STRUCTURAL AND BIOCHEMICAL STUDIES ON THE

ARYL HYDROCARBON RECEPTOR

AND ON THE

IMMUNITY-RELATED GTPASE IRGA6

Dissertation zur Erlangung des akademischen Grades des
Doktors der Naturwissenschaften (Dr. rer. nat.)

eingereicht im Fachbereich Biologie, Chemie, Pharmazie
der Freien Universität Berlin

vorgelegt von

M. Sc. Kathrin Wiebke Schulte

aus Leer (Ostfriesland)

Berlin

2015

Die vorliegende Arbeit wurde im Zeitraum Mai 2011 bis Juni 2015 am

Max-Delbrück-Centrum für Molekulare Medizin in der Helmholz-Gemeinschaft

unter der Leitung von

Prof. Dr. Oliver Daumke

angefertigt.

1. Gutachter: Prof. Dr. Oliver Daumke

2. Gutachter: Prof. Dr. Udo Heinemann

Disputation am 06.10.2015

Contents

Contents

Contents

List of figures

List of tables

Part I

1   Introduction

1.1   bHLH-PAS domain comprising transcription factor family

The basic helix-loop-helix-PER-ARNT-SIM (bHLH-PAS) domain comprising transcription factor family is well conserved from metazoan to humans.[1] Members of this family respond to several external stimuli in order to regulate diverse biological processes ranging from development and cell cycle control over xenobiotic response to circadian rhythm.[2-4] These proteins selectively heterodimerize via the well-conserved bHLH and PAS domains. Differences in C-terminal regulatory motifs control the activity of these complexes.
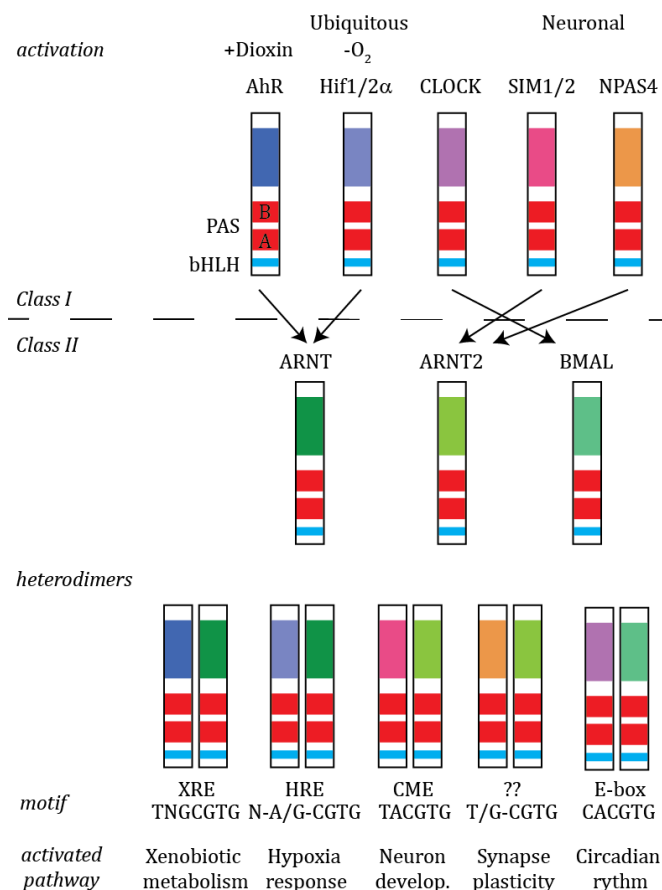
Figure 1: The bHLH-PAS domain comprising transcription factor family.Tasks of different members of the class I and class II bHLH-PAS domain comprising transcription factor family. Dimers are formed between a class I factor (single-minded homologue (SIM), neuronal PAS domain-containing protein (NPAS), aryl hydrocarbon receptor (AHR), hypoxia-inducible factor (HIF) and circadian locomotor output cycles kaput (CLOCK)) and a class II factor (AHR nuclear translocator (ARNT), ARNT2, brain muscle ARNT-like 1 (BMAL)). The heterodimers bind to class I-specific variations of the canonical E-box sequence and regulate specific target genes, thereby mediating various developmental or homeostatic processes and/or responses to environmental or physiological stresses. Class II factors are capable of forming dimers with more than one class I factor. Figure modified from [5]. For explanation see text.

Heterodimerization takes place between Class I and Class II bHLH-PAS proteins. Class I bHLH-PAS proteins are described as cytosolic signal sensors, which are guided to the nucleus upon activation. Class II bHLH-PAS proteins are localized in the nucleus and are capable to heterodimerize with different class I bHLH-PAS proteins. The evolution of these modular regulatory domains and the ability to selectively dimerize with different family members allows fine-tuning of gene expression and thus the ability to keep mammals healthy and in balance with their environment.

1.2   The Aryl hydrocarbon receptor

The aryl hydrocarbon receptor (AhR) is of ancient origin. Phylogenetic analyses reveal that functional orthologues of the AhR gene can be found in living mammals, amphibians, reptiles and birds.[6] AhR belongs to the bHLH-PAS domain containing transcription factor family and is best known for mediating toxicity and influencing tumorigenesis. Recently, a function in normal development and immunology was identified. Interestingly, nearly all vertebrate AhR are the only members of the bHLH-PAS domain comprising transcription factor family known to be activated by ligands.[5] AhR was discovered in the 80s as the receptor which binds 2,3,7,8-tetrachlorodibenzodioxin (TCDD; dioxin) with high affinity thus mediating its toxicity.[7] Nevertheless, the primary function of mammalian AhR is probably related to normal development, since vertebrate AhR arose in biological systems over 450 million years ago when no products of modern industrialization (dioxin etc.) were present. In other words, exogenous ligand binding may be a secondary function of this receptor that arose during evolution.

1.3   Domain organization of AhR

AhR is composed of modular domains eponymous for the subfamily it belongs to.
The basic-helix-loop-helix (bHLH) domain spans the N-terminal region of AhR (hsAhR: residue 27-80). It plays a multifunctional role as it is mediating interactions with HSP90 (90 kDa heat shock protein) and its dimerization partner ARNT (aryl hydrocarbon receptor nuclear translocator). Besides that, the bHLH is crucial for DNA binding (see **Figure 2**). Furthermore, a nuclear import sequence (NLS) was identified, guiding AhR to the nucleus after ligand dependent activation (hsAhR: residues 13-39).[8-10]

The C-terminal half of AhR consists of the transactivation domain (TAD), which is important for AhR-mediated gene transcription via protein-protein interactions.[11-13] The TAD is predicted to be unstructured and comprises three distinct regions: the acidic region, the glutamine (Q)-rich domain and the proline-serine-threonine (P/S/T)-rich domain. All three domains are important for the transactivation (see **Figure 2**).[11,14,15]
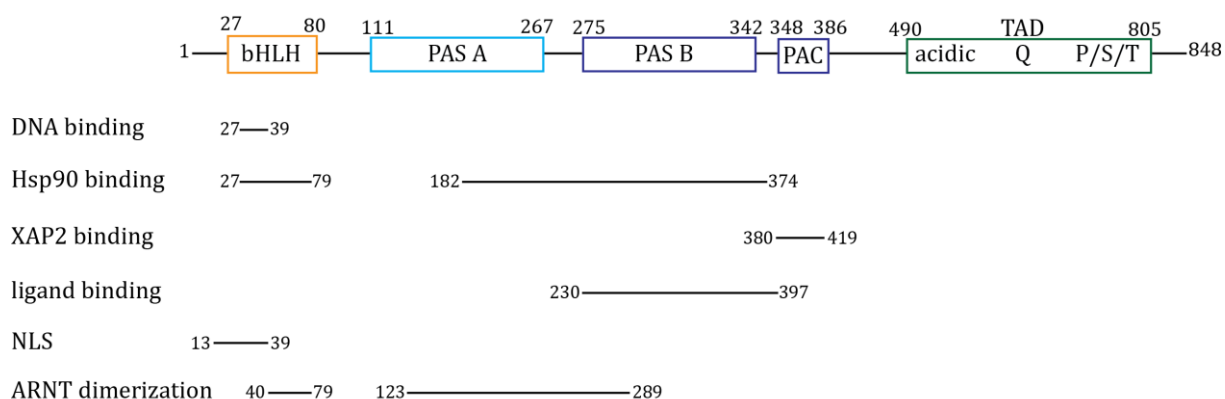
Figure 2: The human aryl hydrocarbon receptor. Schematic view of the domain architecture of the human aryl hydrocarbon receptor (AhR). The first and the last residue of each domain are indicated. Regions of interaction with other proteins are shown as black lines and the first and last residue are indicated. Basic helix-loop-helix (bHLH) domain, Per-ARNT-SIM (PAS) domain, PAS-associated C-terminal domain (PAC), glutamine (Q)-rich region, proline-serine-threonine (P/S/T)-rich region. For explanation see text.

The PAS fold is the most distinguishing feature of the bHLH-PAS domain comprising transcription factor family and shows the highest degree of evolutionary conservation. The canonical PAS fold consists of two domains termed PAS A and PAS B (hsAhR: residue 111-342). Each PAS domain consists of a five-stranded antiparallel β-sheet flanked by two α-helices, shaped in a globular fashion (see **Figure 3**).

Whereas the PAS A domain of AhR is mainly responsible for protein-protein interaction, the PAS B domain additionally has the ability to couple intrinsic ligand binding with transcriptional regulation.[16] The PAS B domain follows a PAC domain (hsAhR: 348-386; PAS associated C-terminal domain), which comprises two helices responsible for the proper folding of the overall PAS fold.[17-19] Ligands such as dioxin are predicted to bind within the hydrophobic pocket of the PAS B domain as an initial step for ligand dependent activation of AhR mediated toxicity.[20,21]
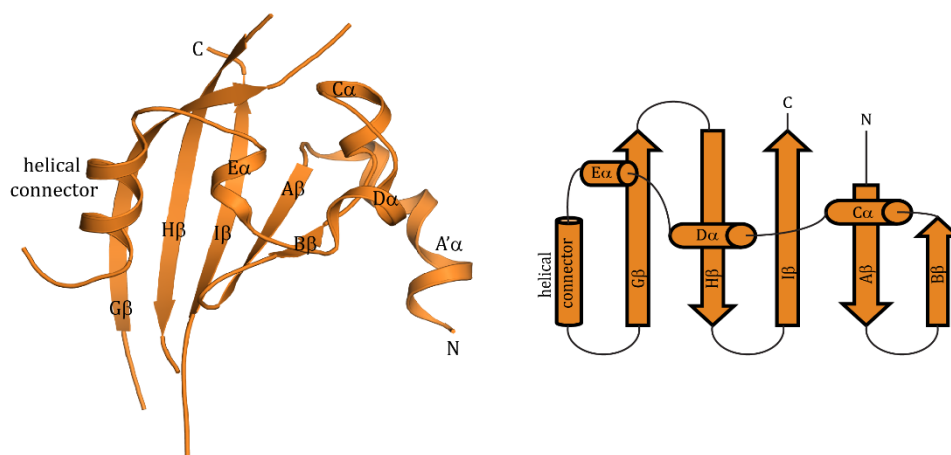
Figure 3: PAS domain fold. Left: Ribbon representation of the PAS A domain of human AhR (PDB: 4M4X). Right: Topology of a typical PAS fold. For explanation see text.

1.4   AhR ligands

Analyses of the ligand binding specificity revealed that AhR is a promiscuous receptor binding structurally diverse chemicals.[22,23]

Exposure to naturally occurring ligands is influenced by our diet. Vegetables are a significant source of AhR ligands as they contain indolce-3-carbinol which in turn can be metabolized to ICZ (indole-[3,2-b]carbazole; see **Figure 4**) and DIM (3,3-diindolylmethane).[24,25] Besides the dietary indoles, flavonoids are another group of naturally occurring dietary AhR ligands. However, they mostly act as antagonists.[23]

Nevertheless, the majority of known high affinity AhR ligands is synthetic in nature and can be divided into two groups.

The planar hydrophobic halogenated aromatic hydrocarbons (HAHs) such as dioxin (TCDD), dibenzofurans and biphenyls are persistent in the environment due to their chemical stability and hydrophobicity and have binding affinities to AhR in the pico-molar to nano-molar range. They are released into the environment during combustion processes and are present in detectable levels in all humans.[26,27] TCDD, which belongs to the HAHs, is the best studied ligand in regard to AhR (see **Figure 4**).

Polycyclic aromatic hydrocarbons (PAHs) such as 3-methylchloranthrene (3MC), benzo[α]pyrene (B[α]P) and benzoflavone are more labile and have a lower affinity to AhR compared to the HAHs (see **Figure 4**).[23,28,29]
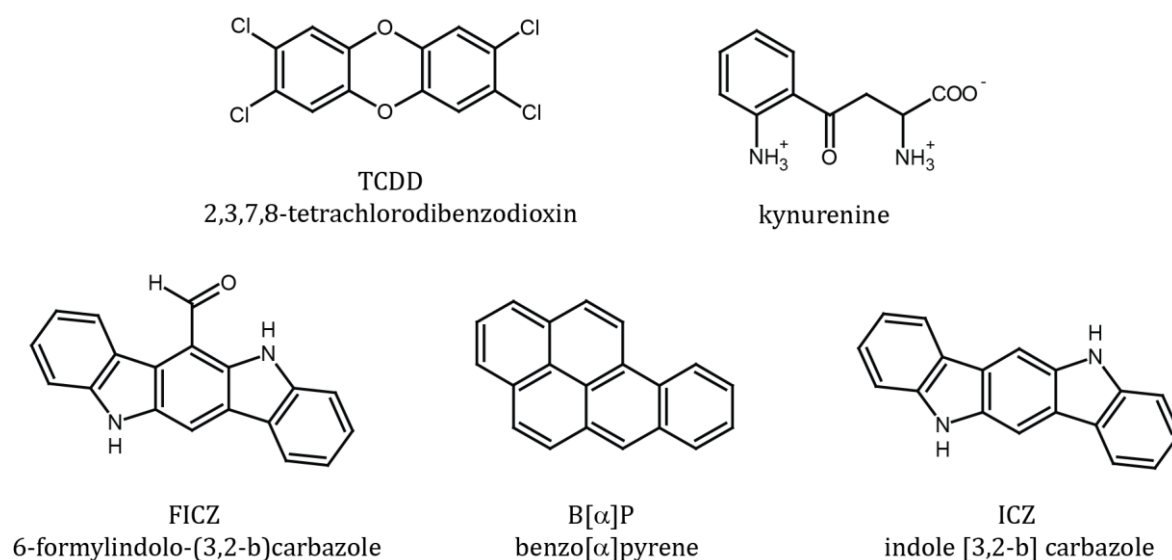
Figure 4: Structural formula of naturally occurring exogenous ligands such as ICZ (indole-[3,2-b]carbazole), endogenous ligands such as kynurenine and FICZ (6-formylindolo-[3,2-b]carbazole) and synthetic ligands such as TCDD (2,3,7,8-tetrachlorodibenzodioxin) and B[α]P (benzo[α]pyrene). For explanation see text.

