity higher than 16 %.

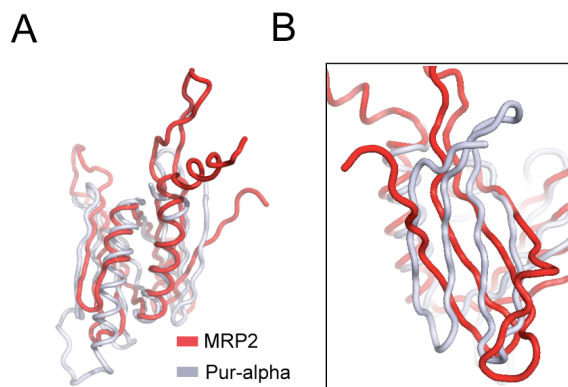
Highest scores were obtained for two hypothetical proteins from cyanobacteria with unknown function (PDB-ID: 2IT9; Z-score: 11.4; RMSD: 3.5 Å for 116  $\alpha$ -carbon pairs / PDB-ID: 2NVN; Z-score: 10.9; RMSD: 3.7 Å for 118  $\alpha$ -carbon pairs).

Among annotated proteins, highest scores were obtained for the Mitochondrial RNA-Binding Protein-2 (MRP-2) (PDB-ID: 2GIA; Z-score: 8.2; RMSD: 4.7 Å for 114  $\alpha$ -carbon pairs) and the Mitochondrial RNA-Binding Protein-1 (MRP-1) (PDB-ID: 2GJE; Z-score: 7.0; RMSD: 4.7 Å for 108  $\alpha$ -carbon pairs) from *Trypanosoma brucei*

(Figure 2-16). A high score was also achieved by Plant Transcriptional Regulator PBF-2 (P24) (PDB-ID: 1L3A; Z-score: 6.9; RMSD: 4.7 Å for 113 α-carbon pairs) from *Solanum tuberosum* (Figure 2-16).



**Figure 2-16. Structures from the PDB with similarity to *D. melanogaster* Pur-α repeats I-II.**  
**A:** Mitochondrial RNA-Binding Protein-1 (MRP1) (PDB-ID: 2GIA) and Mitochondrial RNA-Binding Protein-2 (MRP2) (PDB-ID: 2GJE) from *Trypanosoma brucei*.  
**B:** Plant Transcriptional Regulator PBF-2 (P24) from *Solanum tuberosum* (PDB-ID 1L3A).



**Figure 2-17. Superposition of the PUR domain with a Whirly domain.** A: Structure of Pur-α repeats I-II is depicted in grey, structure of MRP2 is shown in red. B: Close-up showing the orientation of the β-sheets from Pur-α and MRP2.

All three proteins contain a Whirly domain that suits to the overall  $\beta\beta\beta\beta\alpha$ - $\beta\beta\beta\beta\alpha$  topology of the PUR domain and nicely superposes with it (Figure 2-17). The main differences to the PUR domain are longer loop regions and a rotated orientation of the β-sheets. In addition, MRP2 has a longer kinked α-helix at the C-terminus. It is intriguing that the structural similarity is accompanied by a functional correlation, as MRP-2 and MRP-1 are RNA-binding proteins (Schumacher et al., 2006), whereas P24 binds to dsDNA and ssDNA (Desveaux et al., 2002).



To conclude, the systematic structural comparison revealed that Pur- $\alpha$  belongs to the Whirly class of nucleic-acid binding proteins (section 3.1.4).

### Crystallization trials of *D. melanogaster* Pur- $\alpha$ repeat III

To obtain structural information on the PUR repeat III of *D. melanogaster* Pur- $\alpha$ , crystallization trials with protein fragments consisting of the third repeat only were initiated. Again, the structural model made by J. Söding guided the design of expression constructs. Initial crystallization set ups were performed with Pur- $\alpha$  185-260 and crystals formed in 15 out of 384 conditions. For refinement, also Pur- $\alpha$  185-258, Pur- $\alpha$  181-260 and Pur- $\alpha$  188-258 were employed. Best crystals were obtained for Pur- $\alpha$  188-258. Mainly needle-shaped crystals were obtained (Figure 2-18, A). During crystal transfer and cryo-protection, many of them broke or cracked. The needles were tested at the ESRF and diffracted up to 4 Å. In some conditions, additional hexagonal and pentagonal crystal plates formed (Figure 2-18, B).



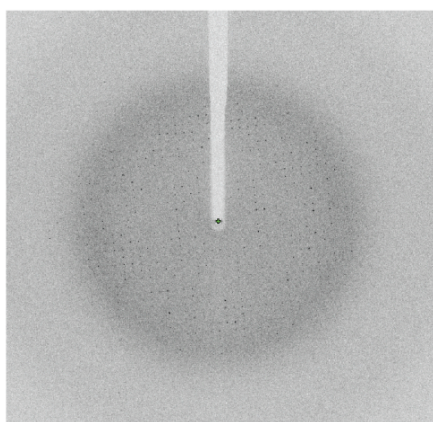
Figure 2-18. Crystals of *D. melanogaster* Pur- $\alpha$  repeat III. A: Pur- $\alpha$  188-258 was crystallized at 4 °C A: Crystallization solution contained 100 mM sodium cacodylate pH 7.5, 0.2 M ammonium acetate, 15 % PEG 3350, and 4 % MPD. B: Crystallization solution contained 100 mM sodium cacodylate pH 7.5, 0.2 M sodium acetate, and 26 % PEG 3350. C: Crystallization solution contained 100 mM sodium cacodylate pH 7.5, 200 M sodium acetate, 22 % PEG 3350.

Most of these plate-shaped crystals grow attached to the plastic surface of the crystallization plate and could not be transferred. However, the few that could be transferred and tested at the ESRF diffracted up to 3 Å. Therefore, it was aimed at refining the conditions in favor of the plate-shaped crystals. It was found that low protein concentration (0.3-0.5 mg/mL) promoted this crystal form and more crystals freely

floating in the crystallization drop were obtained. After optimization, crystals of *D. melanogaster* Pur- $\alpha$  188-258 were grown at a concentrations of 0.3 mg/mL at 4 °C. The crystallization solution contained 22 % PEG 3350, 100 mM sodium cacodylate pH 7.5 and 200 mM sodium acetate. After 1-2 days, needle-shaped crystals appeared with a size of (300 x50 x 50)  $\mu\text{m}$ . After 2-3 weeks, additional pentagonal and hexagonal crystal plates of (100 x50 x 50)  $\mu\text{m}$  size formed, which diffracted up to 2.8 Å (Figure 2-18, C).

A native dataset was recorded at 2.8 Å at the ESRF (ID14-1) (Figure 2-19). Crystals belong to spacegroup  $P2_12_12_1$  with cell constants  $a = 55.3 \text{ \AA}$ ,  $b = 60.3 \text{ \AA}$ ,  $c = 121.7 \text{ \AA}$  and angles  $\alpha = \beta = \gamma = 90.00^\circ$ .

Molecular replacement was attempted using the program PHASER (McCoy et al., 2007). The structural models of *D. melanogaster* Pur- $\alpha$  repeats I-II respective *B. burgdorferi* Pur- $\alpha$  were used as search ensembles, but no convincing solution was found. To tackle the phase problem, the production of selenomethionine-derivatized protein crystals is in progress.



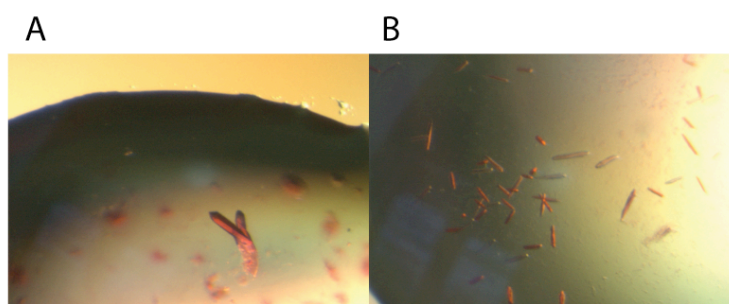
**Figure 2-19. Diffraction image of a native crystal of *D. melanogaster* Pur- $\alpha$  repeat III.**

### **Co-crystallization trials of *D. melanogaster* Pur- $\alpha$ repeats I-II with nucleic acids**

Crystals of *D. melanogaster* Pur- $\alpha$  repeats I-II were soaked with DNA and RNA oligonucleotides, as this fragment proved functional in nucleic acid binding (section 2.2.1.3 and section 5.14). Several datasets were recorded and after indexing, the same space group and unit cell constants were obtained as for the apo-structure. Molecular Replacement with the structural model revealed no extra-density in the unit

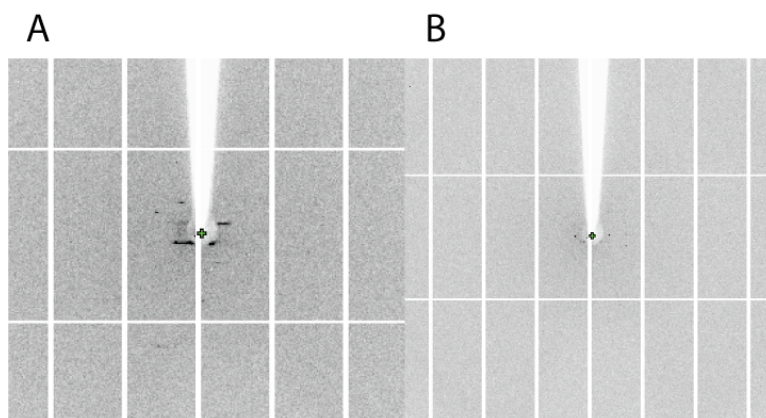
cell, so soaking was not successful. This is not surprising, since for these crystals, a Matthew's coefficient of  $1.9 \text{ \AA}^3/\text{Da}$  was calculated, which corresponds to a solvent content of only 35 %. Therefore the unit cell seems already crowded in the apo-structure, leaving little space left for nucleic acids. In order to obtain a different crystal form, co-crystallization was attempted (section 5.10.4).

The protein fragment of *D. melanogaster* Pur- $\alpha$  repeats I-II, which was crystallized successfully (Figure 2-9), was further employed in co-crystallization trials with nucleic acids. As  $(\text{GGN})_n$  repeats constitute the consensus sequence of Pur- $\alpha$  (section 1.3) (Bergemann et al., 1992),  $(\text{CGG})_n$  repeat oligomers were chosen. To avoid that flexible parts of the nucleic acid disturb crystallization, its lengths was only 7 nucleotides (Table 5-2), as this matches the dimensions of the anticipated binding surface in the crystal structure of Pur- $\alpha$  repeats I-II. In order to distinguish crystals of a complex of protein and DNA from protein crystals, Cy3-labelled DNA was used (section 5.10.4).



**Figure 2-20.** Crystals of *D. melanogaster* Pur- $\alpha$  repeats I-II and DNA. Pur- $\alpha$  40-185 and Cy-3 labeled CGG 7mer ssDNA (1:2).

**A:** Crystals of approximately  $(100 \times 40 \times 20) \mu\text{m}$  size, obtained with 50 mM MES pH 5.9 and 20 % MPD. **B:** Crystals of approximately  $(80 \times 20 \times 20) \mu\text{m}$  size, obtained with 50 mM MES pH 5.9, and 30 % MPD,



**Figure 2-21.** Diffraction image of crystals of *D. melanogaster* Pur- $\alpha$  in complex with DNA. Pur- $\alpha$  40-185 with Cy-3 labelled CGG 7mer ssDNA.

Initial crystals were obtained in 9 out of 480 conditions tested. After preliminary optimization, crystals of *D. melanogaster* Pur- $\alpha$  40-185 and Cy3-CGG 7mer ssDNA (1:2) were grown at a concentration of 2.5-4 mg/mL. Crystallization solutions contained 50 mM MES pH 6.5 and 30 % MPD or 50 mM MES pH 6.5 and 28 % PEG 400 (section 5.10.4). Most crystals had a rice-grain shape of approximately (100 x 25 x 20)  $\mu\text{m}$  size. Individual crystals had a pointed shape with an approximate size of (100 x 40 x 20)  $\mu\text{m}$ . Crystals were tested for diffraction at SLS. The diffraction quality of the crystals is poor, as only individual spots were detected at a maximum resolution of 15 Å, and many spots were enlarged and smeared (Figure 2-21). Refinement of the crystallization conditions is ongoing.

Likewise, co-crystallization with short RNA oligonucleotides was tried. The affinity of *D. melanogaster* Pur- $\alpha$  repeats I-II to CGG 12mer RNA was shown in electrophoretic mobility shift assays (section 2.2.1.3 and Figure 2-33). Initial crystal screens were performed with Cy3-labeled CGG 7mer RNA, but no crystals were obtained.

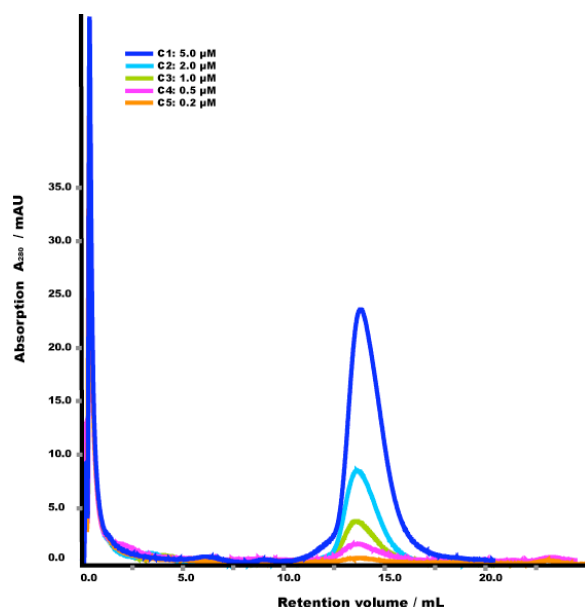
## 2.1.2 Solution structure of Pur- $\alpha$

### 2.1.2.1 Analytical size-exclusion chromatography

To probe the oligomeric state of Pur- $\alpha$ , analytical size-exclusion chromatography was performed. After calibration of the column with proteins of known molecular weights, the molecular weight of the sample was calculated from the elution peak volume.

For *D. melanogaster* Pur- $\alpha$  repeats I-II (amino acid 40 to 185), the peak appeared at 15.0 mL (loaded concentration 5  $\mu$ M), corresponding to a calculated molecular weight of 17.4 kDa. Given the theoretical molecular weight of the protein fragment (16.8 kDa), Pur- $\alpha$  repeats I-II is monomeric in solution.

*D. melanogaster* Pur- $\alpha$  repeats I-III (amino acids 40-255) eluted at 13.79 mL (loaded concentration 5  $\mu$ M), corresponding to a calculated molecular weight of 36.5 kDa. This rather corresponds to a dimer, as the theoretical molecular weight of the protein fragment is 25.2 kDa. The difference between the calculated molecular weight from the calibration curve and the actual molecular weight of a dimer might hint at a monomer-dimer equilibrium. To assess if dimerization is concentration dependent, concentrations in the range of 5  $\mu$ M to 200 nM were loaded (Figure 2-22).



**Figure 2-22.** Analytical size-exclusion chromatography with *D. melanogaster* Pur- $\alpha$  repeats I-III. The column was loaded with 0.5 mL protein solution at five different concentrations as given above.

The estimated protein concentration on the column was calculated by multiplication of the loaded concentration with the dilution factor. The dilution factor was obtained by dividing the volume of the sample loaded by the volume of the elution peak. Leading as

well as trailing portions of the peak were ignored by defining peak volumes to absorption areas above 10 % of the peak maximum. For each concentration, the molecular weight was calculated from the elution peak volume (Table 2-9).

Loaded Protein concentration in nM	Estimated concentration on column in nM	Elution volume at peak in mL	Calculated molecular weight in kDa
5000	1140	13.79	36.5
2000	420	13.56	41.9
1000	150	13.59	41.2
500	75	13.71	38.2
200	30	13.71	38.2

**Table 2-9: Analytical size-exclusion chromatography with *D. melanogaster* Pur- $\alpha$  repeats I-III at various concentrations. Theoretical molecular weight of one peptide chain is 25.2 kDa.**

There was a slight concentration-dependence of the observed elution volumes, as the elution volume was a little smaller for higher concentrations, yielding a higher molecular weight. The exception was the measurement for the highest concentration loaded ( $5 \mu\text{M}$ , corresponding to  $0.126 \text{ mg/mL}$ ), which had a surprisingly high elution volume. However, the calculated molecular weight for each concentration rather corresponds to a dimer than to a monomer. Thus, dimerization occurs at concentrations on the column as low as  $30 \text{ nM}$  and as high as  $1.14 \mu\text{M}$  (Table 2-9). In conclusion, Pur- $\alpha$  repeats I-III dimerizes in solution and possibly exists in a monomer-dimer equilibrium.

The physiological concentration of Pur- $\alpha$  has not been reported, but since dimers also form at a concentration as low as  $30 \text{ nM}$ , Pur- $\alpha$  is likely to dimerize also *in vivo*.

The observation that Pur- $\alpha$  repeats I-II is monomeric, but Pur- $\alpha$  repeats I-III is dimeric, hinted at the fact that repeat III induces dimerization and led to further investigation of the solution structure via small angle X-ray scattering (SAXS).

### 2.1.2.2 Small angle X-ray scattering

To complement the crystal structure with a solution structure, small angle X-ray scattering (SAXS) experiments were performed. The main advantage of solution scattering over crystallography is that the protein is studied in nearly physiological conditions. Moreover, different conformations of flexible parts can be examined (Bernado et al., 2007; Putnam et al., 2007). However, SAXS only allows the

determination of the overall shape of the molecule, as the resolution limit is estimated to be 10 Å (Bernado et al., 2007; Putnam et al., 2007; Svergun et al., 2001). Furthermore, the molecular weight of the sample and thus the oligomeric state of the protein in solution can be determined (Mylonas and Svergun, 2007).

In an SAXS experiment, a solution of the biological macromolecule of interest is exposed to X-rays and the scattered intensity is recorded in dependence of the scattering angle. The intensity is plotted as a function of the momentum transfer or scattering vector  $s$  ( $s = 4\pi\lambda^{-1}\sin\theta$ , with the scattering angle  $2\theta$ ). Due to the random orientation of the protein molecules, the scattering intensity is isotropic and proportional to the scattering intensity of one molecule averaged over all orientations (Svergun et al., 2001). To eliminate the contribution of the solvent to the scattering, blank measurements with buffer only are performed before and after each experiment.

At lowest resolution, the scattering intensity is given by a single parameter, the radius of gyration ( $R_g$ ). The radius of gyration is a measure for the size and shape of the scattering particles, as it is the square root of the average squared distances of each scattering atom from the center of each particle. For scattering angles near zero, the Guinier plot ( $\ln(I(s))$  against  $s^2$ ) gives a linear function (Guinier and Fournet, 1955). Its slope can be used to calculate  $R_g$ . It should be noted that only homogenous samples with low attractive forces between the molecules give a linear Guinier plot. It can thus be used to monitor sample quality. To obtain a linear Guinier plot, only data for small scattering angles were used with the constraint  $R_g * s < 1.3$ .

The intercept of the Guinier plot gives the scattering intensity at zero angle ( $I_0$ ). As  $I_0$  is only dependent on the amount of scattering electrons of the molecule, it can be used to calculate the molecular weight of the sample (Putnam et al., 2007). To this end, scattering curves for standard proteins (bovine serum albumin and lysozyme) with known molecular weight were recorded. Their scattering intensities were extrapolated to zero angle and provided a calibration curve ( $I_0$  versus  $M$ ). The scattering intensity of the sample was extrapolated to zero angle and the molecular weight was calculated from the calibration curve.

At high resolution, the scattering intensity is dependent on the shape of the scattering molecule. For a folded macromolecule, the scattering intensity  $I(s)$  falls off with  $s^{-4}$  (Porod's law). From Porod's law, the Kratky plot ( $s^2 * I(s)$  versus  $s$ ) can be deduced, which shows typical characteristics for unfolded, partially folded and folded

macromolecules. For folded protein, Porod's law applies, and the Kratky plot has a parabola-like shape. For partially unfolded and unfolded proteins, the Kratky plot shows a plateau for higher  $s$  values (Putnam et al., 2007).

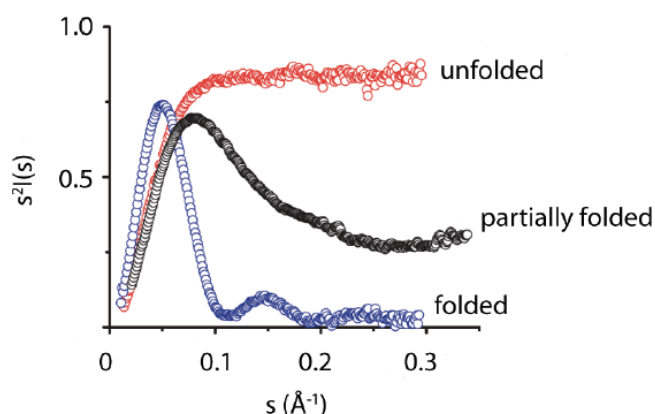


Figure 2-23. Kratky plot for folded, partially folded, and unfolded proteins. Figure taken from (Putnam et al., 2007).

The pair-distribution function  $P(r)$  is the Fourier transform of the scattering curve  $I(s)$  and represents a real-space approximation of the scattering particles. It gives information about all distances ( $r$ ) between the electrons in the protein. To obtain the pair-distribution function, the maximum linear dimension of the particle ( $D_{\max}$ ) has to be iteratively determined. This was done by inspection of the pair-distribution function for different estimates of  $D_{\max}$ .  $P(r)$  should be zero at  $r = D_{\max}$  and at  $r = 0$ . The pair-distribution function also provides information on the shape of the scattering molecule, as it has the form of a parabola for globular proteins and a more skewed shape for elongated molecules (Putnam et al., 2007).

From the scattering curves, a three-dimensional reconstruction of the scattering molecule can be calculated. For *ab initio* reconstruction, the program GASBORp was used (Svergun et al., 2001). It uses a chain-like ensemble of dummy residues that correspond to the distances between  $C\alpha$ -atoms in a peptide chain (Svergun et al., 2001). The obtained shape reconstructions from independent calculations were aligned with the program SUBCOMP to yield an averaged model with the most conceivable and reliable features (Kozin and Svergun, 2000).



## SAXS of *D. melanogaster* Pur- $\alpha$ repeats I-II

In the crystal structure, *D. melanogaster* Pur- $\alpha$  repeats I-II appeared monomeric with both PUR repeats intertwining into each other (Figure 2-12). To assess if this conformation is also found in solution, the fragment that was crystallized (Pur- $\alpha$  40-185) was employed in SAXS measurements.

Synchrotron SAXS data was recorded at beamline X33 at European Molecular Biology Laboratory (EMBL, Hamburg, Germany) and at beamline ID14-3 at ESRF. The measurement was performed at three different concentrations in the same buffer that was used for analytical size-exclusion chromatography (500 mM sodium chloride, 20 mM HEPES) with the addition of 3 mM DTT (Table 2-10 and section 5.15).

Data analysis was done with the ATSAS program package (Konarev et al., 2006).

An  $R_g$  of 2.0-2.1 nm was obtained from the Guinier plot with a  $D_{max}$  of 7.5-8.0 nm (Table 2-10).

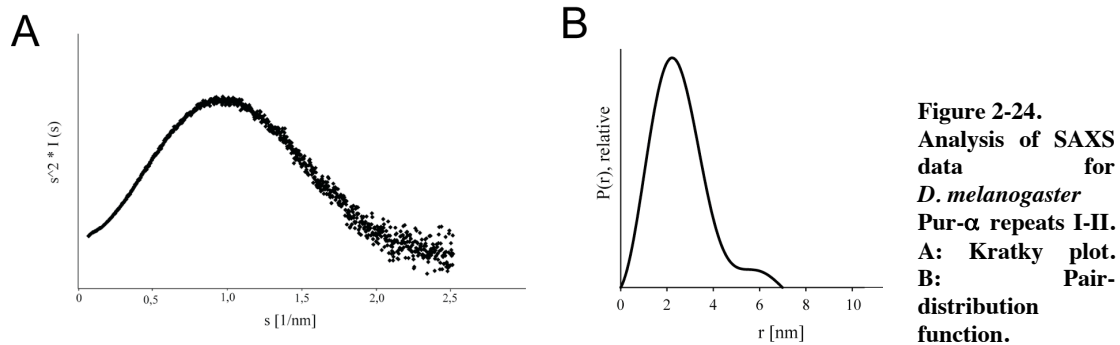
The molecular weight for each sample was obtained from the scattering intensity at zero angle (Table 2-10). An average molecular weight of  $(17.5 \pm 1.2)$  kDa was calculated. Given the theoretical molecular weight of the fragment (16.8 kDa), this confirms that Pur- $\alpha$  repeats I-II is monomeric.

Concentration in mg/mL	Concentration in $\mu$ M	$R_g$ in nm	$D_{max}$ in nm	Mol. weight in kDa	Beamline
4.55	270.8	2.0	7.5	17.4	ID14-3
3.90	232.1	2.1	8.0	18.7	X33
1.76	104.8	2.0	7.5	16.4	X33

Table 2-10. SAXS measurements for *D. melanogaster* Pur- $\alpha$  repeats I-II.  $R_g$  refers to radius of gyration,  $D_{max}$  refers to maximum particle size.

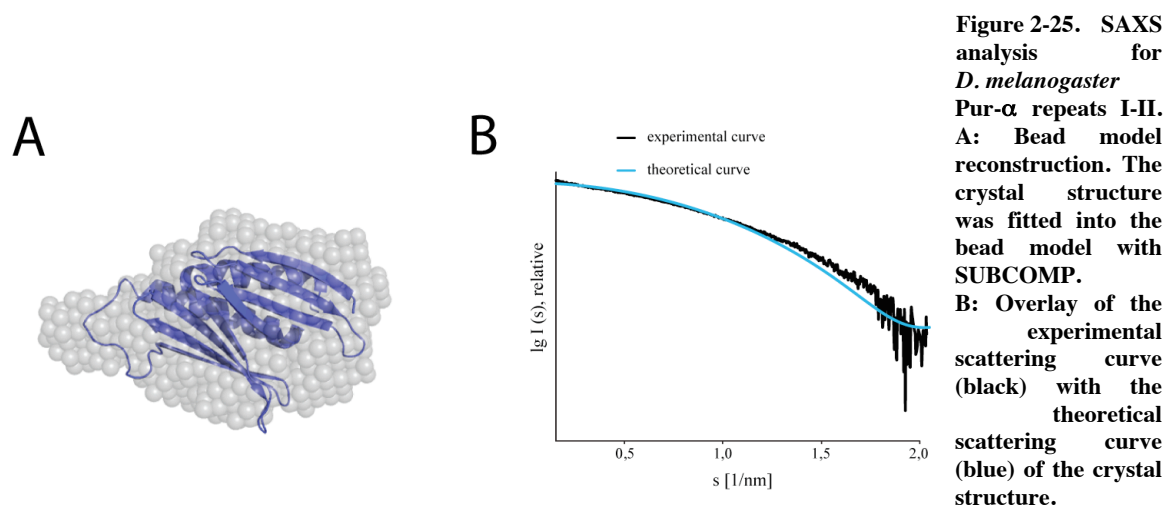
The Kratky plot of *D. melanogaster* Pur- $\alpha$  repeats I-II has the typical bell-shaped form of a folded protein (Figure 2-23 and Figure 2-24, A) (Putnam et al., 2007).

The shape of the Pair-distribution function showed a rather symmetric peak that is typical for a globular molecule (Figure 2-24, B). Since the PUR domain seen in the crystal structure has an overall globular shape, this complies with its existence also in solution.



**Figure 2-24.** Analysis of SAXS data for *D. melanogaster* Pur- $\alpha$  repeats I-II. A: Kratky plot. B: Pair-distribution function.

Using GASBORp, a bead model reconstruction was calculated from the SAXS data (Figure 2-25, A) (Svergun et al., 2001). The crystal structure could be fitted nicely into the envelope, further supporting the assumption that the PUR domain is also stable in solution (Figure 2-25, B) From the crystal structure, a theoretical scattering curve was calculated using CRY SOL (Svergun et al., 1995). The theoretical and the experimental scattering curve superpose nicely (Figure 2-25). This confirms that the structure of *D. melanogaster* Pur- $\alpha$  repeats I-II in aqueous solution equals the structure observed in the crystal.



**Figure 2-25.** SAXS analysis for *D. melanogaster* Pur- $\alpha$  repeats I-II. A: Bead model reconstruction. The crystal structure was fitted into the bead model with SUBCOMP. B: Overlay of the experimental scattering curve (black) with the theoretical scattering curve (blue) of the crystal structure.

### SAXS of *D. melanogaster* Pur- $\alpha$ repeats I-III

By analytical size-exclusion chromatography, it was found that *D. melanogaster* Pur- $\alpha$  repeats I-III forms dimers (section 2.1.2.1 and Figure 2-22). This issue was further examined by SAXS.

Synchrotron SAXS data was recorded at beamline X33 at EMBL and at beamline ID14-3 at ESRF. The measurement was performed at three different concentrations in the same buffer that was used for analytical size-exclusion chromatography (500 mM sodium chloride, 20 mM HEPES) with the addition of 3 mM DTT (Table 2-11 and section 5.15). It should be noted that buffer conditions were not optimal. The examination of the Guinier plot for higher concentrations than 1.53 mg/mL showed that the protein aggregated. Therefore, only the lowest concentration was used for the bead model reconstruction. For future experiments, an extensive screening for improved buffer conditions (e. g. with dynamic light scattering) is recommended, to obtain scattering curves at higher concentrations and with an improved signal to noise ratio.

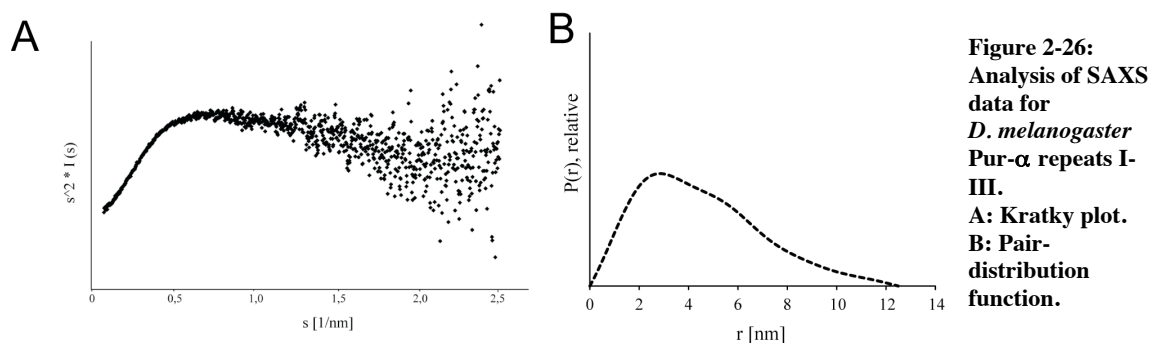
Data analysis was done with the ATSAS program package (Konarev et al., 2006). An  $R_g$  value of 3.6-4.4 nm was obtained from the Guinier plot with a  $D_{max}$  of 13.0-14.0 nm (Table 2-11).

The molecular weight for each sample was obtained from the scattering intensity at zero angle, yielding an average molecular weight of  $(46.0 \pm 4.6)$  kDa (Table 2-11). This corresponds to a dimer, as the theoretical molecular weight of the fragment Pur- $\alpha$  repeats I-III is 25.2 kDa. The dimerization of Pur- $\alpha$  repeats I-III is induced by PUR repeat III, as Pur- $\alpha$  repeats I-II is monomeric (Table 2-10).

Concentration in mg/mL	Concentration in $\mu$ M	$R_g$ in nm	Dmax in nm	Mol. Weight in kDa	Beamline
1.53	60.5	4.4	14.0	50.3	ID14-3
1.11	43.9	3.8	13.5	46.6	ID14-3
0.9	35.6	3.6	13.0	41.1	X33

Table 2-11. SAXS data of *D. melanogaster* Pur- $\alpha$  repeats I-III.  $R_g$  refers to radius of gyration,  $D_{max}$  refers to maximum particle size.

The Kratky plot shows that Pur- $\alpha$  repeats I-III is predominantly folded, but might contain some unstructured regions (Figure 2-23 and Figure 2-26, A) (Putnam et al., 2007). The Pair-distribution function shows the typical skewed shape of an elongated molecule (Figure 2-26, B).

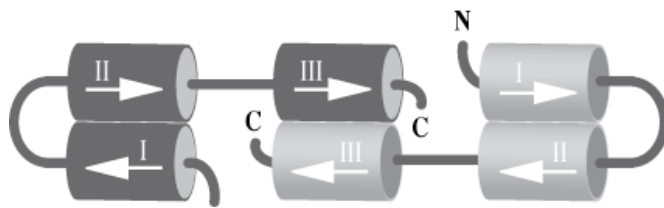


**Figure 2-26:**  
Analysis of SAXS  
data for  
*D. melanogaster*  
Pur- $\alpha$  repeats I-  
III.  
**A:** Kratky plot.  
**B:** Pair-  
distribution  
function.

Taken together, the analysis of the SAXS data confirmed the existence of the dimer, but also suggested that it might not be entirely stable. Firstly, the average molecular weight is lower than expected for a dimer (Table 2-10). This could arise from a monomer-dimer equilibrium, as suggested by the fact that a higher molecular weight was calculated for higher protein concentrations. Secondly, the Kratky plot indicates that there are unstructured regions (Figure 2-26, A). These are most likely due to the PUR repeat III, as the Kratky plot for Pur- $\alpha$  repeats I-II shows the typical form of a folded protein (Figure 2-24, A). An explanation would be that a fraction of Pur- $\alpha$  is monomeric. The third repeat might be unstructured, when it has no interaction partner.

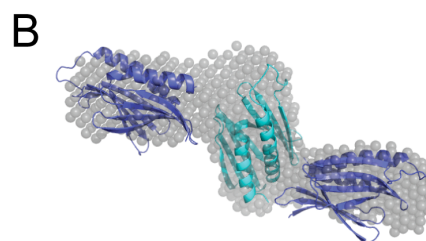
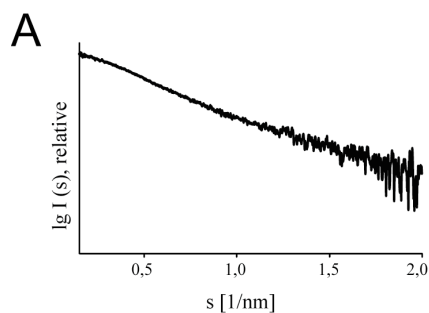
Dimerization is most likely achieved by interaction of PUR repeat III with another PUR repeat III from a second peptide chain (Figure 2-27). In this model, PUR repeats I and II of both Pur- $\alpha$  molecules form the globular PUR domain as seen in the crystal structure. The two PUR repeats III connect both peptide chains, yielding an elongated dimer (Figure 2-27).

PUR repeat III is nearly as homologous to the other two PUR repeats in *D. melanogaster* Pur- $\alpha$ , as they are to each other and to *Borrelia* PUR (section 3.1.6). Given that the latter three all adopt a very similar fold (Figure 2-15), it is tempting to speculate that also PUR repeat III might adopt a similar fold (section 3.1.6). The proposed interaction between two PUR repeats III (Figure 2-27) might then be constituted by PUR domain formation.



**Figure 2-27. Dimerization model of Pur- $\alpha$ .** One molecule is depicted in dark grey, the other in light grey. Two PUR repeats III interact with each other to dimerize Pur- $\alpha$ .

To test this hypothesis, the surface envelope was calculated from the SAXS data using the program GASBORp (Figure 2-28) (Svergun et al., 2001). The calculated bead model had an unexpected Z-shape that can accommodate three PUR domains (Figure 2-28, B). For a good fit into the envelope, the PUR domain in the middle has to be oriented perpendicular to the two flanking PUR domains. Thus, although the overall shape of Pur- $\alpha$  repeats I-III is elongated, the PUR domains are not arranged in a strictly linear way. It should be noted that the SAXS data do not prove that the dimerization model is correct, but they are in accordance with the model (section 3.1.6 and section 3.1.7).



**Figure 2-28. SAXS analysis for *D. melanogaster* Pur- $\alpha$  repeats I-III.** A: Scattering curve. B: Bead model reconstruction for Pur- $\alpha$  repeats I-III. Three copies of the crystal structure of Pur- $\alpha$  repeats I-II were placed into the bead model.

## 2.2 Functional studies

### 2.2.1 Nucleic-acid binding of Pur- $\alpha$

#### 2.2.1.1 Human Pur- $\alpha$

Filter binding assays were performed with the intention to find high-affinity oligonucleotides for co-crystallization trials with human Pur- $\alpha$ . The studies concentrated on previously known targets that included the recognition sequence (GGN)<sub>n</sub>.

Krachmarov and colleagues demonstrated that Pur- $\alpha$  binds the major late promoter of human polyoma virus 2 (JCV) via the Tat-responsive DNA element *upTAR*. At the JCV *upTAR* element, transcription is synergistically enhanced by Tat protein and Pur- $\alpha$  (Krachmarov et al., 1996). The ssDNA oligonucleotide JCV *upTAR* was constructed by D. Niessing in previous work (Table 5-2).

Chepenik and colleagues demonstrated binding of Pur- $\alpha$  to HIV-1 *TAR* RNA and suggested a role of Pur- $\alpha$  in the activation of transcription of HIV-1 (Chepenik et al., 1998). Short RNA stem loops derived from HIV-1 *TAR* RNA (TAR14AU and TAR14GC) were constructed and tested for binding in previous work by D. Niessing (Table 5-2).

The (GGN)<sub>n</sub> repeat motif is also present in the 3'-overhang of human telomeric DNA (Itoh et al., 1998). Thus short sequences derived from human telomeric DNA (hTel12 DNA and hTel6 DNA) were also tested in previous work by D. Niessing (Table 5-2).

Johnson and colleagues demonstrated binding of Pur- $\alpha$  to short sequences from *BC200* RNA and its functional rat homolog, *BCI* RNA (Johnson et al., 2006; Kobayashi et al., 2000). Thus *in vitro* transcribed *BC200* RNA (200 nt) and *BCI* RNA (152 nt), were included in binding experiments (Table 5-2).

#### Filter binding assays

For none of these targets, binding constants were previously determined. To evaluate which oligonucleotides might be most suitable for co-crystallization, filter binding assays with human Pur- $\alpha$  were performed. The oligonucleotides are radioactively labeled at their 5'-end, incubated with increasing amounts of protein and blotted on an activated nitrocellulose membrane (section 5.16.2 and section 5.16.3). The membrane retains the protein, but not the oligonucleotide, unless it is bound to the protein. Radioactive signal on

the membrane is detected using a Phosphorimaging system. By plotting the relative bound fraction of radioactively labeled oligonucleotide against the protein concentration, saturation curves are obtained. The equilibrium dissociation constant ( $K_D$ ) is derived from the Langmuir isotherm.

Filter binding assays were performed with human Pur- $\alpha$  56-287 C272S, because this was the largest fragment of Pur- $\alpha$  that could be purified and it includes the entire nucleic-acid binding domain (White et al., 2009) (Table 2-1 and Table 2-12).

Oligonucleotide	Type	Length	$K_D$ in nM	n
JCVupTAR	ssDNA	12 nt	232 $\pm$ 37	2
hTel12	ssDNA	12 nt	443 $\pm$ 46	3
TAR14AU	Stem loop RNA	16 nt	38 $\pm$ 1	2
TAR14GC	Stem loop RNA	16 nt	85	2
Cntrl1	ssDNA	12 nt	no binding	3
PolyC	ssDNA	12 nt	no binding	3
<i>BC1</i>	Stem loop RNA	152 nt	724 $\pm$ 23	3
<i>BC200</i>	Stem loop RNA	200 nt	621 $\pm$ 95	3

**Table 2-12. Filter binding assays with human Pur- $\alpha$  56-287 C272S.** nt = nucleotides,  $K_D$  = equilibrium dissociation constant, n = number of experiments.

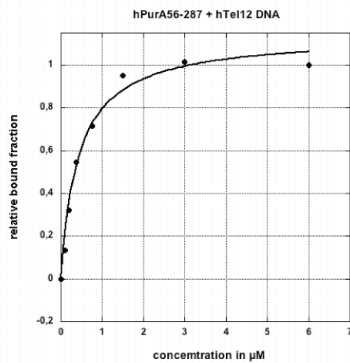
Human Pur- $\alpha$  bound with good affinity to the tested DNA as well as RNA oligonucleotides, while preferring the latter ( $K_D$  in the nanomolar range). The affinity was strongest for TAR14AU stem loop RNA ( $K_D$  = 38 nM), followed by TAR14GC stem loop RNA ( $K_D$  = 85 nM) (Table 2-12 and Figure 2-29).

The affinity to the DNA oligonucleotides was about 5-times weaker, as for JCVupTAR ssDNA ( $K_D$  = 232 nM) and hTel12 ssDNA ( $K_D$  = 443 nM). Binding to the *in vitro* transcribed RNA (*BC1* and *BC200*) was also significantly lower, with a  $K_D$  of 724 nM (612 nM, respectively) (Table 2-12 and Figure 2-29).

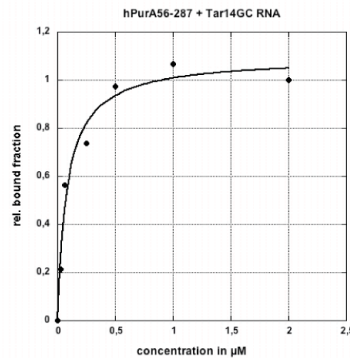
In the following, TAR14AU stem loop RNA and JCVupTAR ssDNA were employed in co-crystallization assays (section 2.1.1.1).

No significant binding was observed to the control oligonucleotides (Cntrl1 and PolyC) (Table 5-2), which do not contain the consensus sequence. This indicates a sequence-specific mode of binding.

A



B



**Figure 2-29. Representative filter binding curves for human Pur- $\alpha$ .**

**A: Binding of Pur- $\alpha$  56-287 C272S to hTel12 DNA,  $K_D = 438$  nM.**

**B: Binding of Pur- $\alpha$  56-287 C272S to TAR14GC RNA,  $K_D = 85$  nM.**

### 2.2.1.2 *B. burgdorferi* Pur- $\alpha$

To assess if the *B. burgdorferi* ortholog of Pur- $\alpha$  is indeed a functional Pur protein, its nucleic-acid binding properties were probed. The oligonucleotides previously employed for human Pur- $\alpha$  were used in filter binding assays with *B. burgdorferi* Pur- $\alpha$  full length (amino acids 1 to 122).

#### Filter binding assays

In filter binding assays, the binding of *B. burgdorferi* Pur- $\alpha$  to ssDNA was confirmed (JCVupTAR ssDNA  $K_D = 480$  nM and hTel12 ssDNA  $K_D = 413$  nM) (Table 2-13). The affinities were similar to those obtained with human Pur- $\alpha$  ( $K_D = 200$ -400 nM) (Table 2-12 and Figure 2-31, A).

The binding seems to be specific, as no significant signal was observed for control ssDNA (Cntrl1) lacking the consensus sequence (Table 5-2). This indicated that the *B. burgdorferi* ortholog shares not only a sequence similarity, but also a functional conservation and encouraged the pursuit of its crystal structure (section 2.1.1.2).



Oligonucleotide	Type	Length	K <sub>D</sub> in nM	n
JCVupTAR	ssDNA	12 nt	480 ± 81	4
hTel12	ssDNA	12 nt	413 ± 18	4
TAR14AU	Stem loop RNA	16 nt	no saturation	3
TAR14GC	Stem loop RNA	16 nt	no saturation	3
Cntrl1	ssDNA	12 nt	no binding	3
MS2	Hairpin RNA	19 nt	no saturation	3
PP7	Hairpin RNA	25 nt	no saturation	3

**Table 2-13. Filter binding assays with *B. burgdorferi* Pur- $\alpha$  full length. K<sub>D</sub> = equilibrium dissociation constant, n = number of experiments.**

Binding to the RNA oligonucleotides could be not confirmed. The signal of bound RNA increased with higher protein concentration, but no saturation was observed even with protein concentrations as high as 8  $\mu$ M (Table 2-13). This indicates a very weak affinity or even an unspecific mode of binding.

Crystal structure determination of *B. burgdorferi* Pur- $\alpha$  revealed a structural similarity to bacteriophage coat proteins from Pseudomonas phage pp7 (Chao et al., 2008) and from enterobacterio phage ms2 (Ni et al., 1995) (section 2.1.1.2). These are RNA binding proteins, which bind a specific hairpin RNA (PP7 and MS2, respectively). It was speculated whether *B. burgdorferi* Pur- $\alpha$  binds to these hairpin RNA, therefore filter binding experiments with the respective oligonucleotides were performed.

*B. burgdorferi* Pur- $\alpha$  full length showed some binding to the PP7 RNA and MS2 RNA oligonucleotides, as the signal of retained radioactivity increased with higher protein concentrations. Alas, no saturation of the signal was obtained with a protein concentration as high as 8  $\mu$ M (Table 2-14). A K<sub>D</sub> would be even higher for this interaction. The binding is therefore very weak and/or unspecific (Table 2-13).

In summary, *B. burgdorferi* Pur- $\alpha$  binds specifically to the PUR consensus sequence in DNA oligonucleotides, but only very weakly and/or unspecifically to the RNA targets tested. It is possible that species-specific RNA targets for *B. burgdorferi* Pur- $\alpha$  exist, but none are reported.

### Filter binding assays with mutant Pur- $\alpha$

To map the binding surface on the crystal structure of *B. burgdorferi* Pur- $\alpha$ , mutational studies were performed. Surface assessment of the crystal structure of *B. burgdorferi* Pur- $\alpha$  led to the identification of candidate amino acids that might be involved in nucleic-acid binding. Surface-exposed candidate amino acids were selected by their positive charge (preferably arginines) and their degree of conservation between species, as the DNA binding properties are conserved. The respective arginines were replaced by alanines (section 5.6.1).

One point mutation (R18A) was introduced in the second  $\beta$ -strand of the  $\beta$ -sheet in the *Borrelia* PUR repeat. Two adjacent arginines (R28A, R29A) were mutated in the connector between the second and third  $\beta$ -strand. One mutation was introduced in the fourth strand of the  $\beta$ -sheet (R49A) (Figure 2-30).

The *B. burgdorferi* Pur- $\alpha$  full length mutant proteins were tested in filter binding assays and the results were compared to the binding affinities of the wild type protein (Table 2-13 and Table 2-14).

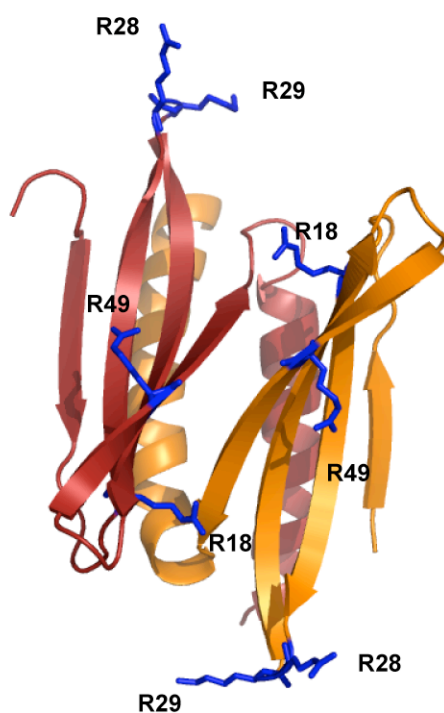


Figure 2-30. Position of the mutations which affect nucleic acid binding in *B. burgdorferi* Pur- $\alpha$ . Backbone Ribbon model of the crystal structure, one monomer is shown in red, the second monomer is depicted in gold. Arginines that were mutated to alanines are highlighted in blue.

Protein	Oligonucleotide	Type	Length	$K_D$	n
bbPurA1-122 wt	JCV upTAR	ssDNA	12 nt	480 ± 81	3
bbPurA1-122 wt	hTel12	ssDNA	12 nt	413 ± 18	3
bbPurA1-122 R18A	JCV upTAR	ssDNA	12 nt	Appr. 1 $\mu$ M	3
bbPurA1-122 R18A	hTel12	ssDNA	12 nt	No saturation	3
bbPurA1-122 R28A	hTel12	ssDNA	12 nt	No saturation	3
bbPurA1-122 R28A, R29A	hTel12	ssDNA	12 nt	no binding	3
bbPurA1-122 R49A	JCV upTAR	ssDNA	12 nt	no binding	3
bbPurA1-122 R49A	hTel12	ssDNA	12 nt	No binding	3

**Table 2-14. Filter binding assays with mutants of *B. burgdorferi* Pur- $\alpha$ .  $K_D$  = equilibrium dissociation constant, n = number of experiments, wt = wild type. No saturation at 8  $\mu$ M.**

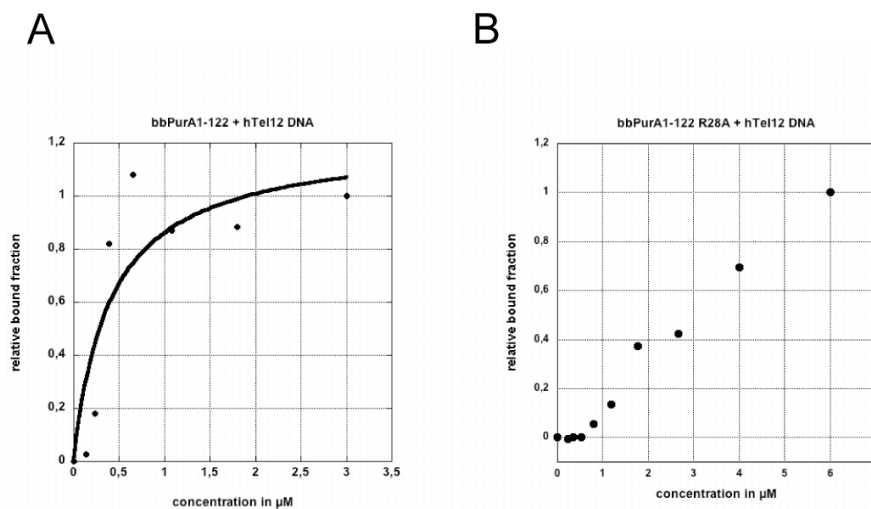
The *B. burgdorferi* Pur- $\alpha$  1-22 R18A bound JCVupTAR ssDNA with an approximate  $K_D$  of 1  $\mu$ M (Table 2-14), thus the mutation R18A only weakly affected binding. The affinity to hTel12 ssDNA was more strongly impaired. The signal of retained oligonucleotide increased with higher protein concentrations, but no saturation of the signal was obtained with a protein concentration as high as 8  $\mu$ M (Table 2-14). A  $K_D$  would be even higher for this interaction.

The same effect was observed for the mutation R28A, as for the binding of *B. burgdorferi* Pur- $\alpha$  1-122 R28A to hTel12 ssDNA, no saturation was achieved at a protein concentration of 8  $\mu$ M (Table 2-14).

Because of their immediate vicinity, arginine 29 was also mutated and the double mutant was assessed. *B. burgdorferi* Pur- $\alpha$  1-22 R28A R29A showed a total loss of affinity to both ssDNA oligomers tested (Table 2-14).

An abolishment of ssDNA binding was also seen for *B. burgdorferi* Pur- $\alpha$  1-22 R49A

In summary, the effects of the point mutations support the hypothesis that the  $\beta$ -sheets are the nucleic-acid binding surfaces of *B. burgdorferi* Pur- $\alpha$ .



**Figure 2-31.**  
**Representative filter binding curves for *B. burgdorferi* Pur- $\alpha$ .**  
**A: Binding to hTel12 DNA with wild type Pur- $\alpha$  full length.**  
**B: Binding to hTel12DNA with mutant protein Pur- $\alpha$  full length R28A.**

### 2.2.1.3 *D. melanogaster* Pur- $\alpha$

The binding affinities of *D. melanogaster* Pur- $\alpha$  to nucleic acid targets previously described for human Pur- $\alpha$  were assessed. Further, two additional sequences were probed.

Bergemann and colleagues assessed the binding of human Pur- $\alpha$  to zones of replication and suggested (GGAGG) ssDNA as the minimal requirement for binding. MF0677 ssDNA is a 24mer containing this consensus sequence and flanking regions (Table 5-2). This oligonucleotide was previously employed in mobility shift competition assays (Bergemann et al., 1992).

Jin and colleagues reported that *D. melanogaster* Pur- $\alpha$  binds to r(CGG)<sub>n</sub> repeats in the *Fmr1* mRNA in a *Drosophila* model of Fragile X Tremor/Ataxia Syndrome. For pull-down experiments, r(CGG)<sub>105</sub> RNA was employed (Jin et al., 2007).

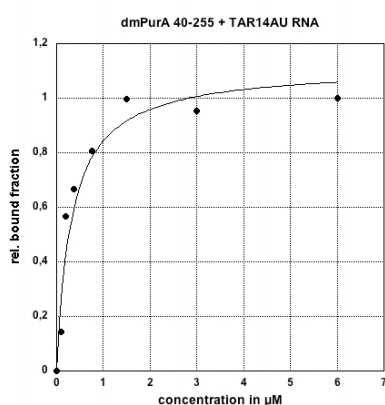
In this study, the binding of Pur- $\alpha$  to shorter repeat oligonucleotides (CGG 12mer ssRNA and CGG 25mer ssRNA) was probed (Table 5-2).

## Filter binding assays with Pur- $\alpha$ repeats I-III

Filter binding assays with *D. melanogaster* Pur- $\alpha$  and the previously described target oligomers of human Pur- $\alpha$  was assessed in filter binding assays. The fragment of Pur- $\alpha$  used contained the core region (PUR repeats I-III) and was comprised of amino acids 40 to 255.

The obtained binding affinities (Table 2-15 and Figure 2-32) were similar compared to those for human Pur- $\alpha$  (Table 2-12).

Noticeably, unlike for human Pur- $\alpha$ , the standard deviations for this set of experiments were high, reaching 28 % of the measured  $K_D$  value (Table 2-15). This reduces the reliance of the data and might arise from precipitation of the protein on the membrane. Thus, another approach to assess nucleic-acid binding was followed, that is electrophoretic mobility shift assay (EMSA).



**Figure 2-32. Representative filter binding curve for *D. melanogaster* Pur- $\alpha$  repeats I-III. Binding of Pur- $\alpha$  40-255 to TAR14AU RNA.**

Oligonucleotide	Type	Length	$K_D$ in nM	n
hTel12	ssDNA	12 nt	$873 \pm 241$	4
CGG 12mer	ssRNA	12 nt	$675 \pm 134$	2
TAR14AU	Stem loop RNA	16 nt	$376 \pm 91$	2
TAR14GC	Stem loop	16 nt	311	1

**Table 2-15. Filter binding assays with *D. melanogaster* Pur- $\alpha$  40-255.  $K_D$  = equilibrium dissociation constant, n = number of experiments.**

### **Electrophoretic mobility shift assays with full length Pur- $\alpha$**

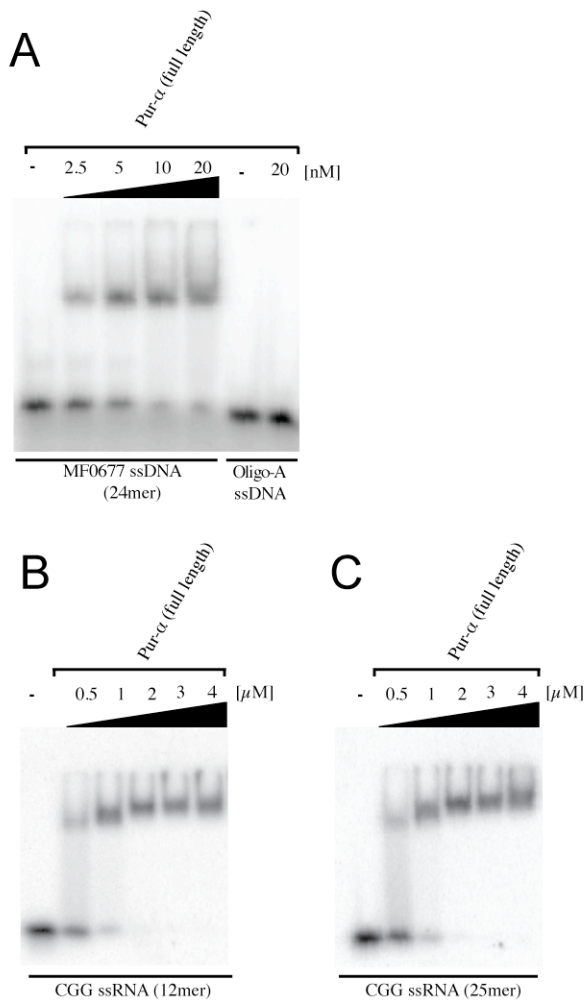
Nucleic-acid binding of full length *D. melanogaster* Pur- $\alpha$  was assessed by EMSA with RNA and DNA oligonucleotides. Oligonucleotides were radioactively labeled at their 5'-end, incubated with increasing amounts of protein and resolved on 6 % TBE gels. Gels were read out using a Phosphorimaging system (section 5.16.2 and section 5.16.4). The  $K_D$  was estimated from the gels as the protein concentration, which shifted half the amount of labeled oligonucleotide.

The nucleic-acid binding properties of *D. melanogaster* Pur- $\alpha$  full length (amino acids 1 to 274) were examined in EMSAs. Full length Pur- $\alpha$  strongly bound to MF0677 ssDNA with a  $K_D$  of about 5 nM. In the control experiment, no binding to Oligo-A ssDNA was observed (Figure 2-33, A).

Binding to the RNA target was considerably weaker. Pur- $\alpha$  full length bound CGG RNA 12mer with a  $K_D$  of about 1  $\mu$ M (Figure 2-33, B).

The binding of Pur- $\alpha$  to CGG 12mer ssRNA was weaker than expected from previous work (Jin et al., 2007). Jin and colleagues pulled down Pur- $\alpha$  from fly extracts with notably longer RNA - r(CGG)<sub>105</sub> RNA – therefore it was speculated if Pur- $\alpha$  would exhibit a more pronounced affinity to longer repeats. To test this, binding experiments were performed with CGG 25mer ssRNA, but no increase in affinity was observed, as the  $K_D$  is approximately 1  $\mu$ M (Figure 2-33, C).

Taken together, the EMSAs showed strong and reproducible binding of *D. melanogaster* full length Pur- $\alpha$  to DNA as well as RNA.



**Figure 2-33.** EMSAs with *D. melanogaster* Pur- $\alpha$  full length. **A:** Binding to MF0677 ssDNA and to Oligo-A control DNA. **B:** Binding to CGG 12mer ssRNA. **C:** Binding to CGG 25mer RNA.

### EMSAs with Pur- $\alpha$ repeats I-II

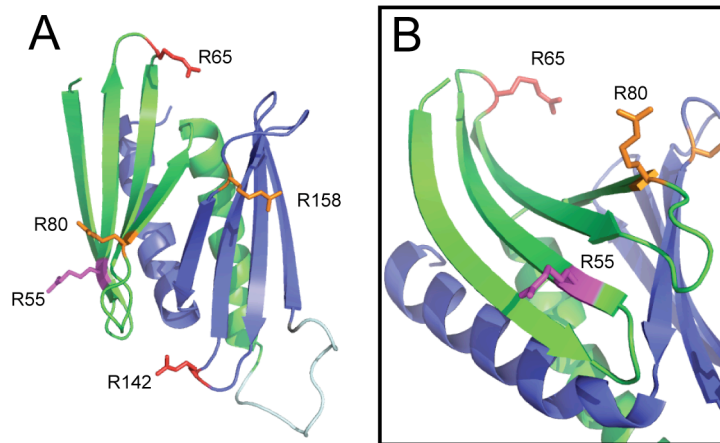
With the motivation to assess whether the crystallized fragment of Pur- $\alpha$  is indeed a functional nucleic-acid binding domain, EMSAs were performed with Pur- $\alpha$  repeats I-II (amino acids 40 to 185) (Figure 2-35).

Strong binding was observed with MF0677 ssDNA, indicating a dissociation constant ( $K_D$ ) of at least 5 nM (Figure 2-35). Binding to CGG RNA 12mer was considerably weaker and indicated a  $K_D$  around 1  $\mu$ M (Figure 2-35).

The measured affinities for Pur- $\alpha$  repeats I-II were analogous to those for full length Pur- $\alpha$ . The crystallized PUR domain is thus indeed a functional nucleic-acid binding domain.

## EMSAs with mutant Pur- $\alpha$ repeats I-II

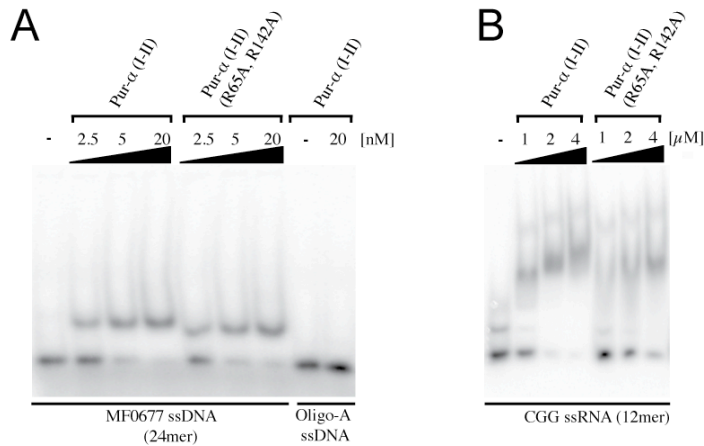
To probe the hypothesis that the  $\beta$ -sheets are the nucleic-acid binding surface of the PUR domain (section 2.1.1.3), point mutations were introduced and their effect on DNA and RNA binding was assessed. As sites of mutagenesis, positively charged, conserved residues on the surface of *D. melanogaster* Pur- $\alpha$  repeats I-II were chosen. Due to the repetitive structure of Pur- $\alpha$ , point mutations were introduced in pairs, replacing both homologous amino acids in each repeat. The respective arginines were replaced by alanines (section 5.6.1). Double mutants were R80A, R158A and R65A, R142A (Figure 2-34).



**Figure 2-34. Position of the mutations which affected nucleic-acid binding in *D. melanogaster* Pur- $\alpha$  repeats I-II. A:** Ribbon backbone model of the PUR domain, the first repeat is depicted in green, the second in blue. Arginines that were mutated to alanines are highlighted in red (R80, R142) or orange (R80, R158). The residue highlighted in magenta (R55) is homologous to a residue in mouse Pur- $\alpha$  that was found to be crucial for nucleic-acid binding in a previous study (Wortman et al., 2005). **B:** Close-up on the concave surface of  $\beta$ -sheet of repeat II.

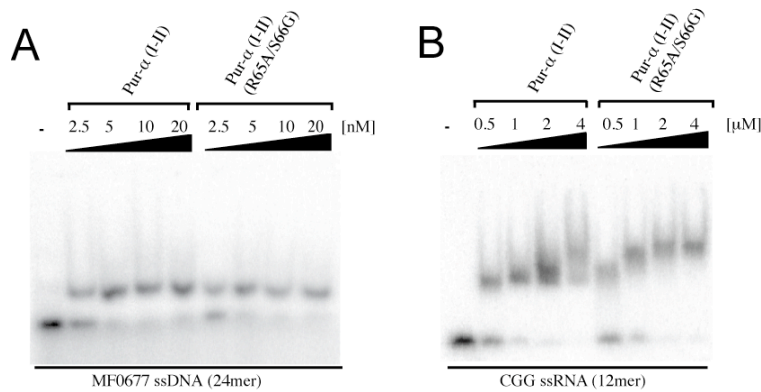


The mutations R65A and R142A are located in the fourth  $\beta$ -strand of the  $\beta$ -sheet of each repeat (Figure 2-34). The double mutant Pur- $\alpha$  40-185 R65A R142A showed the same affinity to ssDNA as the wild type protein, but its affinity to RNA was moderately decreased (Figure 2-35).