Despite the focus on AhR activation by exogenous ligands, developmental and physiological defects in AhR-null mice were the driving force to look for putative endogenous ligands. Besides arachidonic acid metabolites, products of the heme degradation and cholesterol derivatives, the most likely candidates are tryptophan metabolites.[30-33] Two tryptophan metabolizing enzymes named indolamine 2,3-dioxygenase (IDO) and tryptophan 2,3-dioxygenase (TDO) are generating the metabolite kynurenine (see **Figure 4**). It was previously reported, that kynurenine is generated in large amounts in brain tumor cells, associated with AhR-mediated suppression of the antitumor response together with malignant progression.[34] Another physiological ligand for AhR is 6-formylindolo-[3,2-b]carbazole (FICZ), which is formed by photolysis of tryptophan via visible or UV light (see **Figure 4**).

1.5   AhR signal transduction

1.5.1   The canonical pathway

The activity of AhR has to be tightly controlled due to the deleterious effects AhR can mediate in the organism. In general, AhR is thought to reside in the cytoplasm in an inactive multi-protein complex. This complex includes two molecules of the chaperone Hsp90, the aryl hydrocarbon receptor interacting protein (XAP2) and the 23 kDa co-

chaperone p23.[35-37] The interaction with Hsp90 was mapped to two different regions, namely the bHLH (hsAhR: residue 27-79) and the PAS fold (hsAhR: residue 182-374).[38-41] XAP2 was identified to further stabilize the chaperone complex by mediating a contact between Hsp90 and the PAS B domain (hsAhR: residue 380-419) of AhR (see **Figure 2**).[42,43] Lastly, the complex is reinforced by p23, which also interacts with Hsp90 and AhR.[44,45] The influence of the Hsp90 machinery is manifold as it stabilizes AhR in the cytoplasm, inhibits nuclear translocation of AhR by masking its NLS, prevents premature ARNT binding and stabilizes the high affinity ligand-binding conformation of AhR.
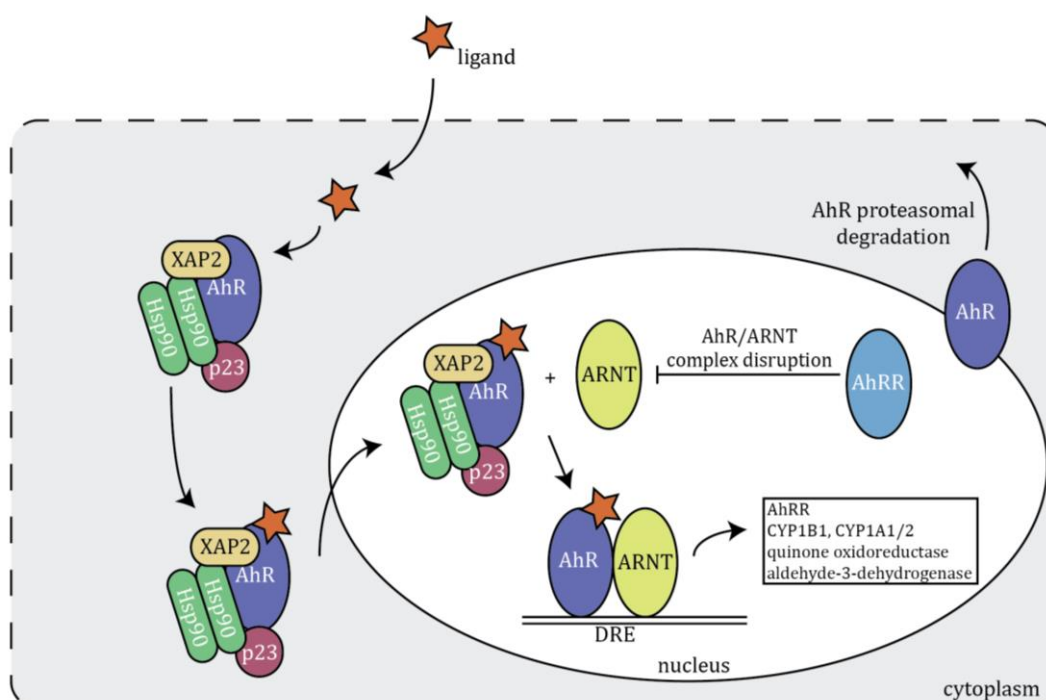
Figure 5: The canonical AhR pathway. In the absence of ligand AhR is present in the cytoplasm bound to two molecules of Hsp90 and one molecule of p23 and XAP2. Upon ligand binding, the chaperoning complex translocates to the nucleus. AhR heterodimerizes with ARNT and is released from the chaperoning complex. The ligand bound AhR/ARNT complex binds to the DRE, inducing the transcription of the target genes such as AhRR, CYP1B1, CYP1A1/2, quinone oxidoreductase or aldehyde-3-dehydrogenase. Figure modiefied from [46] For explanation see text.

Most described AhR ligands simply diffuse into the cell because of their lipophilic nature. Ligand binding to the PAS B domain does not lead to Hsp90 machinery dissociation, but to a conformational change of AhR that allows binding of the NLS to importin β and thus the nucleoplasmatic shuttling of the AhR-chaperone-complex.[10,47-49] Once located in the nucleus, the class II bHLH-PAS protein aryl hydrocarbon receptor nuclear translocator (ARNT), a constitutively nuclear protein, is required to release the ligand bound AhR from its chaperoning complex. This is achieved by heterodimerization (see **Figure 5**).[47]

Although the sequence similarity between AhR and ARNT is less than 20%, both share the same domain architecture. Dimerization takes place via the bHLH domains, which form a stable four-helical bundle. The interaction is further supported by PAS fold dimerization to increase the stability of the heterodimeric complex.

The ligand bound Ahr/ARNT heterodimeric transcription factor complex directs transcription of genes from the dioxin-responsive enhancer elements (DRE). The DNA binding is achieved by the formation of a four-helical bundle of the bHLH domains of the AhR/ARNT complex recognizing the DRE (5'-NGCGTG-3'). The first half of DRE, 5'-NGC-3', is recognized by the DNA binding region within the bHLH of AhR (hsAhR: residues 27-39) and the bases 5'-GTG-3' are recognized by the bHLH of ARNT (hsARNT: residues 94-102).[50-54]

The AhR/ARNT complex bound to the DRE subsequently initiates the recruitment of co-regulatory proteins, which modulate the transcription of target genes. For example, AhR directly controls the positive transcription elongation factor (P-TEFb) and subunits of the mediator complex to regulate transcriptional elongation.[55-57] Besides that, AhR can affect the local chromatin architecture by interacting with a subunit of the SWI/SNF chromatin remodeling complex.[58,59] Local histone hyper-acetylation and methylation can also be affected by AhR via interaction with the steroid receptor coactivator-1 (SRC-1) or by displacing the histone deacetylase complex (HDAC).[60-62] Finally, transcriptional cofactors, such as transacting transcription factor 1 (Sp1) or the receptor interacting protein 140 (RIP140), support the transcriptional activation of the AhR/ARNT heterodimer.[63,64]

Although the list of AhR target genes is still expanding, the transcriptional activation of xenobiotic metabolizing enzymes such as cytochrome P450 1A1 (CYP1A1), CYP1A2, CYP1B1, UDP-glucuronosyl transferase, quinone oxidoreductase, aldehyde-3-dehydrogenase and glutathione-S-transferase is well known. These enzymes belong to the phase I and phase II metabolic pathway and promote the metabolism and clearance of xenobiotics in order to re-establish homeostatic levels of AhR activators.[65]

Further regulation of AhR signaling has been suggested via two distinct mechanisms.

Upon transcriptional activation, the ligand-activated AhR is guided to the 26S proteasome for degradation by the ubiquitin-proteasome pathway thereby limiting the AhR activation (see **Figure 5**).[66-68]

Secondly, the AhR signaling includes a negative feedback loop which involves the aryl hydrocarbon receptor repressor (AhRR). The AhRR is a DRE-regulated gene and its expression is rapidly increased upon AhR transcriptional activation.[69] The AhRR is

structurally similar to AhR. It consists of an additional repressor domain and dimerizes with ARNT without any need of an agonist. The exact mechanism of AhRR mediated inhibition of the AhR canonical pathway remains unclear, but it is proposed that AhRR competes with AhR for dimerization with ARNT and thus for transcriptional activation (see **Figure 5**).[69-71]

1.5.2   The adaptive response

Altogether, the AhR signal transduction can lead to the adaptive response which is defined as the transcriptional expression of the xenobiotic metabolizing enzymes.[72] AhR binding to exogenous ligands such as dioxin leads to the transcriptional activation of a gene battery that promote the metabolism of these substances.[65]

1.5.3   The toxic response

Toxicity can be a side effect of the adaptive response, if the activated metabolizing enzymes in turn process toxic metabolites. Dioxin toxicity commonly includes thymic involution, immune suppression, hepatocellular damage, chloracne, epithelial hyperplasia, teratogenesis and tumor promotion.[73-75]

1.6   The role of AhR in development

The physiological and developmental importance of AhR was identified by characterization of AhR null mice. For example, the response to PAHs and dioxins did not lead to an up-regulation of the metabolic enzymes known for the adaptive pathway, but rather to a resistance to most aspects of dioxin toxicity.[76-78] Furthermore, a number of surprising pathologies have been identified in AhR null mice, indicating a role in normal development. They have a markedly smaller liver and abnormal vasculature in liver, kidney and eye.[76,78,79] Cardiac hypertrophy and hypertension have also been seen in AhR null mice.[80] The most consistent phenotype in AhR null mice is a patent ductus venosus, which is part of the fetal circulatory system.[81,82] In 100% of the AhR null mice the ductus venosus remains patent into adulthood. Although several phenotypes of AhR knock out mice are known, the mechanism which involves AhR in development and whether endogenous ligands are required remains unclear.

1.7   Role of AhR in cell proliferation

During the last 20 years, it has become evident that AhR is also involved in promotion of cell cycle progression. Interestingly, the influence of AhR on the cell cycle control can take place in a ligand independent manner. Cyclins, cyclin dependent kinases (CDKs) and cyclin dependent kinase inhibitors (CKI) orchestrating the cell proliferation. They are responsible for a controlled transition of the cell cycle phases and their expression is controlled by retinoblastoma (Rb) and the E2F family of transcription factors. AhR has been shown to interact directly with CDKs as well as with Rb in a ligand-independent fashion (see **Figure 6**).[83-86]
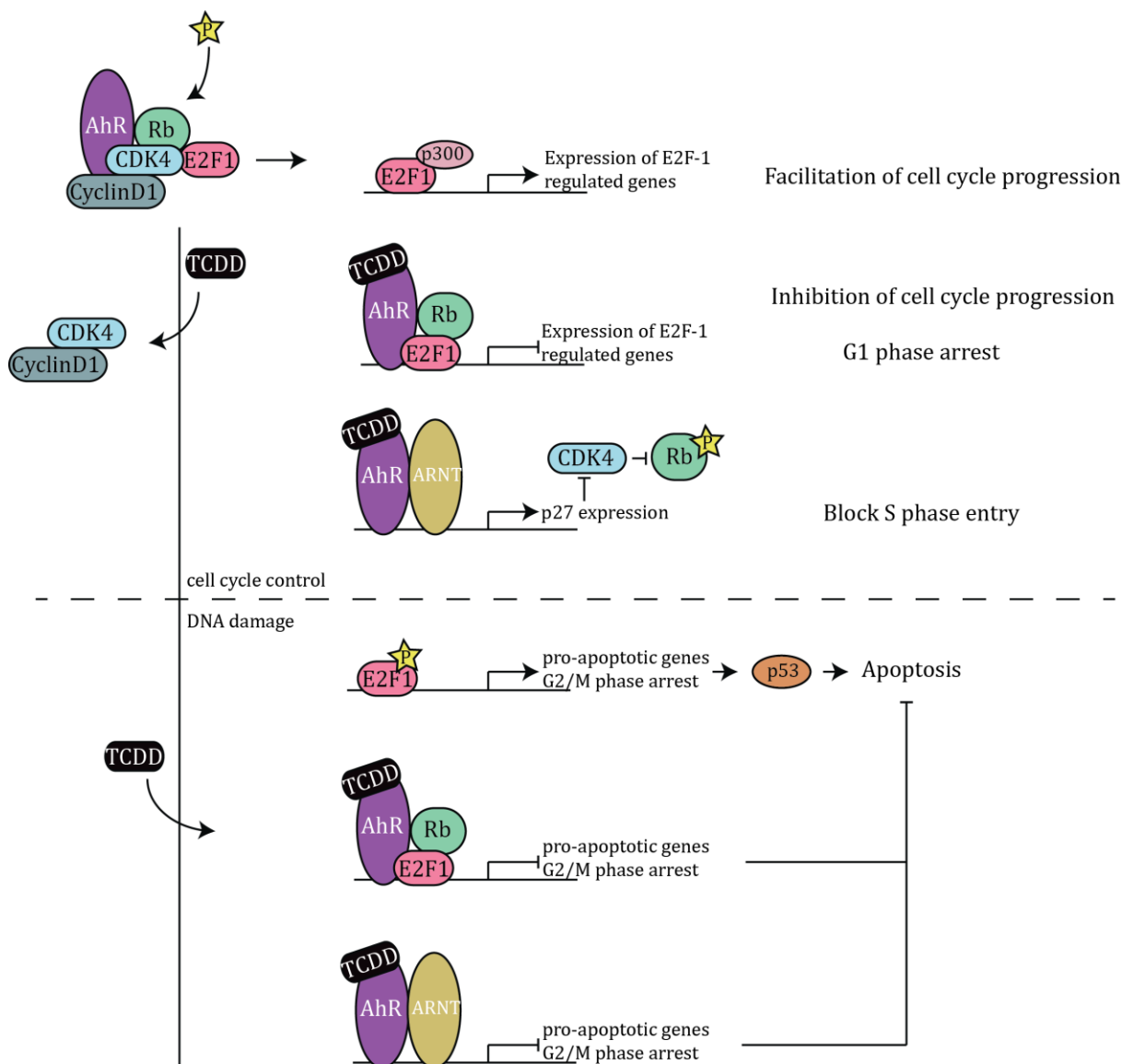
Figure 6: Potential pathways by which AhR is involved in regulation of cell cycle progression and apoptosis in presence and absence of ligands. AhR (aryl hydrocarbon receptor), CDK4 (cyclin dependent kinase 4), Rb (retinoblastoma), TCDD (2,3,7,8-tetrachlorodibenzodioxin), ARNT (aryl hydrocarbon receptor nuclear translocator). For explanation see text.

9

Furthermore, embryonic fibrobalsts, isolated from AhR-null mice, show accumulation of cells in the G_2/M phase, reduced proliferation rates and increased apoptosis.[87] Moreover, adult AhR-null mice display tissue specific effects such as hyper-proliferation of hair follicles and liver blood vessels, as well as increased levels of apoptosis in liver tissue.[88]

The first evidence that AhR is involved in the control of cell cycles in a ligand-dependent manner was given by the fact that, upon treatment with dioxins or PCBs, the up-regulation of immediate early genes, causing proliferation, was induced. This was mainly studied in the liver, where AhR has been suggested to be a tumor promoter.[83,89-91]

TCDD has also been shown to influence pro-mitotic signals such as the Ras-dependent signaling cascade. Upon TCDD binding to AhR, the Src kinase, which is in complex with Hsp90, is released from the chaperone and translocates to the nucleus. There, c-Src can cause phosphorylation of EGFR (epidermal growth factor receptor) and thus activation of TGF-α and further downstream effectors as Ras.[89,92-94]

A recent study showed that TCDD-bound AhR cannot interact with CDK4, leading to reduced CDK-dependent phosphorylation of Rb and thus to G_1 arrest (see **Figure 6**).[86] In another described mechanism, TCDD binding induces expression of the CDK inhibitor p27, which in turn blocks activity of CDK4/6 in order to prevent S phase entry (see **Figure 6**).[95] AhR bound to TCDD was also identified to be involved in the suppression of apoptosis by inhibition of p53, a powerful tumor suppressor protein (see **Figure 6**).[96]

Altogether, AhR cross-talk with a variety of cellular signal transduction cascades influences the cell cycle control. Proliferation, cell cycle arrest or apoptosis can be induced. However, more research is required to understand the exact mechanisms and signaling pathways of AhR in cell-cycle control.

1.8 Role of AhR in the immune system

It has also been shown that AhR regulates the immune system in a steady state as well as in the case of infection and inflammation.

Recent studies revealed that AhR is involved in balancing the levels of regulatory T cells (T_{reg}) and T-helper (Th) 17 cells. The differentiation of T_{reg} cells correlates with the initiation of pro-inflammatory Th17 cell development and is activated by AhR expression, which is in turn induced by the transforming growth factor (TGF)-signaling pathway.[97] Altogether, this contributes to the pathogenesis of autoimmune diseases as autoimmune encephalitis [98], collagen-induced arthritis [99] and colitis [100].

Dendritic cells play a key role in antigen presentation and in the initiation of T-cell response. AhR is expressed at high levels in dendritic cells, and the addition of an AhR agonist affects the normal dendritic cell differentiation. AhR interacts with RelB, which is a transcription factor necessary for the differentiation and function of dendritic cells. This interaction induces transcriptional activation and leads to abnormal differentiation of dendritic cells.[101]

Macrophages are important for the innate immune response as they produce pro-inflammatory cytokines, among others, by activation via bacterial lipopolysaccharides (LPS). Here, AhR plays an essential role in inhibition of the LPS signaling pathway via interaction with the signal transduction activator transcriptional factor 1 (STAT-1).[102]

Furthermore, an influence of AhR in infection diseases has been observed in several experimental models. AhR is able to control inflammation in parasitic diseases, such as Toxoplasmosis or Leishmaniasis.[103,104]

Also viral infection is connected with AhR activation. For example, in case of influenza virus, activation of AhR by TCDD significantly decreases the survival rate in animal models.[105]

Recently, it was found that AhR as an immune regulator modulates the development of cancer, which will be discussed in the next section in more detail.

Altogether, the role of AhR-induced immunomodulation is multi-dimensional. Ligand induced activation of AhR leads to immunosuppression via boosted T_{reg} cell differentiation. Opposed to this, activation of AhR can induce Th17 cell polarization, which increases the severity of autoimmune disease. Besides T cells, AhR also appears to play a vital role in macrophages, dendritic cells and virus infection. It is well accepted that the magnitude and the type of immune response is dependent on the local cytokine milieu and AhR appears to be one of the key factors involved in the fine turning of this cytokine balance.[106]

1.9   Role of AhR in cancer

Ahr is known to influence the major stages of tumorigenesis: initiation, promotion, progression and metastasis. Several studies on aggressive tumors and tumor cell lines showed an increased level of AhR in the nucleus, which indicates that AhR is chronically activated in tumors. Immunohistochemical analysis of liver, breast, gastric, prostate and small cell lung tumors confirmed an increased level of AhR expression.[107-112] In case of

prostate cancer tissue and lung squamous cell carcinoma, the increased expression and transcriptional activity of AhR was proven by analysis of the cytoplasmic versus the nuclear localization of AhR.[108,111] In case of prostate cancer, enhanced nuclear localization of AhR correlates with the number of poorly differentiated cells, whereas in case of the lung carcinoma, the levels of nuclear AhR are associated with a higher motility rate. Furthermore, the analysis of the expression level of AhR target genes reveals that in case of glioma an increased CYP1B1 expression level correlates with a decreased survival rate of patients.[34]
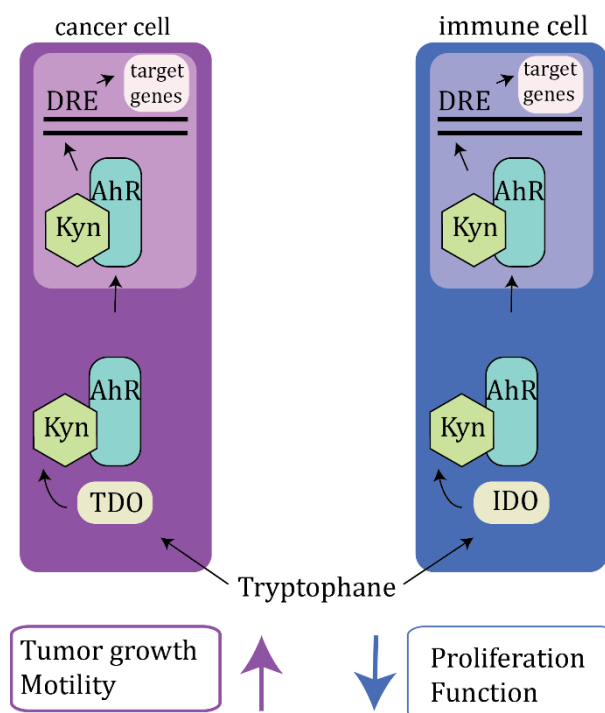
Figure 7: Synoptic scheme visualizing how kynurenine (Kyn), produced by the tryptophan 2,3-dioxygenase (TDO) and indoleamine 2,3-dioxygenase 1 (IDO), influences tumor growth and motility in AhR dependent manner. For explanation see text. Figure modified from [34].

However, the complex inflammatory microenvironment is understood in more detail only in glioma. In this case, glioma consist of a number of cell types expressing various endogenous ligands, acting in an autocrine and paracrine pathway and leading to sustained AhR activation. The key players in that regard are the tryptophan 2,3-dioxygenase (TDO2) and the indoleamine 2,3-dioxygenase 1 (IDO1) pathways, metabolizing tryptophan to kynurenine, a weak AhR ligand. TDO2 and IDO1 can be expressed in diverse types of tumors, which leads to kynurenine production.[113] TDO2 expression, kynurenine production and AhR-regulated CYP1B1 activation correlate in human glioblastoma tissue, which supports growth and motility of glioblastomas.[34]

Immune cells within the tumor microenvironment have increased levels of IDO1 and thus increased levels of kynurenine interacting with AhR. This contributes to immune tolerance and altogether further supports the motility and growth of glioblastoma.[114]

In general, AhR function with regard to tumor cell proliferation seems to be dependent of the model tissue. Nevertheless, it is obvious that AhR does affect tumor cell proliferation through diverse contradictory mechanisms which increases the interest on AhR as a theoretical target for suppression of tumor growth.

1.10 Structural insights into the bHLH-PAS domain containing transcription factor family

Recently, the crystal structure of a member of the bHLH-PAS domain containing transcription factor family, homologous to AhR/ANRT, was solved (PDB: 4F3L). The crystal structure consists of the bHLH domain and the PAS fold of mouse CLOCK and mouse BMAL1.[115]
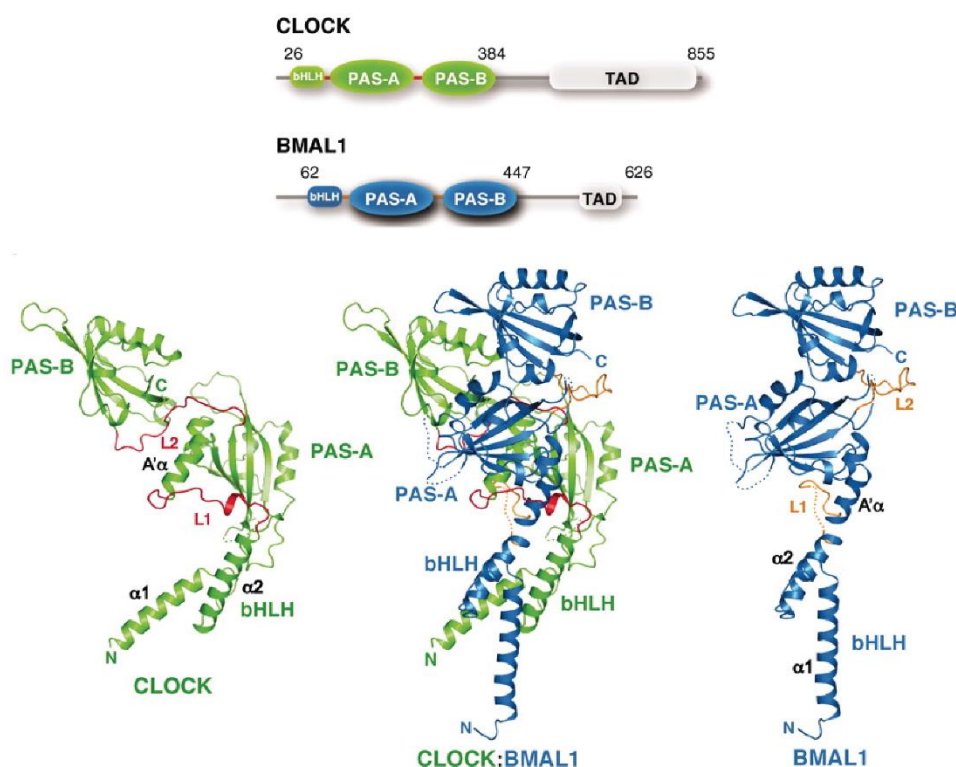
Figure 8: Overall structure of mouse CLOCK/BMAL1 (PDB: 4F3L). Top: Domain organization of CLOCK and BMAL1. Bottom: Ribbon representation of the CLOCK/BMAL1 heterodimer (center) and the individual subunits of CLOCK (left) and BMAL1 (right). CLOCK is colored in green, BMAL1 in blue. Figure taken from [115]. For explanation see text.

The structure revealed an intertwined asymmetric heterodimer with three domains in each of the two subunits. The bHLH, PAS A and PAS B domain of one subunit are interacting primarily with the corresponding domain of the other subunit, thus forming three distinct interfaces (see **Figure 8**).[115]

Although sharing the same overall fold, the bHLH domains crystallized rather differently. Whereas the α2 helix of BMAL1 is nearly continuous with the A'α helix of the PAS A domain, the α2 helix and the A'α helix of CLOCK are around 23 Å apart. Nevertheless, the bHLH domains form a stable four-helical bundle (see **Figure 8**).[115] Additionally, the crystal structure of the isolated bHLH domains of human CLOCK/BMAL1 (PDB: 4H10) bound to a sticky-end 15mer DNA was solved.[116] It revealed minor structural changes within the bHLH domain compared to the mentioned mouse CLOCK/BMAL complex.[116] The N-terminal helix α1 of CLOCK is extended by three helical turns and is slightly shifted outwards probably due to DNA binding (see **Figure 9**).[116]
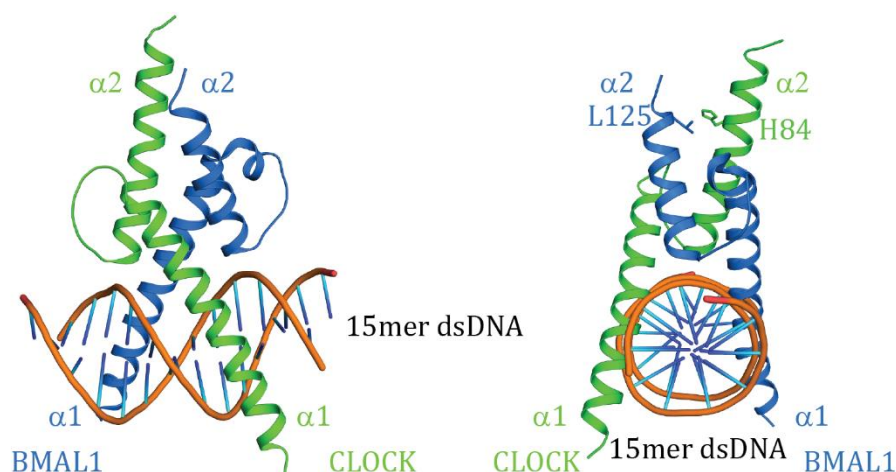
Figure 9: Ribbon representation of the bHLH domains of human BMAL1 and human CLOCK bound to a sticky end 15mer dsDNA (left, PDB: 4H10). Site view of PDB: 4H10. Residues crucial for the formation are labelled. Figures taken from [116]. For explanation see text.