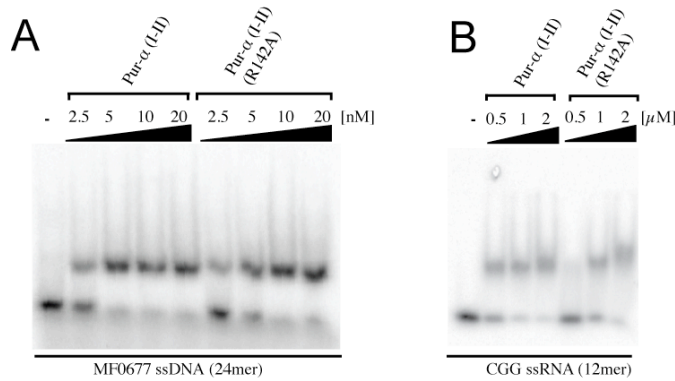
**Figure 2-35. EMSA with *D. melanogaster* Pur- $\alpha$ .** A: Wild type Pur- $\alpha$  repeats I-II and double mutant Pur- $\alpha$  repeats I-II R65A, R142A with MF0677 ssDNA, and control Oligo-A ssDNA. B: Wild type Pur- $\alpha$  repeats I-II and double mutant Pur- $\alpha$  repeats I-II R65A, R142A with CGG 12mer ssRNA.

To further examine the contribution of single amino acids, fragments of Pur- $\alpha$  with single mutations were constructed. The point mutation R65A had no considerable effect on ssDNA or ssRNA binding (Figure 2-36). It shall be mentioned that the clone used had a spontaneous mutation at position 66 (S66G). However, since it showed no effect on binding, it can nevertheless be concluded that the mutation R65A has no effect on binding, either.



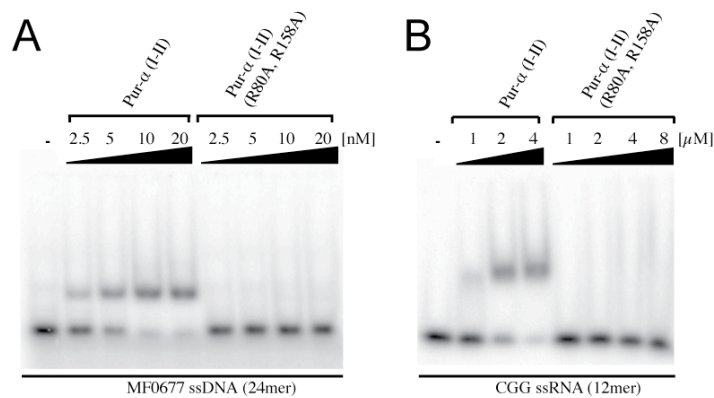
**Figure 2-36. EMSA with *D. melanogaster* Pur- $\alpha$ .** A: Wild type Pur- $\alpha$  repeats I-II and double mutant Pur- $\alpha$  repeats I-II R65A, S66G with MF0677 ssDNA, and control Oligo-A ssDNA. B: Wild type Pur- $\alpha$  repeats I-II and double mutant Pur- $\alpha$  repeats I-II R65A, S66G with CGG 12mer ssRNA.

The point mutation R142A had no effect on ssDNA binding, but moderately impaired ssRNA binding (Figure 2-37).



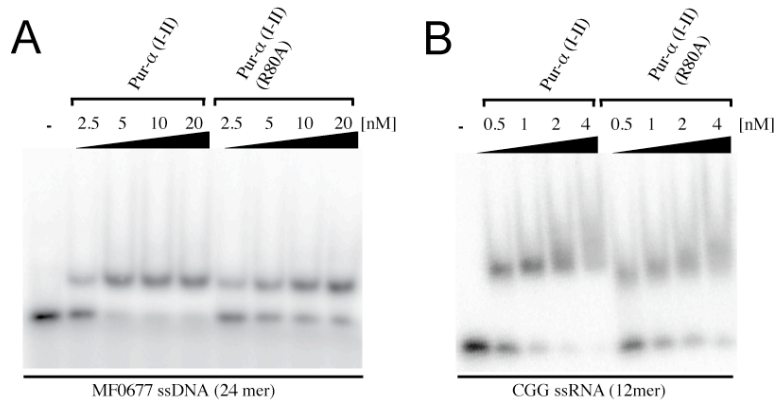
**Figure 2-37.** EMSA with *D. melanogaster* Pur- $\alpha$ . **A:** Wild type Pur- $\alpha$  repeats I-II and point mutant Pur- $\alpha$  repeats I-II R142A, with MF0677 ssDNA, and control Oligo-A ssDNA. **B:** Wild type Pur- $\alpha$  repeats I-II and point mutant Pur- $\alpha$  repeats I-II R142A with CGG 12mer ssRNA.

The mutations R80A and R158A are located in the loop that connects  $\beta$ -strand 2 with  $\beta$ -strand 3 in both  $\beta$ -sheets (Figure 2-34). Pur- $\alpha$  40-185 R80A R158A showed a total loss of affinity to ssDNA as well as to ssRNA binding (Figure 2-38).

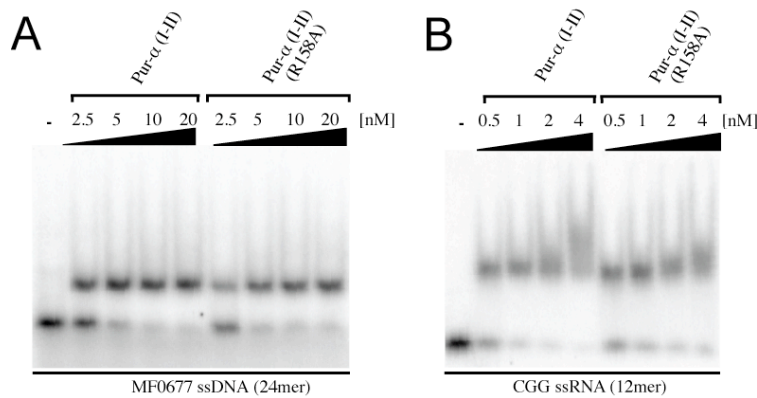


**Figure 2-38.** EMSA with *D. melanogaster* Pur- $\alpha$ . **A:** Wild type Pur- $\alpha$  repeats I-II and double mutant Pur- $\alpha$  repeats I-II R80A, R158A with MF0677 ssDNA. **B:** Wild type Pur- $\alpha$  repeats I-II and double mutant Pur- $\alpha$  repeats I-II R80A, R158A with CGG 12mer ssRNA.

The point mutation R80A decreased both ssDNA as well as ssRNA binding (Figure 2-39). The point mutation R158A had no effect on ssDNA binding, but ssRNA binding was decreased (Figure 2-40).



**Figure 2-39. EMSA with *D. melanogaster* Pur- $\alpha$ .**  
**A:** Wild type Pur- $\alpha$  repeats I-II and point mutant Pur- $\alpha$  repeats I-II R80A with MF0677 ssDNA, and control Oligo-A ssDNA.  
**B:** Wild type Pur- $\alpha$  repeats I-II and point mutant Pur- $\alpha$  repeats I-II R80A with CGG 12mer ssRNA.



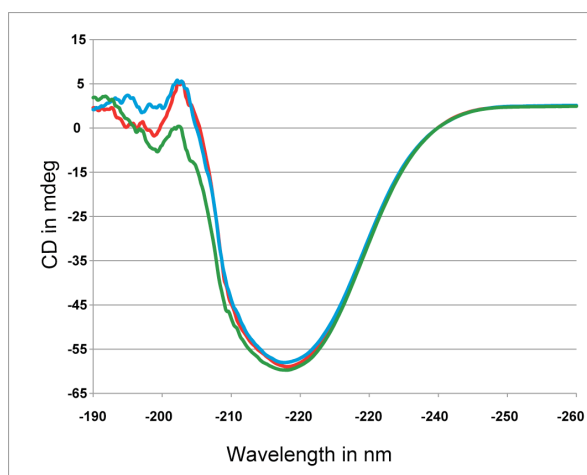
**Figure 2-40. EMSA with *D. melanogaster* Pur- $\alpha$ .**  
**A:** Wild type Pur- $\alpha$  repeats I-II and point mutant Pur- $\alpha$  repeats I-II R158A with MF0677 ssDNA, and control Oligo-A ssDNA.  
**B:** Wild type Pur- $\alpha$  repeats I-II and point mutant Pur- $\alpha$  repeats I-II R158A with CGG 12mer ssRNA.

In summary, the mutational studies show that surface-exposed amino acids located in the  $\beta$ -sheets are involved in nucleic-acid binding. This is further confirmed by the previously described mutation of arginine 71 to glutamic acid in mouse Pur- $\alpha$  (Wortman et al., 2005). This mutation reduced binding to ssDNA as well as unwinding of dsDNA. The homologous residue in *D. melanogaster* Pur- $\alpha$  is arginine 55, which is located in the second  $\beta$ -strand of the  $\beta$ -sheet of PUR repeat I (Figure 2-34).

## Circular dichroism spectra of mutant Pur- $\alpha$ repeats I-II

Circular dichroism (CD) was used to confirm that the mutant proteins employed in binding studies are not defective in folding. CD measures the difference in absorption of left-handed polarized light and right-handed polarized light. In the far-UV region (190-250 nm), a peptide bond gives rise to characteristic shape and magnitude of CD spectra, dependent on its structural environment (e.g.  $\alpha$ -helix,  $\beta$ -sheet or random coil).

CD spectra were recorded for native *D. melanogaster* Pur- $\alpha$  repeats I-II and for double mutants Pur- $\alpha$  repeats I-II R65A, R142 as well as Pur- $\alpha$  repeats I-II R80A, R158A (Figure 2-34). Overlay of the CD spectra for native protein and mutant proteins demonstrates a nice superposition (Figure 2-41). Thus the conformation of the mutant proteins and the native protein seems to be equivalent. It is concluded that the reduced nucleic-acid binding affinities of the mutant proteins are due to loss of functional amino acids that constitute the binding.



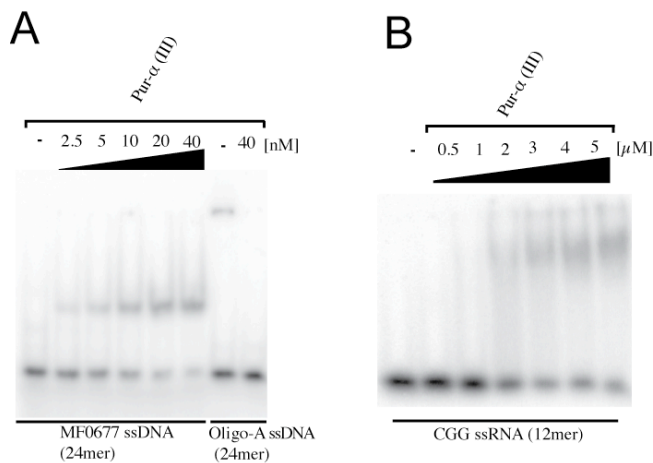
**Figure 2-41.** CD spectra of *D. melanogaster* Pur- $\alpha$  repeats I-II. The spectra for the wild type protein is depicted in green, for the double mutant R80A, R158A in red, and for the double mutant R65A, R142A in blue.

### EMSAs with Pur- $\alpha$ repeat III

The observation that Pur- $\alpha$  repeats I-II does not show any significant difference in binding affinity compared to full length Pur- $\alpha$ , suggests that repeat III does not strongly contribute to nucleic-acid binding (Figure 2-33 and Figure 2-35)

To examine this, EMSAs were performed with a fragment of Pur- $\alpha$  comprising the third repeat only (amino acids 185 to 260).

Pur- $\alpha$  repeat III does bind with approximately 8-fold lower affinity to MF0677 ssDNA ( $K_D = 40$  nM) (Figure 2-42) than Pur- $\alpha$  repeat I-II. For Pur- $\alpha$  repeat I-II, the  $K_D$  was 5 nM (Figure 2-35). Binding to CGG 12mer ssRNA was also drastically reduced ( $K_D = 5 \mu\text{M}$ ) (Figure 2-42). For Pur- $\alpha$  repeats I-II, it was 1  $\mu\text{M}$  (Figure 2-35). The third repeat thus only weakly contributes to nucleic-acid binding (section 3.2.2).



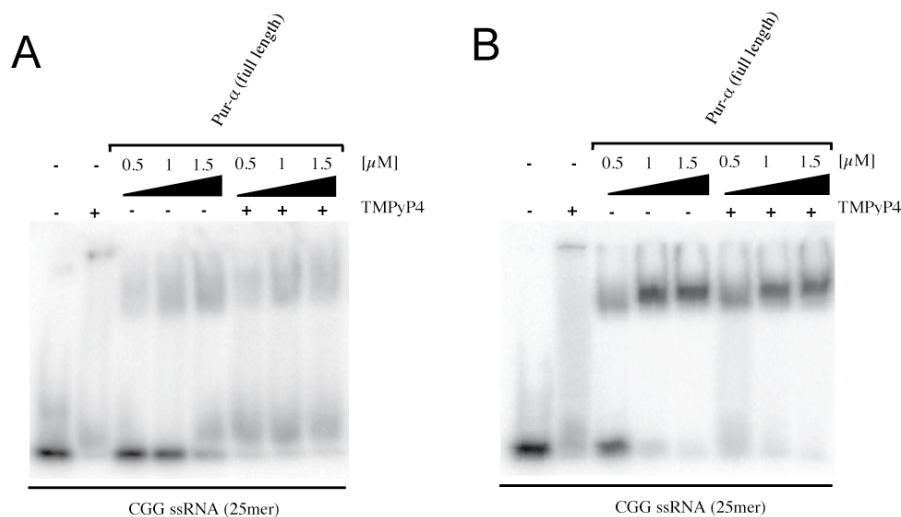
**Figure 2-42. EMSA with *D. melanogaster* Pur- $\alpha$ .**  
**A: Pur- $\alpha$  repeat III with MF0677 ssDNA. B: Pur- $\alpha$  repeat III with CGG 12mer ssRNA.**

## EMSA with quadruplex RNA and full length Pur- $\alpha$

In the 5'-untranslated region of *Fmr1* mRNA, r(CGG)<sub>n</sub> repeats were reported to form quadruplex structures, which are recognized and bound by Fmr protein (Bole et al., 2008; Schaeffer et al., 2001). Therefore it was speculated if Pur- $\alpha$  prefers quadruplex structures as well.

Quadruplexes can form in DNA and RNA sequences rich in guanine (G) content. They are formed by planar G tetrads, which are held together by reverse Hoogsteen base pairing. Several G tetrads can stack together and form quadruplex structures, which depend on the presence of monovalent cations. Potassium and sodium ions stabilize quadruplexes, whereas lithium ions do not, due to their small size (Burge et al., 2006; Williamson, 1994).

It was examined if Pur- $\alpha$  prefers quadruplex structures using EMSAs under quadruplex-promoting and -destabilizing conditions. Besides their dependence on potassium, quadruplexes can be destabilized by intercalation of a cationic porphyrin. In gel shift experiments, the cationic porphyrin 5,10,15,20-tetra(N-methyl-4-pyridyl)porphin (TMPyP4) was used as described (section 5.16.4) (Ofer et al., 2009).



**Figure 2-43. EMSA with *D. melanogaster* Pur- $\alpha$  full length and CGG 25mer ssRNA. A: Potassium containing conditions. B: Potassium-free conditions. TMPyP4 refers to the quadruplex-destablizing drug 5,10,15,20-tetra(N-methyl-4-pyridyl)porphin.**

EMSAs were performed with *D. melanogaster* Pur- $\alpha$  full length and CGG RNA 25mer with gels and protein buffer containing potassium chloride. The observed  $K_D$  is about 1-1.5  $\mu\text{M}$  (Figure 2-43, A). The addition TMPyP4 in the control lane without protein altered the running behavior of the RNA and no deposit band was observed. It is thus not

possible to compare the  $K_D$  with and without the quadruplex-destablizing drug (Figure 2-43, A).

Therefore, the same experiment was performed with gels and protein buffer containing no potassium, but lithium chloride. Quadruplexes cannot form under these conditions. The binding Pur- $\alpha$  to the CGG RNA 25mer exhibited a  $K_D$  of about  $0.5 \mu\text{M}$  (Figure 2-43, B). With the addition of TMPyP4, the shifted band remained, whereas the deposit was smeared as in the control lane containing TMPyP4 and RNA (Figure 2-43, B).

If the binding affinities of Pur- $\alpha$  between quadruplex-promoting (Figure 2-43, A) and quadruplex-preventing (Figure 2-43, B) conditions are compared, the binding appears stronger in the latter case. This could hint a preference of Pur- $\alpha$  for non-quadruplex RNA.

## 2.2.2 Protein binding of Pur- $\alpha$

### 2.2.2.1 Yeast-two-hybrid screen

To identify protein binding partners of Pur- $\alpha$ , a yeast-two-hybrid screen was commissioned to Hybrigenics (Paris, France). This technique allows for the assessment of non-covalent protein-protein interactions and is based on the modular structure of transcription factors. A bait protein is fused to the DNA-binding domain of a transcription factor and a number of prey proteins are fused to its activation domain. Yeast strains containing plasmids carrying either one of the two hybrid proteins are mated with each other. When bait and prey proteins interact with each other, the domains of the transcription factor come together. This stimulates specific transcription of the reporter gene, allowing for growth under selective conditions. Positive clones are then sequenced to reveal the identity of the prey protein.

As bait protein, a fragment of *D. melanogaster* Pur- $\alpha$  comprised of amino acids 40 to 254 was cloned in vectors pB27 (LexA system) and pB66 (Gal4 system) (Formstecher et al., 2005; Fromont-Racine et al., 1997). Prey proteins were derived from a complex library of randomly primed cDNA fragments from poly(A)<sup>+</sup>RNA that was isolated from adult fly heads (Formstecher et al., 2005). No auto-activation with the bait alone was observed in either system. In the LexA system, 103 million interactions were tested and in the Gal4 system 71 million interactions were probed. This refers to 10-fold (respectively 7-fold) the library complexity and indicates that screening was exhaustive. Positive clones were then sequenced at their 5'- and 3'-ends. With the longest continuous sequence a BLAST search was performed against GenBank, yielding the protein reference (Altschul et al., 1997; Benson et al., 1994). A positive interaction is often detected with several preys that correspond to fragments of the same protein. The common sequence shared by overlapping fragments is referred to as selected interacting domain (SID) (Rain et al., 2001). For each interaction, Hybrigenics calculates a biological score (PBS), which annotates its reliance. The PBS is based on a statistical model for the competition for bait-binding between fragments and represents the probability that the interaction is due to random noise or false positives (Rain et al., 2001). Interactions that are most probable to be specific are ranked in category A. Interactions with a decreased reliance are ranked in categories B and C. Category D includes interactions most probable due to artifacts. Category E reflects unspecific interaction due to highly connected prey domains that have a tendency to yield false positives.



## Interactions with known proteins

The yeast-two-hybrid screen identified seven interaction partners, which correspond to known proteins. These are Arrestin1 (Arr1), Btb VII, CG5758, Connector of AP to kinase (Cka), Eye, Lamin C (LamC), and Pur- $\alpha$  (Table 2-16). Among these proteins, it is noticeable that all interactions except one are based on only one prey fragment. This results in the classification in category D for most of them and hampers the correct assignment of the interaction domain (Table 2-16).

Protein	AA full length	SID	PBS	Number of clones	Screen
Arr1	364	183-347	D	1	Gal4
Btb VII	743	460-743	E	1	LexA
CG5758	625	223-426	D	1	Gal4
Cka	730	407-730	D	1	Gal4
Eye	838	300-493	D	1	Gal4
LamC	621	555-621	C	2	Gal4
Pur- $\alpha$	260	146-260	D	1	LexA

**Table 2-16. Interactions of Pur- $\alpha$  with known proteins found in the yeast-two-hybrid screen. *D. melanogaster* Pur- $\alpha$  40-254 was used as a bait. AA refers to amino acids, SID refers to selected interaction domain, PBS is the biological score assigned by Hybrigenics.**

### Arrestin1

Arrestin1 (Arr1) was found as an interaction partner in the Gal4 system and was assigned to category D (Table 2-16). Arrestin1 is specifically expressed in the retina and functions as a multifunctional adaptor and transducer molecule (Lefkowitz and Shenoy, 2005). It regulates and mediates the signal transduction via G-protein coupled receptors. Arrestin1 silences the receptor by binding to the light-activated phosphorylated rhodopsin, leading to receptor endocytosis and desensitizing of the cell (Sato and Ready, 2005). Arrestins shuttle between nucleus and cytoplasm. They have a dipartite structure with an N-terminal and a C-terminal domain. The found interacting region for Pur- $\alpha$  (Table 2-16) maps to the C-terminal domain (amino acids 179-352), which is also involved in rhodopsin binding (Hirsch et al., 1999) (section 3.3.2.2).

## **Btb VII**

Btb VII was found in the LexA screen and was listed in category E, since it represents a highly connected, “sticky” protein (Table 2-16). The Btb protein-protein interaction domain is conserved throughout eukaryotes and exposes a large, versatile interaction surface. Btb proteins exert diverse functions, and are frequently included in transcriptional regulation and protein degradation (Perez-Torrado et al., 2006) (section 3.3.2.1).

## **CG5758**

In the Gal4 screen, an interaction to CG5758 was found and assigned to category D (Table 2-16). The protein CG5758 was first discovered in a DNA microarray screen searching for genes with potential roles in the development of the wing disc of the *Drosophila* embryo (Butler et al., 2003). Butler and colleagues found that the CG5758 transcript was enriched in the dorsal hinge and predicted a role in cell adhesion for the gene product (Butler et al., 2003; Jacobsen et al., 2006). The selected interaction domain of CG5758 with Pur- $\alpha$  (Table 2-16) is a predicted extracellular Fas domain (section 3.3.2.1).

## **Cka**

The interaction with Connector of kinase to AP-1 (Cka) was found in the Gal4 screen and assigned to category D (Table 2-16). Cka is a regulatory kinase involved in the JUN N-terminal kinase signal transduction pathway in *Drosophila* embryogenesis. This pathway controls the epithelial cell sheet movement during dorsal closure in the embryo. Cka regulates AP-1 activity and thus controls the spatial-temporal expression of AP-1 regulated genes (Chen et al., 2002). The found fragment (Table 2-16) corresponds to a known protein-protein interaction domain that is a WD40-repeat domain (section 3.3.2.2).

## **Eye**

The interaction to Eye protein was observed in the Gal4 screen and scored in category D (Table 2-16). The *eyeless (eye)* gene is a master regulatory gene for eye development in the *Drosophila* embryo. Eye, which is the *Drosophila* homolog of human Pax6, is also essential to control differentiation and function of neurons of the adult brain (Callaerts et al., 2001). It acts as a transcription factor and is suggested to regulate several genes involved in learning and memory (Callaerts et al., 2001). Eye is composed of four domains: at the N-terminus the paired domain, followed by the linker region connecting to the homeodomain, and the C-terminal domain. The paired domain and the homeodomain are DNA-binding regions, whereas the C-terminal domain is the transactivating domain of the transcription factor (Clements et al., 2009; Walther and Gruss, 1991). The fragment interacting with Pur- $\alpha$  in the yeast-two-hybrid screen (Table 2-16) contains the homeodomain (amino acids 398-470) (section 3.3.2.2).

## **Lamin C**

In the Gal4 screen, the interaction to Lamin C (LamC) was scored in category C, because it was observed for two clones. However, the two LamC fragments were identical (Table 2-16). Lamin C belongs to the lamin family of intermediate filaments, which localize to the inner surface of the nuclear envelope (Schulze et al., 2005). It was previously believed that lamins do solely serve as mechanical support, but today there are hints that lamins also regulate tissue-specific gene expression (Worman and Courvalin, 2002). Of particular interest are laminopathies, which are caused by mutations in lamin proteins and cause disorders of muscle, fat and nerve cells (Worman and Courvalin, 2002). A-type lamins (Lamin C and Lamin A) are composed of a N-terminal rod domain followed by a C-terminal globular domain. The latter exhibits an Ig-like structure and contains mutation sites for several inheritable laminopathies (Krimm et al., 2002). The proposed interaction domain for Pur- $\alpha$  (Table 2-16) corresponds to a part of the C-terminal domain (amino acids 465 to 582), but is not identical with it (section 3.3.2.2).

## Pur- $\alpha$

The self-interaction of Pur- $\alpha$  was confirmed in the LexA screen and assigned to category D (Table 2-16). Only one clone was positive, bearing a fragment of Pur- $\alpha$  comprised of amino acids 146 to 260. This corresponds to a part of PUR repeat II in addition to the entire PUR repeat III (section 2.1.1.3 and section 3.3.2.2).

## Interactions with non-existing proteins

A significant proportion of positive preys does not correspond to existing proteins (Table 2-17). Sequencing of those clones yields out-of-frame fragments, antisense fragments, and fragments located in untranslated regions of the mRNA. Because such preys do not correspond to existing proteins, no confidence score was given by Hybrigenics.

Protein	minimal Start/Stop (nt)	Comment	Number of clones	Screen
Argk	1675-345	antisense	4	Gal4
Cac	7659-6257	antisense, fully in 3'-UTR	5	Gal4
Camta	351-32	out-of-frame	1	Gal4
Ced-6	1839-1362	antisense	3	Gal4
E(Pc)	5225-6849	out-of-frame	1	Gal4
GstS1	998-1232	out-of-frame, fully in 3'-UTR	1	Gal4
inaF	674-1363	out-of-frame	9	Gal4
ND75	1469-2303	out-of-frame	1	Gal4
Pcm	2140-1856	antisense	1	Gal4
Pde1c	3910-3355	out-of-frame, fully in 3'-UTR	2	Gal4
Rab5	789-128	out-of-frame	1	Gal4
scrt	610-1004	out-of-frame	1	Gal4
Socs16D	...-2476	antisense	1	Gal4

**Table 2-17. Interactions of Pur- $\alpha$  with non-existing proteins found in the yeast-two-hybrid screen. *D. melanogaster* Pur- $\alpha$  40-254 was used as a bait. nt refers to nucleotides.**

Interaction with so-called non-existing proteins do not yield a coding region when their sequence is used as a query for a BLAST search against GenBank (Altschul et al., 1997; Benson et al., 1994). In some cases, the sequence aligns with chromosomal DNA. These

are referred to as GenMatches by Hybrigenics (Table 2-18). The highest PBS (category A) was achieved for a GenMatch (GID 158246), which was found in as much as 32 different fragments (Table 2-18).

Gene	minimal Start/Stop (nt)	Number of clones	Screen
GID 112980799	1-...	1	Gal4
GID 113193577	1-107	3	Gal4
GID 113194556	1-638	1	Gal4
GID 113194865	1-551	1	Gal4
GID 113194944	1-314	1	Gal4
GID 113194944	1-352	1	Gal4
GID 158246	36-231	32	Gal4
GID 55380579	1-391	1	Gal4
GID 55380579	1-856	1	Gal4

**Table 2-18. GenMatch interaction partners of Pur- $\alpha$  found in the yeast-two-hybrid screen. *D. melanogaster* Pur- $\alpha$  40-254 was used as a bait. nt refers to nucleotide.**

## Chapter 3 Discussion

### 3.1 Structure of Pur- $\alpha$

#### 3.1.1 PUR repeats are highly conserved structural units

Amino acid sequence analysis of metazoan Pur- $\alpha$  reveals that it is composed of three repetitive units, which are referred to as PUR repeats. J. Söding, employing the HHrepID server, did the assignment of the repeats, whose accuracy is confirmed by structural characterization. In the crystal structure of *B. burgdorferi* Pur- $\alpha$  as well as in the crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II, the folded domains superpose exactly with the predicted repeat boundaries (Figure 2-6, Figure 2-12, and Appendix 7-2).

The PUR repeats are located in the core region of Pur- $\alpha$ , which is required and sufficient for nucleic-acid binding and mediates many protein interactions (White et al., 2009). However, they are not identical to the previously described three class I and two class II repeats, which were also mapped in this region (Gallia et al., 2000). The detection of PUR repeats was made possible by advances in bioinformatic analysis, such as the HHrepID server. This program is very sensitive to distant homology, especially for repeat proteins (Biegert and Söding, 2008). The PUR repeats in *D. melanogaster* share an overall amino acid sequence identity of approximately 30 % (Table 2-5). Amino acid sequence alignment shows that the PUR repeats are conserved in orthologs from different species. Pur- $\alpha$  from *Caenorhabditis elegans*, *Arabidopsis thaliana*, and human bear also three PUR repeats, whereas *B. burgdorferi* Pur- $\alpha$  has only one PUR repeat (Figure 2-2 and Appendix 7-2).

#### 3.1.2 *B. burgdorferi* Pur- $\alpha$ is a functional PUR protein

The detection of PUR repeats enabled the identification of a distant bacterial homolog, *B. burgdorferi* Pur- $\alpha$ . Its core region (*Borrelia* PUR) shares an identity of about 20 % with the PUR repeats of *D. melanogaster* Pur- $\alpha$  (Figure 2-2).

In addition to sequence identity, experimental evidence confirmed that *B. burgdorferi* Pur- $\alpha$  is indeed a functional PUR protein. Nucleic-acid binding properties were examined in filter binding assays with DNA and RNA targets that were previously reported to bind human Pur- $\alpha$  (section 2.2.1). Binding to DNA oligomers was specific ( $K_D = 400$ - $500$  nM), and showed affinities comparable to human Pur- $\alpha$  ( $K_D = 200$ - $400$  nM). This suggests a conserved functionality and therefore encouraged the pursuit of the

*B. burgdorferi* Pur- $\alpha$  crystal structure (section 2.1.1.2). However, no binding was detected to the RNA oligomers. It cannot be judged if this result means that *B. burgdorferi* Pur- $\alpha$  has no RNA-binding ability, since there may exist unknown, species-specific RNA targets.

Because *B. burgdorferi* Pur- $\alpha$  is annotated as a hypothetical protein, no comment on its *in vivo* functionality is attempted. However, one might speculate if parasitic *Borrelia* has adopted this gene from a metazoan host, as PUR proteins are generally conserved in metazoan, but only found in very few bacterial families.

*B. burgdorferi* Pur- $\alpha$  was crystallized and its structure was solved at 1.9 Å ( $R_{\text{work}} = 18.7\%$ ,  $R_{\text{free}} = 25.0\%$ ) (section 2.1.1.2 and Figure 2-6). Based on the crystal structure and the sequence alignment, a homology model of *D. melanogaster* Pur- $\alpha$  was build by J. Soeding, employing the HHpred server (Appendix 7-1) (Söding et al., 2005). The domain boundaries predicted by the model allowed for the design of crystallizable metazoan Pur- $\alpha$  fragments, apparently for the first time.

### 3.1.3 Two PUR repeats form a PUR domain

A fragment of *D. melanogaster* Pur- $\alpha$  composed of the first two PUR repeats was crystallized and its structure was solved at 2.1 Å ( $R_{\text{work}} = 22.3\%$ ,  $R_{\text{free}} = 24.0\%$ ). The comparison with the structural model of *B. burgdorferi* Pur- $\alpha$  reveals a highly conserved fold with a  $\beta\beta\beta\alpha$ -topology for each PUR repeat. In both crystal structures, two PUR repeats intertwine into each other to form a globular PUR domain. The PUR domain is structurally conserved in both homologs with an RMSD of 2.3 Å (Figure 2-15). In the case of *B. burgdorferi* Pur- $\alpha$ , PUR domain formation is achieved by inter-molecular dimerization, whereas the fragment of *D. melanogaster* Pur- $\alpha$  bears two PUR repeats and forms an intra-molecular dimer (Figure 2-6 and Figure 2-12).

To distinguish between a nonspecific crystal contact and a specific interaction with biological relevance, the interaction surface of proteins in a crystal is analyzed. The combination of size, shape and complementary of the contact area allows for the prediction, if the crystallized state reflects the solution structure of the protein (Bahadur and Zacharias, 2008). The buried surface interface is 1830 Å<sup>2</sup> per repeat in the *D. melanogaster* case, and 2058 Å<sup>2</sup> in the *B. burgdorferi* structure (section 2.1.1.2 and 2.1.1.3). These values are typical for heterodimers and significantly exceed those obtained by average crystal packing (Bahadur and Zacharias, 2008). Moreover, the shape

of the contact surface is typical for a specific interaction, as it is formed by one large patch without cavities and does not include water molecules. Typical is also the high number of aromatic and aliphatic residues on the buried surface and its depletion in charged residues, with a clear separation of hydrophobic core residues and polar rim residues (Bahadur and Zacharias, 2008). Taken together, the nature of the interaction surface indicates that the PUR domain is not a crystallographic artifact, but stable and specific.