Mutational studies showed, that the residues H84 (CLOCK) and L125 (BMAL1) are crucial for forming the heterodimeric four helical bundle (see **Figure 9**, right).[116] In similar studies, the residues important for the recognition of the distinct E-box element by the human CLOCK/BMAL1 heterodimer were characterized. The residues R39, E43 and R47 in CLOCK and the residues H77, E81 and R85 in BMAL1 are crucial for DNA recognition. Additionally, residue I80 from BMAL1 interacts with the thymine at position 7 (see **Figure 10**).[116]
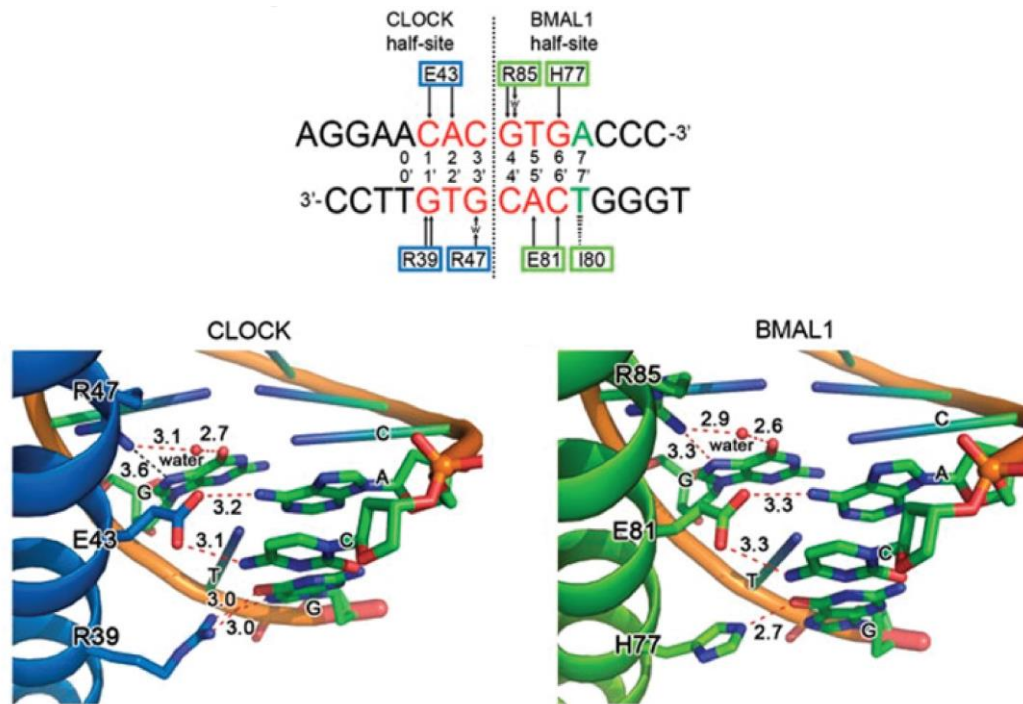
Figure 10: CLOCK and BMAL1 read DNA bases asymmetrically. Top: Schematic recognition diagram of human CLOCK/BMAL1 to the DNA in the crystal structure. E-box element is colored in red. Residues involved in E-box recognition are highlighted. Black arrows represent hydrogen bonds. Lower-case 'w' means water molecule. Zebra triangle represents hydrophobic contact. DNA is divided into the CLOCK half-site and the BMAL1 half-site by a dashed line.(PDB: 4H10). Bottom: The recognition mechanism of CLOCK-BMAL1 bHLH domains to canonical E-box. Detailed interactions between CLOCK and E-box (left) and between BMAL1 and E-box (right). The red dashed lines represent the hydrogen-bonding contacts and the black dashed line represents the distance between two atoms. The bond lengths or distance are shown in Angstroms (Å). Figure taken from [116].

The PAS A domains of mouse CLOCK/BMAL1 (PDB: 4F3L) adopt a typical PAS fold and are structurally very similar (see **Figure 8** and **Figure 11**). The N-terminal flanking helices (A'a) of both subunits, which are external to the PAS domain, pack against each other and face the β-sheet side of the opposite PAS A domain.[115] This domain-swapped helical interface is also known for other PAS domain containing proteins. Recently, the crystal structure of the PAS A domain of a human AhR homodimer was published (PDB: 4M4X) and revealed a similar dimerization mode (see **Figure 11**).[117] The functional importance of the A'α helix for AhR homodimerization was investigated by mutational studies. These could confirm that the A'α helix is a critical dimerization component. Furthermore, cell-based assays reveal that the A'α helix of AhR is also necessary for heterodimerization with ARNT and thus crucial for transcriptional activation.[117]

PAS A structures
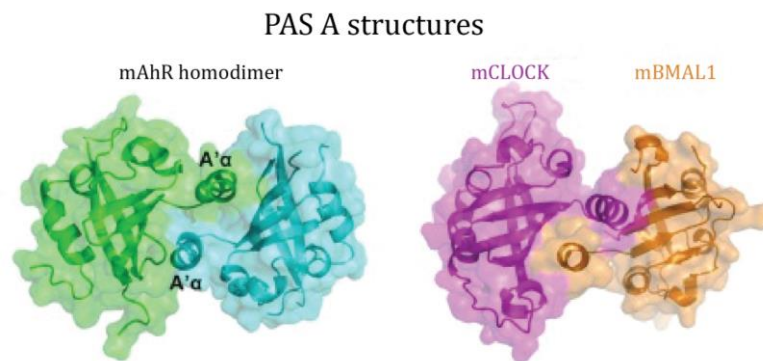
Figure 11: Comparism of the PAS A dimerization interface. Ribbon representation of the homodimer mouse AhR (left; PDB: 4M4X) and the heterodimer of mouse CLOCK/BMAL1 (right; PDB: 4F3L). The helix A'α is involved in interface formation. Figure taken from [117]. For explanation see text.

In contrast to that, the PAS B domain dimerization in the CLOCK/BMAL1 complex (PDB: 4F3L) is completely different.[115] An external helix (as helix A'α in PAS A domain dimerization) is not existing in PAS B dimerization. The PAS B domains are oriented in a rather parallel fashion. The β sheet of BMAL1 PAS B interacts with the α-helical face of the CLOCK PAS B domain (see **Figure 12**, right). Another crystal structure of the isolated PAS B domain containing complex of human HIF-2α and ARNT (PDB: 3F1P; see **Figure 12**, left) revealed a completely different orientation.[118] Here, the β sheet faces of both PAS B domains are interacting with each other and the α helical regions are pointing outwards. Thus, the α helical regions are not involved in this PAS B domain formation.[119] Whether this is an artefact of single domain crystallization or whether different modes of PAS B domain dimerization within the bHLH-PAS domain comprising transcription factor family are necessary for individual transcriptional activation, remains elusive.

PAS B structures
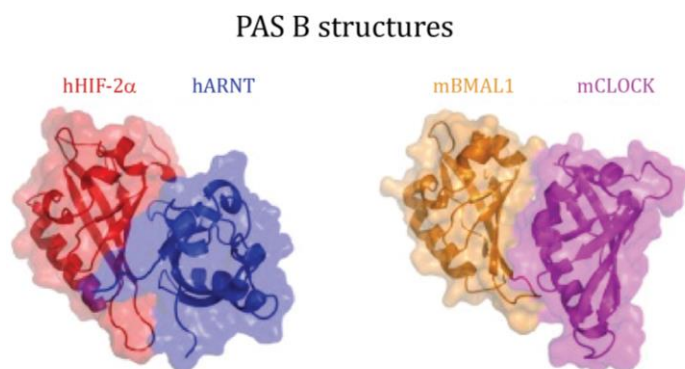
Figure 12: Comparism of the PAS B dimerization interface. Ribbon representation of the heterodimerization of human ARNT/HIF-2α (left; PDB: 3F1P) via the β sheets and the heterodimerization of mouse CLOCK/BMAL1 (right; PDB: 4F3L) via the β sheet/ α helical face of the domains. Figure taken from [117]. For explanation see text.

Although ligand binding to the PAS B domain has only been shown to have a physiological effect for AhR within the bHLH-PAS domain comprising transcription factor family, the isolated PAS B domain of human HIF-2α was used for ligand binding studies. A ligand named THS-044 was identified to bind to the PAS B domain of HIF-2α with an affinity of 1-10 µM. Structural analysis of the ligand bound PAS B domain of HIF-2α in complex with human ARNT showed that the otherwise water-filled cavity within the PAS fold of HIF-2α was occupied by the hydrophobic ligand (PDB: 3H7W; see **Figure 13**).[119] Several computational analyses predicted a similar mode of ligand binding for AhR.[120]
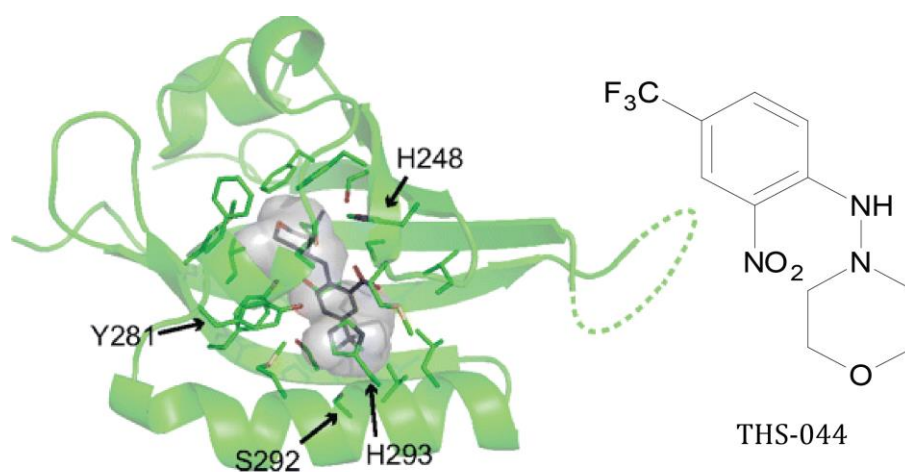
Figure 13: Ribbon representation of Hif-2α PAS B domain (PDB: 3H7W). Hydrophobic residues within the cavity are shown in stick representation (green). Internal cavity is shown as grey spheres. Ligand THS-044 is shown in stick representation (grey). Labeled residues are directly interacting with the ligand. Figure taken from [119]. For explanation see text.

Altogether, during recent years there was great progress in understanding the DNA binding, dimerization and ligand binding of the bHLH-PAS domain containing transcription factor family. Nevertheless, more structural studies are necessary to clarify the heterodimerization mode of the homologous AhR/ARNT complex, its DNA binding and ligand recognition as well as potential rearrangements within the heterodimeric assembly upon DNA binding.

2   **Scope of this work**

The aryl hydrocarbon receptor (AhR) is a ligand-activated transcription factor. AhR comprises the bHLH domain at the N-terminus followed by a PAS fold and the TAD domain. AhR is a class I bHLH-PAS transcription factor and heterodimerize with a class II bHLH-PAS transcription factor (Aryl hydrocarbon receptor nuclear translocator; ARNT) in order to induce gene transcription. AhR is the only member of the bHLH-PAS domain comprising transcription factor family that is able to bind ligands. It is very well studied in the pharmacology field for its role in mediating the toxicity of xenobiotics such as 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD; dioxin).

However, a large scale expression and purification protocol for the AhR/ARNT complex has not been published so far. The objective of this thesis is to establish a purification protocol of the AhR/ARNT complex for further biochemical and structural analysis. Via CD-spectroscopy and RALS analysis the proper folding and oligomerization state of the AhR/ARNT complex shall be analyzed.

Cell-based assays revealed the dioxin responsive element (DRE; 5'- NGCGTG -3') the AhR/ARNT complex bind to in order to initiate the adaptive response. Via ITC experiments, I want to determine the exact affinity of the AhR/ARNT complex to the DRE and to analyze the importance of the bases up- and downstream the DRE in regard to DNA binding affinity.

A recent crystal structure of the homologous heterodimeric transcription factor of mouse CLOCK/BMAL1 (PDB: 4F3L) showed an asymmetric intertwined heterodimerization involving the bHLH and the PAS-fold. Structural studies of the AhR/ARNT complex could reveal whether this mode of heterodimerization is conserved within the bHLH-PAS comprising transcription factor family.

Slight rearrangements within the bHLH upon DNA binding could be observed in the crystal structure of an isolated bHLH domain bound to DNA (human CLOCK/BMAL1; PDB: 4H10) which revealed slight rearrangements within the bHLH domain. It became not clear whether the rearrangements in the bHLH domain upon DNA binding mediate also the loosening of the PAS domain interactions. Moreover, the disruption of the AhR/ARNT complex at transcriptional termination is not understood. Structural studies of the AhR/ARNT complex bound to DNA could reveal possible rearrangements upon DNA binding and thus could help to understand the mechanism of transcriptional termination in more detail.

3 Materials and Methods

3.1 Materials

A detailed list of instruments and chemicals can be found in the appendices.

3.1.1 Enzymes

DNAseI	Roche
DpnI	New England Biolabs
KOD polymerase	Novagen
Precission Protease™	GE Healthcare
T4 DNA ligase	New England Biolabs
BamHI	New England Biolabs
NotI	New England Biolabs

3.1.2 Kits

Mark12™ unstained ps	Life Technologies
Unstained protein MWM	Thermo Scientific
2-log DNA ladder	New England Biolabs
100-bp DNA ladder	New England Biolabs
QIAprep™ spin mini prep kit	Qiagen
innuPREP Plasmid mini kit	analytic Jena
QIAquick gel extraction kit	Quiagen
GeneAmp© dNTPs	Life Technologies
JBS Classic suite	Jena Bioscinece
JBS ClassicII suite	Jena Bioscience
JBScreen Plus HTS	Jena Bioscience
NucPro suite	Jena Bioscience
PACT suite	Quiagen
PEG suite	Quiagen
PEGII suite	Quiagen
JBS JCSG	Jena Bioscience

3.1.3   Cell lines

E.coli TG1	K12, supE, hsdΔ5, thi, Δ(lac-proAB), F'[traD36, proAB+, lacq, lacZΔM15] (Promega)
E.coli BL21 (DE3) Rosetta	F- ompT hsdsSB(rB- mB-) gal dcm (DE3) pRARE (CmR) pRARE containing the tRNA genes argU, argW, leX, glyT, leuW, proL, met, thrT, tyrU, thrU (Novagen)

3.1.4   Vectors

pETDuet-1	Novagen
pET30-EK/LIC	Novagen
pET32-EK/LIC	Novagen
pET46-EK/LIC	Novagen

3.1.5   Clones

The cDNA for mouse ARNT with NCBI gene bank code BC012870.1 was purchased from Source Bioscience. The first 500bp of human AhR (NM_001621.4) were synthesized with respect to optimized *E.coli* expression by MWG eurofins.

3.1.6   Primers

mmARNT 85-360 pET30-EK/LIC	fwd: 5'-GAA GTT CTG TTC AGG GGC CCA TG GAT AAA GAG AGA CTT-3'
	rev: 5'-GAG GAG AAG CCC GGT CA TAC CTG CAG CCT GCC-3'
mmARNT Δ274-303	fwd: 5'-AGG TGT GGT ACT AGC TCC GAA GGA GAA CCT CAC TTT-3'
	rev: 5'-AAA GTG AGG TTC TCC TTC GGA GCT AGT ACC ACA CCT-3'
mmARNT C256S	fwd: 5'-CAG TCT TCC ATG AGG ATG AGT ATG GGC TCA CGA AGG-3'
	rev: 5'-CCT TCG TGA GCC CAT ACT CAT CCT CAT GGA AGA CTG-3'
cohsAhR 23-273 pET32-EK/LIC	fwd: 5'-GAA GTT CTG TTC AGG GGC CCA TG GTC AAA CCG ATT CCC-3'

rev: 5'- GAG GAG AAG CCC GGT TCA ATG CTC GGT ACC TTC-3'

cohsAhR Δ182-210   fwd: 5'-TGG GCA CTG AAT CCG TCA GAA AAT TCC CCG TTA ATG-3'

rev: 5'-CAT TAA CGG GGA ATT TTC TGA CGG ATT CAG TGC CCA-3'

3.1.7   Media

Luria-Bertani (LB)   10g/L tryptone/peptone; 10g/L NaCl; 5g/L yeast extract

Terrific-Broth (TB)   powder 50g/L + 4mL glycerol

SeMet-minimal medium   1xM9 salt, 1x trace elements, 1x vitamins, amino acid mix, $MgSO_4$ 1mM, $CaCl_2$ 0.3 mM, glucose 2g/L, NH_4Cl 0.5 g/L

3.1.8   Antibiotics

Carbenicillin   100 g/L in 40% (v/v) EtOH (1000x)

Chloramphenicol   34 g/L in 100% EtOH (1000x)

Kanamycinsulfate   50 g/L in H_2O (1000x)

3.1.9   Buffers

Buffer A   25 mM HEPES/NaOH pH 7.5 - 8.5, 200 mM NaCl, 30 mM Imidazol, 5 mM β-Mercaptoethanol, 0.1 mM Pefabloc, 1 mM DNAseI

Buffer B   25 mM HEPES/NaOH pH 7.5 - 8.5, 200 mM NaCl, 500 mM Imidazol, 5 mM β-Mercaptoethanol

Buffer C   25 mM HEPES/NaOH pH 7.5 - 8.5, 200 mM NaCl, 5 mM DTT

10x TBE buffer   108g/L TRIS-base, 55g/L boric acid; 9.3g/L Na_2EDTA pH 8.0

4x SDS sample buffer   0.25 mM Tris-HCl pH 6,8, 1.6mM EDTA, 8% (w/v) SDS, 40% (v/v) Glycerol, 100mM DTT, Bromphenolblau

10x SDS running buffer   250 mM TRIS-base, 1.92 M Glycin, 1% (w/v) SDS

TSS buffer	85% (v/v) LB-medium, 10% (v/v) PEG 3350. 5% (v/v) DMSO, 50mM $MgCl_2$, pH 6.5
100x trace elements	EDTA 5 g/L, $FeSO_4$ 0.5 g/L, $ZnCl_2$ 0.05 g/L, $CoSO_4$ 0.01 g/L, $CuCl_2$ 0.01g/L, H_3BO_3 0.01 g/L, $MnCl_2$ x $4H_2O$ 1.6 g/L, pH 7.0, autoclaved
1000x vitamins	Riboflavin (B2) 1 g/L, Niacinamide (B3) 1 g/L, Pyridoxine monohydrochloride 0.1 g/L, Thiamine (B1) 1 g/L
Amino acid mix	L-Lysine, L-Phenylalanine, L-Threonine, L-Isoleucine, L-leucine, L-Valine 0.1 g/L each, L-Selenomethionine 0,05 g/L
5xM9 salt	Na_2HPO_4 x $2H_2O$ 80g/L, KH_2PO_4 20g/L, NaCl 5g/L, pH 7.2

3.2   Molecular biology methods

3.2.1   Polymerase chain reaction

DNA from AhR and ARNT was amplified with gene specific primers using KOD-polymerase. [121]

3.2.2   DNA digestion

PCR products or methylated template-DNA were treated using enzymes from New England Biolabs according to manufacturer's protocol.

3.2.3   Agarose gel electrophoresis

Agarose gels were prepared and run according to standard procedures using 1xTBE buffer with 0.3 μg/mL ethidium bromide.[121]

3.2.4   DNA purification

Desired DNA bands were excised from gel and subsequently purified using QIAquick gel extraction kit according to manufacturer's protocol.

3.2.5   Ligation

The amount of plasmids and PCR products were quantified by absorption measurements at 260 nm with a Nanodrop 2000 spectrometer. 10 ng of plasmid was ligated with six-fold molar excess of PCR product using T4 DNA ligase at RT for one hour.

3.2.6   Ligase-independent-cloning (LIC)

PCR product with specific 5' overhang was amplified, treated with T4 polymerase and incubated with distinct vectors according to manufacturer's protocol.[122]

3.2.7   Site directed mutagenesis

Site-directed mutagenesis was carried out using KOD polymerase using the following protocol.

>   1 µL Primer 1 (10 pmol/µL)
>   1 µL Primer 2 (10 pmol/µL)
>   0.5 µL template plasmid (50-100 ng)
>   1.6 µL MgSO4 (25 mM)
>   2.0 µL dNTP (2 mM each)
>   0.4 µL KOD (1U/µL)
>   2 µL 10x buffer
>   11.5 µL ddH$_2$O

The following thermocycle protocol was applied:

>   94°C   5'
>   94°C   30s
>   58°C   30s   25x
>   68°C   1'/5000bp
>   68°C   7'

3.2.8   Preparation of chemically competent *E.coli*

Competent cells were prepared as described.[123] 1 L LB-medium was inoculated with 10 mL pre-culture and grown at 37 °C until an OD$_{600}$= 0.4. Bacteria were incubated on ice for

20 min, pelleted for 5 min at 1200 x g at 4 °C and resuspended in 100 mL ice-cold TSS buffer, flash frozen and stored at -80 °C.

3.2.9   Transformation of chemically competent *E.coli*

Transformation of bacteria via heat shock was done according to standard protocol.[124]

3.2.10   Isolation of plasmid DNA

Plasmids were isolated from bacteria using the QIAprep Miniprep Kit according to the manufacturer's protocol. Elution of the DNA was done with 40 µL ddH$_2$O.

3.3   Biochemical methods

3.3.1   SDS PAGE

Analytical separation of proteins with different molecular weight was performed using a denaturing, discontinuous SDS-polyacrylamide gel electrophoresis at a neutral pH.[125,126]

3.3.2   Protein solubility test

For a protein solubility test, 1 L TB medium supplemented with respective antibiotics was inoculated 1:100 with an over night (O/N) LB-medium *E.coli* BL21 (DE3) Rosetta culture carrying the respective expression plasmid. Cell culture was grown at 37°C till an OD$_{600}$ of 0.5 was reached. Temperature was decreased to 17 °C and protein expression induced by addition of 200 µM IPTG. After additional 20 hrs of incubation, bacteria were harvested by centrifugation at 4000 x g for 15 min at 4 °C and pellets were resuspended in 45 mL buffer A of appropriate pH per 1 L bacteria culture pellet. The bacteria suspension was passed four times through a Fluidizer to disrupt the cells. The lysate was cleared at 35000 x g at 4 °C for 45 min in an Optima-L100 K ultracentrifuge. The supernatant was filtered and applied to a 1 mL HisTrap HP chromatography column equilibrated with 10 CV buffer A. The column was washed again with buffer A. The protein was eluted with stepwise increasing imidazole concentration via mixing buffer A and B. Samples taken at every step of the test purification were analyzed via SDS-PAGE.

3.3.3   Large scale protein overexpression

Large scale protein overexpression was carried out using 10 – 20 L TB medium supplemented with the respective antibiotics analogous to protein solubility tests. Induction of protein expression was done at OD_{600}= 0.3-0.5 with 200 μM IPTG. The bacterial cell pellets were stored at -80 °C. For incorporation of selenomethionine, protein was expressed in minimal medium as described here [127].

3.3.4   Protein purification

All further steps were carried out at 4 °C. The filtered supernatant was applied to a column loaded with 15 mL Ni Sepharose HP beads equilibrated with 10 CV buffer A. The column was washed with buffer A until a baseline at A_{280} was reached. Bound protein was eluted by increasing the imidazole concentration stepwise via mixing buffer A and B. Eluted protein was dialyzed overnight against buffer A without imidazole in the presence of a catalytic amount of GST-PreScission protease (1:200) to cleave the N-terminal His_6-Tag. Protein was reapplied to a Ni-NTA connected in line to a 1mL GST-Trap column equilibrated with buffer A without imidazole. Protein was eluted with buffer A plus 20 mM Imidazole while the cleaved His_6-Tag and the GST-PreScission stuck to the columns. The peak fractions were pooled, concentrated and applied to a Superdex 200 SEC column, equilibrated with buffer C. Respective fractions were pooled, concentrated and flash frozen in liquid nitrogen. The purification were analyzed via SDS-PAGE.

3.3.5   Protein concentration

Protein solutions were concentrated using Amicon centrifugal devices with a 30 kDa cutoff according to the manufacturers' protocol.

3.3.6   Determination of protein concentration

Absorption at wavelength λ= 280 nm was measured using a Nanodrop 2000. The extinction coefficient ε of different protein constructs were determined via ProtParam (http://web.expasy.org/protparam).[128] The protein concentration was calculated according to the Lambert-Beer law.

$$A = -\log\left(\frac{I}{I_0}\right) = \varepsilon \cdot c \cdot l$$

3.3.7 Protein storage

Protein was flash frozen in appropriate aliquots in liquid nitrogen and stored at -80 °C.

3.3.8 Isothermal titration calorimetry (ITC)

ITC experiments were carried out at 20 °C in an ITC-200 with the protein/DNA in buffer C. Protein concentration in the cell was 30-100 μM whereas the dsDNA concentration in the syringe was 400 – 550 μM. Double-stranded DNA was prepared by premixing same molecular amounts (800-110 μM) of single-stranded DNA. Heating to 95 °C for 30 min and slow cooling to room temperature oN resulted in dimerization of the DNA. Binding isotherms were fitted and equilibrium dissociation constants calculated using the Microcal ORIGIN software.

3.3.9 Analytical gel filtration and right angle light scattering (RALS)

A coupled RALS-refractive index detector (Malvern) was connected in line to an analytical gel filtration column Superdex S75 10/300 (flowrate 0.6 ml/min). To determine absolute mass of applied proteins, data were analyzed with the OmniSec software. 100 μL of protein solution with a concentration of 1–3 mg/mL was applied to the column. Buffer C was the running buffer.

3.3.10 Circular dichroism spectroscopy

To analyze, whether the purified proteins are folded, circular dichroism (CD) spectra were recorded on a Jasco J-720 spectropolarimeter (Jasco Corp., Hachioji, Japan) equipped with a 450 W lamp. The spectra were recorded at room temperature from 260-190 nm using a 1 mm quartz cuvette. The CD buffer contained 20 mM Na-phosphate buffer and 150 mM NaF and the protein was dialyzed O/N against it and a total concentration of 0.2 mg/mL protein solution was used for the measurements. The collected spectra were analyzed using the CDNN software.[129] The plane-polarized light is made up of equal parts of left-handed and right-handed polarized light. Chiral particles as amino acids absorb these two components unequally.[130] The CD spectrometer measures the differences and reports it in degrees of ellipticity (θ) as a function of wavelength. Highly asymmetric secondary structure elements such as β-sheets and α-helices have characteristic CD spectra. The

CDNN software deconvoluted the measured spectra and predicted the amounts of secondary structure elements within the protein.

3.4   Crystallographic and computational methods

3.4.1   Protein crystallization

Initial crystallization trials were carried out with native AhR/ARNT complex or ARNT protein alone. The frozen protein solution was gently thawed on ice and used at a concentration of 9 mg/mL. For crystallization trials in complex with DNA, different lengths of respective double stranded DNA were added in a slight molar excess and incubated for 60 min on ice, followed by a purification via SEC (S200 10/300). Complexes were concentrated to 9 mg/mL. Crystallization trials of complexes were carried out at 4 and 20 °C. Automated trials were done in a 96 well format using the sitting drop vapor diffusion method. 70 μL of the premixed reservoir solution were automatically pipetted to the 96 well crystallization plate. The sitting drop consisted of 200 nL protein solution and 200 nL reservoir solution. The workflow was carried out by the Art Robbins Gryphon system. Non-automated crystallization trials were carried out in a 24 well format using the hanging drop vapor diffusion approach. For this method, 500 μL of reservoir solution was pipetted into a 24-well crystallization plate. The hanging drops consisted of different ratios and amounts of reservoir and protein solution. Three final crystallization conditions were optimized to 1: 20% w/v PEG 3350, 200 mM ammonium formate pH 6.6, 2: 20% PEG 3350, 200mM ammonium formate pH 6.3 and 3: 20% PEG 8000, 10 mM calcium chloride and 10 mM Tris-HCl pH 7.5 in a 24-well format.

3.4.2   Cryo-protection of crystals

Crystals were treated with suitable cryo-protectant prior to data collection to prevent or alleviate radiation damage. Cryo-solutions were composed of equal volumes of protein buffer and reservoir solution from the corresponding crystallization condition plus varying concentrations of PEG 200, 400, 3350, 8000, MPD, Glycerol or, Ethylenglycol. Crystals were soaked into 4 μL cryo-solution at 4 °C for at least 5 s before flash freezing in liquid nitrogen.

For slight dehydration, crystals were incubated in cryo-solution for 0.5-18 hrs as a hanging drop over 100μL of the reservoir solution plus 400 mM NaCl.

The final cryo-solution for crystallization condition 1 and 2 was 25 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, 35 % (w/v) PEG 3350, 230 mM ammonium formate. The cryo-solution for crystallization condition 3 contained 25 mM HEPES pH 8.0, 200 mM NaCl, 2 mM DTT, 30 % w/v PEG 4000, 5 mM calcium chloride, 10 mM Tris-HCl pH 7.5. Crystals were tested at BESSY II, Berlin.

3.4.3   Data collection

All data were recorded at BL 14.1 at BESSY II, Berlin using a PILATUS 6M detector. A native data set from a single crystal with a φ increment of 0.1° at a temperature of 100 K was used. The used wavelength was 0.91841 Å and the detector distance 650 mm. 3600 images were recorded with an exposure time of 1 s. Initial indexing and calculation of an optimal data collection strategy was done with the Mosflm software.[131] The dataset was processed using XDS program suite.[132] Data collection of anomalous diffracting crystals were done at different wavelengths and detector distances.

3.4.4   Protein structure solution

Crystallographic symmetry operations as reflections, inversions or rotations of the asymmetric unit describe the unit cell. The latter makes up the crystal via three-dimensional translation. The space group in a crystal is defined by the dimension of the unit cell and the symmetry operations applied to the asymmetric unit. Since inversions and reflections cannot be applied on chiral molecules, protein crystals can only crystallize in 65 of 230 possible space groups.[133]

In a given crystal, electrons diffract an impinging X-ray beam in a way described as elastic or Thomson scattering. The symmetry of the X-ray diffraction contains the information about the unit cell dimension and its crystallographic symmetry, whereas the intensities of the diffracted X-ray beam contain the information of the electron density contribution in the crystal. The electron density explains the spatial arrangement of the atoms in the protein crystal and thus the crystal structure of the protein.

The sum of the scattering by all individual atoms in a unit cell is given by its structure factor $\vec{F}(S)$, where j is the sum of all atoms in the unit cell, f_j is the atomic scattering factor, $\vec{r_j}$ the position of the atom j with respect to the origin and $\vec{S} = \vec{s} - \vec{s_0}$ with $\vec{s}$ as the scattered wave factor and $\vec{s_0}$ as incident wave vector.

$$\vec{F}(S) = \sum_{j=1}^{n} f_j \cdot e^{(2\pi \vec{r_j} \vec{S})}$$

Equation 1: Calculation of the structure factor $\vec{F}(S)$ of a given cell.