This conclusion is further confirmed by the assessment of the oligomeric state of Pur- $\alpha$  in solution. In analytical size-exclusion chromatography, the crystallized fragment of *D. melanogaster* Pur- $\alpha$  repeats I-II eluted at a volume corresponding to a monomer. From the calibration curve, a molecular weight of 17.4 kDa was calculated. This is in good accordance with the theoretical molecular weight (16.8 kDa) of the fragment. Since the size-exclusion column was calibrated with globular protein, the good accordance also hints at a globular shape of the fragment, as given by PUR domain formation.

The monomeric state was also confirmed by molecular weight calculation from SAXS data, yielding 17.2 kDa (section 2.1.2.2 and Table 2-10). Moreover, the experimental scattering curve nicely superposes with the theoretical scattering curve calculated from the structural model. Consistently, the crystal structure fits into the bead model calculated from SAXS data (section 2.1.2.2 and Figure 2-25).

It is thus concluded that the PUR domain also constitutes the solution state of PUR repeat I and PUR repeat II of *D. melanogaster* Pur- $\alpha$ .

### **3.1.4 The PUR domain is similar to bacteriophage coat and Whirly proteins**

The crystal structure of a protein often reveals previously unknown relationships to other proteins, from which a functional correlation can be deduced.

Systematic structural comparison for the *B. burgdorferi* Pur- $\alpha$  crystal structure yielded a similarity to bacteriophage coat proteins from *Pseudomonas* phage pp7 and *Enterobacterio* phage ms2. Both coat proteins show a similar dimerization mode as *B. burgdorferi* Pur- $\alpha$ . The dimers have  $\alpha$ -helices on one side and an extended  $\beta$ -sheet on the opposing side (Figure 2-8). The bacteriophage coat proteins have a similar topology as PUR repeats. However, a pronounced difference is that they form a flat  $\beta$ -sheet surface, not a roof-like structure as a PUR domain. Interestingly, the  $\beta$ -sheet sides of both



bacteriophage coat proteins have been previously shown to bind RNA (Chao et al., 2008; Ni et al., 1995). The present study found that the  $\beta$ -sheet side of the PUR domain is required for DNA and RNA binding. Both bacteriophage proteins are evolutionary related, as reflected by their amino acid sequence identity (similarity) of 24 % (50 %). The lack of significant sequence similarity to *B. burgdorferi* Pur- $\alpha$  speaks for a convergent evolution of the nucleic-acid binding surfaces.

Surprisingly, the DaliLite server (Holm et al., 2008) identified different structural neighbors for the highly homologous structure of *D. melanogaster* Pur- $\alpha$  repeats I-II. This might reflect a limitation of the DaliLite search method, as the only significant difference between both structures is that one is formed by an inter-molecular dimer and the other one by an intra-molecular dimer (Figure 2-15).

The identified structural neighbors revealed that Pur- $\alpha$  belongs to the Whirly class of nucleic-acid binding proteins. This structural relationship is more pronounced than the one to the bacteriophage coat proteins, as the PUR domain shares topology as well as overall orientation with the Whirly domain and nicely superposes with it (Figure 2-15). However, Pur- $\alpha$  shows no significant sequence homology to the Whirly proteins MRP1/MRP2 and P24. Furthermore, the proteins differ in their quaternary structure. Whereas the Whirly proteins tetramerize via interactions of their  $\alpha$ -helical sides, such an orientation is not seen in the crystal lattice of Pur- $\alpha$  repeats I-II. The analysis of the oligomeric state of *D. melanogaster* Pur- $\alpha$  demonstrates that it does not tetramerize in a comparable fashion, but dimerizes via PUR repeat III (section 2.1.2).

Intriguingly, Pur- $\alpha$  also shares a functional similarity with Whirly proteins. MRP1/MRP2 was previously reported to bind ssRNA, whereas P24 has been shown to bind ssDNA as well as dsDNA (Desveaux et al., 2002; Schumacher et al., 2006). Pur- $\alpha$  combines the functions of both proteins as it binds both types of nucleic acids (section 2.2.1). As a recurrent feature, the  $\beta$ -sheets of the Whirly proteins have been shown to mediate nucleic-acid binding. However, the mode of binding is presumably different in Pur- $\alpha$ . On the one hand, both  $\beta$ -sheets of the PUR domain are involved in nucleic-acid binding, whereas in MRP1/MRP2 only one of the two  $\beta$ -sheets is a functional RNA-binding motif. Moreover, MRP1/MRP2 binds RNA unspecifically, and in the co-complex with gRNA (PDB-ID 2GJE), the nucleobases are oriented away from the protein surface. For sequence-specific binding of Pur- $\alpha$ , an orientation of the bases towards the protein surface is more likely.

As for Pur- $\alpha$ , ATP-independent nucleic acid unwinding properties were reported for the Whirly proteins (Darbinian et al., 2001a). As seen in the crystal structure, the binding of MRP1/MRP2 to hairpin RNA requires unwinding of the latter (Schumacher et al., 2006). P24 has been suggested to exploit melted regions of DNA and to intercalate the entire tetramer between the two strands (Desveaux et al., 2002). Similarly, one might speculate that the Pur- $\alpha$  dimer intercalates into a DNA duplex by binding each strand with one of its binding domains (section 3.2.1 and Figure 3-2, A).

### **3.1.5 Dimerization of Pur- $\alpha$ is mediated by PUR repeat III**

This study demonstrates that *D. melanogaster* Pur- $\alpha$  form dimers in solution, as established by analytical size-exclusion chromatography and SAXS experiments. The use of truncation mutants of *D. melanogaster* Pur- $\alpha$  demonstrated that dimerization requires PUR repeat III.

A fragment of Pur- $\alpha$  composed of repeats I-II is monomeric in solution, as demonstrated by analytical size-exclusion chromatography and SAXS analysis (section 2.1.2 and section 3.1.3). In contrast, analytical size-exclusion chromatography with a fragment bearing PUR repeats I-III yielded an elution volume more consistent to a dimer. From the elution peak volume, a molecular weight of 36.5 kDa was calculated and theoretical molecular weight is 25.2 kDa. Dimerization was observed over a wide range of concentrations (5  $\mu$ M to 200 nM) (Figure 2-22). The deviation from the ideal value for a dimer is most likely due to the fact that the protein exists in a dynamic monomer-dimer equilibrium. Equivalent results were obtained when the molecular weight of the fragment was calculated from SAXS analysis, giving a molecular weight of 46.0 kDa (Table 2-11). The deviation from the ideal volume might be due to a slight instability of the dimer. A portion of Pur- $\alpha$  might be present in monomeric form. This could also explain that the Kratky plot of the Pur- $\alpha$  repeats I-III hints at some unstructured regions (Figure 2-23 and Figure 2-26). One might speculate that the third repeat is unstructured, unless it has an interaction partner (section 3.1.6).

In summary, size-exclusion chromatography and SAXS analysis showed that the third repeat induces dimerization of Pur- $\alpha$ . Therefore, a dimerization model was proposed that links two Pur- $\alpha$  molecules via the third repeat (Figure 2-15).

In the present study, dimerization of Pur- $\alpha$  was independent of nucleic acid and its absence was confirmed by the absorption spectra at wavelengths 260 nm and 280 nm (section 5.8).

Gallia and colleagues reported that self-association of Pur- $\alpha$  requires RNA (Gallia et al., 1999). They used GST-tagged human Pur- $\alpha$  coupled to glutathione-sepharose beads to pull down Pur- $\alpha$ . They showed that RNase, but not DNase, abolished self-association of Pur- $\alpha$  and that binding was restored by the addition of specific *PU-RNA*. The disagreement could be explained by the dimerization of Pur- $\alpha$ . One might assume that Pur- $\alpha$  could not be pulled down in the absence of nucleic acid, because all Pur- $\alpha$  molecules in the pull down assay are already dimeric. There are no free PUR repeats III left to form additional interactions. The putative fraction of monomeric Pur- $\alpha$  that is discussed above is presumably too small to pull down significant amounts of the protein. One might speculate that the self-association they observed in the presence of *PU-RNA* (Gallia et al., 1999) is indeed due to the cross-linking of Pur- $\alpha$  by RNA. The present study shows that one PUR domain is sufficient to bind nucleic acids hence one Pur- $\alpha$  dimer has (at least) two independent binding surfaces (section 2.2.1). Large complexes could form by the simultaneous interaction of the RNA with two or more different Pur- $\alpha$  dimers (Figure 3-2, C). In accordance, the 290 nt *PU-RNA* includes at least two putative Pur- $\alpha$  binding sites with the consensus sequence (GGN) (Tretiakova et al., 1998). It should be noted that Gallia and colleagues do not comment on the oligomeric state of the observed self-association (Gallia et al., 1999).

Ramsey and colleagues investigated the quaternary structure of recombinant mouse Pur- $\beta$ , which shares a sequence identity of 70 % with Pur- $\alpha$  (Ramsey et al., 2007). They subjected nucleic-acid free Pur- $\beta$  to dynamic light scattering and analytical ultracentrifugation and ascertained a dynamic monomer-dimer equilibrium. Moreover, their molecular shape calculations resulted in an elliptical shape for the homodimer (Ramsey et al., 2007). This is in good accordance with the SAXS analysis and the proposed dimerization model of the present work (section 2.1.2.2 and Figure 2-27).

### 3.1.6 PUR repeat III is predicted to form a PUR domain

Although there is no direct proof, several lines of evidence suggest that the third PUR repeat in *D. melanogaster* Pur- $\alpha$  also forms a PUR domain – potentially with another repeat III from a different peptide chain.

Firstly, the amino acid sequence of *D. melanogaster* PUR repeat III is homologous to the three other PUR repeats described in this study. Its identity (similarity) to *D. melanogaster* PUR repeat I, and *D. melanogaster* repeat II and *B. burgdorferi* PUR is 24 % (53 %), 34 % (56 %), respective 13 % (38 %). The degree of conservation between the latter three is similar, ranging from 19 % (47 %) between PUR repeat I and *B. burgdorferi* PUR to a maximum of 33 % (61 %) between PUR repeat I and PUR repeat II (section 2.1.1.3 and 2.1.1.2). Since *D. melanogaster* PUR repeat I, *D. melanogaster* PUR repeat II and *B. burgdorferi* PUR adopt the same fold with an RMSD of only 2.3 Å (section 2.1.1.3 and Figure 2-15), it seems reasonable to assume that PUR repeat III also complies with their overall fold. Consistently, secondary structure prediction with the NPS server (Combet et al., 2000) confirms the assumption of a  $\beta\beta\beta\beta\alpha$ -topology for the third repeat of *D. melanogaster* Pur- $\alpha$ . The position of the  $\alpha$ -helix is predicted exactly at the position that is expected by the sequence homology to PUR repeat I and II. The number of  $\beta$ -strands is less clear, however their predicted start site coincides with the beginning of the conserved region.

Secondly, the third repeat bears conserved hydrophobic residues that could stabilize a PUR domain. The anticipated driving force of PUR domain formation is hydrophobic interaction and a number of hydrophobic residues was identified in the respective buried surface in the crystal structures of *B. burgdorferi* Pur- $\alpha$  and *D. melanogaster* Pur- $\alpha$  repeats I-II (section 2.1.1.2 and section 2.1.1.3). Amino acid sequence alignment shows that the positions of these residues are highly conserved (Figure 3-1). In total, 39 hydrophobic residues are located in the buried surfaces of both crystal structures. The great majority, 33 residues, is found in corresponding positions in each repeat, yielding 11 conserved amino acid positions likely involved in dimerization. It is striking that the third PUR repeat of *D. melanogaster* Pur- $\alpha$  offers hydrophobic residues in all of these 11 positions (Figure 3-1).

```

dm PUR-I  EQE L A T K M - Q I Q S K R F Y I D V K Q N R R S R - F I K V A E I G A D G - - - - R R N P L G S F - T A A E R D H L S S F S D Y Y A S L
dm PUR-II DGK L K S - E M M I K D Y R R Y I D V K E N A R G R - F L R V S Q T I T - - R G G - P R S Q I A L P A Q G M I E R D A L T D L L E E F G A N
dm PUR-III - - - L P E E R H M K V D N K N F Y F D I G Q N N R G - V Y M R I S E V K N N - - - - F R T S I T I P E K C W I R F R D I F N D Y C E K M K K S
bb PUR    R G E V Y S E K L F T E S E R T V F F N V K E N R K G D Y F I N I V E S K R S P S G D F E R H S I F V Y E E N I N E E S N L L K A I A V I K Q K

```

**Figure 3-1. Amino acid sequence alignment of *D. melanogaster* PUR repeats I, II and III and *Borellia* PUR. Hydrophobic amino acids are depicted in light green, hydrophobic residues located in the buried surface of the respective PUR domain are shown in dark green.**

Finally, the experimentally observed dimerization of Pur- $\alpha$  repeats I-III hints at PUR domain formation of repeat III. Dimerization of Pur- $\alpha$  in dependence of the third repeat was confirmed by analytical size-exclusion chromatography and SAXS analysis (section 2.1.2 and 3.1.5). In this study, the most intuitive and simple dimerization mode is proposed (section 2.1.2.2 and Figure 2-27). In the model, PUR repeats I and II interact with each other as seen in the crystal structure, whereas PUR repeat III interacts with another PUR repeat III from a second Pur- $\alpha$  molecule (Figure 2-27). The hypothesis that both repeats III attach to each other by PUR domain formation is in accordance with the surface envelope calculated from SAXS data. The Z-shaped bead model is able to accommodate three PUR domains, while for a good fit, the one potentially formed by the repeats III has to be placed perpendicular to the other two (Figure 2-28).

In conclusion, the sequence homology to the other PUR repeats, the conservation of residues of presumed structural importance, and the experimentally observed dimerization speak for a PUR domain formation of PUR repeat III in *D. melanogaster* Pur- $\alpha$ .

### 3.1.7 Interaction of the PUR repeats is likely specific

Given the above elaborations about the homology of the third repeat (section 3.1.6), the immediate question arises if PUR domain formation is specific. It should be noted that there is no direct evidence for specific interaction of two repeats III and for their PUR domain formation.

The PUR repeats in *D. melanogaster* Pur- $\alpha$  may also have some promiscuity in binding, meaning that the interaction observed in the crystal structure is not the only possible combination of repeats. In the crystal structure, PUR repeat II interacts with PUR repeat I (Figure 2-12), thus the proposed dimerization model includes their interaction and the interaction of two PUR repeats III (Figure 2-27). One could speculate that PUR repeat II

may also interact with PUR repeat III, leaving PUR repeat I to bridge to another peptide chain by binding to another PUR repeat I. By analytical size-exclusion chromatography or SAXS analysis, this dimerization mode would be indistinguishable from the proposed one (section 2.1.2).

However, if the repeats were freely interchangeable, also cross-links of more than two Pur- $\alpha$  molecules could be possible. In the analytical size exclusion chromatography (maximum concentration 0.126 mg/mL) and in the SAXS analysis (maximum concentration 1.53 mg/mL) no species with a higher molecular weight than the dimer were observed. Therefore mutually exclusive interaction of the PUR repeats and the proposed dimerization mode appear most plausible (Figure 2-27).

Specific PUR domain formation may be driven by some structural peculiarity of the third repeat. To gain understanding of this issue, crystallization trials of the *D. melanogaster* Pur- $\alpha$  repeat III are ongoing (section 2.1.1.3). There might well be specific contacts on the interface between the PUR repeats III that help to discriminate between the other repeats. It is also possibly that the length of the linker between the PUR repeats has an influence on specificity. In *D. melanogaster* Pur- $\alpha$ , the linker between PUR repeats I and PUR repeat II has a length of 24 amino acids, whereas between PUR repeat II and PUR repeat III, it is only 3 amino acids long (Appendix 7-2).

### **3.1.8 The PUR domain is likely conserved in different orthologs**

Based on the high degree of sequence conservation, it can be speculated that Pur- $\alpha$  from different species share a common fold, as it is the case for *B. burgdorferi* and *D. melanogaster* Pur- $\alpha$  (Figure 2-6 and Figure 2-12).

The amino acid sequence alignment for *D. melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana* and human Pur- $\alpha$  shows the high degree of conservation between them (Appendix 7-2). The secondary structure prediction is based on the sequence alignment and the crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II (Appendix 7-2). Notably, sequence conservation follows a typical pattern. As expected, the degree of conservation correlates with structured regions and is high in the predicted PUR domains and low in linker regions. This may indicate that the domain boundary assignment is correct for the orthologs. Moreover, in a PUR repeat, sequence conservation is typically higher around the  $\beta$ -sheets than in the  $\alpha$ -helices. On the one hand, this could denote that

the  $\beta$ -sheets are of functional importance, as confirmed by nucleic-acid binding experiments and mutational studies (section 2.2.1). On the other hand, this could hint at the point that the  $\beta$ -sheets are required for fold stabilization, as supported by the assessment of the buried surface interfaces (section 2.1.1).

## **3.2 Interaction with nucleic acids**

### **3.2.1 The PUR domain binds nucleic acids**

The crystallized PUR domain (*D. melanogaster* Pur- $\alpha$  repeats I-II) was found to be a functional nucleic-acid domain by EMSAs and filter binding assays (section 2.2.1.3).

Assessment of the degree of conservation and the electrostatic potential of the surface of the PUR domain led to the hypothesis that the  $\beta$ -sheets form the nucleic-acid binding surface (Figure 2-13 and Figure 2-14). For the PUR domain of *B. burgdorferi* Pur- $\alpha$ , this was confirmed by filter binding assays. It was found that ssDNA binding was abolished when arginines located in the fourth  $\beta$ -strand (R49) of the  $\beta$ -sheet or in the connector between its second and third  $\beta$ -strand (R28, R29) were replaced by alanines (Figure 2-30 and Table 2-14). Thus the  $\beta$ -sheets of *B. burgdorferi* Pur- $\alpha$  are involved in DNA-binding. The corresponding amino acids conserved in *D. melanogaster* Pur- $\alpha$  were subjected to mutagenesis and EMSAs were performed with wild type and mutant PUR domains (see 2.3.1.3). In the  $\beta$ -sheet formed by the first PUR repeat, the mutation of arginine 80 to alanine (R80) had only moderate effect on ssDNA and ssRNA binding. The same was observed for the corresponding mutation (R158A) in the  $\beta$ -sheet of the second PUR repeat. In contrary, the double mutant R80A, R158A showed a total loss of affinity to the nucleic acid probes (section 2.2.1.3). This demonstrates that both  $\beta$ -sheets are functionally important.

However, one might speculate if both  $\beta$ -sheets bind simultaneously to one nucleic acid, or if they independently bind two distinct nucleic acids. In the first scenario, the loss of binding affinity of the double mutant can be explained with a cooperative binding that needs both arginines for full affinity. In the second scenario, the observed higher  $K_D$  would indeed not refer to a loss in affinity, but to a reduced number of active binding sites.

The separation of the two conserved  $\beta$ -sheets by an unconserved rim rather suggest that they act as independent binding surfaces (Figure 2-14). Furthermore, the differences of PUR repeat I and PUR repeat II in surface conservation and electrostatic surface potential suggest that they might have evolved to bind distinct nucleic acids (section 2.1.1.3, Figure 2-13 and Figure 2-14). Even if both  $\beta$ -sheets denote independent binding surfaces, it is still unknown if the PUR domain can bind two nucleic-acid chains simultaneously. Provided this is the case, this would mean that the Pur- $\alpha$  dimer has even four nucleic-acid binding surfaces (section 3.2.4, Figure 3-2). One might speculate if DNA duplex destabilization by Pur- $\alpha$  (Darbinian et al., 2001a) is achieved by intercalation of the PUR domain into the dsDNA. In this scenario, each  $\beta$ -sheet might bind one strand of the DNA (section 3.2.4 and Figure 3-2, A).

One should note that even if both  $\beta$ -sheets have autonomous binding properties, it is still likely that one PUR repeat alone is not functional, but requires an interaction partner for fold stabilization by PUR domain formation. The curved  $\beta$ -sheets of a PUR domain have a shape that seems to be well suited to accommodate a nucleic-acid stretch. Shape complementarity is a common feature of RNA-protein recognition, with the typical interplay between a convex RNA that fits into a concave protein surface (Bahadur et al., 2008).

For a deeper insight into the nucleic-acid binding mode of Pur- $\alpha$ , a more detailed mutational analysis and, importantly, the assessment of binding stoichiometry are required. A crystals structure of a co-complex with RNA or DNA is highly desired. Therefore, crystallization trials with *D. melanogaster* Pur- $\alpha$  repeats I and II and DNA oligonucleotides (section 2.1.1.3) are ongoing. A co-structure would also contribute to the understanding of the determinants of specific affinity of Pur- $\alpha$ .

### **3.2.2 PUR repeat III only weakly contributes to nucleic-acid binding**

It was then assessed if the PUR repeat III, which was not in the crystal structure, has also nucleic-acid binding properties. For *D. melanogaster* Pur- $\alpha$  repeat III, binding affinities for ssDNA and ssRNA were 8fold, respectively 5fold decreased compared to Pur- $\alpha$  repeats I-II (Figure 2-42). Since Pur- $\alpha$  repeats III was stable in solution and also crystallized (section 2.1.1.3 and Figure 2-18), it seems unlikely that the weaker affinity is



due to a folding defect. Furthermore, full length Pur- $\alpha$  had no increased affinities compared to the fragment containing only PUR repeats I-II (section 2.2.1.3).

Interestingly, the previously mapped nucleic-acid binding region of human Pur- $\alpha$  (amino acid 54-250) (White et al., 2009) does not include the entire PUR repeat III (amino acids 218-278) (Appendix 7-2). This suggests that the decreased binding capacity of PUR repeat III is not due to the limited number of oligonucleotides tested in this study, but is a general feature and conserved throughout species. It is likely that PUR repeat III is not required for nucleic acid binding, but may have adopted a different function. As shown in this study, the third repeat is required for dimerization (section 2.1.2). In addition, the assessment of previously mapped binding regions (White et al., 2009) reveals that the third repeat mediates many protein interactions (section 3.3.1).

### 3.2.3 Pur- $\alpha$ binds DNA as well as RNA

One intriguing feature of Pur- $\alpha$  is its affinity to DNA as well as to RNA. By filter binding assays, strong affinities of human Pur- $\alpha$  to ssRNA as well as ssDNA oligonucleotides were measured (Table 2-12). A comparable affinity ( $K_D = 15$  nM) was previously reported for human Pur- $\alpha$  and ssDNA from the bovine papilloma virus (BPV-1) minimal origin of replication (Jurk et al., 1996).

For *B. burgdorferi* Pur- $\alpha$ , a functional conservation was observed, as it bound to the same DNA oligomers as human Pur- $\alpha$ , though its affinity was decreased (Table 2-13). No affinity was observed for *B. burgdorferi* Pur- $\alpha$  to the RNA oligomers tested (Table 2-13). This does not imply that *B. burgdorferi* Pur- $\alpha$  has no RNA binding capacity, as there might exist unknown, species-specific RNA targets (section 2.2.1.2).

Filter binding assays with *D. melanogaster* Pur- $\alpha$  resulted in similar equilibrium dissociation constants as for human Pur- $\alpha$ , but their reproducibility was reduced (Table 2-15). This might be due to protein precipitation on the membrane (section 5.16.3). Thus another experimental set up was chosen, that is EMSA. It should be said that at this point of the thesis, the motivation to perform nucleic-acid binding studies changed and therefore different oligomers were tested, enabling direct comparison of both techniques. In the beginning, it was searched for small oligonucleotides with strong affinity, which could be used for co-crystallization assays. But when the crystal structure of *D. melanogaster* Pur- $\alpha$  was solved, it became critical to

confirm functionality of the crystallized fragment with previously described oligomers, e. g. MF0677 ssDNA 24mer (Bergemann et al., 1992).

Somewhat surprisingly, the affinity to r(CGG) repeat RNA ( $K_D = 1 \mu\text{M}$ ) was lower than expected. Jin and colleagues pulled down Pur- $\alpha$  from cell lysate with bead-coupled r(CGG)<sub>105</sub> RNA (Jin et al., 2007). Furthermore, they propose that in the pathogenesis of FXTAS, cellular Pur- $\alpha$  is sequestered from its normal functions by its binding to the r(CGG)<sub>55-200</sub> repeats in premutation *Fmr1* mRNA (Jin et al., 2007). It seems unlikely that such a weak affinity can account for these effects. Maybe there is an unknown factor involved that stabilizes the interaction of Pur- $\alpha$  and the RNA. Possibly the length of the r(CGG) repeats is crucial and the 12mer and 25mer tested in this study were too short for full affinity. A longer stretch of RNA could increase the affinity by back-folding (Figure 3-2, B) or by cross-linking of Pur- $\alpha$  (Figure 3-2, C).

There are other transcription factors known that bind both types of nucleic acids, but little is known about the structural basis for their dual nucleic-acid specificity (Cassiday and Maher, 2002). It would be interesting to address if Pur- $\alpha$  binds DNA and RNA via the same binding mode, that means if the same binding surface and the same amino acid contacts are employed. Structure-guided mutagenesis of *D. melanogaster* Pur- $\alpha$  (section 2.2.1.3) suggests that both DNA and RNA are bound via the  $\beta$ -sheet side of the PUR domain, as the double mutant R80A, R158A abolished binding for both types of nucleic acids. On the other hand, the double mutant R65A, R142A did not affect DNA binding, but reduced the affinity to RNA. Possibly more extended mutational studies and the assessment of various oligomers of each type of nucleic acid can pin down the individual amino acids involved. Furthermore, it is interesting to examine if RNA and DNA are bound competitively. Consistent with this speculation, Tretiakova and colleagues reported that interaction of Pur- $\alpha$  with a specific RNA, named *PU-RNA*, prevented its binding to the *MBI* regulatory motif of the *myelin basic protein* gene (Tretiakova et al., 1998).

The secondary and tertiary structure of RNA and DNA strands with the same sequence are expected to be very different. Thus the question arises whether specificity is achieved in the same way. One might speculate that the RNA mimics the secondary structure of the DNA to exploit the same binding mode (Cassiday and Maher, 2002). For a different binding mode speaks the comparison of protein data bank entries of protein interactions

with DNA or RNA, which reveals similarities, but also differences in binding (Bahadur et al., 2008). Bahadur and colleagues reported that unique features of the RNA are essential in their contribution to protein binding. Typical RNA-protein interactions are formed by the 2'-OH of the sugar, the main chain NH groups, and arginine as well as lysine side chains. In DNA binding, the phosphate backbone is more frequently involved, but not the sugar (Bahadur et al., 2008). Notably, in sequence-specific RNA binding, interactions are often mediated by stacking of aromatic residues with the nucleobases (Auweter et al., 2006). The solvent-accessible surface of the crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II bears the conserved tyrosines 57 and 134 as well as the conserved phenylalanines 68 and 145. It would be interesting to assess if mutation of these amino acids would affect specifically RNA binding, but not DNA binding.

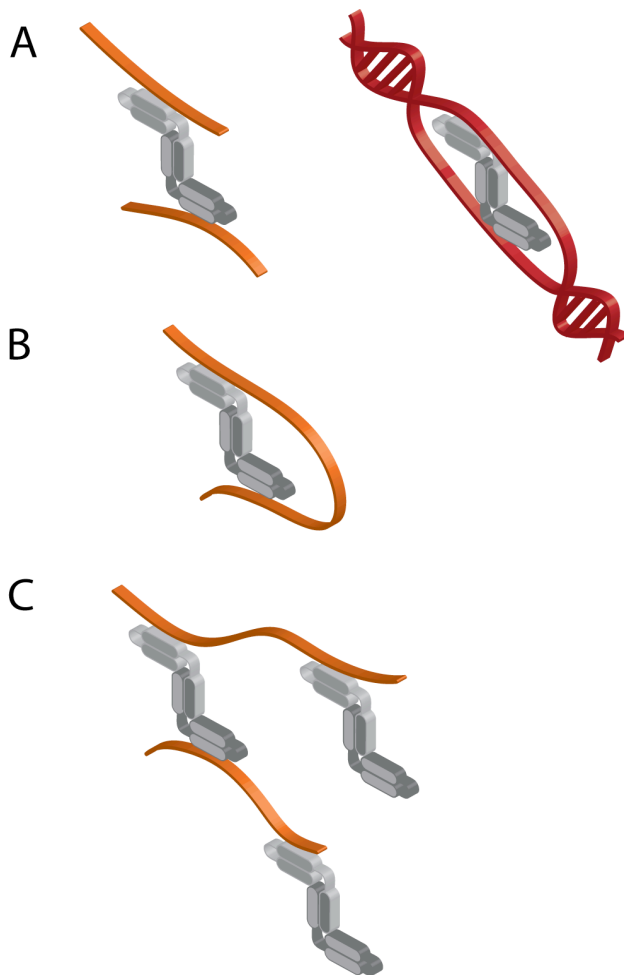
### **3.2.4 The Pur- $\alpha$ dimer bears two nucleic-acid binding domains**

Dimerization of Pur- $\alpha$  (section 3.1.5) implies that each Pur- $\alpha$  molecule possesses two nucleic-acid binding PUR domains. This allows for a variety of potential binding modes for nucleic acids (Figure 3-2).

The Pur- $\alpha$  dimer may bind two nucleic-acid strands simultaneously, each with one PUR domain (Figure 3-2, A). It is tempting to speculate that this binding mode might account for DNA duplex destabilizing properties of Pur- $\alpha$  (Darbinian et al., 2001a). The dimer might exploit melted regions of DNA and intercalate between the strands (Figure 3-2, A). A similar intercalation was previously proposed for the tetramer of the structural homolog of Pur- $\alpha$ , P24 (section 2.1.2.3) (Desveaux et al., 2002). Admittedly, in this scenario, not only a PUR consensus sequence, but likewise the complement strand would have to be bound. As only a preference for (GGN)<sub>n</sub>, not for the complement (CCN)<sub>n</sub> is reported, this interaction would be presumably unspecific. Interestingly, Darbinian and colleagues report that Pur- $\alpha$  destabilizes DNA helices independently of its recognition motif (Darbinian et al., 2001a).

The Pur- $\alpha$  dimer could also employ both its PUR domains to bind one nucleic-acid strand bidentally with both binding surfaces and thereby increasing its affinity (Figure 3-2, B). This seems to be a well suiting binding mode for the interaction with long repetitive sequences, as Pur- $\alpha$  has been shown to bind to repetitive single-stranded nucleic acids, such as (GGN)<sub>n</sub> or r(CGG)<sub>n</sub> repeats (Gallia et al., 2000; Jin et al., 2007). Alas, if the Z-

shaped arrangement of the Pur- $\alpha$  dimer were stable also when nucleic acid is bound, multidendate binding would require extreme bending of the nucleic acid (Figure 3-2, B). Consistent with this idea, Chao and colleagues recently proposed that binding of  $\beta$ -Actin mRNA to ZBP1 induces a 180° loop in the RNA (Chao et al., 2010). Still, the required flexibility might be well possible for RNA, but it is more unlikely in the case of dsDNA. However, conformational flexibility may be achieved by the duplex unwinding properties of Pur- $\alpha$  (Darbinian et al., 2001a). On the other hand, it cannot be ruled out that Pur- $\alpha$  itself undergoes conformational changes upon nucleic-acid binding.