The total wave $\vec{K}(S)$ scattered by the sum of all unit cells in a crystal can be written as

$$\vec{K}(S) = \vec{F}(S) \cdot \sum_{t=0}^{n_1} e^{(2\pi i t \vec{a} \vec{S})} \cdot \sum_{u=0}^{n_2} e^{(2\pi i u \vec{b} \vec{S})} \cdot \sum_{v=o}^{n_3} e^{(2\pi i v \vec{c} \vec{S})}$$

Equation 2: Calculation of total scattering wave $\vec{K}(S)$

n_{1-3} are the number of unit cells in direction $\vec{a}, \vec{b}$ and $\vec{c}$ and the position of each unit cell is: $t \cdot \vec{a} + u \cdot \vec{b} + v \cdot \vec{c}$. $\vec{K}(S)$ equals almost always zero, unless $\vec{a} \cdot \vec{S}$, $\vec{b} \cdot \vec{S}$ and $\vec{c} \cdot \vec{S}$ are integers of h, k and l. This phenomenon is known as Laue condition with hkl as so-called Miller indices. The equivalent to the Laue conditions in real space is known as Bragg's Law,

$$n\lambda = 2d\sin\theta$$

Equation 3: Bragg's Law describes the condition for constructive interference for successive crystallographic planes in a crystal lattice in real space.

where λ is the wavelength of the incident x-ray beam, d is the distance between the planes of the crystal lattice, and θ describes the angle between the incident x-ray and the scattering planes.

The structure factor $\vec{F}(S)$ can be also written as the integration over all electrons in the unit cell, where $\rho(\vec{r})$ is the electron density at position $\vec{r}$.

$$\vec{F}(S) = \int_{cell} \rho(\vec{r}) e^{(2\pi i \vec{r_j} \vec{S})} dv$$

Equation 4: Calculation of the structure factor $\vec{F}(S)$ of unit cell as an integration over all atoms.

Written in fractional coordinates (x,y,z) for the unit cell and given V as the unit cell volume, one can write $dv = V \cdot dx\, dy\, dz$ and $\vec{r} \cdot \vec{S} = (\vec{a} \cdot x + \vec{b} \cdot y + \vec{c} \cdot z) \cdot \vec{S} = hx + ky + lz$. Due to this the structure factor $\vec{F}(S)$ can be written as a function of h,k,l.

$$\vec{F}(h,k,l) = V \int\limits_{x=0}^{1} \int\limits_{y=0}^{1} \int\limits_{z=0}^{1} \rho(xyz) \times e^{(-2\pi i(hx+ky+lz))} dxdydz$$

Equation 5: Structure factor $\vec{F}(S)$ as a function of h k l.

$\vec{F}(h,k,l)$ is the Fourier transform of $\rho(h,k,l)$ and vice versa. The Laue condition restricts scattering to occur only in discrete directions, therefor the integration can be written as a summation.

$$\rho(x\,y\,z) = \frac{1}{V} \sum_h \sum_k \sum_l |F(h\,k\,l)| \cdot e^{-2\pi i(hx+ky+lz)+ia(h\,k\,l)}$$

Equation 6: Electron density according to the Laue conditions.

The structure factors $|F(h\,k\,l)|$ are proportional to the measured intensities on the detector and thus experimentally accessible. This is not true for the phase angles (h k l). This crucial information gets lost during measurement. Nevertheless, there are several common ways to determine the phases for structure determination of macromolecules. [134]

Solving the phase problem via the isomorphous replacement method requires incorporation of heavy atoms into the protein of interest. If the crystals of the derivative protein and the native one are isomorphous, the heavy atom structure factor amplitudes can be approximated with regard to the differences of the reflection amplitudes of both crystals.

Furthermore, the phase problem can be solved via single or multiple wavelength anomalous dispersion (SAD or MAD) utilizing anomalous scattering from just one crystal. The so called anomalous difference describes the occurring difference in the reflection pairs $\vec{F}(hkl)$ and $\vec{F}(-h-k-l)$ if the incident X-ray beam is adjusted to an energy close to the absorption edge of the given element. The anomalous difference can be used to determine the location of the anomalous scatterer in the crystal. Very often selenium in incorporated in the protein as selenomethionine as the production is very well established.

A third approach requires the knowledge of a homologues structure, which can be used to approximate the phase angles of the crystallized protein and to solve the structure. For this, both proteins should have a sequence identity > 25% and an r.m.s.d. of the C_α positions < 2 Å. This method is named molecular replacement.

4   Results

4.1   Approaches to purifying the Aryl hydrocarbon receptor

Scientists in the fields of cancer, immunity, medicine and pharmacy have been highly interested in studying the Aryl hydrocarbon receptor in detail *in vitro* due to its high potential as a therapeutic target. However, no successful high yield expression and purification protocols have been reported to date.

Numerous approaches used in the context of this study attempting to overexpress and purify the aryl hydrocarbon receptor will be outlined briefly here. Protein overexpression and solubility tests for all mentioned attempts were carried out as described in 3.3.2. Protein purification trials were done as described in 3.3.4. All constructs were designed based on a homology model to the known structures of mouse CLOCK and mouse BMAL1. A clone list and sequence alignments can be found in the appendices.

Early on, it was recognized that the DNA constructs of the human aryl hydrocarbon receptor (**h**omo **s**apiens, **hs**AhR) were not over-expressed in *E.coli* Rosetta 2 (DE3). One potential reason for this failure could be that the codons used in the hsAhR gene are not compatible with *E.coli* codon usage, despite the presence of genes expressing rare tRNAs in *E.coli* Rosetta 2 (DE3). To circumvent this problem, the DNA sequence of the protein was optimized for the codon usage of *E.coli*. This strategy was successful and DNA constructs comprising the **c**odon **o**ptimized hsAhR (**cohs**AhR) showed overexpression in *E.coli* Rosetta 2 (DE3). However, all tested constructs were insoluble. Also the optimization of the protein purification buffers such as varying the buffer, the NaCl concentration, and the pH, adding glycerol or detergents such as Triton or Tween20 did not lead to soluble protein.

Another strategy to improve the solubility of cohsAhR was to fuse different solubility tags like Glutathione-S-Transferase (GST), Maltose binding protein (MBP) or Thioredoxin (Trx) to the protein constructs of cohsAhR. However, the yield of purified protein was too low for further biochemical characterization due to the fact that the protein constructs were still insoluble and precipitated after tag-cleavage or earlier.

Subsequently, I attempted to enrich and purify cohsAhR from the inclusion bodies of *E.coli* Rosetta 2 (DE3) after over-expression. The enrichment of unfolded cohsAhR protein constructs purified from inclusion bodies was possible, but various refolding protocols (rapid refolding, slow refolding, on column refolding, refolding via dialysis) using 6 M Guanidinium hydrochloride or 8 M Urea led to protein precipitation.

As mentioned in the introduction, Ahr is able to bind ligands. Thus another approach was the expression and/or purification of cohsAhR in the presence of different ligands (Kynurein, Tryptophan, VAF-34, ITE (2-(1'H-indole-3'-carbonyl)-thiazole-4-carboxylic acid methyl ester) and Indole-3-carbinole). However, the different ligands did not help to stabilize the cohsAhR constructs by binding to the PAS B domain.

As AhR is associated with Hsp90 and XAP2 in the cell, co-expression with these chaperones in *E.coli* Rosetta 2 (DE3) was attempted. Although co-expression of all components was achieved, the chaperoning partners could not stabilize AhR as such that purification of AhR in complex with its chaperoning partners was possible.

If activated upon ligand binding, AhR translocate to the nucleus and forms a heterodimeric complex with the well-known and best studied interaction partner of AhR, the Aryl hydrocarbon receptor nuclear translocator (ARNT). For this study, ARNT from the organism **m**us **m**usculus was used (**mm**ARNT) with a sequence identity to human ARNT of 92%. After construct optimization, single-expression in *E.coli* Rosetta 2 (DE3) and purification of different mmARNT constructs containing the bHLH and/or the PAS fold was successful.

In order to support folding of cohsAhR, the latter was fused directly to its interaction partner mmARNT. Fusion of different protein constructs of cohsAhR and mmARNT via glycine-serine-linkers of different lengths was successful and over-expression in *E.coli* Rosetta 2 (DE3) could be obtained. But this approach did not stabilize AhR in complex with ARNT. The protein purification revealed just aggregated fusion-protein running in the void volume of gel filtration runs.

4.2   Protein purification of the AhR/ARNT protein complex

For biochemical and structural analysis, milligram amounts of protein are required. In order to fulfill this criterion, co-expression of the interaction partners AhR and ARNT in *E.coli* Rosetta 2 (DE3) was optimized and successful using the following constructs. CohsAhR 23-273 (**c**odon **o**ptimized; **h**omo **s**apiens) was cloned via LIC cloning into the bacterial expression vector pET-32EK/LIC. The protein construct mmARNT 85-360 (**m**us **m**usculus) was cloned via LIC cloning into the bacterial expression vector pET-30-EK/LIC and further optimized by inserting a deletion (Δ274-303) and a cysteine mutation (C256S). The constructs comprise the bHLH and the PAS A domain of AhR and ARNT respectively and the complex will be named AhR/ARNT complex in the following. The

bacterial strain *E. coli* Rosetta 2 (DE3), optimized for IPTG-inducible eukaryotic protein expression, was subsequently co-transformed with sequence-verified clones. The vectors provide different resistances to antibiotics so that co-transformation and co-expression could be ensured. The constructs were expressed as N-terminally hexahistidine(His$_6$)-tagged proteins with a PreScission protease cleavage site in the linker region between the His$_6$-tag and the protein sequence. The protein construct cohsAhR carried an additional Thioredoxin(Trx)-tag localized between the His$_6$-tag and the PreScission cleavage site. Protein over-expression and solubility was analyzed as described in 3.3.2. It can be seen in **Figure 14** that expression of the desired constructs resulted in additional bands of approximately 43 kDa (tagged cohsAhR) and 32 kDa (tagged mmARNT) compared to the non-induced sample.
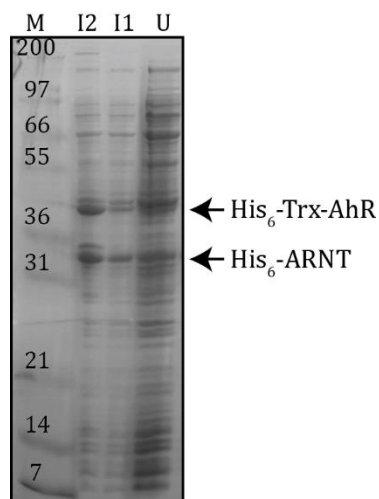
Figure 14: SDS-PAGE of a protein over-expression experiment. M: protein marker with molecular weight in kDa, I2: Overexpression of His$_6$-Trx-cohsAhR and His$_6$-mmARNT after 18hrs, I1: Overexpression of His$_6$-Trx-cohsAhR and His$_6$-mmARNT after 6 hrs, U: None induces *E.coli* Rosetta 2 (DE3) cultures.

After successful small-scale tests of protein over-expression, the constructs were over-expressed in 20L *E.coli* Rosetta 2 (DE3) cultures as described in 3.3.3. After O/N expression at 17 °C, cells were broken and the soluble material was separated from the cell pellet by ultracentrifugation. The supernatant containing the His$_6$-tagged AhR/ARNT complex was purified via affinity chromatography. Bound complex was eluted from the Ni-NTA-material and dialyzed overnight at 4°C in the presence of PreScission protease to cleave the N-terminal tags. Afterwards, the complex was purified via a second affinity chromatography and a subsequent size exclusion chromatography (SEC) in order to remove the His$_6$-tag, the PreScission protease and the excess of ARNT (see 3.3.4). The

molecular weight of cohsAhR 23-273 and mmARNT 85-360 (Δ274-303, C256S) after tag-cleavage is 26 and 29 kDa, respectively. Samples of each purification step were collected and analyzed using SDS-PAGE (see 3.3.1). The gel filtration peak fractions containing the desired AhR/ARNT complex were pooled and concentrated (see 3.3.5). The protein concentrations were determined photometrically at λ=280nm (see 3.3.6) with the assumption that the AhR/ARNT complex is formed in a 1:1 stoichiometry. Appropriate protein aliquots were snap-frozen for long term storage (see 3.3.7). The yield of AhR/ARNT complex was up to 4 mg per liter expression culture.
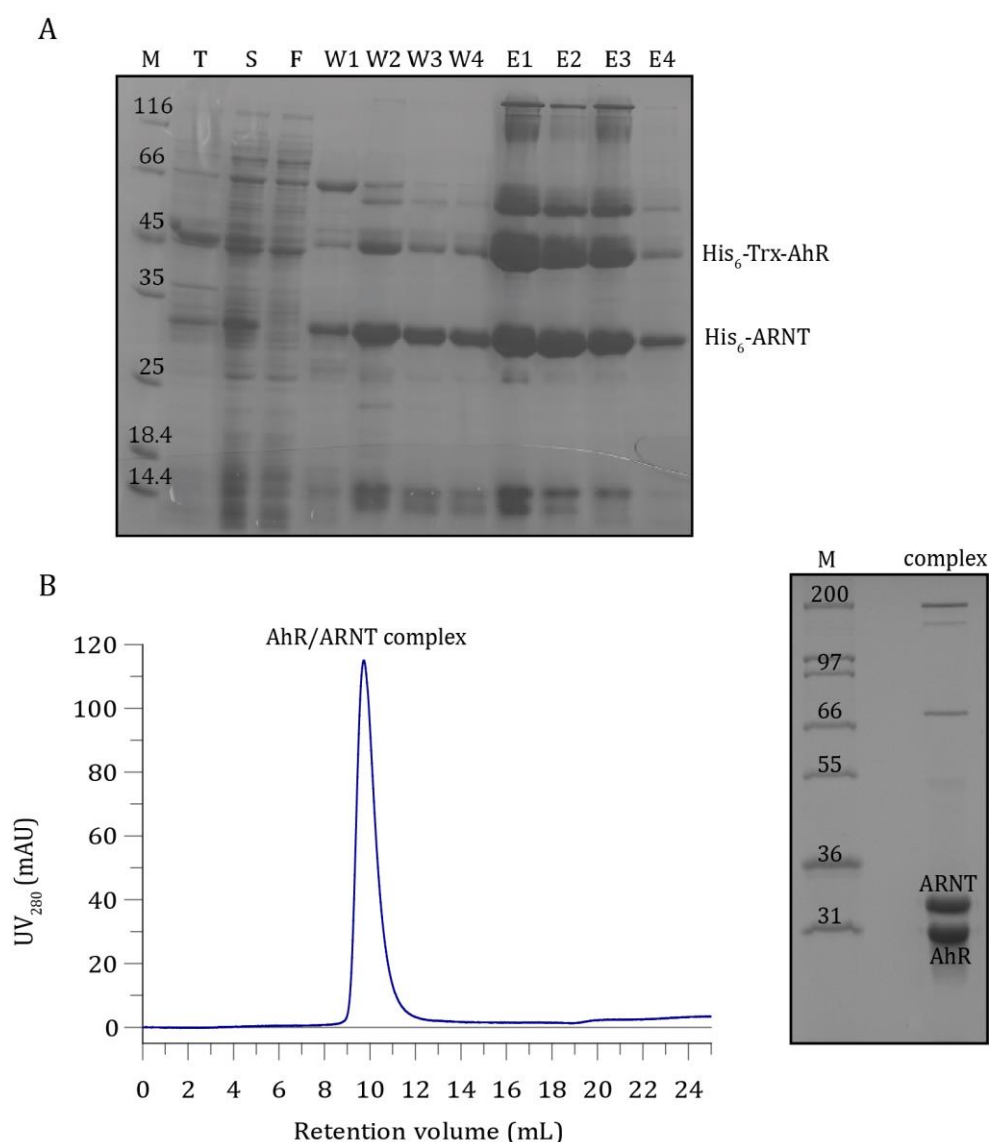
Figure 15: Co-purification of the AhR/ARNT complex. Co-purification of cohsAhR 23-273 and mmARNT 85-360 (Δ274-303, C256S). A: SDS-PAGE showing the expression and affinity purification of the protein complex. M: marker, T: total cell extract, S: supernatant, F: flow through, W1-4: Ni-NTA column wash fractions, E1-4: Ni-NTA elution fractions. B: SEC elution profile of the AhR/ARNT complex (Superdex S75 10/300) and SDS-PAGE of the AhR/ARNT complex after SEC and concentration.

Figure 15 illustrates that most of the proteins in the soluble cell fraction do not bind to the Ni-NTA beads (lane F). Wash fractions (lane W1-W4) contained nonspecifically bound material and traces of the AhR/ARNT complex. In the SEC, the AhR/ARNT complex elutes in one sharp peak corresponding to a molecular weight of approximately 60 kDa. The SDS-PAGE of pooled and concentrated SEC peak fractions showed only minor protein contamination (lane complex).

4.3 CD (circular dichroism) measurements

As mentioned in 4.1, overexpression and purification of AhR turned out to be highly challenging. If expressed by itself, AhR precipitated after tag-cleavage or earlier, depending on purification conditions. Thus, the proper folding of the AhR/ARNT complex was analyzed using CD-spectroscopy. To this end, the purified complex (see 4.2) was dialyzed over night against NaF and phosphate buffer and a CD spectrum was collected (see 3.3.10). For comparison, mmARNT 85-360 (Δ274-303, C256S) was expressed and purified separately as described in 3.3.4 and also analyzed via CD spectroscopy.
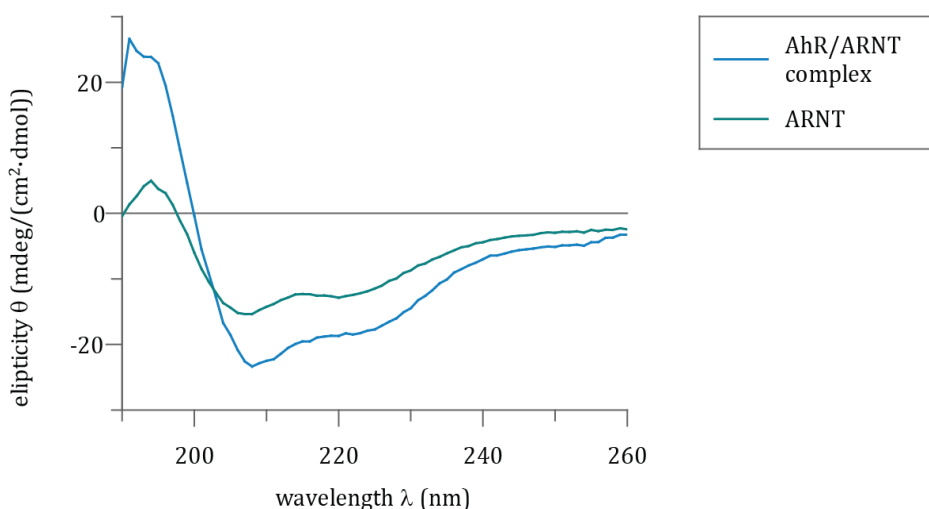
Figure 16: Circular dichroism spectra from 190 – 260 nm of the purified AhR/ARNT complex and of ARNT protein alone. The ellipticity is plotted against the wavelength. For explanation see text.

The CD-spectra indicate (see **Figure 16**) that both the AhR/ARNT complex and ARNT alone are properly folded. Minima at λ= 208 nm and λ= 222 nm, characteristic for the α-helical content, are seen for both samples. The deconvolution of the spectra, calculated with the CDNN software, revealed the secondary structure composition of the AhR/ARNT

complex and ARNT (see Table 1). The striking difference in α-helical content between AhR/ARNT complex (33.4%) and ARNT alone (23.3%) led to the hypothesis that in the AhR/ARNT complex, the bHLH domains of AhR and ARNT folded properly into a four helical bundle, whereas ARNT is a monomer in solution and thus does not form this four helical bundle.

Most importantly, these results indicate that AhR within the AhR/ARNT complex is properly folded as there are no larger unstructured regions detected compared to ARNT alone.

	AhR/ARNT complex	ARNT
α-helices	33.4	23.3
antiparallel β-sheets	10.3	17.3
parallel β-sheets	7.9	10.0
β-turns	17.3	19.7
coiled-coil	28.3	23.6
unstructured	2.8	6.1

Table 1: Calculated secondary structure compositions for the AhR/ARNT complex and ARNT in (%). Calculations were done with CDNN. For explanation see text.

4.4   ITC measurements

In order to evaluate DNA binding and quantitatively determine the affinity of purified AhR/ARNT complex towards double stranded DNA (dsDNA), ITC experiments were carried out (see 3.3.8). Based on the experiments done in [135], dsDNA containing the identified DRE element 5'-GCGTG-3' flanked with up to six further bases on each site was used. The list of dsDNA sequences used and representative ITC data are shown in **Figure 17**. Binding constants are listed in **Table 2**.
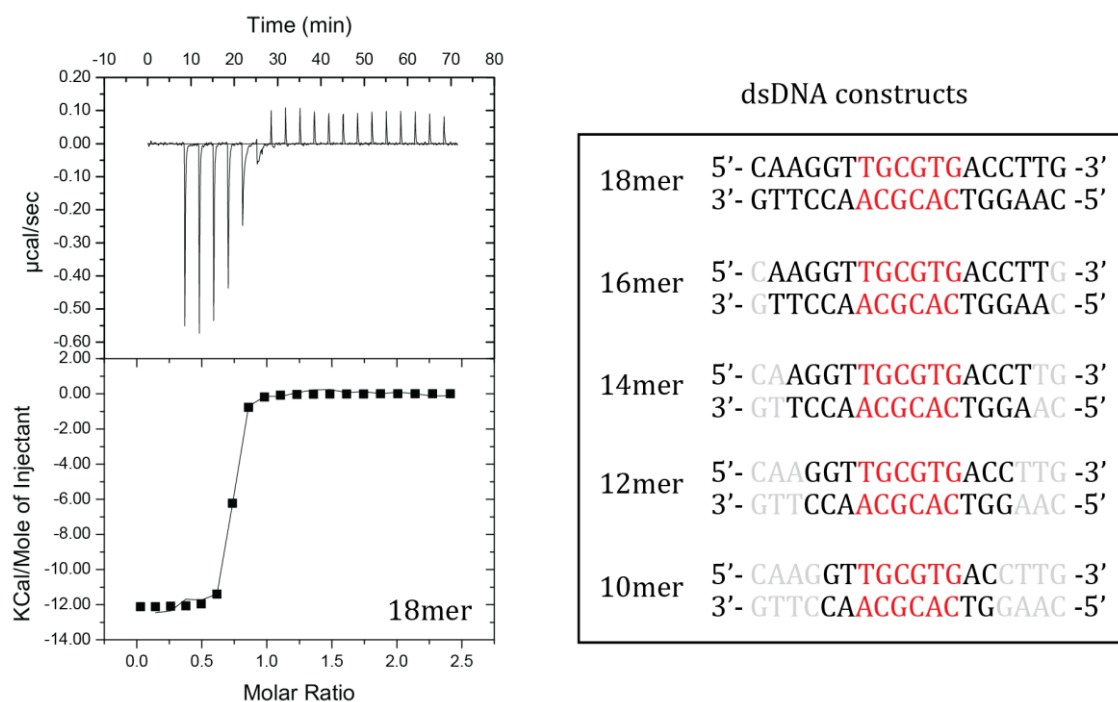
Figure 17: ITC experiments. DNA binding affinities of different dsDNA constructs were determined by isothermal titration calorimetry. Left: ITC measurement of the 18mer DNA (435 μM), which was titrated to the AhR/ARNT complex (40 μM) Right: schematic view of used blunt end dsDNA constructs. Red: XRE element, grey: bases which are not present in the dsDNA construct compared to the 18mer.

As can be seen in **Figure 17** and **Table 2**, the 18mer dsDNA construct bound to the AhR/ARNT complex with high affinity (38 nM). In order to analyze the necessity of the DNA bases flanking the XRE element for binding, the 18mer was shortened sequentially (see **Figure 17** and **Figure 18**) and the K_D to the AhR/ARNT complex was determined for each dsDNA construct via ITC measurements. Interestingly, the K_D remained approximately constant in the range of ~ 100 nM if titrating the 18mer – 12mer to the AhR/ARNT complex.

DNA construct	K_D (nM)	N
18mer	38 ± 10	0.7
16mer	104 ± 26	0.7
14mer	131 ± 40	0.7
12mer	93 ± 28	0.7
10mer	787 ± 70	0.9

Table 2: Experimentally determined dissociation constants (K_D) of different dsDNA constructs and stoichiometry (N) of DNA/complex formation.

However, further shortening of the 12mer dsDNA to a 10mer dsDNA led to a drop in the binding affinity (e.g. increase of the K_D; see **Table 2**). The K_D of the 10mer to the

AhR/ARNT complex is 8-fold higher than the K_D of the 12mer to the AhR/ARNT complex. This indicates that bases flanking the XRE element, beyond the 12mer, do not influence the binding affinity.

The stoichiometry (N) of binding was also calculated from ITC experiments. N describes how many molecules of the dsDNA construct bind to one molecule of the AhR/ARNT complex. For different dsDNA constructs, values of N= 0.7-0.9 were obtained (see **Table 2**). Given that one molecule AhR/ARNT complex binds one dsDNA, a stoichiometry N= 1 was expected.
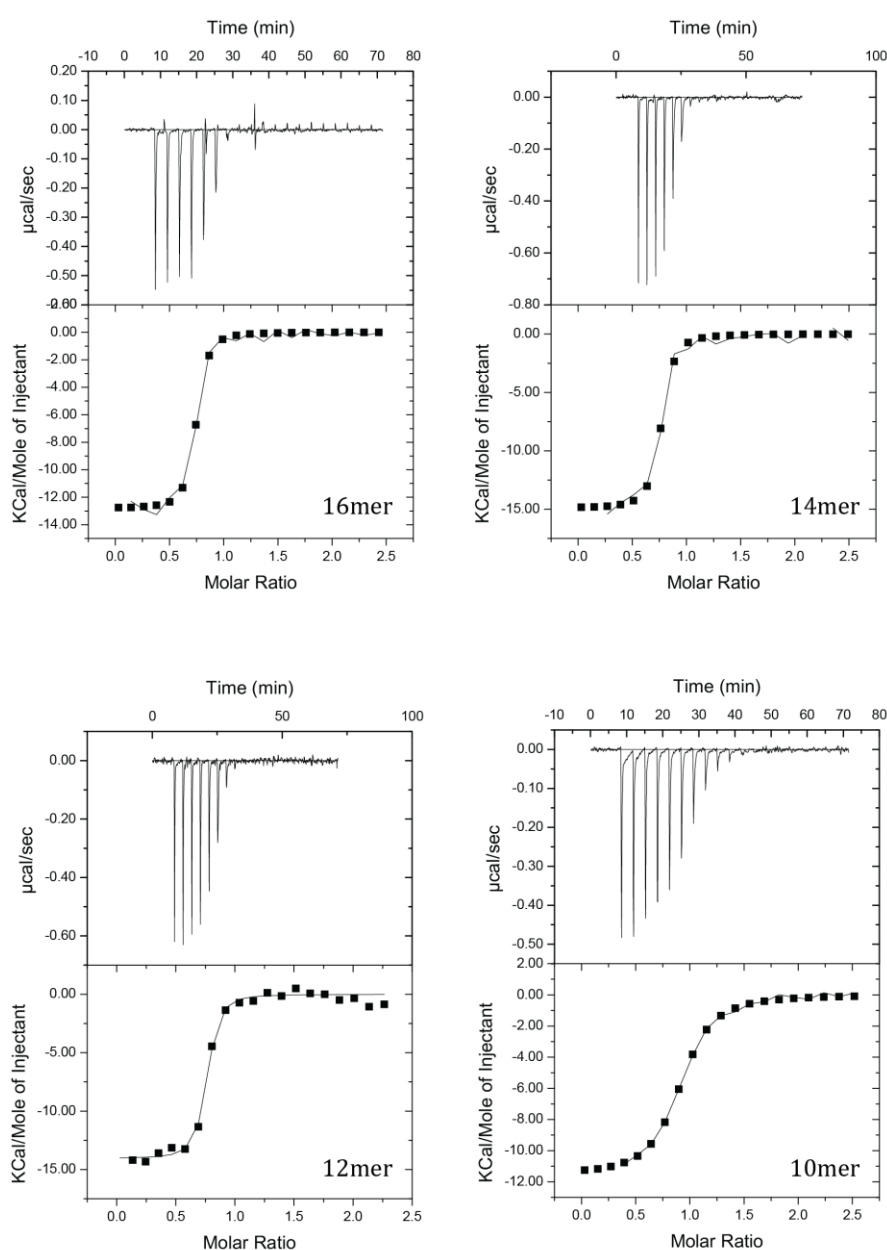
Figure 18: ITC experiments. DNA binding affinities of different dsDNA constructs were determined by isothermal titration calorimetry. ITC measurements of the 16mer, 14mer, 12mer and 10mer vs. the AhR/ARNT complex are shown.

These deviations can most probably be ascribed to slight errors in the calculation of the molar concentration of the AhR/ARNT complex and/or the dsDNA. These calculations were based on the assumption that the AhR/ARNT proteins form a 1:1 complex and that the single stranded DNA fragments hybridize 1:1 without any excess of either binding partner. Nevertheless, a high affinity binding of dsDNA comprising the DRE sequence could be identified.

Moreover, the observed dsDNA binding confirms that the AhR/ARNT complex formation is functional. Furthermore, this result emphasizes that DNA binding to the complex requires neither the PAS B domain nor the unstructured C-terminal half of the proteins. Nevertheless, further methods will have to be applied to confirm the stoichiometry of the AhR/ARNT complex bound to dsDNA.

4.5   RALS measurements

Via RALS measurements, the molecular weight of proteins or protein-DNA complexes can be calculated. In order to determine the stoichiometry of binding more accurately, RALS measurements were performed with the AhR/ARNT complex alone and in combination with the 12mer dsDNA as described in 3.3.9. For the apo AhR/ARNT complex, a molecular weight of 55.86 kDa was calculated from the experimental data using the OmniSec software. Assuming a 1:1 stoichiometry, the theoretical molecular weight of the AhR/ARNT complex is 55.09 kDa. This result underlines that the AhR/ARNT complex comprises one molecule AhR (26 kDa) and one molecule ARNT (29 kDa), forming a stable heterodimeric complex also in absence of dsDNA (see **Figure 19**).