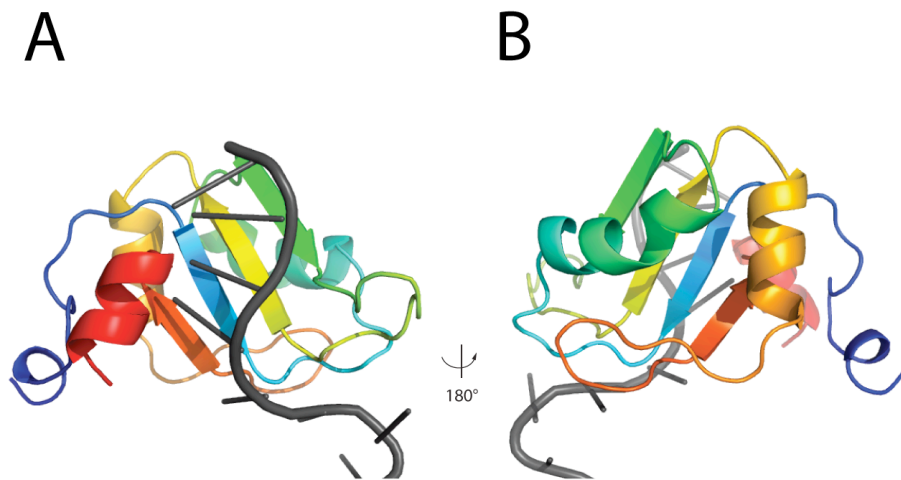
**Figure 3-2. Potential binding modes for Pur- $\alpha$  and nucleic acids. The Z-shaped Pur- $\alpha$  dimer is depicted in grey, with one peptide chain in light grey, the other in dark grey. Single stranded nucleic acid is shown in orange, double stranded DNA is shown in red. A: Simultaneous binding of two distinct nucleic acid chains by one Pur- $\alpha$  dimer. B: Bidentate binding of the Pur- $\alpha$  dimer to an individual nucleic acid. C: Cross-linking of several Pur- $\alpha$  dimers by an elongated nucleic acid.**

The Pur- $\alpha$  dimer might also be connected by long nucleic acids that have multiple interaction sites, and interact with several Pur- $\alpha$  molecules (Figure 3-2, C). This would lead to cross-linked Pur- $\alpha$  complexes. This binding mode might play a role in the pathogenesis of the neurodegenerative disease Fragile X Tremor/Ataxia syndrome (FXTAS) (section 1.3). The patients carry a premutation that leads to prolonged stretches of r(CGG) repeats (55-200) in the *Fmr1* mRNA (Berry-Kravis et al., 2007). Jin and colleagues have shown that Pur- $\alpha$  binds r(CGG) repeat RNA and propose that the loss of functional Pur- $\alpha$  by this affinity is the underlying cause for FXTAS (Jin et al., 2007). Interestingly, a pathological characteristic of FXTAS is the formation of inclusions in nerve cells, which contain *Fmr1* premutation RNA and Pur- $\alpha$  together with additional proteins (Greco et al., 2002; Iwahashi et al., 2006; Jin et al., 2007). In these inclusions, Pur- $\alpha$  might be cross-linked by the *Fmr1* premutation RNA (Figure 3-2, C).

An even higher complexity of binding modes and cross-linking can be envisioned, if both  $\beta$ -sheets of a PUR domain can bind two distinct nucleic acids simultaneously (section 3.2.1).

### **3.2.5 The nucleic-acid binding surface of a PUR domain is reminiscent of an RRM domain**

The modular structure of Pur- $\alpha$  is a common feature of RNA-binding proteins. Many RNA-binding proteins are composed of small RNA binding domains, which perform sequence-specific binding. Examples are Pumilio (PUF) repeats, zinc-finger domains, K-homology (KH) domains, and RNA-recognition motifs (RRM). They have a length of 35 to 90 amino acids and are often found in multiple copies in one protein (Auweter et al., 2006). By their versatile combination, these RNA-binding domains can satisfy diverse functional needs (Lunde et al., 2007). Affinity, as well as specificity of the binding can be increased by the interaction of one nucleic-acid strand with two or more binding modules (Lunde et al., 2007).



**Figure 3-3. Example of an RRM domain. Crystal structure of hnRNPA1 (UP1) RRM2 in complex with single stranded telomeric DNA (PDB-ID 2UP1) (Ding et al., 1999).**

**A:** The RRM domain is depicted in a color gradient from blue (N-Terminus) to red (C-terminus), the ssDNA is depicted in gray. **B:** View from A, rotated by 180° around the vertical axis.

Intriguingly, the proposed nucleic-acid binding surface of a PUR domain is reminiscent of the binding surface of an RRM domain. The RRM domain, also known as RNA-binding domain (RBD) or ribonucleoprotein domain (RNP) is found in all kingdoms of life (Maris et al., 2005). The RRM fold is one of the most abundant protein domains in eukaryotes and found in a variety of RNA- and DNA-binding proteins with a wide range of specificities and affinities (Maris et al., 2005). The RRM fold is composed two  $\alpha$ -helices that are sandwiched against a four-stranded antiparallel  $\beta$ -sheet (Figure 3-3). As shown in this study, one PUR repeat bears an  $\alpha$ -helix that is packed against a four stranded antiparallel  $\beta$ -sheet. RRM domains typically bind nucleic acids via the central  $\beta$ -strands of the  $\beta$ -sheet (Maris et al., 2005) (Figure 3-3). Likewise, this study identified the  $\beta$ -sheet side of a PUR domain as its nucleic-acid binding surface (section 3.2.1). Another functional similarity between RRM domains and the PUR domain is that they do not only bind both types of nucleic acids, but also interact with proteins (Maris et al., 2005) (White et al., 2009). However, there is a striking discrepancy in topology. A PUR domain is constituted of two  $\beta\beta\beta\alpha$ -units, whereas a RRM domain is build from a  $\beta\alpha\beta\beta\alpha$ -topology. Therefore, albeit a drastically different topology, a similar binding surface might be achieved. One might speculate if sequence-specific nucleic-acid binding follows a similar mechanism for PUR domains and RRM domains.

### 3.3 Interaction with proteins

#### 3.3.1 Protein interactions are mediated by PUR repeats

Structural characterization of Pur- $\alpha$  gives insight not only into nucleic-acid binding, but also in protein-protein interactions. The functional role of the PUR repeats is highlighted by previous efforts to map the binding regions of human Pur- $\alpha$  to various protein interaction partners. Interestingly, many binding regions overlap with the PUR repeats and predicted structural domains in human Pur- $\alpha$  (Figure 1-3 and Appendix 7-2). For human Pur- $\alpha$ , PUR repeat I comprises amino acids 59 to 123. It interacts with the polyomavirus large T-antigen, as the mapped binding region includes amino acids 72 to 123. (Gallia et al., 1998). Similarly, the minimal binding region of YB1 and CyclinT1 (amino acids 167-216) is part of the second PUR repeat (amino acids 137-209) (Darbinian et al., 2001c; Safak et al., 1999). Intriguingly, the mapped binding region for Rb, Cdk2, and Cdk9 superposes almost exactly with the PUR repeat III (amino acids 218-278), with a deviation of not more than four amino acids (Darbinian et al., 2001c; Johnson et al., 1995; Liu and Johnson, 2002).

It was previously known that the binding region for many proteins coincides with the nucleic-acid binding region of Pur- $\alpha$  (Gallia et al., 2000). From a structural point of view, one might wonder if also the same surface areas are involved. Ma and colleagues reported that the protein-protein binding surface of a protein can be predicted by the degree of conservation of its solvent-exposed residues (Ma et al., 2003). Especially conserved tryptophanes, and, to a lesser extent, phenylalanines and methionines, are frequently responsible for protein-protein contacts (Ma et al., 2003). Surface assessment of the crystal structure of *D. melanogaster* Pur- $\alpha$  repeats I-II reveals that the  $\beta$ -sheet side of the PUR domain shows more pronounced conservation than the  $\alpha$ -helical side (Figure 2-14). This suggests that the  $\beta$ -sheet side might also be employed in protein binding. In the crystal structure, the conserved and solvent-accessible residues phenylalanine 68 and phenylalanine 145 located in the  $\beta$ -sheets are putative candidates to probe this hypothesis with mutational studies.

### 3.3.2 Yeast-two hybrid screen

The yeast-two hybrid with *D. melanogaster* Pur- $\alpha$  (amino acids 40 to 254) yielded seven interaction partners corresponding to known proteins. After their inspection, only five of them (Arr1, LamC, Eye, Cka and Pur- $\alpha$ ) appear credibly.

A yeast-two hybrid system bears inherent limitations due to false-positive and false-negative results. In general, an observed interaction requires further examination, e. g. by pull-down assays, to confirm its specificity. A specific positive interaction is ideally expected to arise in several different protein fragments of the library. However, in the screen of this study, only interactions based on unique fragments were found. This means an increased probability that the results are due to random noise and leads to low confidence scores (section 2.2.2). Therefore the data has to be handled with caution, unless confirmed by supplementing techniques. In addition to background noise, false positive results can also derive from interactions without any biological relevance, e. g. if the interacting proteins reside in distinct cellular compartments.

The inconsistency between the Gal4 screen and the LexA system might hint at a systematic problem. The more sensitive Gal4 screen yielded 5 putative protein interactions, 13 interactions with non-existing proteins, and 9 GenMatches (section 2.2.2). On the other hand, only 2 putative protein interactions were found in the LexA system (section 2.2.2). The fact that the expected self-interaction of Pur- $\alpha$  was only observed in the LexA system, but not in the Gal4 system, suggests that the bait protein in the latter screen might have a folding problem. This idea is supported by the high number of interactions that refer to non-existing proteins and GenMatches, which were found in the Gal4 system. If the bait is not correctly folded, these interactions are likely false positives. Improper folding is possibly due to steric hinderance by the N-terminal DNA-binding domain. Therefore an independent confirmation of the interactions is absolutely required. The lack of trust in the Gal4 system is exceptionally unfortunate, since all of the putative new binding partners for Pur- $\alpha$  were found in this screen (section 2.2.2 and Table 2-16).

The failure to observe physiologic interactions is referred to as false negative results. As a general principle, one should note that the absence of proof is not proof of absence. Consequently, the absence of previously known interaction partners in the yeast-two-hybrid screen does not necessarily discredit them. The demonstration of direct physical interaction of two proteins, e. g. by pull-down assays, seems to be more reliable than a



yeast-two-hybrid system. False negative results can be caused by improper folding of the bait as discussed above and by the fact that the N-terminal part of the protein of interest is not available for the interaction because it is fused to the DNA-binding domain. Furthermore, the protein fragment chosen as a bait might not be optimal. In the yeast-two-hybrid screens, *D. melanogaster* Pur- $\alpha$  40-254 was employed because it was stable in *in vitro* experiments and contains the main protein interaction region of Pur- $\alpha$  (White et al., 2009).

Another reason for false negative results might be that the targeting of the transcription factor domains to the nucleus might be prevented by the bait or prey proteins fused to them. Kinetics might also explain missing interactions, as an interaction can not be detected if its  $K_D$  is significantly higher than the hybrid protein concentration in the yeast cell.

### 3.3.2.1 Potential artifacts

The observed interaction of Pur- $\alpha$  and BtbVII is likely to be an artifact, as this protein is known to be highly promiscuous in its protein interactions. Therefore, BtbVII tends to produce false positive results (Table 2-16).

The interaction to CG5758 is most likely an artifact, since the mapped interaction domain is a predicted extracellular Fas domain, which is inapplicable to the intra-cellular localization of Pur- $\alpha$  (Table 2-16). The biological relevance of this interaction has thus to be questioned.

The found interactions with so-called non-existing protein does not inevitably go back to random noise (Table 2-17). The preys do not represent existing proteins, but they can be helpful to identify a consensus motif for protein binding. If multiple independent clones are observed for one given prey (as it is the case for antisense Cac and out-of-frame inaF), then a cryptic peptide is likely to bind Pur- $\alpha$  (Table 2-17). In contrast, an interaction based on only one prey fragment has a higher probability to be caused by random noise, as it is the case for most other found interactions with non-existing proteins (Tab. 2-17).

The interactions with proteins whose sequences do not correspond to translated regions of the genome, but to chromosomic regions are referred to as GenMatches by Hybrigenics (Table 2-18). Possibly, they represent so far uncharacterized genes whose

expression is not known. Alas, this is increasingly unlikely with today's high quality of genome annotation, especially for *Drosophila*. They may also correspond to untranslated poly-adenylated RNA or may derive from poly-adenylated mRNA not yet fully spliced. As a consequence, the clone contains both exonic and intronic regions and does not align with previously identified mRNA, but rather with chromosomal regions. An interaction observed in multiple independent clones (as it is for GID 158246) likely represents a cryptic peptide preferred by Pur- $\alpha$ . The other GenMatches found are only represented by unique clones and are probably caused by residual background (Table 2-18).

### **3.3.2.2 Arr1, LamC, Eye and Cka are putative binding partners of Pur- $\alpha$**

The fact that the yeast-two-hybrid screen yielded only one fragment per protein interaction partner, indicates a low reliability and hampers the correct assignment of the interaction domain. Since a full domain is most likely properly folded and functional, the reliance of an interaction is increased, when the fragment corresponds to a known protein domain. This is the case for Arrestin1 and Cka, where the interaction fragment is identical with the C-terminal domain or the protein-binding WD40-repeat domain, respectively (Table 2-16). For LaminC, Eye, and Pur- $\alpha$  the fragments are longer but contain a known domain that is likely to be folded. Of course, in the evaluation of one interaction, the biological context has to be considered.

#### **Arrestin1**

The putative interaction of Pur- $\alpha$  with Arrestin1 (Table 2-16) leads to tempting speculations, since upon activation of photosensitive cells, arrestins translocate from the cell cytoplasm to the plasma membrane to activate G-protein coupled receptors (GPCRs) (Nair et al., 2005). The interaction of Arrestin1 and Pur- $\alpha$  might hint at yet another role of Pur- $\alpha$  in active transport. Arrestin1 translocates from the soma of the rod cell to the light-sensitive microvilli. There it is recruited to photoactivated rhodopsin in order to quench the excitation of the photoreceptor (Nair et al., 2005). The nature of this translocation is not sufficiently elucidated. A diffusion mechanism was proposed based on the observation that the transport is energy independent (Nair et al., 2005). Satho and colleagues supported the diffusion model by demonstrating that the translocation of

Arrestin1 is independent of the motor protein Eye-enriched Myosin-III (NINAC). On the other hand, Lee and Montell showed a requirement of NINAC for the transport of Arrestin2, which is the major arrestin with a 7-fold higher abundance than Arrestin1 (Lee and Montell, 2004; Satoh and Ready, 2005). Furthermore, Marszalek and colleagues found that Kinesin-II is essential for the selective transport of arrestin in mammalian photoreceptors (Marszalek et al., 2000). If Arrestin1 is indeed actively transported, it is intriguing to speculate if there is a conserved core complex that is responsible for a variety of different transport processes. Kanai and colleagues proposed a functional core complex might exist for mRNA localization in dendrites and named Pur- $\alpha$  as a likely key factor (Kanai et al., 2004).

### **LaminC**

The interaction of LaminC and Pur- $\alpha$  might also have some functional relevance (Table 2-16). LaminC is located at the inner membrane of the nucleus (Worman and Courvalin, 2002). Pur- $\alpha$  is shuttling between cytoplasm and nucleus, and is therefore accessible to LaminC (Itoh et al., 1998; Osugi et al., 1996). A functional link between Pur- $\alpha$  and LaminC might be transcriptional regulation, as both proteins are involved in this process (White et al., 2009). The role of lamins in tissue-specific gene expression is not yet understood (Worman and Courvalin, 2002). It was observed that mutations in lamins have an effect on gene expression pattern and RNA-Polymerase II-dependent transcription (Spann et al., 2002; Tsukahara et al., 2002). A further link between both proteins is Fragile X Temor/Ataxia Syndrome (FXTAS), as the inclusions present in the brains of FXTAS patients contain LaminC together with 19 other proteins (Arocena et al., 2005; Iwahashi et al., 2006). Aroncena and colleagues showed that r(CGG)-repeat RNA disrupts the lamin A/C structure and leads to the formation of inclusions (Arocena et al., 2005). In a *Drosophila* model of FXTAS, Jin et al. demonstrated that Pur- $\alpha$  binds to CGG-repeat RNA and is also present in these inclusions (Jin et al., 2007). Possibly Pur- $\alpha$  is precipitated together with its interaction partner LaminC in these inclusions.

## **Eye**

The possible interaction with Eye (Table 2-16) is interesting because of its functional link to Pur- $\alpha$  in brain development. Eye is a transcription factor and the *Drosophila* homolog of Pax6. A role for Eye was suggested in terminal differentiation and the maintenance of cell-type specific expression profiles in neurons. Eye was shown to control the expression of insulin-like peptides (Dilps) in insulin-producing neurons (Clements et al., 2008). By its expression-pattern overlap with genes involved in learning and memory, a role of Eye in the transcription of those genes was also suggested (Callaerts et al., 2001). Similarly, Pur- $\alpha$  was also shown to be essential for brain development and neuronal differentiation (Khalili et al., 2003).

## **Cka**

The kinase Cka is involved in the signal transduction pathway regulating the dorsal closure of the *Drosophila* embryo (Chen et al., 2002). Interestingly, the examination of Cka mutant flies hinted at an additional role for Cka in neuronal differentiation during eye development (Weng and Wei, 2002). This might constitute a functional link to Pur- $\alpha$ , as it is essential for neuronal differentiation (Khalili et al., 2003). Furthermore, Pur- $\alpha$  is involved in the pathogenesis of FXTAS (Berry-Kravis et al., 2007; Jin et al., 2007) and in the *Drosophila* model of FXTAS, an eye neurodegeneration phenotype is observed (Jin et al., 2007). The fragment of Cka that interacted with Pur- $\alpha$  in the yeast-two-hybrid screen (Table 2-16) comprises a WD40 repeat region. WD40 repeats are conserved in a variety of proteins (Smith et al., 1999). They form a  $\beta$ -propeller structure which serves as a versatile interaction platform for protein-protein contacts (Smith et al., 1999).

## **Pur- $\alpha$**

As demonstrated by this study, Pur- $\alpha$  is dimeric in solution (section 2.2.2). Size-exclusion chromatography and SAXS analysis support the model that dimerization is induced via the third PUR repeat in full length Pur- $\alpha$ . In accordance with this model, the self-interaction of Pur- $\alpha$  in the yeast-two-hybrid screen involved a fragment of Pur- $\alpha$  that included PUR repeat III (Table 2-16). Since the interaction was only observed with a unique clone, a precise mapping of the interaction domain was not possible. As discussed

above, when taking the Pur- $\alpha$  self-interaction as a positive control, the quality of the yeast-two-hybrid screen seems poor. Most prominently, the interaction was only found in the LexA system, but not in the Gal4 system, reducing the confidence of the latter screen. Since the Gal4 screen identified the putative interaction partners Arrestin, LamC, Eye, and Cka, it is absolutely required to revise them, e. .g. by pull-down assays.

#### **Chapter 4 Conclusion**

This study structurally and functionally characterizes Pur- $\alpha$ . It shows that Pur- $\alpha$  adopts a Whirly-like fold, identifies a novel nucleic-acid binding surface and proposes a dimerization model with direct functional implications.

Furthermore, the present work represents a case study how the bottleneck of X-ray structure analysis, that is crystallization, was overcome by crystal structure determination of a bacterial homolog and iterative use of sensitive bioinformatic tools.

## Chapter 5 Materials and Methods

### 5.1 Consumables and chemicals

All chemicals were purchased in reagent grade or better from Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany), Roche (Mannheim, Germany), or Sigma Aldrich (Hamburg and Seezle, Germany) unless stated otherwise. Enzymes for molecular biology and electrophoresis products were obtained from Fermentas (St. Leon-Rot, Germany) or New England Biolabs (Frankfurt, Germany). Media for bacterial cultures were prepared from reagents from Becton, Dickinson & Co (Heidelberg, Germany) and Sigma Aldrich. Media for insect cell culture were obtained from Sigma Aldrich. Chromatographic materials and columns were purchased from GE Healthcare (Munich, Germany), with the exception of amylose resin, which was obtained from New England Biolabs. Crystallization tools and screens were obtained from Hampton Research (Aliso Viejo, USA) and Qiagen (Hilden, Germany). DNA oligonucleotides were synthesized by Thermo Fisher Scientific (Schwerte, Germany). RNA oligonucleotides were ordered from Dharmacon/Perbio Science (Bonn, Germany) or Metabion (Martinsried, Germany). Radioactive nucleotides were purchased from Hartmann Analytik (Braunschweig, Germany).

### 5.2 Oligonucleotides

#### 5.2.1 DNA oligonucleotides for molecular biology

No.	Name	Sequence 5'-3'
1	hPurA60 BamHI for	ATG GGA TCC GAG CTG GCC TCC AAG CGG GTG G
2	hPurA56 BamHI for	ATG GGA TCC CAC GAG ACG CAG GAG CTG GCC
4	bbPur1 BamHI for	ATG GGA TCC ATG GGA GAG AGA GGG GAA GTA
7	bbPurA3 BamHI for	ATG GGA TCC GAG AGA GGG GAA GTA TAC
8	bbPurA5 BamHI for	ATG GGA TCC GGG GAA GTA TAC TCT GAA AAA C
11	bbPurA1 L12M, F22M for	ATG GGA GAG AGA GGG GAA GTA TAC TCT GAA AAA ATG TTT ACA GAG TCT GAG AGA ACT TAT TTT
12	bbPurA55 I59M for	TAT GAA GAG AAT ATG AAT GAG TTT GAA TCC
16	hPurA 245 XhoI rev	GCA CTC GAG CTA TCG CAT AAA CAC GCC GTA CTT
17	hPurA287 XhoI rev	GCA CTC GAG CTA CTC CCT CTG CTT CTC TTG AAT
18	hPurA 274 XhoI rev	GCA CTC GAG CTA GTA CTT GCA GAA GGT GTG TCC
22	bbPur122 XhoI rev	GCA CTC GAG CTA ATA ATC CTT CTT TTT AAA TCG

25	bbPurA120 XhoI rev	GCA CTC GAG CTA CTT CTT TTT AAA TCG
26	bbPurA105 XhoI rev	GCA CTC GAG CTA TCT AGA ATC GTC TAA TTT
27	bbPurA100 XhoI rev	GCA CTC GAG CTA TTT AGA TCT CTC CCC ATA
28	bbPurA110 XhoI rev	GCA CTC GAG CTA AGA TTT CTT ATC AAA TCT
29	bbPurA64 I59M rev	GGA TTC AAA CTC ATT CAT ATT CTC TTC ATA
30	bbPurA95 XhoI rev	GCA CTC GAG CTA ATA TTC ACC ATA CCC TTT
45	bbPurA70 XhoI rev	GCA CTC GAG CTA TAT TGC CTT AAG CAA
46	bbPurA74 XhoI rev	GCA CTC GAG CTA CTT AAT AAC CGC TAT
47	bbPurA78XhoIrev	GCA CTC GAG CTA AGA TAC TTT CTG CTT
48	bbPurA82 XhoI rev	GCA CTC GAG CTA AAC GGA CCC CGT AGA
51	bbPurA90 XhoI rev	GCA CTC GAG CTA ATA CCC TTT ATT ATG
52	hPurA42 BamHI for	AAA GGA TCC GGC AGT GGC GGC GGC
53	hPurA259 XhoI rev	GCA CTC GAG CTA CAC GGT GAT GGA GTT
55	hPurA191 BamHI for	AAA GGA TCC CCC GCG CAG GGG CTC
60	hPurA105 XhoI rev	GCA CTC GAG CTA TGA CAT GGA GAG
61	hPurA186 XhoI rev	GCA CTC GAG CTA ATG GCC CTG CGT
62	bbPurA23 R28A, K29A for	AAT GTC AAG GAA AAT GCA GCA GGA GAT TAT TTT
63	bbPurA34 R28A, K29A rev	TAA AAA ATA ATC TCC TGC TGC ATT TTC CTT GAC
64	bbPurA44 R49A for	AGT GGA GAT TTT GAA GCA CAT TCT ATT TTT GTA
65	bbPurA54 R49A rev	TAC AAA AAT AGA ATG TGC TTC AAA ATC TCC ACT
66	bbPurA1 R18A for	AAAGGATCCATGGGAGAGAGAGGGGAAGTATACTCT GAAAAACTATTT
78	dmPurA40 BamHI for	AAA GGA TCC GTC GAA CAG GAA TTG GCT
79	dmPurA115 BamHI for	AAA GGA TCC CCG GAA GAT GGT AAA CTT
82	dmPurA105 XhoI rev	GCA CTC GAG CTA AGA AGC GTA GTA ATC
83	dmPurA185 XhoI rev	GCA CTC GAG CTA ATC ATT AGC TCC AAA C
84	dmPurA245 XhoI rev	GCA CTC GAG CTA AAT ATC ACG AAA GCG
85	hPurA140 BamHI for	AAA GGA TCC CGC CGG GCG CTC AAA AGC
87	hPurA202 XhoI rev	GCA CTC GAG CTA CAG AGC GTC ACG GAA
88	hPurA268 XhoI rev	GCA CTC GAG CTA TCC GAA CTT GGC CCA
93	dmPurAa70 R65A XhoI rev	GCA CTC GAG CTA CTT TAT AAA ACG GCT TGC TCT ATT TTG TTT TAC
95	dmPurAa1 BamHI for	AAA GGA TCC ATG TCC GAT TTG GGA ATG
96	dmPurAa274 XhoI rev	GCA CTC GAG CTA TTT AAG ACT ATT TGA AGA
97	bbPurA24 R28A BamHI for	AAA GGA TCC GTC AAG GAA AAT AGA GCA GGA
98	dmPurAa152 R158A BamHI for	AAA GGA TCC ACA AGA GGG GGG CCT GCA TCT CAA

100	dmPurAa136 R142A BamHI for	AAA GGA TCC TTA AAA GAA AAT GCG GCT GGC CGA
102	bbPurA34 R28A XhoI rev	GCA CTC GAG CTA TAA AAA ATA ATC TCC TGC TCT ATT TTC CTT GAC
103	dmPurA162 R158A XhoI rev	GCA CTC GAG CTA ATC ATT AGC TCC AAA GAT TTG AGA TGC AGG CCC CCC TCT TGT
106	dmPurA146 R142A XhoI rev	GCA CTC GAG CTA CCG TAA AAA TCG GCC AGC CGC ATT TTC TTT TAA
117	dmPurAa36 BamHI for	AAA GGA TCC GGA AGA AGC GGC GTC GAA CAG
118	dmPurAa46 BamHI for	AAA GGA TCC ACG AAA ATG TTG CAA ATA CAA
124	dmPurA260 XhoI rev	GCA CTC GAG CTA AAT TGA GTC GGA CGA TTT
129	dmPurA60 R65A BamHI for neu	AAA GGA TCC AAA CAA AAT AGA GCA GCC CGT TTT
130	dmPurA85 R80A XhoI rev neu	GCA CTC GAG CTA CAA GTA AAT TTG ACT TGC TCT ACC ATC AGC GCC
131	dmPurA75 R80A BamHI for neu	AAA GGA TCC GGC GCT GAT GGT AGA GCA AGT CAA ATT TAC TTG
159	dmPurA181 BamHI for	AAA GGA TCC GAG TTT GGA GCT AAT GAT GGA
160	dmPurA188 BamHI for	AAA GGA TCC GGG TTT AAA GGA GAT TTA CCG
161	dmPurA258 XhoI rev	GCA CTC GAG CTA GTC GGA CGA TTT TTT CAT
162	dmPurA185 BamHI for neu	AAA GGA TCC AAT GAT GGA GGG TTT AAA GGA

**Table 5-1. Oligonucleotides for molecular biology.** dmPurA refers to *D. melanogaster* Pur- $\alpha$ , bbPurA refers to *B. burgdorferi* Pur- $\alpha$ , hPurA refers to human Pur- $\alpha$ . The numbers in the column “name” indicate the start site respective stop site of the amino acid sequence.

#### 4.2.2 DNA and RNA oligonucleotides for binding assays and crystallization

Name	Sequence 5'-3'	Length in nt	Type
CGG 6 mer DNA	CGG CGG	6	ssDNA
CGG 7 mer DNA	G CGG CGG	7	ssDNA
CGG RNA 12mer	r(CGG) <sub>4</sub>	12	ssRNA
CGG RNA 25mer	r(CGG) <sub>8</sub> C	25	ssRNA
Cntrl1	CCT CCG CCT CCG	12	ssDNA
Cy3-CGG DNA	Cy3-G CGG CGG	7	ssDNA labeled with Cy3
Cy3-CGG RNA	Cy3-G CGG CGG	7	ssRNA labeled with Cy3



hTel12	AGG GTT AGG GTT	12	ssDNA
hTel12-bromo	AGG GTT AGG GXT , X=Bromo-dU	12	ssRNA
hTel6	GTT AGG	6	ssDNA
JCVupTAR	GGA GGG GGA GGC	12	ssDNA
MF0677	GGA GGT GGT GGA GGG AGA GAA AAG	24	ssDNA
MS2	ACA UGA GGA UUA CCC AUG U	19	Hairpin RNA
Oligo-A DNA	A <sub>24</sub>	24	ssDNA
Oligo-A RNA	A <sub>12</sub>	12	ssRNA
Poly C	C <sub>12</sub>	12	ssDNA
pp7	GGC ACA GAA GAU AUG GCU UCG UGC C	25	Hairpin RNA
TAR14AU	AGA GCC UGG GAG CUC U	16	Stemloop RNA
TAR14GC	GGA GCC UGG GAG CUC C	16	Stemloop RNA

**Table 5-2. RNA and DNA oligonucleotides for binding assays and crystallization.**

## 5.3 Plasmids

### 5.3.1 Commercial plasmids

Name	Application	Tag	Antibiotic	Source
pGEX-6P-1	Recombinant protein expression in <i>E. coli</i>	GST	Amp	GE Healthcare (Munich, Germany)
pET28a	Recombinant protein expression in <i>E. coli</i>	His <sub>6</sub>	Kan	Novagen (Schwalbach, Germany)
PETM40	Recombinant protein expression in <i>E. coli</i>	MBP	Kan	EMBL (Heidelberg, Germany)
pFastBac1-GST	Bac-to-Bac Baculovirus expression system	GST	Amp, Gen	Invitrogen (Karlsruhe, Germany) modified (GST-tag) by D. Niessing (Gene Center, Munich, Germany)

**Table 5-3. Commercial plasmids for recombinant protein expression.**

### 5.3.2 Plasmids for recombinant protein expression in *E. coli*

No.	Insert	Vector	Restriction sites	Primer	Template	Note
AG25	bbPurA1-100	pGEX-6P-1	BamHI/XhoI	4, 27	AG21	
AG28	bbPurA1-120	pGEX-6P-1	BamHI/XhoI	4, 25	AG21	
AG21	bbPurA1-122	pGEX-6P-1	BamHI/XhoI	4, 22		a)

AG97	bbPurA1-122 R18A	pGEX-6P-1	BamHI/XhoI	66, 22	AG21
AG130	bbPurA1-122 R28A	pGEX-6P-1	BamHI/XhoI	97, 102, 4, 22	AG21
AG94	bbPurA1-122 R28A, K29A	pGEX-6P-1	BamHI/XhoI	62, 63, 4, 22	AG21
AG96	bbPurA1-122 R49A	pGEX-6P-1	BamHI/XhoI	64, 65, 4, 22	AG21
AG32	bbPurA3-100	pGEX-6P-1	BamHI/XhoI	7, 27	AG21
AG41	bbPurA3-100 L12M, F22M, I59M	pGEX-6P-1	BamHI/XhoI	7, 27, 11, 12, 29, 22	AG21
AG33	bbPurA3-105	pGEX-6P-1	BamHI/XhoI	7, 26	AG21
AG34	bbPurA3-110	pGEX-6P-1	BamHI/XhoI	7, 28	AG21
AG29	bbPurA3-120	pGEX-6P-1	BamHI/XhoI	7, 25	AG21
AG75	bbPurA3-70	pGEX-6P-1	BamHI/XhoI	7, 45	AG21
AG76	bbPurA3-74	pGEX-6P-1	BamHI/XhoI	7, 46	AG21
AG77	bbPurA3-78	pGEX-6P-1	BamHI/XhoI	7, 47	AG21
AG78	bbPurA3-82	pGEX-6P-1	BamHI/XhoI	7, 48	AG21
AG81	bbPurA3-90	pGEX-6P-1	BamHI/XhoI	7, 51	AG21
AG31	bbPurA3-95	pGEX-6P-1	BamHI/XhoI	7, 30	AG21
AG42	bbPurA3-95 L12M, F22M, I59M	pGEX-6P-1	BamHI/XhoI	7, 30, 11, 12, 29, 22	AG21
AG59	bbPurA5-100	pGEX-6P-1	BamHI/XhoI	8, 27	AG21
AG35	bbPurA5-95	pGEX-6P-1	BamHI/XhoI	8, 30	AG21
AG104	dmPurA1-274	pGEX-6P-1	BamHI/NotI	-	- b)
AG156	dmPurA1-274	pGEX-6P-1	BamHI/XhoI	95, 96	AG104
AG110	dmPurA115-185	pGEX-6P-1	BamHI/XhoI	79, 83	AG156
AG172	dmPurA181-260	pGEX-6P-1	BamHI/XhoI	159, 124	AG104
AG146	dmPurA185-260	pGEX-6P-1	BamHI/XhoI	162, 124	AG156
AG176	dmPurA188-258	pGEX-6P-1	BamHI/XhoI	160, 161	AG156
AG128	dmPurA36-185	pGEX-6P-1	BamHI/XhoI	117, 83	AG156
AG108	dmPurA40-105	pGEX-6P-1	BamHI/XhoI	78, 82	AG156
AG109	dmPurA40-185	pGEX-6P-1	BamHI/XhoI	78, 83	AG156
AG155	dmPurA40-185 R142A	pGEX-6P-1	BamHI/XhoI	15, 96, 78, 83, 100, 106	AG156
AG158	dmPurA40-185 R158A	pGEX-6P-1	BamHI/XhoI	95, 96, 78, 83, 98, 103	AG156
AG154	dmPurA40-185 R65A, R142A	pGEX-6P-1	BamHI/XhoI	95, 96, 78, 83, 129, 93, 100, 106	AG156

AG162	dmPurA40-185 S66G	R65A,	pGEX-6P-1	BamHI/XhoI	95, 96, 78, 83, 129, 93	AG156	C
AG161	dmPurA40-185	R80A	pGEX-6P-1	BamHI/XhoI	95, 96, 78, 83, 131, 130	AG156	
AG155	dmPurA40-185 R158A	R80A,	pGEX-6P-1	BamHI/XhoI	95, 96, 78, 83, 98, 103, 131, 130	AG156	
AG111	dmPurA40-245		pGEX-6P-1	BamHI/XhoI	78,79	AG104	
AG13	dmPurA40-255		pGEX-6P-1	BamHI/XhoI	-	-	d)
AG145	dmPurA40-274		pGEX-6P-1	BamHI/XhoI	78, 96	AG156	
AG129	dmPurA46-185		pGEX-6P-1	BamHI/XhoI	118, 83	AG156	
AG54	hPurA1-322		pGEX-6P-1	EcoRI/XhoI		-	e)
AG115	hPurA140-268		pGEX-6P-1	BamHI/XhoI	85, 88	AG54	
AG103	hPurA191-259		pGEX-6P-1	BamHI/XhoI	55, 53	AG54	
AG91	hPurA42-186		pGEX-6P-1	BamHI/XhoI	52, 61	AG54	
AG19	hPurA56-245		PETM40	BamHI/XhoI	2, 16	AG19	
AG19	hPurA56-245		pGEX-6P-1	BamHI/XhoI	-	-	d)
AG17	hPurA56-274		pET28a	BamHI/XhoI	2, 18	AG9	
AG9	hPurA56-274		pGEX-6P-1	BamHI/XhoI	-	--	d)
AG138	hPurA56-287		pGEX-6P-1	BamHI/XhoI	2, 17	AG54	
AG20	hPurA56-287 C272S		PETM40	BamHI/XhoI	2, 17	AG10	
AG10	hPurA56-287 C272S		pGEX-6P-1	BamHI/XhoI	-	-	d)
AG113	hPurA60-105		pGEX-6P-1	BamHI/XhoI	1, 60	AG54	
AG114	hPurA60-202		pGEX-6P-1	BamHI/XhoI	1, 87	AG54	
AG16	hPurA60-245		pET28a	BamHI/XhoI	1, 16	AG10	
AG6	hPurA60-245		pGEX-6P-1	BamHI/XhoI	-	b	d)
AG4	hPurA96-287		pGEX-6P-1	BamHI/XhoI	-	-	d)

**Table 5-4. Plasmids for recombinant protein expression in *E. coli*.** a) Obtained from U. Schulte-Spechtle, Nationales Referenzzentrum für Borrelien (Munich, Germany), b) obtained from K. Förstemann, Gene Center (Munich, Germany), c) spontanenou mutation S66G, d) obtained from D. Niessing, Gene Center (Munich, Germany), e) obtained from E. Johnson, Mount Sinai School of Medicine (New York, USA).

### 5.3.3 Plasmids for baculovirus-based recombinant protein expression in insect cells

No.	Insert	Vektor	Restriction sites	Primer	Template
AG60	hPurA56-245	pFastBac1-GST	BamHI/XhoI	2, 16	AG54
AG61	hPurA56-287	pFastBac1-GST	BamHI/XhoI	2, 17	AG54

**Table 5-5. Plasmids for baculovirus-based recombinant protein expression in insect cells.**

### 5.3.4 Plasmids for *in vitro* RNA transcription

Name	Gene	Organism	Length in nt	Antibiotic	Source
pBCX607	BC1 RNA	rat	152 nt	Amp	Obtained from J. Brosius (University Münster, Germany)
pBC200	BC200 RNA	human	200 nt	Amp	Obtained from J. Brosius (University Münster, Germany)

Table 5-6. Plasmids for *in vitro* RNA transcription.

### 5.4 Bacterial strains

Name	Essential genotype	Source
XL-1 Blue	recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F <sup>+</sup> proAB lacI <sup>q</sup> ZΔM15Tn10 (Tet <sup>r</sup> )]	Stratagene (La Jolla, USA)
BL21 Star (DE3)	B F <sup>-</sup> ompT hsdS(r <sub>B</sub> <sup>-</sup> m <sub>B</sub> <sup>-</sup> ) dcm+ Tet <sup>r</sup> galλ (DE3) EndA Hte [argU ileY leuW Cam <sup>r</sup> ]	Stratagene (La Jolla, USA)
B834 (DE3)	F <sup>-</sup> ompT gal met r <sub>B</sub> m <sub>B</sub>	Novagen (Schwalbach, Germany)
DH10Bac	F- <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) φ80 <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>endA1</i> <i>araD139</i> Δ( <i>ara, leu</i> )7697 <i>galU galK</i> λ- <i>rpsLnupG/bMON14272/pMON7124</i>	Invitrogen (Karlsruhe, Germany)

Table 5-7. Bacterial strains.

### 5.5 Media for cell culture

Medium	Composition
Luria-Bertani medium	1 % (w/v) bacto tryptone, 0.5 % (w/v) bacto yeast extract, 0.5 % (w/v) NaCl for plates + 1.5 % (w/v) agar
DH10Bac medium	1 % (w/v) bacto tryptone, 0.5 % (w/v) bacto yeast extract, 0.5 % (w/v) NaCl, 1 M IPTG, 100 μg/mL bromo-chloro-galactopyranoside (X-Gal), 7 mg/mL gentamycin, 10 mg/mL tetracycline, 50 mg/mL kanamycin for plates + 1.5 % (w/v) agar
SOC	2 % (w/v) bacto tryptone, 0.5 % (w/v) bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl <sub>2</sub> , 10 mM MgSO <sub>4</sub> , pH 7.0
Minimal medium	7.5 mM (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> , 8.5 mM NaCl, 55 mM KH <sub>2</sub> PO <sub>4</sub> , 100 mM K <sub>2</sub> HPO <sub>4</sub> , 1 mM MgSO <sub>4</sub> , 20 mM glucose, 1 mg/L CaCl <sub>2</sub> , 1 mg/L FeCl <sub>2</sub> , 1 mg/L Thiamine, 1 mg/L Biotin, 1 μg/L trace elements as follows: Cu <sup>2+</sup> , Mn <sup>2+</sup> , Zn <sup>2+</sup> , MoO <sub>4</sub> <sup>2-</sup> , 100 mg/L amino acids as follows: 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

Table 5-8. Media for cell culture.

## **5.6 Molecular biology**

### **5.6.1 Standard cloning and site-directed mutagenesis**

Genes of interest were cloned by standard methods as described (Sambrook and Russell, 2001). The protocol includes amplification of the target gene by PCR, digestion of the PCR product with restriction enzymes, digestion of vector DNA with restriction enzymes and ligation of DNA fragments. DNA fragments were separated by agarose gelelectrophoresis and purified with the NucleoSpin-Extract II kit (Macherey-Nagel, Düren, Germany). Point mutations were inserted by PCR-based site-directed mutagenesis (Ho et al., 1989). Cloning strategies for all plasmids and the oligonucleotides employed are summarized above (Table 5-1 and Table 5-4).

### **5.6.2 Transformation of *E. coli* and isolation of plasmid DNA**

Plasmid DNA was transformed in chemically competent *E. coli* XL-1 Blue cells as indicated (Hanahan, 1983; Sambrook and Russell, 2001). Isolation of plasmid DNA was performed with the NucleoSpin Plasmid kit (Macherey-Nagel, Düren, Germany). DNA sequencing was commissioned to Eurofins MWG Operon (Ebersberg, Germany).

## **5.7 Protein expression**

### **5.7.1 Recombinant protein expression in *E. coli***

Affinity-tagged Pur- $\alpha$  fusion proteins were recombinantly expressed in *E. coli* BL21 Star (DE3) cells. Typically, a 3 L scale of LB media was used. Media was supplemented with the appropriate antibiotic and cells were grown at 37 °C to an OD<sub>600</sub> of 0.3-0.4. The medium was cooled to 18 °C within 1 h and then recombinant protein expression was induced by the addition of 0.25 mM IPTG. After 12-16 h, cells were harvested by centrifugation (15 min, 4000 rpm, 3500 g) at 4 °C, partitioned in 2 aliquots, flash frozen in liquid nitrogen, and stored at 80 °C.

## **Expression of selenomethionine-substituted protein**

L-selenomethionine-substituted protein was expressed as described (Doublet, 1997). The *E. coli* methionine auxotrophic strain B834 (DE3) was grown in minimal medium essentially as elaborated above.

### **5.7.2 Baculovirus-based recombinant protein expression in insect cells**

The Bac-to-Bac baculovirus expression system (Anderson et al., 1996) was applied according to the manufacturer's instructions (Invitrogen, Karlsruhe, Germany). The commercial pFastBac1 vector was modified to carry the gene for a GST-tag, yielding pFastBac1-GST vector (D. Niessing, Gene Center, Munich, Germany). The gene of interest was cloned into pFastBac1-GST and the plasmid was amplified in XL-1 Blue cells (section 5.6).

For transposition, the pFastBac1-GST plasmid was transformed into DH10Bac *E. coli* cells (Anderson et al., 1996; Luckow et al., 1993) and regenerated in 1 mL SOC medium for 6 h. DH10Bac cells were streaked on DH10Bac medium agar plates and grown at 37 °C for 48 h. Successful clones were identified using blue/white selection. Liquid cultures (5 mL LB medium with 7 mg/mL gentamycin, 10 mg/mL tetracycline, 50 mg/mL kanamycin) were inoculated with colonies with a white phenotype. Bacmid DNA was isolated from 5 mL culture with a Nucleospin Plasmid kit (Macherey-Nagel, Düren, Germany), but using isopropanol precipitation of DNA instead of the provided columns. High Five insect cells (Davis et al., 1992; Wickham et al., 1992) were cultured at 27 °C in HyClone SF medium (Thermo Fisher Scientific, Schwerte, Germany) according to the provider's instructions (Invitrogen, Karlsruhe, Germany). For transfection, 2 µg bacmid DNA and 3 µL FuGeneHD reagent (Roche, Basel, Switzerland) were mixed with 300 µL Hyclone SF medium and incubated for 45 min at room temperature. The formed complex was added to 2 mL of High Five cell suspension (0.25 million cells per mL). Cells were incubated at 27° C for 3-4 days and then inspected for signs of infection under the microscope. After infection, cells were centrifuged (10 min, 3000 rpm, 1000 g) and the supernatant (P1 virus) was stored at 4 °C.

Virus amplification was done in 50 mL cell culture (1 million cells per mL) with the addition of 2 mL P1 virus. After incubation at 27 °C for 48 h, cell density was determined with the help of a counting chamber and the culture was diluted to 1 million

cells per mL. This procedure was repeated until cells stopped growing. After additional 24 h, the cells were centrifuged (10 min, 3000 rpm, 1000 g) and the supernatant (P2 virus) was stored at 4 °C.

For protein expression, 1 L of cell culture (1 million cells per mL) was infected with 100 mL of P2 virus and cultured at 27 °C for 3-4 days. The culture was diluted every day to 1 million cells per mL and checked for protein expression *via* SDS-PAGE (section 5.9). At the peak of protein production, cells were harvested by centrifugation (15 min, 3000 rpm, 1000 g) and frozen in liquid nitrogen.

## **5.8 Protein purification**

### **5.8.1 Purification of GST-tagged proteins**

Cell pellets were thawed and resuspended in 25 mL buffer A containing 500 mM NaCl, 50 mM HEPES pH 7.5. After the addition of ½ tablet Roche complete protease inhibitor and 10 mM PMSF, cells were disrupted by sonication on ice. The lysate was cleared by centrifugation (45 min, 15 000 rpm, 27000 g) at 4 °C. All subsequent purification steps were carried out at 4 °C. The lysate was loaded on a GSTrap 5 mL column equilibrated with buffer A. After washing with 100 mL buffer A, the fusion protein was eluted from the column with 5 column volumes buffer B (500 mM NaCl, 20 mM HEPES pH 8.0, 25 mM Glutathion). Protein containing fractions were identified by their absorption at wavelength 280 nm ( $A_{280}$ ) and pooled. GST was cleaved off by the addition of 50 µg PreScission Protease (GE Healthcare, Munich) prior to dialysis against buffer C (250 mM NaCl, 20 mM HEPES pH 8.0) for 12 h at 4 °C. GST was subtracted using a GSTrap column and contaminating nucleic acids were removed by a HiTrapQ column. Pur- $\alpha$  was further purified by a Heparin column and was eluted with buffer D (2 M NaCl, 20 mM HEPES pH 8.0). After concentration using a centrifugal filter device (Amicon Ultra, Millipore, Billerica, USA), the protein was centrifuged (20 min, 13 000 rpm, 16 000 g, 4 °C) and finally purified by size-exclusion chromatography.

For *B. burgdorferi* Pur- $\alpha$  and *D. melanogaster* Pur- $\alpha$  repeats I-II, size-exclusion chromatography was performed in buffer C with a Superose 12 10/300 GL column.

For *D. melanogaster* full-length protein, size-exclusion chromatography was carried out on a Superdex S200 16/60 column in buffer C.

*D. melanogaster* Pur- $\alpha$  repeats I-II double mutant (R80A, R158A) did not bind to Heparin column, thus flow-through from GSTrap and HiTrapQ column was directly subjected to size-exclusion chromatography on Superdex S200 16/60 column with buffer C.

Since *D. melanogaster* Pur- $\alpha$  repeats I-III and fragments of human Pur- $\alpha$  were unstable in low salt conditions, dialysis and all further purification steps for these protein fragments were carried out in buffer E (500 mM NaCl, 20 mM HEPES pH 8.0).

Selenomethionine-substituted protein was purified analogous to native protein with the addition of 2 mM DTT in all buffers.

Protein fragments of human Pur- $\alpha$  were purified analogously, with buffer A (500 mM KCl, 100 mM TRIS pH 8.4), buffer B (500 mM KCl, 100 mM TRIS pH 8.4, 25 mM Glutathion), buffer C same as buffer A, buffer D (2 M KCl, 100 mM TRIS pH 8.4) and size-exclusion chromatography on a Superdex S200 16/60 column in buffer A.

### **5.8.2 Purification of His-tagged proteins**

The cell pellet was resuspended in 30 mL buffer N1 containing 50 mM potassium phosphate buffer pH 8.0, 300 mM NaCl, 0.5 mM PMSF. Cells were disrupted by sonication followed by centrifugation as described (section 5.8.1). The lysate was loaded on a 5 mL HisTrap column pre-equilibrated with buffer N1. The column was washed with buffer N2 (50 mM potassium phosphate buffer pH 8.0, 2 M NaCl) until the  $A_{280}$  absorption baseline appeared stable. The protein was eluted by buffer N3 (50 mM potassium phosphate buffer pH 8.0, 1 M NaCl, 500 mM Imidazol) with a gradient over 10 column volumes. Protein containing fractions were identified *via* the  $A_{280}$  absorption. The fractions were pooled and dialyzed for 12 h against buffer N4 (50 mM potassium phosphate buffer pH 7.5, 200 mM NaCl, 0.5 mM EDTA, 2 mM DTT). The solution was loaded on a Heparin column pre-equilibrated with buffer N4 and eluted with buffer N5 (50 mM potassium phosphate buffer pH 7.5, 2 M NaCl, 0.5 mM EDTA, 2 mM DTT). Finally, size-exclusion chromatography was carried out on a Superdex S200 16/60 column in buffer N4.



### **4.8.3 Purification of MBP-tagged proteins**

Cells were resuspended in 30 mL buffer M1 containing 200 mM NaCl, 20 mM TRIS pH 7.4 and 1 mM EDTA. After the addition of ½ tablet Roche complete protease inhibitor and 10 mM PMSF, cells were disrupted and centrifuged as described (section 5.8.1). Pre-equilibrated (buffer M1) amylose resin (2 mL) was incubated with the lysate for 1 h while shaken gently. The resin was washed with 150 mL buffer M1 and loaded on a column. The protein was eluted with 10 mL buffer M2 (200 mM NaCl, 20 mM TRIS pH 7.4, 1 mM EDTA 20 mM maltose). The solution was diluted with 20 mL buffer M0 (20 mM TRIS pH 7.4 and 30 mM MgCl) and 100 µL RNase A (10 mg/mL) as well as 5 µL DNaseI (10 U/µL) were added and incubated for 14 h. The solution was passed through a HiTrapQ column and directly loaded on a Heparin column pre-equilibrated with buffer M1. The protein was eluted from the Heparin column with buffer M3 (2 M NaCl, 20 mM TRIS pH 7.4, 1 mM EDTA). The final purification step was carried out on a Superdex S200 16/60 column in buffer M4 containing 100 mM NaCl and 10 mM TRIS pH 7.4.

## **5.9 Protein analysis**

Efficiency of purification steps and purity of the final sample was monitored by sodium dodecylsulfate polyacrylamide gelelectrophoresis (SDS-PAGE) as described (Laemmli, 1970). Electrophoretic separation was carried out on 16-19 % polyacrylamide gels followed by staining with Coomassie Blue staining solution. For purified protein samples, absence of contaminating nucleic acids was confirmed by measuring the ratio of absorption at wavelengths 260 nm and 280 nm. Only samples with a ratio  $A_{260}/A_{280}$  of 0.6 or lower were used.

### **5.9.1 Analytical size-exclusion chromatography**

Analytical size-exclusion chromatography was performed with a Superose 12 10/300 GL column standardized with a size-exclusion calibration kit (BioRad, Munich, Germany). The column was equilibrated with buffer E and 0.5 mL of Pur-α was loaded with a flow

rate of 0.9 mL/min. Peak fractions were identified by  $A_{280}$  and the presence of Pur- $\alpha$  was confirmed by SDS-PAGE.

### **5.9.2 Circular dichroism spectroscopy**

To assess proper folding of the mutant proteins, Circular Dichroism (CD) spectra were recorded using a Jasco 810 Spectropolarimeter. Protein samples at a concentration of 0.5 mg/mL in buffer containing 150 mM KCl and 20 mM HEPES pH 8.0 were measured in a 0.1-cm cuvette at 4 °C. Spectra were recorded from wavelength 190 nm to 260 nm with a scan speed of 200 nm per minute.

### **5.10 Crystallization**

For all crystallization experiments, only freshly prepared protein samples were used and centrifuged before set ups (20 min, 13 000 rpm, 16 000 g, 4 °C). Initial crystallization conditions were screened robotically by using the crystallization facility of the Max-Planck-Institute for Biochemistry (Martinsried, Germany). Equipment included a Phoenix nanodispenser robot (Art Robbins Instruments, Sunnyvale, USA) and an Xtal-focus visualization system (Tecan, Männedorf, Switzerland and Leica, Wetzlar, Germany). The system used a 0.5  $\mu$ L drop size and a 1:1 mixture of protein and crystallization solution. Refinement was done manually using hanging-drop vapor diffusion technique with a drop size of 2  $\mu$ L and a 1:1 mixture of protein and crystallization solution.

#### **5.10.1 Crystallization and structure determination of *B. burgdorferi* Pur- $\alpha$**

Diverse protein constructs were concentrated in buffer C (250 mM NaCl, 20 mM HEPES pH 8.0) and initial screening was performed with a protein solution of 3 mg/mL. Initial crystals appeared in various conditions. After optimization, crystals of a fragment of *B. burgdorferi* Pur- $\alpha$  comprising amino acids 3 to 100 were grown at 4 °C from a protein solution of 2.2 mg/mL and a crystallization solution containing 100 mM HEPES pH 7.2 and 20 % PEG 3350. Cuboid crystals of approximately (200 x 100 x 30)  $\mu$ m size appeared after 2-4 days.

For selenomethionine-substituted protein, 1 mM DTT and 1 mM TCEP were added prior to crystal set up. After optimization, crystals were grown at 4 °C in 2.8 M sodium formate with a protein concentration of 1.2 mg/mL and the stoichiometric addition of a short DNA oligomer (hTel12 ssDNA), albeit the latter was not visible in the structural model. Cubic and spindle shaped crystals of (100 x 50 x 50)  $\mu\text{m}$  size appeared within 2-5 days.

Prior to data collection, crystals were cryo-protected in mother liquor plus 20 % ethylene glycol in 3 steps and frozen in liquid nitrogen.

Multiple wavelength anomalous dispersion (MAD) experiments were recorded at beamline X06SA/PXI (SLS, Villigen) and native datasets were recorded at beamline ID23-1 (ESRF, Grenoble). The data were integrated and scaled with the XDS program package (Kabsch, 1993). Matthews coefficient was calculated with MATTHEWS\_COEF (Kantardjieff and Rupp, 2003; Matthews, 1968). Phases were obtained by SAD using SHELX (Sheldrick, 2008). The model was build manually from the selenomethionine-dataset using COOT (Emsley and Cowtan, 2004). Refinement was performed with REFMAC (Murshudov et al., 1997; Terwilliger, 2002). The final model was analyzed using SFCHECK (Vaguine et al., 1999).

### **5.10.2 Crystallization and structure determination of *D. melanogaster* Pur- $\alpha$ repeats I-II**

Diverse protein constructs were concentrated in buffer C (250 mM NaCl, 20 mM HEPES pH 8.0) and initial screening was performed with protein solutions of 1-5 mg/mL, yielding initial crystals in various conditions. After optimization, crystals of a fragment of *D. melanogaster* Pur- $\alpha$  comprising amino acids 40 to 185 were grown at 21 °C with a protein solution of 3-5 mg/mL. The crystallization solution contained 50 mM MES pH 5.9, 200 mM  $\text{MgCl}_2$ , and 25 % PEG 3350. Cuboid crystals of approximately (100 x 50 x 50)  $\mu\text{m}$  size appeared within 2-4 days. For the selenomethionine-derivatized protein, 1 mM DTT and 1 mM TCEP were added to the protein solution. After optimization, crystals of selenomethionine-derivatized *D. melanogaster* Pur- $\alpha$  40-185 were grown at a concentration of 2.7 mg/mL. The crystallization solution contained 200 mM  $\text{MgCl}_2$ , 100 mM HEPES pH 7.8, and 22 % PEG 3350. The crystals had a cuboid shape and a size of approximately (150 x 50 x 50)  $\mu\text{m}$ .

For cryo-protection, native crystals were shortly incubated in reservoir solution containing 20% glycerol in 2 steps. Selenomethionine-derivatized crystals were shortly incubated in reservoir solution containing 20% ethylene glycol in 2 steps. Then crystals were frozen in liquid nitrogen.

Multiple wavelength anomalous dispersion (MAD) experiments were performed at beamline ID14-4 (ESRF, Grenoble) and native datasets were collected at beamline ID14-1 (ESRF, Grenoble). The data were integrated and scaled with the XDS program package (Kabsch, 1993). Phases were obtained with Crank (Crunch2/BP3/Solomon) (Collaborative Computational Project, 1994; Ness et al., 2004), and extended to 2.1 Å resolution. Parts of the final model were build automatically with Buccaneer (Cowtan, 2006). The final model was build from the native dataset and manually completed using COOT (Emsley and Cowtan, 2004). Refinement was performed with REFMAC (Murshudov et al., 1997; Terwilliger, 2002). The final model was analyzed using SFCHECK (Vaguine et al., 1999).

### **5.10.3 Crystallization of *D. melanogaster* Pur- $\alpha$ repeat III**

Several protein fragments were concentrated in buffer C (250 mM NaCl, 20 mM HEPES pH 8.0) and initial screening was performed with protein solutions of 3.0-5.7 mg/mL. Initial crystals were obtained in several conditions. After optimization, crystals of a fragment of *D. melanogaster* Pur- $\alpha$  comprising amino acids 188 to 258 were grown at a concentrations of 0.3 mg/mL at 4 °C. The crystallization solution contained 22 % PEG 3350, 100 mM sodium cacodylate pH 7.5 and 200 mM sodium acetate. After 1-2 day, needle-shaped crystals appeared. After 2-3 weeks, additional pentagonal and hexagonal crystal plates of (100 x 50 x 50) $\mu$ m size appeared. This crystal form had improved diffraction properties and individual crystals diffracted up to 2.8 Å.

### **5.10.4 Co-crystallization with nucleic acids**

A fragment of *D. melanogaster* Pur- $\alpha$  comprising amino acids 40 to 185 was mixed with Cy3-CGG 7mer DNA in a ratio 1:2 in buffer C (250 mM NaCl, 20 mM HEPES pH 8.0). After incubation on ice for 30 min, size-exclusion chromatography was performed in buffer C using a S12 10/300 GL column. The co-elution of protein and DNA was

confirmed by absorption spectra as well as SDS-PAGE. The complex was concentrated and initial screens were set up with solution of approximately 3.6 mg/mL. Initial crystals were obtained in a few and related conditions.

After preliminary optimization, crystals diffracted up to 15 Å. Crystals of approximately (20 x 20 x 40) μm size were grown at 4 °C at a concentration of 3.1 mg/mL in 0.05 M MES pH 6.5 and 30 % MPD or 28 % PEG 400, resp. Rice grain shaped crystals appeared after 2-4 days.

Initial crystallization trials with Cy3-CGG 7mer RNA (3.3 mg/mL) yielded no crystals.

### **5.11 Limited proteolysis**

Protein samples in buffer containing 125 mM NaCl and 100 mM TRIS pH 8.0 were supplied with protease (trypsin or chymotrypsin) in a ratio of 50:1 (*w/w*) and incubated at room temperature or on ice. Probes of 15 μL were taken at time points 0 s, 30 s, 1 min, 2 min, 5 min, 10 min, 30 min, 60 min, 90 min and the reaction stopped by the addition of 5 μL SDS-gel loading dye. Samples were resolved by SDS-PAGE, defined bands were cut out, blotted on a PVDF membrane (Immobilon P, Millipore, Billerica, USA) *via* passive absorption and subjected to Edman sequencing (Messer et al., 1997).

### **5.12 Lysine methylation**

Reductive methylation of lysine residues was performed as described (Walter et al., 2006). Freshly prepared 1 M dimethylborane complex (ABC) and 1 M formaldehyde was cooled on ice. 20 μL of 1 M ABC and 40 μL 1 M formaldehyde were added to 1 mL protein solution (1 mg/mL) in buffer containing 50 mM HEPES pH 7.5 and 250 mM NaCl. After incubation at 4 °C for 1.5 h, additional 20 μL of 1 M ABC and 40 μL 1 M formaldehyde were added and the reaction was carried out 14 h at 4 °C. Then, 100 μL 1 M ammonium sulfate were added followed by size-exclusion chromatography on a Superdex S200 16/60 column in buffer containing 20 mM TRIS pH 7.5, 200 mM NaCl, and 1 mM DTT.

### 5.13 Heavy atom derivatization by soaking

Solutions of heavy atom salts with a concentration of 10 mM, 5 mM and 1 mM were prepared. 0.1  $\mu\text{L}$  of the respective solution was added to the crystallization drop and incubated for 10 min, 30 min, 1 h and 12 h.

Heavy atom salts tested for crystals of a fragment of *B. burgdorferi* Pur- $\alpha$  comprising amino acids 3 to 100 included  $\text{K}_2\text{PtCl}_4$ ,  $\text{K}_2\text{Pt}(\text{CN})_4$ ,  $(\text{NH}_4)_2\text{PtCl}_4$ ,  $\text{K}_2\text{PtCl}_6$ , and  $\text{K}_2\text{PtI}_6$ .

For crystals of a fragment of *D. melanogaster* Pur- $\alpha$  comprising amino acids 40 to 185, additionally 5-Amino-2,4,6-triiodoisophthalic acid (JBS Magic Triangle) and Hexatantalum tetradecabromide (JBS Tantalum Cluster) were tested (Beck et al., 2008; Knablein et al., 1997) (Jena Bioscience, Jena, Germany).

### 5.14 DNA soaking experiments

Crystals of a fragment of *B. burgdorferi* Pur- $\alpha$  comprising amino acids 3 to 100 were grown as described (section 5.10.1). DNA oligonucleotides Bromo-hTel12 DNA, hTel6 DNA or JCVupTar DNA (Table 5-2) were added to the crystallization drop to a final concentration of 0.3-0.8  $\mu\text{g}/\text{mL}$ . Crystals were incubated for 1 or 2 days before they were cryo-protected and frozen in liquid nitrogen.

Crystals of *D. melanogaster* Pur- $\alpha$  repeats I-II were grown as described (section 5.10.2). In a hanging-drop vapor diffusion chamber, a drop of mother liquor with the addition of 1.95  $\mu\text{g}/\mu\text{L}$  CGG7mer DNA was prepared and equilibrated against the reservoir for 1 h. Crystals were transferred to the soaking solution and incubated for 2 h, 3 h, and 12 h. Crystals were transferred, shortly incubated in 20 % glycerol and flash frozen in liquid nitrogen.

### 5.15 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) data were collected at the ID14-3 beamline (ESRF, Grenoble, France) and the X33 beamline (EMBL/DESY, Hamburg, Germany). Scattering curves were measured in 20 mM HEPES pH 8.0, 500 mM NaCl, and 3 mM DTT with exposure times of 2 min (X-33) and ten times 30 seconds (ID14-3), respectively. Protein concentration was 0.9-4.55 mg/mL. Before and after each measurement, a blank

measurement with sample buffer was taken. Primary analysis of the data was done with PRIMUS (Konarev et al., 2003) and data evaluation and processing was performed using the ATSAS program package (Konarev et al., 2006). The averaged blank measurements were subtracted from the scattering curves. The measurements were scaled to their respective protein concentration and merged in one dataset. The Kratky plot was used to inspect correct folding of the protein sample. For molecular-mass determination, scattering intensities were extrapolated to zero angle ( $I_0$ ), using bovine serum albumine and lysozyme as references. The Guinier approximation was used with the constraint  $s \cdot R_g < 1.3$  to calculate the radius of gyration  $R_g$  (Guinier and Fournet, 1955). The maximum dimension of the particle  $D_{\max}$  was iteratively determined by evaluating the resulting  $R_g$  value and the shape of the pair-distribution function  $P(r)$  using GNOM (Svergun, 1992). *Ab initio* models were calculated with GASBORp (Svergun et al., 2001) assuming P2 symmetry for Pur- $\alpha$  repeats I-III and P1 symmetry for Pur- $\alpha$  repeats I-II. In each case, 10 bead models were averaged using DAMAVER (Volkov and Svergun, 2003). For Pur- $\alpha$  repeats I-II, the final bead model was overlaid with the crystal structure using SUPCOMB (Kozin and Svergun, 2000). CRY SOL (Svergun et al., 1995) was used to determine the theoretical scattering curve based on the Pur- $\alpha$  repeats I-II crystal structure.

## **5.16 Nucleic-acid binding assays**

### **5.16.1 *In vitro* transcription of RNA**

All RNA experiments were carried out in buffers made from RNase free water and using RNase free chemicals and equipment. RNase free water was produced by adding 0.05 % diethylpyrocarbonate (DEPC) to Millipore-purified water and incubating for 8-12 h at 38 °C, then autoclaving for 20 min at 121 °C.

*In vitro* transcription was performed with the MEGAshortscript T7 Kit (Ambion, Austin, USA) according to the manufacturer's instruction. Plasmid DNA was linearized with *DraI* and purified by ammonium acetate/ethanol precipitation. In a 20  $\mu$ L standard reaction, 2 pmol of template DNA were used. The transcription reaction was carried out at 37 °C for 3-4 h. Then, 1 U of calf intestine alkaline phosphatase (CIAP) was added and incubated for additional 30 min. The template was digested by the addition of 2 U Turbo DNase (Ambion, Austin, USA) prior to purification by native polyacrylamide

electrophoresis (PAGE). Electrophoresis was carried out at room temperature at 80 V using a 5 %-gel and 1 x TBE running buffer (90 mM TRIS pH 8.3, 90 mM boric acid, 2 mM EDTA). RNA was visualized by UV shadowing and then extracted from the gel by crush & soak technique. The band was cut into small pieces and incubated with 400  $\mu$ L crush & soak solution (0.3 M NaCl, 0.5 mM EDTA) for 12 h at 37 °C. The probe was filtered through an Ultrafree MC centrifugal filter device (Ambion, Austin, USA). RNA was concentrated and buffer was exchanged using a Ultrafree-0.5 centrifugal filter device (Millipore, Billerica, US). Concentration was determined by the absorption at wavelength 260 nm ( $A_{260}$ ). Sample integrity was confirmed by analytical native PAGE.

### **5.16.2 Isotopic labeling of oligonucleotides**

DNA and RNA oligonucleotides were radioactively labeled at their 5'-ends using  $\gamma$ - $^{32}$ P-ATP and T4 Polynucleotide Kinase (PNK) following the manufacturer's protocol (Fermentas, St. Leon-Rot, Germany). 5 pmol of the oligonucleotide were incubated with 30  $\mu$ Ci  $\gamma$ - $^{32}$ P-ATP, 10 units PNK and the supplied buffer A for 45 min at 37 °C. The reaction was stopped by incubation at 70 °C for 10 min. RNA oligonucleotides were purified by a Sephadex G-25 Quick Spin column (Roche, Basel, Switzerland). DNA oligonucleotides were purified with the Qiaquick Nucleotide Removal Kit (Qiagen, Hilden, Germany). Radiolabelled oligonucleotides were stored at -20 °C for up to 1 week.

### **5.16.3 Filter binding assays**

Nitrocellulose filter binding assays were performed essentially as described (Wong and Lohman, 1993). The protein was transferred into binding buffer (100 mM NaCl, 10 mM HEPES pH 8.0, 2.5 mM MgCl<sub>2</sub>, 1 mM DTT) and serial protein dilutions (0-10  $\mu$ M) were incubated with a constant amount of radioactively labeled oligonucleotide (0.5 nM) for 20 min at room temperature. A nitrocellulose filter (Optitran BA-S85 reinforced NC, Whatman/GE Healthcare, Munich, Germany) was activated by incubating in 0.4 M KOH for 10 min followed by washing 8 times with 200 mL water. The nitrocellulose filter and a nylon membrane (Roti-Nylon Plus, Roth, Karlsruhe, Germany) were equilibrated in binding buffer for 1 h. A Bio-Dot microfiltration apparatus (BioRad, Munich, Germany)



was equipped with both membranes and each well was washed with 50  $\mu$ L binding buffer. 75  $\mu$ L of each binding reaction were applied on the membranes, followed by washing with 75  $\mu$ L binding buffer. A phosphor imager system was used to measure the retained radioactively labeled oligonucleotides on the nicrocellulose filter. The storage phosphor screen (GE Healthcare, Munich, Germany) was exposed to the filter for 1-1.5 h before read out on a Storm Scanner (Molecular Dynamics, Sunnyvale, USA). KaleidaGraph (Synergy software, Reading, USA) was used to plot of the fraction of bound oligonucleotide versus the protein concentration. The equilibrium-dissociation constant  $K_D$  was derived from the Langmuir isotherm (Smith, 1998).

#### **5.16.4 Electrophoretic mobility shift assays**

Electrophoretic mobility shift assays (EMSAs) were essentially performed as previously described (Sambrook and Russell, 2001). Protein was transferred to DEPC-treated binding buffer (50 mM Tris-HCl pH 7.4, 150 mM KCl, 1 mM EDTA, 1 mM  $MgCl_2$ , and 1 mM DTT) and monitored for proper folding by size-exclusion chromatography. Serial protein dilutions and a constant amount of radiolabeled DNA respective RNA oligomer (2.5 nM) were incubated in binding buffer on ice for 20 min. DNA-binding experiments always contained 25  $\mu$ g/ml PolydI/PolydC competitor, and RNA-binding experiments always contained 25  $\mu$ g/ml yeast tRNA as well as 200 U/mL RiboLock RNase inhibitor (Fermentas). Reaction mixtures (15  $\mu$ l) were loaded onto 6 % TBE polyacrylamide gels and analyzed after electrophoresis (35 minutes, 110 V) by a phosphor imaging system (section 6.16.3).

Assessment of RNA quadruplex binding was performed on 6 % TBE polyacrylamide gels with the addition of 20 mM KCl or LiCl, respectively. The protein was transferred in buffer containing 50 mM Tris-HCl pH 7.4, 1 mM EDTA, 1 mM  $MgCl_2$ , 1 mM DTT and 150 mM KCl or 150 mM LiCl, respectively. CGG 25mer RNA was transferred to the respective buffer and heated for 10 min at 65° C, then slowly cooled to room temperature. A constant amount of RNA (0.5 nM) was incubated with serial protein dilutions in the respective binding buffer. Binding reactions contained 25  $\mu$ g/ml yeast tRNA as well as 200 U/mL RiboLock RNase inhibitor (Fermentas). The cationic porphyrin 5,10,15,20-tetra(*N*-methyl-4-pyridyl)porphin (TMPyP4) was added to a

concentration of 5 nM and incubated in the dark. Binding reactions were incubated on ice for 20 min before resolved in the dark.

## **5.17 Bioinformatics**

### **5.17.1 Protein parameters**

The ExPaSy Proteomics Server (<http://www.expasy.ch>) provided tools for physical and chemical parameter calculation of the recombinant proteins. ProtParam (Wilkins et al., 1999) was used to determine the isoelectric point, molecular weight and extinction coefficient of the protein fragments.

### **5.17.2 Sequence alignment and secondary structure prediction**

DNA and protein sequences were obtained from the NCBI database (Benson et al., 1994) (<http://www.ncbi.nlm.nih.gov>). Homology searches were accomplished by the BLAST server (Altschul et al., 1997) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Sequence alignment and scoring was done with BLASTp (protein-protein BLAST). Multiple sequence alignment was done with ClustalW, (Larkin et al., 2007) (<http://www.ebi.ac.uk/Tools/clustalw2/index.html>) and visualized using Jalview (Waterhouse et al., 2009). Repeat detection in metazoan Pur- $\alpha$  was done by J. Söding (Gene Center, Munich, Germany) using the HHRepID server (Biegert and Söding, 2008) (<http://toolkit.tuebingen.mpg.de/hhrepid>).

Secondary structure prediction was performed with the Network Protein Sequence Analysis server from the Pole Bioinformatique Lyonnais using the consensus prediction option combining the programs DSC, MLRC, and PHD (Combet et al., 2000) ([http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=/NPSA/npsa\\_server.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html)).

The computational model of *D. melanogaster* Pur- $\alpha$  was accomplished by J. Söding (Gene Center, Munich, Germany), using the Structural Prediction Server HHpred (Söding, 2005; Söding et al., 2005) (<http://toolkit.tuebingen.mpg.de/hhpred>).

### **5.17.3 Structure visualization and analysis**

Images of the crystal structures were prepared with PyMol (DeLano Scientific, San Carlos, USA). Electrostatic surface potentials were calculated and represented with CCP4mg, and buried surface areas were calculated by Areaimol (Collaborative Computational Project, 1994). Sequence alignments was performed using T-coffee (Notredame et al., 2000) and a surface plot of sequence conservation was prepared with Chimera (Pettersen et al., 2004) Systematic structural comparison was accomplished using the DaliLite server (Holm et al., 2008) ([http://ekhidna.biocenter.helsinki.fi/dali\\_server/](http://ekhidna.biocenter.helsinki.fi/dali_server/)).

## Chapter 6 References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *PNAS* 25, 13.
- Anderson, D., Harris, R., Polayes, D., Ciccarone, V., Donahue, R., Gerard, G., and Jessee, J. (1996). Rapid generation of recombinant baculoviruses and expression of foreign genes using the Bac-to-Bac virus expression system. *Focus* 17, 53-58.
- Arocena, D.G., Iwahashi, C.K., Won, N., Beilina, A., Ludwig, A.L., Tassone, F., Schwartz, P.H., and Hagerman, P.J. (2005). Induction of inclusion formation and disruption of lamin A/C structure by premutation CGG-repeat RNA in human cultured neural cells. *Hum Mol Genet* 14, 3661-3671.
- Auweter, S.D., Oberstrass, F.C., and Allain, F.H. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res* 34, 4943-4959.
- Bahadur, R.P., and Zacharias, M. (2008). The interface of protein-protein complexes: analysis of contacts and prediction of interactions. *Cell Mol Life Sci* 65, 1059-1072.
- Bahadur, R.P., Zacharias, M., and Janin, J. (2008). Dissecting protein-RNA recognition sites. *Nucleic Acids Res* 36, 2705-2716.
- Barbarese, E., Koppel, D.E., Deutscher, M.P., Smith, C.L., Ainger, K., Morgan, F., and Carson, J.H. (1995). Protein translation components are colocalized in granules in oligodendrocytes. *J Cell Sci* 108 ( Pt 8), 2781-2790.
- Barbee, S.A., Estes, P.S., Cziko, A.M., Hillebrand, J., Luedeman, R.A., Coller, J.M., Johnson, N., Howlett, I.C., Geng, C., Ueda, R., *et al.* (2006). Staufen- and FMRP-containing neuronal RNPs are structurally and functionally related to somatic P bodies. *Neuron* 52, 997-1009.
- Barr, S.M., and Johnson, E.M. (2001). Ras-induced colony formation and anchorage-independent growth inhibited by elevated expression of Puralpha in NIH3T3 cells. *J Cell Biochem* 81, 621-638.
- Becalska, A.N., and Gavis, E.R. (2009). Lighting up mRNA localization in *Drosophila* oogenesis. *Development* 136, 2493-2503.
- Beck, T., Krasauskas, A., Gruene, T., and Sheldrick, G.M. (2008). A magic triangle for experimental phasing of macromolecules. *Acta Crystallogr D Biol Crystallogr* 64, 1179-1182.
- Benson, D.A., Boguski, M., Lipman, D.J., and Ostell, J. (1994). GenBank. *Nucleic Acids Res* 22, 3441-3444.
- 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. *Mol Cell Biol* 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. *Mol Cell Biol* 12, 5673-5682.
- Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129, 5656-5664.

- Berry-Kravis, E., Abrams, L., Coffey, S.M., Hall, D.A., Greco, C., Gane, L.W., Grigsby, J., Bourgeois, J.A., Finucane, B., Jacquemont, S., *et al.* (2007). Fragile X-associated tremor/ataxia syndrome: clinical features, genetics, and testing guidelines. *Mov Disord* 22, 2018-2030, quiz 2140.
- Biegert, A., and Söding, J. (2008). De novo identification of highly diverged protein repeats by probabilistic consistency. *Bioinformatics* 24, 807-814.
- Bole, M., Menon, L., and Mihailescu, M.R. (2008). Fragile X mental retardation protein recognition of G quadruplex structure per se is sufficient for high affinity binding to RNA. *Mol Biosyst* 4, 1212-1219.
- Bramham, C.R., and Wells, D.G. (2007). Dendritic mRNA: transport, translation and function. *Nat Rev Neurosci* 8, 776-789.
- Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K., and Neidle, S. (2006). Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res* 34, 5402-5415.
- Butler, M.J., Jacobsen, T.L., Cain, D.M., Jarman, M.G., Hubank, M., Whittle, J.R., Phillips, R., and Simcox, A. (2003). Discovery of genes with highly restricted expression patterns in the *Drosophila* wing disc using DNA oligonucleotide microarrays. *Development* 130, 659-670.
- Callaerts, P., Leng, S., Clements, J., Benassayag, C., Cribbs, D., Kang, Y.Y., Walldorf, U., Fischbach, K.F., and Strauss, R. (2001). *Drosophila* Pax-6/eyeless is essential for normal adult brain structure and function. *J Neurobiol* 46, 73-88.
- Cassiday, L.A., and Maher, L.J., 3rd (2002). Having it both ways: transcription factors that bind DNA and RNA. *Nucleic Acids Res* 30, 4118-4126.
- Chang, C.F., Gallia, G.L., Muralidharan, V., Chen, N.N., Zoltick, P., Johnson, E., and Khalili, K. (1996). Evidence that replication of human neurotropic JC virus DNA in glial cells is regulated by the sequence-specific single-stranded DNA-binding protein Pur alpha. *J Virol* 70, 4150-4156.
- Chao, J.A., Patskovsky, Y., Almo, S.C., and Singer, R.H. (2008). Structural basis for the coevolution of a viral RNA-protein complex. *Nat Struct Mol Biol* 15, 103-105.
- Chao, J.A., Patskovsky, Y., Patel, V., Levy, M., Almo, S.C., and Singer, R.H. (2010). ZBP1 recognition of beta-actin zipcode induces RNA looping. *Genes Dev* 24, 148-158.
- Chen, H.W., Marinissen, M.J., Oh, S.W., Chen, X., Melnick, M., Perrimon, N., Gutkind, J.S., and Hou, S.X. (2002). CKA, a novel multidomain protein, regulates the JUN N-terminal kinase signal transduction pathway in *Drosophila*. *Mol Cell Biol* 22, 1792-1803.
- Chepenik, L.G., Tretiakova, A.P., Krachmarov, C.P., Johnson, E.M., and Khalili, K. (1998). The single-stranded DNA binding protein, Pur-alpha, binds HIV-1 TAR RNA and activates HIV-1 transcription. *Gene* 210, 37-44.
- Cinque, P., Koralnik, I.J., Gerevini, S., Miro, J.M., and Price, R.W. (2009). Progressive multifocal leukoencephalopathy in HIV-1 infection. *Lancet Infect Dis* 9, 625-636.
- Clements, J., Hens, K., Francis, C., Schellens, A., and Callaerts, P. (2008). Conserved role for the *Drosophila* Pax6 homolog Eyeless in differentiation and function of insulin-producing neurons. *Proc Natl Acad Sci U S A* 105, 16183-16188.
- Clements, J., Hens, K., Merugu, S., Dichtl, B., de Couet, H.G., and Callaerts, P. (2009). Mutational analysis of the eyeless gene and phenotypic rescue reveal that an intact

Eyeless protein is necessary for normal eye and brain development in *Drosophila*. *Dev Biol* 334, 503-512.

Collaborative Computational Project, N. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr D Biol Crystallogr* 50, 760-763.

Combet, C., Blanchet, C., Geourjon, C., and Deleage, G. (2000). NPS@: network protein sequence analysis. *Trends Biochem Sci* 25, 147-150.

Cowieson, N.P., Kobe, B., and Martin, J.L. (2008). United we stand: combining structural methods. *Curr Opin Struct Biol* 18, 617-622.

Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr D Biol Crystallogr* 62, 1002-1011.

Darbinian, N., Cui, J., Basile, A., Del Valle, L., Otte, J., Miklossy, J., Sawaya, B.E., Amini, S., Khalili, K., and Gordon, J. (2008). Negative regulation of AbetaPP gene expression by pur-alpha. *J Alzheimers Dis* 15, 71-82.

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

Darbinian, N., Gallia, G.L., King, J., Del Valle, L., Johnson, E.M., and Khalili, K. (2001b). Growth inhibition of glioblastoma cells by human Pur(alpha). *J Cell Physiol* 189, 334-340.

Darbinian, N., Gallia, G.L., Kundu, M., Shcherbik, N., Tretiakova, A., Giordano, A., and Khalili, K. (1999). Association of Pur alpha and E2F-1 suppresses transcriptional activity of E2F-1. *Oncogene* 18, 6398-6402.

Darbinian, N., Sawaya, B.E., Khalili, K., Jaffe, N., Wortman, B., Giordano, A., and Amini, S. (2001c). Functional interaction between cyclin T1/cdk9 and Puralpha determines the level of TNFalpha promoter activation by Tat in glial cells. *J Neuroimmunol* 121, 3-11.

Darbinian, N., White, M.K., and Khalili, K. (2006). Regulation of the Pur-alpha promoter by E2F-1. *J Cell Biochem* 99, 1052-1063.

Davis, T.R., Trotter, K.M., Granados, R.R., and Wood, H.A. (1992). Baculovirus expression of alkaline phosphatase as a reporter gene for evaluation of production, glycosylation and secretion. *Biotechnology (N Y)* 10, 1148-1150.

Desveaux, D., Allard, J., Brisson, N., and Sygusch, J. (2002). A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nat Struct Biol* 9, 512-517.

Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., and Xu, R.M. (1999). Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* 13, 1102-1115.

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. *J Neurochem* 105, 1979-1995.

Doublet, S. (1997). Preparation of Selenomethionyl Proteins for Phase Determination. In *Meth Enzymol*, J. Charles W. Carter, and R.M. Sweet, eds., pp. 523-537.

Eberwine, J., Miyashiro, K., Kacharmina, J.E., and Job, C. (2001). Local translation of classes of mRNAs that are targeted to neuronal dendrites. *Proc Natl Acad Sci U S A* 98, 7080-7085.

- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* *60*, 2126-2132.
- Fletcher, D.A., and Mullins, R.D. (2010). Cell mechanics and the cytoskeleton. *Nature* *463*, 485-492.
- Formstecher, E., Aresta, S., Collura, V., Hamburger, A., Meil, A., Trehin, A., Reverdy, C., Betin, V., Maire, S., Brun, C., *et al.* (2005). Protein interaction mapping: a *Drosophila* case study. *Genome Res* *15*, 376-384.
- Fromont-Racine, M., Rain, J.C., and Legrain, P. (1997). Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet* *16*, 277-282.
- Gallia, G.L., Darbinian, N., Johnson, E.M., and Khalili, K. (1999). Self-association of Puralpha is mediated by RNA. *J Cell Biochem* *74*, 334-348.
- Gallia, G.L., Johnson, E.M., and Khalili, K. (2000). Puralpha: a multifunctional single-stranded DNA- and RNA-binding protein. *Nucleic Acids Res* *28*, 3197-3205.
- Gallia, G.L., Safak, M., and Khalili, K. (1998). Interaction of the single-stranded DNA-binding protein Puralpha with the human polyomavirus JC virus early protein T-antigen. *J Biol Chem* *273*, 32662-32669.
- Gao, X., Bain, K., Bonanno, J.B., Buchanan, M., Henderson, D., Lorimer, D., Marsh, C., Reynes, J.A., Sauder, J.M., Schwinn, K., *et al.* (2005). High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. *J Struct Funct Genomics* *6*, 129-134.
- Gao, Y., Tatavarty, V., Korza, G., Levin, M.K., and Carson, J.H. (2008). Multiplexed dendritic targeting of alpha calcium calmodulin-dependent protein kinase II, neurogranin, and activity-regulated cytoskeleton-associated protein RNAs by the A2 pathway. *Mol Biol Cell* *19*, 2311-2327.
- 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.
- Greco, C.M., Hagerman, R.J., Tassone, F., Chudley, A.E., Del Bigio, M.R., Jacquemont, S., Leehey, M., and Hagerman, P.J. (2002). Neuronal intranuclear inclusions in a new cerebellar tremor/ataxia syndrome among fragile X carriers. *Brain* *125*, 1760-1771.
- Guinier, A., and Fournet, F. (1955). *Small Angle Scattering of X-rays*. Wiley Interscience *New York*.
- Gupta, M., Sueblinvong, V., Raman, J., Jeevanandam, V., and Gupta, M.P. (2003). Single-stranded DNA-binding proteins PURalpha and PURbeta bind to a purine-rich negative regulatory element of the alpha-myosin heavy chain gene and control transcriptional and translational regulation of the gene expression. Implications in the repression of alpha-myosin heavy chain during heart failure. *J Biol Chem* *278*, 44935-44948.
- Haas, S., Gordon, J., and Khalili, K. (1993). A developmentally regulated DNA-binding protein from mouse brain stimulates myelin basic protein gene expression. *Mol Cell Biol* *13*, 3103-3112.
- Habiger, C., Stelzer, G., Schwarz, U., and Winnacker, E.L. (1992). Two cellular single-strand-specific DNA-binding proteins interact with two regions of the bovine papillomavirus type 1 genome, including the origin of DNA replication. *J Virol* *66*, 5988-5998.

- 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.
- Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* 166, 557-580.
- Hirokawa, N., and Takemura, R. (2005). Molecular motors and mechanisms of directional transport in neurons. *Nat Rev Neurosci* 6, 201-214.
- Hirsch, J.A., Schubert, C., Gurevich, V.V., and Sigler, P.B. (1999). The 2.8 Å crystal structure of visual arrestin: a model for arrestin's regulation. *Cell* 97, 257-269.
- Ho, S.N., Hunt, H.D., Horton, R.M., Pullen, J.K., and Pease, L.R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77, 51-59.
- Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics* 24, 2780-2781.
- Holm, L., and Park, J. (2000). DaliLite workbench for protein structure comparison. *Bioinformatics* 16, 566-567.
- Huang, C.C., Chiribau, C.B., Majumder, M., Chiang, C.M., Wek, R.C., Kelm, R.J., Jr., Khalili, K., Snider, M.D., and Hatzoglou, M. (2009). A bifunctional intronic element regulates the expression of the arginine/lysine transporter Cat-1 via mechanisms involving the purine-rich element binding protein A (Pur alpha). *J Biol Chem* 284, 32312-32320.
- Huang, Y.S., Carson, J.H., Barbarese, E., and Richter, J.D. (2003). Facilitation of dendritic mRNA transport by CPEB. *Genes Dev* 17, 638-653.
- 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. *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 Differ* 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* 129, 256-271.
- Jacobsen, T.L., Cain, D., Paul, L., Justiniano, S., Alli, A., Mullins, J.S., Wang, C.P., Butchar, J.P., and Simcox, A. (2006). Functional analysis of genes differentially expressed in the *Drosophila* wing disc: role of transcripts enriched in the wing region. *Genetics* 174, 1973-1982.
- 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. *Mol Cell Biol* 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.
- Job, C., and Eberwine, J. (2001). Identification of sites for exponential translation in living dendrites. *Proc Natl Acad Sci U S A* *98*, 13037-13042.
- 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. *J Biol Chem* *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. *J Neurosci Res* *83*, 929-943.
- Johnstone, O., and Lasko, P. (2001). Translational regulation and RNA localization in *Drosophila* oocytes and embryos. *Annu Rev Genet* *35*, 365-406.
- Jurk, M., Weissinger, F., Lottspeich, F., Schwarz, U., and Winnacker, E.L. (1996). Characterization of the single-strand-specific BPV-1 origin binding protein, SPSF I, as the HeLa Pur alpha factor. *Nucleic Acids Res* *24*, 2799-2806.
- Kabsch, W. (1993). Automatic Processing of Rotation Diffraction Data from Crystals of Initially Unknown Symmetry and Cell Constants. *J Appl Crystallogr* *26*, 795-800.
- Kaminski, R., Darbinian, N., Sawaya, B.E., Slonina, D., Amini, S., Johnson, E.M., Rappaport, J., Khalili, K., and Darbinyan, A. (2008a). Puralpha as a cellular co-factor of Rev/RRE-mediated expression of HIV-1 intron-containing mRNA. *J Cell Biochem* *103*, 1231-1245.
- Kaminski, R., Darbinyan, A., Merabova, N., Deshmane, S.L., White, M.K., and Khalili, K. (2008b). Protective role of Puralpha to cisplatin. *Cancer Biol Ther* *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.
- Kantardjieff, K., and Rupp, K. (2003). Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals. *Protein Science* *12*, 1865-1871.
- Kelm, R.J., Jr., Cogan, J.G., Elder, P.K., Strauch, A.R., and Getz, M.J. (1999a). Molecular interactions between single-stranded DNA-binding proteins associated with an essential MCAT element in the mouse smooth muscle alpha-actin promoter. *J Biol Chem* *274*, 14238-14245.
- Kelm, R.J., Jr., Elder, P.K., and Getz, M.J. (1999b). The single-stranded DNA-binding proteins, Puralpha, Purbeta, and MSY1 specifically interact with an exon 3-derived mouse vascular smooth muscle alpha-actin messenger RNA sequence. *J Biol Chem* *274*, 38268-38275.
- Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., and Phillips, D.C. (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* *181*, 662-666.
- Kenneson, A., Zhang, F., Hagedorn, C.H., and Warren, S.T. (2001). Reduced FMRP and increased FMR1 transcription is proportionally associated with CGG repeat number in intermediate-length and premutation carriers. *Hum Mol Genet* *10*, 1449-1454.

- 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. *Mol Cell Biol* 23, 6857-6875.
- Kiebler, M.A., Hemraj, I., Verkade, P., Kohrmann, M., Fortes, P., Marion, R.M., Ortin, J., and Dotti, C.G. (1999). The mammalian stau protein localizes to the somatodendritic domain of cultured hippocampal neurons: implications for its involvement in mRNA transport. *J Neurosci* 19, 288-297.
- 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. *J Biol Chem* 283, 9113-9126.
- Kindler, S., Wang, H., Richter, D., and Tiedge, H. (2005). RNA transport and local control of translation. *Annu Rev Cell Dev Biol* 21, 223-245.
- Knablein, J., Neufeind, T., Schneider, F., Bergner, A., Messerschmidt, A., Lowe, J., Steipe, B., and Huber, R. (1997). Ta6Br(2+)12, a tool for phase determination of large biological assemblies by X-ray crystallography. *J Mol Biol* 270, 1-7.
- Knapp, A.M., Ramsey, J.E., Wang, S.X., Godburn, K.E., Strauch, A.R., and Kelm, R.J., Jr. (2006). Nucleoprotein interactions governing cell type-dependent repression of the mouse smooth muscle alpha-actin promoter by single-stranded DNA-binding proteins Pur alpha and Pur beta. *J Biol Chem* 281, 7907-7918.
- Knowles, R.B., Sabry, J.H., Martone, M.E., Deerinck, T.J., Ellisman, M.H., Bassell, G.J., and Kosik, K.S. (1996). Translocation of RNA granules in living neurons. *J Neurosci* 16, 7812-7820.
- Kobayashi, M., Kubota, M., and Matsuura, Y. (1999). Crystallization and improvement of crystal quality for x-ray diffraction of maltotriose trehalose synthase by reductive methylation of lysine residues. *Acta Crystallogr D Biol Crystallogr* 55, 931-933.
- 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. *Biochem Biophys Res Commun* 277, 341-347.
- Koehl, P., and Levitt, M. (2002). Sequence variations within protein families are linearly related to structural variations. *J Mol Biol* 323, 551-562.
- 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., Volkov, 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.
- Kong, T., Scully, M., Shelley, C.S., and Colgan, S.P. (2007). Identification of Pur alpha as a new hypoxia response factor responsible for coordinated induction of the beta 2 integrin family. *J Immunol* 179, 1934-1941.
- Kosik, K.S., and Krichevsky, A.M. (2002). The message and the messenger: delivering RNA in neurons. *Sci STKE* 2002, pe16.
- Kozin, M.B., and Svergun, D.I. (2000). Automated matching of high- and low-resolution structural models. *J Appl Crystallogr* 34, 33-41.

- Krachmarov, C.P., Chepenik, L.G., Barr-Vagell, S., Khalili, K., and Johnson, E.M. (1996). Activation of the JC virus Tat-responsive transcriptional control element by association of the Tat protein of human immunodeficiency virus 1 with cellular protein Pur alpha. *Proc Natl Acad Sci U S A* *93*, 14112-14117.
- Krichevsky, A.M., and Kosik, K.S. (2001). Neuronal RNA granules: a link between RNA localization and stimulation-dependent translation. *Neuron* *32*, 683-696.
- Krimm, I., Ostlund, C., Gilquin, B., Couprie, J., Hossenlopp, P., Mornon, J.P., Bonne, G., Courvalin, J.C., Worman, H.J., and Zinn-Justin, S. (2002). The Ig-like structure of the C-terminal domain of lamin A/C, mutated in muscular dystrophies, cardiomyopathy, and partial lipodystrophy. *Structure* *10*, 811-823.
- Kundrot, C.E. (2004). Which strategy for a protein crystallization project? *Cell Mol Life Sci* *61*, 525-536.
- Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* *227*, 680-685.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettingan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). ClustalW and ClustalX version 2. *Bioinformatics* *23*, 2.
- Lasham, A., Lindridge, E., Rudert, F., Onrust, R., and Watson, J. (2000). Regulation of the human fas promoter by YB-1, Puralpha and AP-1 transcription factors. *Gene* *252*, 1-13.
- Lawrence, C.J., Dawe, R.K., Christie, K.R., Cleveland, D.W., Dawson, S.C., Endow, S.A., Goldstein, L.S., Goodson, H.V., Hirokawa, N., Howard, J., *et al.* (2004). A standardized kinesin nomenclature. *J Cell Biol* *167*, 19-22.
- Lee, S.J., and Montell, C. (2004). Light-dependent translocation of visual arrestin regulated by the NINAC myosin III. *Neuron* *43*, 95-103.
- Lefkowitz, R.J., and Shenoy, S.K. (2005). Transduction of receptor signals by beta-arrestins. *Science* *308*, 512-517.
- 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.
- Limesand, S.W., Jeckel, K.M., and Anthony, R.V. (2004). Puralpha, a single-stranded deoxyribonucleic acid binding protein, augments placental lactogen gene transcription. *Mol Endocrinol* *18*, 447-457.
- Lin, D., Pestova, T.V., Hellen, C.U., and Tiedge, H. (2008). Translational control by a small RNA: dendritic BC1 RNA targets the eukaryotic initiation factor 4A helicase mechanism. *Mol Cell Biol* *28*, 3008-3019.
- Liu, H., Barr, S.M., Chu, C., Kohtz, D.S., Kinoshita, Y., and Johnson, E.M. (2005). Functional interaction of Puralpha with the Cdk2 moiety of cyclin A/Cdk2. *Biochem Biophys Res Commun* *328*, 851-857.
- 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 Res* *30*, 2417-2426.

- Liu, X., Gomez-Pinillos, A., Johnson, E.M., and Ferrari, A.C. (2010). Induction of bicalutamide sensitivity in prostate cancer cells by an epigenetic Puralpha-mediated decrease in androgen receptor levels. *Prostate* 70, 179-189.
- Luckow, V.A., Lee, S.C., Barry, G.F., and Olins, P.O. (1993). Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J Virol* 67, 4566-4579.
- Lunde, B.M., Moore, C., and Varani, G. (2007). RNA-binding proteins: modular design for efficient function. *Nat Rev Mol Cell Biol* 8, 479-490.
- Ma, B., Elkayam, T., Wolfson, H., and Nussinov, R. (2003). Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc Natl Acad Sci U S A* 100, 5772-5777.
- Mallik, R., and Gross, S.P. (2004). Molecular motors: strategies to get along. *Curr Biol* 14, R971-982.
- Maris, C., Dominguez, C., and Allain, F.H. (2005). The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272, 2118-2131.
- Marszalek, J.R., Liu, X., Roberts, E.A., Chui, D., Marth, J.D., Williams, D.S., and Goldstein, L.S. (2000). Genetic evidence for selective transport of opsin and arrestin by kinesin-II in mammalian photoreceptors. *Cell* 102, 175-187.
- Martin, K.C., and Ephrussi, A. (2009). mRNA localization: gene expression in the spatial dimension. *Cell* 136, 719-730.
- Martin, K.C., and Zukin, R.S. (2006). RNA trafficking and local protein synthesis in dendrites: an overview. *J Neurosci* 26, 7131-7134.
- Matthews, B.W. (1968). Solvent content of protein crystals. *J Mol Biol* 33, 491-497.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J Appl Crystallogr* 40, 658-674.
- Messer, M., Griffiths, M., Rismiller, P.D., and Shaw, D.C. (1997). Lactose synthesis in a monotreme, the echidna (*Tachyglossus aculeatus*): isolation and amino acid sequence of echidna alpha-lactalbumin. *Comp Biochem Physiol B Biochem Mol Biol* 118, 403-410.
- Muller, M., Heuck, A., and Niessing, D. (2007). Directional mRNA transport in eukaryotes: lessons from yeast. *Cell Mol Life Sci* 64, 171-180.
- Muralidharan, V., Sweet, T., Nadraga, Y., Amini, S., and Khalili, K. (2001). Regulation of Puralpha gene transcription: evidence for autoregulation of Puralpha promoter. *J Cell Physiol* 186, 406-413.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53, 240-255.
- Mylonas, E., and Svergun, D.I. (2007). Accuracy of molecular mass determination of proteins in solution by small angle X-ray scattering. *J Appl Crystallogr* 40, s245-s249.
- Nair, K.S., Hanson, S.M., Mendez, A., Gurevich, E.V., Kennedy, M.J., Shestopalov, V.I., Vishnivetskiy, S.A., Chen, J., Hurley, J.B., Gurevich, V.V., *et al.* (2005). Light-dependent redistribution of arrestin in vertebrate rods is an energy-independent process governed by protein-protein interactions. *Neuron* 46, 555-567.

- Ness, S.R., de Graaff, R.A., Abrahams, J.P., and Pannu, N.S. (2004). CRANK: new methods for automated macromolecular crystal structure solution. *Structure* 12, 1753-1761.
- Ni, C.Z., Syed, R., Kodandapani, R., Wickersham, J., Peabody, D.S., and Ely, K.R. (1995). Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure* 3, 255-263.
- 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.
- O'Donnell, W.T., and Warren, S.T. (2002). A decade of molecular studies of fragile X syndrome. *Annu Rev Neurosci* 25, 315-338.
- Ofer, N., Weisman-Shomer, P., Shklover, J., and Fry, M. (2009). The quadruplex r(CG<sub>n</sub>) destabilizing cationic porphyrin TMPyP4 cooperates with hnRNPs to increase the translation efficiency of fragile X premutation mRNA. *Nucleic Acids Res* 37, 2712-2722.
- 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. *J Neurochem* 75, 1781-1790.
- Ohashi, S., Koike, K., Omori, A., Ichinose, S., Ohara, S., Kobayashi, S., Sato, T.A., and Anzai, K. (2002). Identification of mRNA/protein (mRNP) complexes containing Pur-alpha, mStaufen, fragile X protein, and myosin Va and their association with rough endoplasmic reticulum equipped with a kinesin motor. *J Biol Chem* 277, 37804-37810.
- Oostra, B.A., and Willemsen, R. (2009). FMR1: a gene with three faces. *Biochim Biophys Acta* 1790, 467-477.
- Osugi, T., Ding, Y., Tanaka, H., Kuo, C.H., Do, E., Irie, Y., and Miki, N. (1996). Involvement of a single-stranded DNA binding protein, ssCRE-BP/Pur alpha, in morphine dependence. *FEBS Lett* 391, 11-16.
- Ouyang, Y., Rosenstein, A., Kreiman, G., Schuman, E.M., and Kennedy, M.B. (1999). Tetanic stimulation leads to increased accumulation of Ca<sup>2+</sup>/calmodulin-dependent protein kinase II via dendritic protein synthesis in hippocampal neurons. *J Neurosci* 19, 7823-7833.
- Penberthy, W.T., Zhao, C., Zhang, Y., Jessen, J.R., Yang, Z., Bricaud, O., Collazo, A., Meng, A., and Lin, S. (2004). Pur alpha and Sp8 as opposing regulators of neural gata2 expression. *Dev Biol* 275, 225-234.
- Perez-Torrado, R., Yamada, D., and Defossez, P.A. (2006). Born to bind: the BTB protein-protein interaction domain. *Bioessays* 28, 1194-1202.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 25, 1605-1612.
- Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 40, 191-285.

- Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., *et al.* (2001). The protein-protein interaction map of *Helicobacter pylori*. *Nature* *409*, 211-215.
- Ramsey, J.E., Daugherty, M.A., and Kelm, R.J., Jr. (2007). Hydrodynamic studies on the quaternary structure of recombinant mouse Purbeta. *J Biol Chem* *282*, 1552-1560.
- Sadakata, T., Kuo, C., Ichikawa, H., Nishikawa, E., Niu, S.Y., Kumamaru, E., and Miki, N. (2000). Puralpha, a single-stranded DNA binding protein, suppresses the enhancer activity of cAMP response element (CRE). *Brain Res Mol Brain Res* *77*, 47-54.
- Safak, M., Gallia, G.L., and Khalili, K. (1999). Reciprocal interaction between two cellular proteins, Puralpha and YB-1, modulates transcriptional activity of JCVCY in glial cells. *Mol Cell Biol* *19*, 2712-2723.
- Sambrook, J., and Russell, D.W. (2001). *Molecular cloning: a laboratory manual*, Third Edition edn (Cold Spring Harbor, Cold Spring Harbor Laboratory Press).
- Satoh, A.K., and Ready, D.F. (2005). Arrestin1 mediates light-dependent rhodopsin endocytosis and cell survival. *Curr Biol* *15*, 1722-1733.
- Savchenko, A., Yee, A., Khachatryan, A., Skarina, T., Evdokimova, E., Pavlova, M., Semesi, A., Northey, J., Beasley, S., Lan, N., *et al.* (2003). Strategies for structural proteomics of prokaryotes: Quantifying the advantages of studying orthologous proteins and of using both NMR and X-ray crystallography approaches. *Proteins* *50*, 392-399.
- Schaeffer, C., Bardoni, B., Mandel, J.L., Ehresmann, B., Ehresmann, C., and Moine, H. (2001). The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* *20*, 4803-4813.
- Schulze, S.R., Curio-Penny, B., Li, Y., Imani, R.A., Rydberg, L., Geyer, P.K., and Wallrath, L.L. (2005). Molecular genetic analysis of the nested *Drosophila melanogaster* lamin C gene. *Genetics* *171*, 185-196.
- Schumacher, M.A., Karamooz, E., Zikova, A., Trantirek, L., and Lukes, J. (2006). Crystal structures of *T. brucei* MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell* *126*, 701-711.
- Schuman, E.M., Dynes, J.L., and Steward, O. (2006). Synaptic regulation of translation of dendritic mRNAs. *J Neurosci* *26*, 7143-7146.
- Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr A* *64*, 112-122.
- Shelley, C.S., Da Silva, N., and Teodoridis, J.M. (2001). During U937 monocytic differentiation repression of the CD43 gene promoter is mediated by the single-stranded DNA binding protein Pur alpha. *Br J Haematol* *115*, 159-166.
- Shelley, C.S., Teodoridis, J.M., Park, H., Farokhzad, O.C., Bottinger, E.P., and Arnaout, M.A. (2002). During differentiation of the monocytic cell line U937, Pur alpha mediates induction of the CD11c beta 2 integrin gene promoter. *J Immunol* *168*, 3887-3893.
- Singer, R.H. (1993). RNA zipcodes for cytoplasmic addresses. *Curr Biol* *3*, 719-721.
- Smith, C.W.J. (1998). *RNA:Protein Interactions - A practical approach* (Oxford, Oxford University Press).
- Smith, T.F., Gaitatzes, C., Saxena, K., and Neer, E.J. (1999). The WD repeat: a common architecture for diverse functions. *Trends Biochem Sci* *24*, 181-185.
- Smyth, D.R., Mrozkiewicz, M.K., McGrath, W.J., Listwan, P., and Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. *Protein Sci* *12*, 1313-1322.

- Söding, J. (2005). Protein homology detection by HMM-HMM comparison. *Bioinformatics* 21, 951-960.
- Söding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Research* 33.
- Spann, T.P., Goldman, A.E., Wang, C., Huang, S., and Goldman, R.D. (2002). Alteration of nuclear lamin organization inhibits RNA polymerase II-dependent transcription. *J Cell Biol* 156, 603-608.
- St Johnston, D. (2005). Moving messages: the intracellular localization of mRNAs. *Nat Rev Mol Cell Biol* 6, 363-375.
- Stacey, D.W., Hitomi, M., Kanovsky, M., Gan, L., and Johnson, E.M. (1999). Cell cycle arrest and morphological alterations following microinjection of NIH3T3 cells with Pur alpha. *Oncogene* 18, 4254-4261.
- Steward, O., and Levy, W.B. (1982). Preferential localization of polyribosomes under the base of dendritic spines in granule cells of the dentate gyrus. *J Neurosci* 2, 284-291.
- Steward, O., and Schuman, E.M. (2003). Compartmentalized synthesis and degradation of proteins in neurons. *Neuron* 40, 347-359.
- Steward, O., Wallace, C.S., Lyford, G.L., and Worley, P.F. (1998). Synaptic activation causes the mRNA for the IEG Arc to localize selectively near activated postsynaptic sites on dendrites. *Neuron* 21, 741-751.
- Steward, O., and Worley, P.F. (2001). Selective targeting of newly synthesized Arc mRNA to active synapses requires NMDA receptor activation. *Neuron* 30, 227-240.
- Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J Appl Crystallogr* 25, 495-503.
- Svergun, D.I., Barberato, C., and Koch, M.H.J. (1995). CRY SOL- a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28, 768-773.
- Svergun, D.I., Petoukhov, M.V., and Koch, M.H.J. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal* 80, 2948-2953.
- Tassone, F., Hagerman, R.J., Taylor, A.K., Gane, L.W., Godfrey, T.E., and Hagerman, P.J. (2000). Elevated levels of FMR1 mRNA in carrier males: a new mechanism of involvement in the fragile-X syndrome. *Am J Hum Genet* 66, 6-15.
- Terwilliger, T.C. (2002). Automated structure solution, density modification and model building. *Acta Cryst D* 58, 1937-1940.
- Thatikunta, P., Sawaya, B.E., Denisova, L., Cole, C., Yusibova, G., Johnson, E.M., Khalili, K., and Amini, S. (1997). Identification of a cellular protein that binds to Tat-responsive element of TGF beta-1 promoter in glial cells. *J Cell Biochem* 67, 466-477.
- Thompson, K.R., Otis, K.O., Chen, D.Y., Zhao, Y., O'Dell, T.J., and Martin, K.C. (2004). Synapse to nucleus signaling during long-term synaptic plasticity; a role for the classical active nuclear import pathway. *Neuron* 44, 997-1009.
- Tiedge, H., Fremeau, R.T., Jr., Weinstock, P.H., Arancio, O., and Brosius, J. (1991). Dendritic location of neural BC1 RNA. *Proc Natl Acad Sci U S A* 88, 2093-2097.
- Tiruchinapalli, D.M., Oleynikov, Y., Kelic, S., Shenoy, S.M., Hartley, A., Stanton, P.K., Singer, R.H., and Bassell, G.J. (2003). Activity-dependent trafficking and dynamic

localization of zipcode binding protein 1 and beta-actin mRNA in dendrites and spines of hippocampal neurons. *J Neurosci* 23, 3251-3261.

Tretiakova, A., Gallia, G.L., Shcherbik, N., Jameson, B., Johnson, E.M., Amini, S., and Khalili, K. (1998). Association of Puralpha with RNAs homologous to 7 SL determines its binding ability to the myelin basic protein promoter DNA sequence. *J Biol Chem* 273, 22241-22247.

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. *J Cell Physiol* 181, 160-168.

Tsukahara, T., Tsujino, S., and Arahata, K. (2002). CDNA microarray analysis of gene expression in fibroblasts of patients with X-linked Emery-Dreifuss muscular dystrophy. *Muscle Nerve* 25, 898-901.

Vaguine, A.A., Richelle, J., and Wodak, S.J. (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* 55, 191-205.

Vale, R.D. (2003). The molecular motor toolbox for intracellular transport. *Cell* 112, 467-480.

Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J Appl Crystallogr* 36, 860-864.

Walter, T.S., Meier, C., Assenberg, R., Au, K.F., Ren, J., Verma, A., Nettleship, J.E., Owens, R.J., Stuart, D.I., and Grimes, J.M. (2006). Lysine methylation as a routine rescue strategy for protein crystallization. *Structure* 14, 1617-1622.

Walther, C., and Gruss, P. (1991). Pax-6, a murine paired box gene, is expressed in the developing CNS. *Development* 113, 1435-1449.

Wang, H., Iacoangeli, A., Lin, D., Williams, K., Denman, R.B., Hellen, C.U., and Tiedge, H. (2005). Dendritic BC1 RNA in translational control mechanisms. *J Cell Biol* 171, 811-821.

Wang, H., Iacoangeli, A., Popp, S., Muslimov, I.A., Imataka, H., Sonenberg, N., Lomakin, I.B., and Tiedge, H. (2002). Dendritic BC1 RNA: functional role in regulation of translation initiation. *J Neurosci* 22, 10232-10241.

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 Biol Ther* 6, 596-602.

Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M., and Barton, G.J. (2009). Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189-1191.

Weng, L., and Wei, D. (2002). Role of Cka in imaginal disc growth and differentiation. *Dros Inf Serv* 85, 5.

White, M.K., Johnson, E.M., and Khalili, K. (2009). Multiple roles for Puralpha in cellular and viral regulation. *Cell Cycle* 8, 1-7.

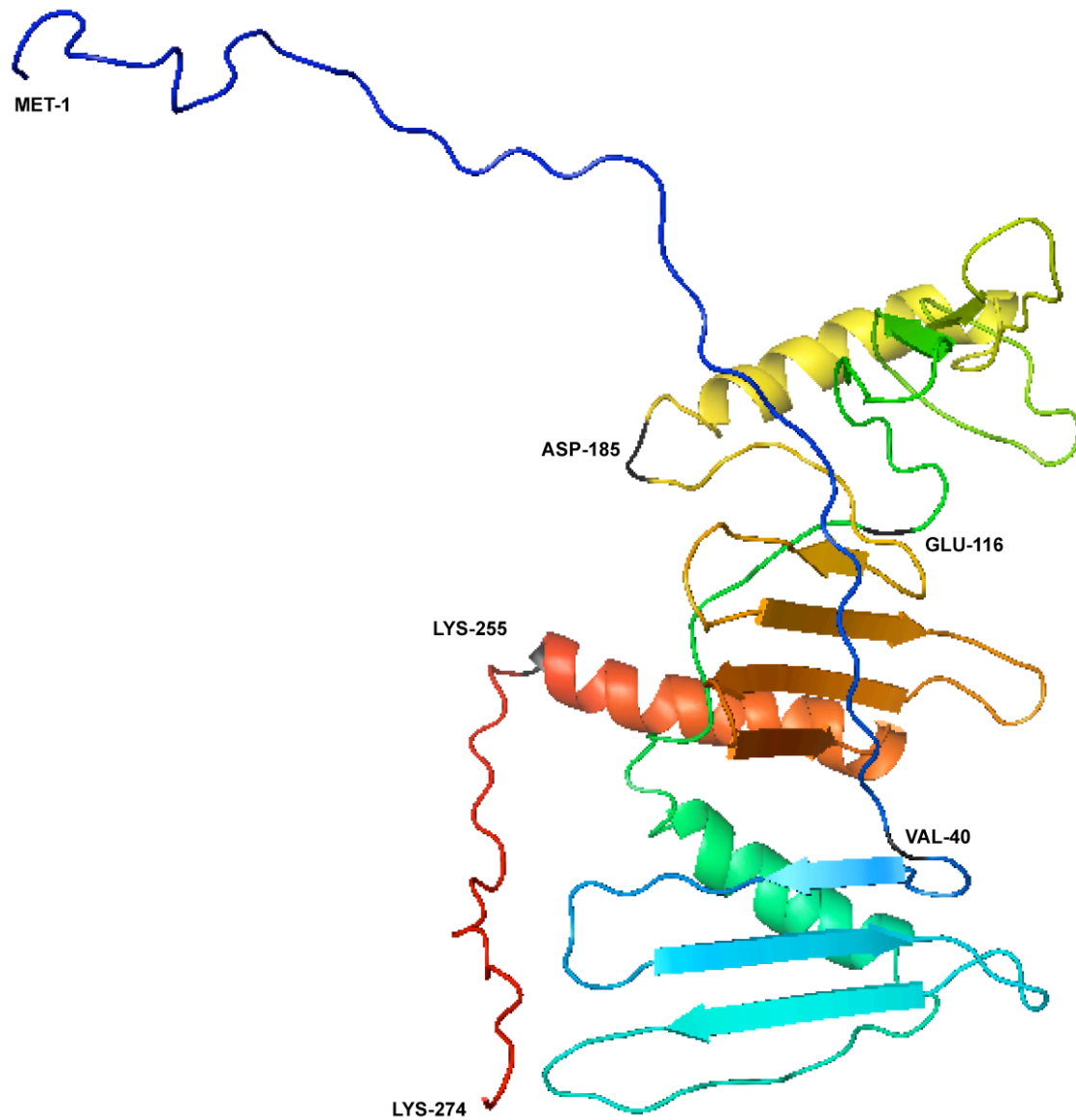
Wickham, T.J., Davis, T., Granados, R.R., Shuler, M.L., and Wood, H.A. (1992). Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system. *Biotechnol Prog* 8, 391-396.



- Wilhelm, J.E., and Vale, R.D. (1993). RNA on the move: the mRNA localization pathway. *J Cell Biol* *123*, 269-274.
- Wilkins, M.R., Gasteiger, E., Bairoch, A., Sanchez, J.C., Williams, K.L., Appel, R.D., and Hochstrasser, D.F. (1999). Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* *112*, 531-552.
- Williamson, J.R. (1994). G-quartet structures in telomeric DNA. *Annu Rev Biophys Biomol Struct* *23*, 703-730.
- Winn, M., Isupov, M., and Mushudov, G.N. (2001). Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D* *57*, 122-133.
- Wong, I., and Lohman, T.M. (1993). A double-filter method for nitrocellulose-filter binding: application to protein-nucleic acid interactions. *Proc Natl Acad Sci U S A* *90*, 5428-5432.
- Worman, H.J., and Courvalin, J.C. (2002). The nuclear lamina and inherited disease. *Trends Cell Biol* *12*, 591-598.
- Wortman, M.J., Johnson, E.M., and Bergemann, A.D. (2005). Mechanism of DNA binding and localized strand separation by Pur alpha and comparison with Pur family member, Pur beta. *Biochim Biophys Acta* *1743*, 64-78.
- Zambrano, N., De Renzis, S., Minopoli, G., Faraonio, R., Donini, V., Scaloni, A., Cimino, F., and Russo, T. (1997). DNA-binding protein Pur alpha and transcription factor YY1 function as transcription activators of the neuron-specific FE65 gene promoter. *Biochem J* *328 ( Pt 1)*, 293-300.
- Zhang, Q., Pedigo, N., Shenoy, S., Khalili, K., and Kaetzel, D.M. (2005). Puralpha activates PDGF-A gene transcription via interactions with a G-rich, single-stranded region of the promoter. *Gene* *348*, 25-32.

## Chapter 7 Appendix

### 7.1 Computational model of *D. melanogaster* Pur- $\alpha$



Appendix 7-1. Computational model for *D. melanogaster* Pur- $\alpha$ . The model was built by J. Soeding using the program HHpred (Söding et al., 2005). The N-terminus is depicted in blue, the C-terminus is shown in red. Amino acids that were used for expression construct design are highlighted in black.



### 7.3 Index of figures and tables

Table 1-1. Examples of mRNAs transported in dendrites .....	4
Figure 1-1. Schematic drawing of the mRNA transport mechanism in neurons.....	5
Figure 1-2. Schematic drawing of localized translation in neurons.....	8
Figure 1-3. Schematic representation of human Pur- $\alpha$ .....	10
Table 2-1. Protein fragments of human Pur- $\alpha$ for crystallization trials..	19
Figure 2-1. Crystals of human Pur- $\alpha$ .....	21
Figure 2-2. Modular Structure of Pur- $\alpha$ .....	23
Table 2-2. Protein fragments of <i>B. burgdorferi</i> Pur- $\alpha$ for crystallization trials. ....	25
Figure 2-3. Crystal of <i>B.burgdorferi</i> Pur- $\alpha$ .....	26
Figure 2-4. Crystals of selenomethionine-derivatized <i>B. burgdorferi</i> Pur- $\alpha$ .....	27
Figure 2-5. Diffraction images of <i>B. burgdorferi</i> Pur- $\alpha$ .....	28
Table 2-3. Data collection for the crystal structure of <i>B. burgdorferi</i> Pur- $\alpha$ .....	29
Table 2-4. Refinement statistics for the crystal structure of <i>B. burgdorferi</i> Pur- $\alpha$ . ....	30
Figure 2-6. Crystal structure of <i>B. burgdorferi</i> Pur- $\alpha$ .....	31
Figure 2-7. Electrostatic surface potential of <i>B. burgdorferi</i> Pur- $\alpha$ .....	32
Figure 2-8. Structures from the PDB with similarity to <i>B. burgdorferi</i> Pur- $\alpha$ .....	33
Table 2-5. Sequence conservation between the three PUR repeats of <i>D. melanogaster</i> Pur- $\alpha$ .....	35
Table 2-6. Protein fragments of <i>D. melanogaster</i> Pur- $\alpha$ for crystallization trials.....	36
Figure 2-9. Crystals of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	37
Figure 2-10. Crystal of selenomethionine-derivatized <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	37
Table 2-7. Data collection for the crystal structure of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	39
Figure 2-11. Diffraction images of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	39
Table 2-8. Refinement statistics for the crystal structure of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	40
Figure 2-12. Crystal structure of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II, PDB-ID 3K44.....	41
Figure 2-13. Electrostatic surface potential of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II. ....	43
Figure 2-14: Surface conservation of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II. ....	44
Figure 2-15. Superposition of the crystal structures of <i>B. burgdorferi</i> Pur- $\alpha$ (red) and <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II (blue).....	45
Figure 2-16. Structures from the PDB with similarity to <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	46
Figure 2-17. Superposition of the PUR domain with a Whirly domain. ....	46
Figure 2-18. Crystals of <i>D. melanogaster</i> Pur- $\alpha$ repeat III.....	47
Figure 2-19. Diffraction image of a native crystal of <i>D. melanogaster</i> Pur- $\alpha$ repeat III.....	48
Figure 2-20. Crystals of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II and DNA .....	49
Figure 2-21. Diffraction image of crystals of <i>D. melanogaster</i> Pur- $\alpha$ in complex with DNA. ....	49
Figure 2-22. Analytical size-exclusion chromatography with <i>D. melanogaster</i> Pur- $\alpha$ repeats I-III. ....	51
Table 2-9: Analytical size-exclusion chromatography with <i>D. melanogaster</i> Pur- $\alpha$ repeats I-III at various concentrations.....	52
Figure 2-23. Kratky plot for folded, partially folded, and unfolded proteins.....	54
Table 2-10. SAXS measurements for <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II. ....	55
Figure 2-25. SAXS analysis for <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	56
Figure 2-24. Analysis of SAXS data for <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II. ....	56
Table 2-11. SAXS data of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-III.....	57
Figure 2-26: Analysis of SAXS data for <i>D. melanogaster</i> Pur- $\alpha$ repeats I-III. ....	58
Figure 2-27. Dimerization model of Pur- $\alpha$ .....	58
Figure 2-28. SAXS analysis for <i>D. melanogaster</i> Pur- $\alpha$ repeats I-III. ....	59
Table 2-12. Filter binding assays with human Pur- $\alpha$ 56-287 C272S. ....	61
Figure 2-29. Representative filter binding curves for human Pur- $\alpha$ .....	62
Table 2-13. Filter binding assays with <i>B. burgdorferi</i> Pur- $\alpha$ full length.....	63
Figure 2-30. Position of the mutations which affect nucleic acid binding in <i>B. burgdorferi</i> Pur- $\alpha$ .....	64
Table 2-14. Filter binding assays with mutants of <i>B. burgdorferi</i> Pur- $\alpha$ .....	65
Figure 2-31. Representative filter binding curves for <i>B. burgdorferi</i> Pur- $\alpha$ .....	66

Table 2-15. Filter binding assays with <i>D. melanogaster</i> Pur- $\alpha$ 40-255. ....	67
Figure 2-32. Representative filter binding curve for <i>D. melanogaster</i> Pur- $\alpha$ repeats I-III. ....	67
Figure 2-33. EMSAs with <i>D. melanogaster</i> Pur- $\alpha$ full length. ....	69
Figure 2-34. Position of the mutations which affected nucleic-acid binding in <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II. ....	70
Figure 2-35. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ ....	71
Figure 2-36. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ .....	71
Figure 2-37. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ .....	72
Figure 2-38. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ .....	72
Figure 2-39. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ .....	73
Figure 2-40. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ .....	73
Figure 2-41. CD spectra of <i>D. melanogaster</i> Pur- $\alpha$ repeats I-II.....	74
Figure 2-42. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ .....	75
Figure 2-43. EMSA with <i>D. melanogaster</i> Pur- $\alpha$ full length and CGG 25mer ssRNA. ....	76
Table 2-16. Interactions of Pur- $\alpha$ with known proteins from the yeast-two-hybrid screen.....	79
Table 2-17. Interactions of Pur- $\alpha$ with non-existing proteins from the yeast-two-hybrid screen.....	82
Table 2-18. GenMatch interaction partners from the yeast-two-hybrid screen.....	83
Figure 3-1. Amino acid sequence alignment of <i>D. melanogaster</i> PUR repeats I, II and III and <i>Borellia</i> PUR. ....	91
Figure 3-2. Potential binding modes for Pur- $\alpha$ and nucleic acids.....	98
Figure 3-3. Example of an RRM domain.....	100
Table 5-1. Oligonucleotides for molecular biology .....	110
Table 5-2. RNA and DNA oligonucleotides for binding assays and crystallization.....	111
Table 5-3. Commercial plasmids for recombinant protein expression.....	111
Table 5-4. Plasmids for recombinant protein expression in <i>E. coli</i> .....	113
Table 5-5. Plasmids for baculovirus-based recombinant protein expression in insect cells. ....	113
Table 5-6. Plasmids for <i>in vitro</i> RNA transcription.....	114
Table 5-7. Bacterial strains. ....	114
Table 5-8. Media for cell culture. ....	114
Appendix 7-1. Computational model for <i>D. melanogaster</i> Pur- $\alpha$ .....	144
Appendix 7-2. Amino acid sequence alignment and secondary structure prediction for different Pur- $\alpha$ orthologs.....	145

## 7.4 Abbreviations

°C	degree Celsius
A, Å	alanine, adenine, Ångström
A <sub>280</sub> , A <sub>260</sub>	absorption at wavelength 260 nm/ 280 nm
Amp	ampicillin
ATP	Adenosine triphosphate
bb, <i>B. burgdorferi</i>	<i>Borrelia burgdorferi</i>
C	Cysteine, cytosine
CD	circular dichroism
Ci	Curie
D	aspartic acid
Da	Dalton
deg	degree
dm, <i>D. melanogaster</i>	<i>Drosophila melanogaster</i>
DNA	deoxyribonucleic acid
ds	double stranded
E	glutamic acid
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EMSA	electrophoretic mobility shift assay
F	phenylalanine
for	forward
g	gram, standard acceleration
G	guanine
Gen	gentamycin
GST	glutathione S-transferase
h	hour, human
HEPES	(4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
I	isoleucine
K	lysine
k	kilo
Kan	kanamycin
KIF	kinesin superfamily protein
L	Liter, leucine
M	mol per Liter, methionine
m	milli, meter
MBP	maltose binding protein
MES	2-( <i>N</i> -morpholino)ethanesulfonic acid
min	minute
MPD	3-Methyl-1,5pentadiol
P	proline
PAGE	polyacrylamide gel electrophoresis
PEG	poly ethylene glycol
PMSF	phenylmethanesulphonylfluoride
PVFD	polyvinylidene fluoride
R	arginine
rev	reversed
RNA, mRNA	(messenger) ribonucleic acid
RNP	ribonucleoprotein particle
rpm	runs per minute
s	second
S	serine
SAXS	small angle X-ray scattering
SDS	sodium docecyl sulfate
ss	single stranded
TRIS	tris(hydroxymethyl)aminomethane
Y	tyrosine
μ	Micro

## Curriculum vitae

Name	Almut Graebisch
Place and date of birth	Berlin, Germany, Aug 13 <sup>th</sup> , 1981
Nationality	German
Family status	Unmarried
PhD thesis Nov 2006 – Mar 2010	Gene Center Munich and Helmholtz Zentrum München (Munich, Germany) Supervisor Dierk Niessing University supervisor Karl-Peter Hopfner „Crystal structure, nucleic-acid binding properties, and dimerization model of Pur- $\alpha$ “
Internships Mai-Nov 2006	GIT Verlag (Darmstadt, Germany) editorial office Nature (Munich, Germany) editorial office Laborjournal (Freiburg, Germany) freelancer since 2007
Diploma thesis Oct-Mar 2006	Max-Planck-Institute of Molecular Physiology, Dortmund Department Prof. H. Waldmann Supervisor Prof. R. P. Breinbauer „Synthesis and examination of ligands and inhibitors of enzymes involved in the biosynthesis of phenazine“ (Sehr gut, Grade A)
Stay abroad Mai-Jul 2005	University Bergen, Norway Laboratory, Prof. G. W. Francis „Proanthocyanins from khat“ (Grade A)
Higher education Oct 2001 - Mar 2006	University Dortmund, Diploma (Sehr gut, Grade A) Main subject: chemistry, subsidiary subject: life sciences
1992-2001 Secondary school	Goethe-Gymnasium (Dortmund, Germany) Abitur (1,4)
1988-1992 Elementary school	Lieberfeld-Grundschule (Dortmund, Germany)
Date and Signature	Mar 31 <sup>st</sup> , 2010

## **Publications**

### **This study:**

Graebisch, 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. *Proc Natl Acad Sci U S A* *106*, 18521-18526.

Graebisch, A. Roche, S., Kostrewa, D., Burley, S. K., Söding, J. and Niessing, D. Computational construct design and X-ray structure determination of Pur- $\alpha$  – a case study (in preparation).

### **Diploma thesis:**

Ahuja, E.G., Janning, P., Mentel, M., Graebisch, A., Breinbauer, R., Hiller, W., Costisella, B., Thomashow, L.S., Mavrodi, D.V., and Blankenfeldt, W. (2008). PhzA/B catalyzes the formation of the tricycle in phenazine biosynthesis. *J Am Chem Soc* *130*, 17053-17061.

### **Journalistic publications:**

Graebisch, A. (2006) Biotech in the People's Republic of China. *Regions and Locations Guide 1*, 64.

Graebisch, A., and Schiermeier, Q. (2006). Anti-evolutionists raise their profile in Europe. *Nature* *444*, 406-407.

Graebisch, A. (2006). Polish scientists fight creationism. *Nature* *443*, 890-891.

Graebisch, A. (2008) Lost in Transcription. *Laborjournal* *5*, 38-39.

Graebisch, A. (2008) Allegorie der Allergie. *Laborjournal* *11*, 38-39.

Graebisch, A. (2008) Allegorie der Allergie. *Biotechnologie online*, 21.11.2008

Graebisch, A. (2009) Evolution at work. *Laborjournal* *3*, 38-39.

Graebisch, A. (2009) Münchner Doktoranden unter sich: Interact 2009. *Laborjournal online*, 10.05.2009



## **Acknowledgements**

I wish to thank my supervisor Dierk Niessing. I truly appreciate his constant support and the opportunity to work at the Gene Center with its excellent equipment, inspiring retreats and informative speakers.

The members of my thesis advisory committee, Klaus Förstemann and Karl-Peter Hopfner, provided valuable advice. Johannes Söding did the bioinformatic analysis that boosted the project forward. Dirk Kostrewa and Stephane Roche generously shared their experience in X-ray crystallography. Alan Cheung collected data at SLS. Gregor Witte helped with SAXS data analysis. All of whom I gratefully acknowledge here with my sincere thanks.

I wish to thank all of Niessing lab for their help - especially Sigrun for cloning and keeping the lab running, Marisa for introducing me to RNA work and nucleic-acid binding assays, Stephane for help with X-ray crystallography and for carefully reading the manuscript. Thank you, Hermine and Christian, for your friendship and best coffee at Gene Center.

I acknowledge the ESRF and the crystallization facility of the MPI for biochemistry.

My mentors Rüdiger Schmidt, Rolf Breinbauer, Wulf Blankenfeldt, and Anna Seidinger accompanied various steps of my career. The organizing committee of Interact PhD symposium 2009 was a great experience. To all of you my sincere thanks.

I am grateful to Debattierclub München e.V. and debaters around the world for broadening my horizons. “Debate is not a hobby, but a lifestyle...”

I am deeply grateful to my dear parents. Thank you, Roman, Martin, Tiina, Moritz and Rasmus – for making Munich a home. Simon – for your patience and back-up. Leonie and Vanessa – for your love. Oli – the sky is the limit for your support. Malte - KADS. Inka and Anna - for being my best friends.

Thank you, Lukas...for giving me a hand.