To load the complex with DNA, the AhR/ARNT complex was pre-incubated 30 min on ice with a slight molar excess of the 12mer dsDNA prior to RALS measurement. The experimentally determined mean molecular weight was 66.54 kDa (see **Figure 20**).

The theoretical molecular weight of one AhR/ARNT complex (55.09 kDa) bound to one 12mer dsDNA is 62.38 kDa. The theoretical molecular weight and the experimentally determined molecular weight were in good agreement, indicating that the AhR/ARNT complex establishes a stable interaction with one molecule 12mer dsDNA.
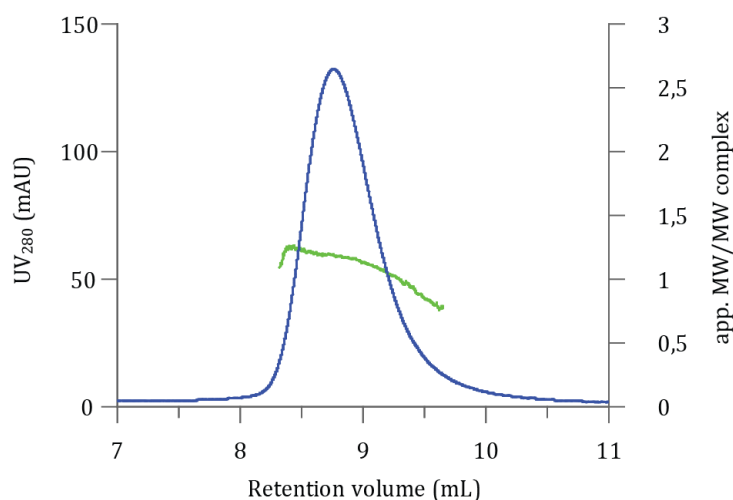
Figure 19: RALS data of the AhR/ARNT complex. The blue curves indicate the UV absorbance (see left y-axis). The calculated relative molecular masses as determined by RALS are shown in green (see right y-axis). App. MW / MW complex corresponds to the quotient of the apparent molecular mass divided by the molecular weight of the AhR/ARNT complex. For explanation see text.

Altogether, the RALS measurements underlined that the AhR/ARNT complex itself forms a heterodimer with a 1:1 stoichiometry and that this heterodimeric complex, in turn, forms a stable interaction with the 12mer dsDNA construct in a 1:1 stoichiometry, suitable for crystallization.
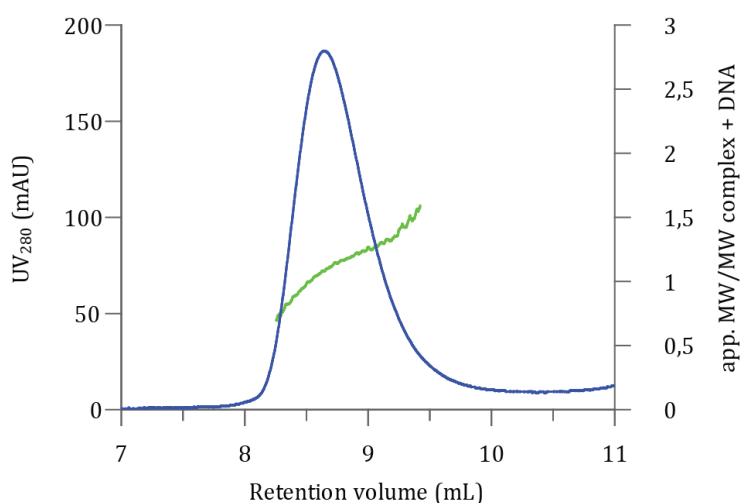
Figure 20: RALS data of the AhR/ARNT complex bound to 12mer dsDNA. The blue curves indicate the UV absorbance (see left y-axis). The calculated relative molecular masses as determined by RALS are shown in green (see right y-axis). app. MW / MW complex + DNA corresponds to the quotient of the apparent molecular mass divided by the theoretical molecular weight of the AhR/ARNT complex bound to 12mer dsDNA. For explanation see text

4.6   Structural analysis of the AhR/ARNT complex

4.6.1   Crystallization

To obtain insights into the atomic structure of the AhR/ARNT complex, purified AhR/ARNT complex comprising the bHLH and the PAS A domain of AhR and ARNT respectively (cohsAhR 23-394; mmARNT 85-360 Δ274-303, C256S) were subjected to crystallization trials. Despite extensive screening efforts, no crystals were obtained for the apo AhR/ARNT complex.

4.6.2   Sample preparation and crystallization of the AhR/ARNT complex bound to dsDNA

Since crystallization trials for the apo AhR/ARNT complex were unsuccessful, crystallization of the complex bound to DNA was attempted. To this end, the AhR/ARNT complex was incubated with a set of double stranded DNA (dsDNA) constructs (18mer – 10mer) with blunt ends (see **Figure 17**) for 30 min on ice and subsequent preparative gel filtration was used to remove the excess of dsDNA. The gel filtration peak containing the AhR/ARNT complex bound to dsDNA was pooled and concentrated to 8 mg/mL, suitable for crystallization (see **Figure 21**).
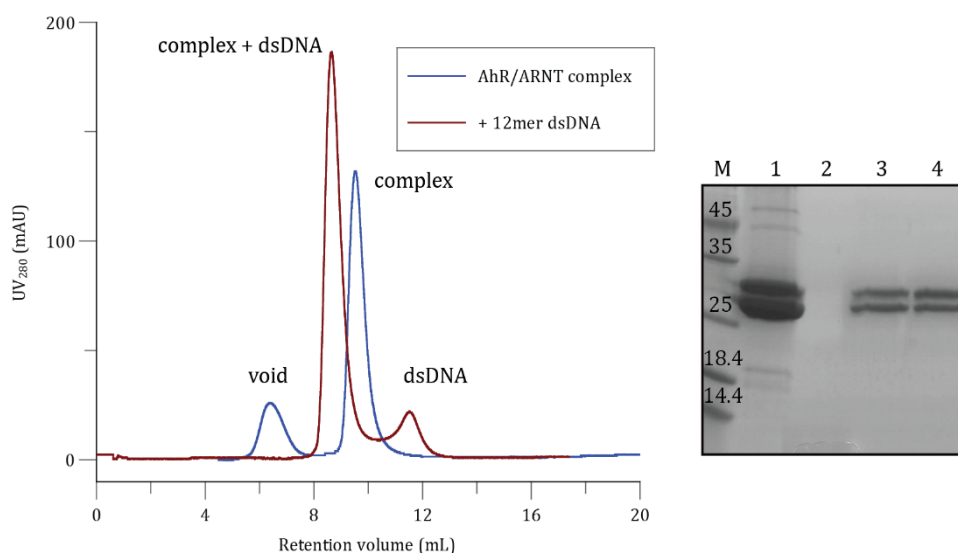
Figure 21: Sample preparation of the AhR/ARNT complex bound to dsDNA. Left: SEC elution profile of the AhR/ARNT complex with/without 12mer dsDNA on a S75 10/300. Right: SDS-gel of the preparative gel filtration. Lane 1: Input AhR/ARNT complex, lane 2: peak fraction of 12mer dsDNA, lane 3: peak fraction of AhR/ARNT complex, lane 4: peak fraction of AhR/ARNT complex with 12mer dsDNA.

The AhR/ARNT complex bound to the different dsDNA constructs was subjected to crystallization trials. Interestingly, crystal hits could only be obtained at 4°C and only if bound to the 12mer dsDNA. In total, 14 crystallization conditions revealed crystals, most of them appearing within 3-5 days after crystallization set up. Three crystallization conditions could be reproduced and where further optimized due to robustness and size of the crystals (see **Figure 22**). Optimization lead to crystals with a maximum size of 50 µm x 20 µm x 300 µm. The most important parameters were found to be the drop size (3µL) and the ratio of protein solution (2µL) to crystallization solution (1µL) within the crystallization drop.

initial crystal hits in 96 well format (dropsize 300 nL)

| 20% (w/v) PEG 3350 200 mM ammoniume formate pH 6.6 | 20% (w/v) PEG 3350 200 mM ammonium formate pH 6.3 | 20% (w/v) PEG 8000 10 mM calcium chloride 10 mM Tris-HCl pH 7.5 |

24 well hand screen
dropsize 3 µL
18% (w/v) PEG 3350
230 mM ammonuim formate
pH 6.6
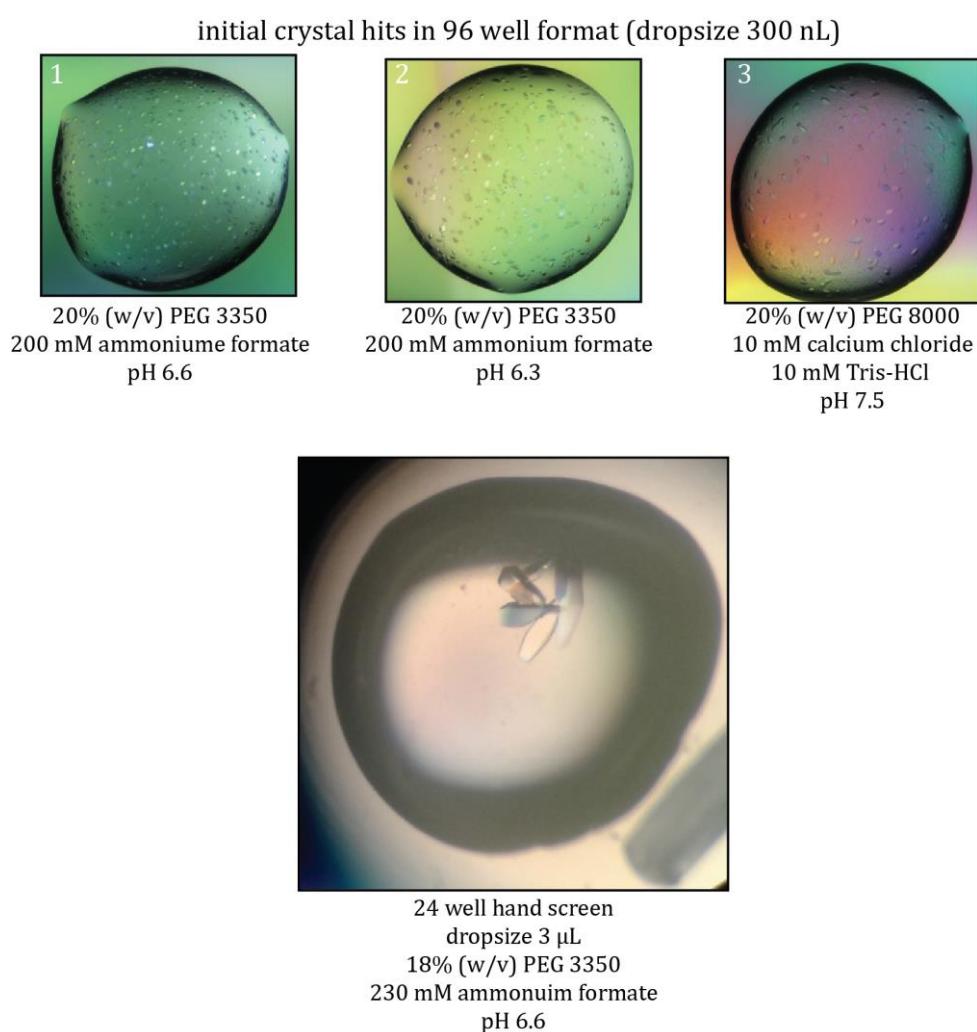
Figure 22: Crystals of the AhR/ARNT complex bound to the 12mer dsDNA. Top: Initial hits 1-3 which could be reproduced successfully. Bottom: Further optimization of the crystallization conditions lead to much bigger and sharper crystals in the 24 well hand screen with a maximum size of 50 µm x 20 µm x 300 µm.

To reduce radiation damage during X-ray exposure, intensive screening for cryo-conditions was done as described in 3.4.2. Only increasing the concentration of PEG, which was already a component of the crystallization conditions 1-3 (PEG 3350 and PEG 8000, respectively), was suitable for cryo protection. The finally optimized cryo conditions are listed in **Table 3** and led to diffraction of the crystals up to 8 Å.

optimized cryo ingredients	crystallization condition		
	1	**2**	**3**
HEPES pH 8.0 (mM)	25		
NaCl (mM)	200		
DTT (mM)	2		
PEG 3350 w/v (%)	35	35	
PEG 8000 w/v (%)			30
calcium chloride (mM)			5
ammonium formate (mM)	230	230	
Tris-HCl pH 7.5 (mM)			10

Table 3: Optimized cryo protection conditions for the crystallization conditions 1-3 (see **Figure 22**) which were used for the cryo protection and dehydration of the crystals as described in the text.

Slight dehydration of the crystals was attempted in order to increase the order in the crystal lattice and thus to increase the resolution the crystals diffract to. Crystals were dehydrated over 0.5-18 hrs at 4°C. Dehydration was carried out by incubating the crystals in the optimized cryo solution as hanging drop. This hanging drop was incubated in a closed environment over the initial crystallization condition plus 400 mM NaCl. The latter should support the hygroscopic effect. After an optimized dehydration time of 1.5 hrs, crystals were directly mounted into cryo-loops and flash frozen.

Screening of nearly 1000 crystals resulted in approximately 5 crystals diffracting to 4 Å and better. Native datasets were collected at BESSY II BL 14.1 in Berlin (for data collection strategy see 3.4.3). The data statistics for the best native dataset can be seen in **Table 4**. The AhR/ARNT complex crystallized in a hexagonal space group $P6_122$ (178) or $P6_522$ (179) with the cell dimensions a= 92.31Å, b= 92.31Å, c= 461.92Å and the crystal diffracted up to 3.5 Å.

Data collection	native data set
Beamline	BESSY II BL 14.1
Wavelength [Å]	0.918
Space group	$P6_522$
Cell dimensions	
a, b, c [Å]	92.31
	92.31
	461.92
α, β, γ [°]	90.00, 90.00, 120.00
Resolution [Å][a]	50.00 (3.71)- 3.5
R_{meas} (%)	25.0 (97.8)
I/σ	8.12 (1.86)
Completeness (%)	96.8 (82.7)
Redundancy	8.7 (7.6)

Table 4: Data collection statistics of the best native datasets of the AhR/ARNT complex bound to 12mer dsDNA. [a] Numbers in parentheses apply to the highest resolution shell.

The major problem with the data obtained from these crystals is the anisotropic diffraction which can be seen in the diffraction images and in the data statistics (see **Figure 23** and **Table 4**, completeness in highest resolution shell). Diffraction anisotropy was analyzed (services.mbi.ucla.edu/anisoscale/) and the mtz-file was ellipsoidally truncated and scaled. The predicted diffraction resolution limits in the directions of the three crystal axes were a= 3.5 Å; b= 4.3 Å; c= 4.5 Å.
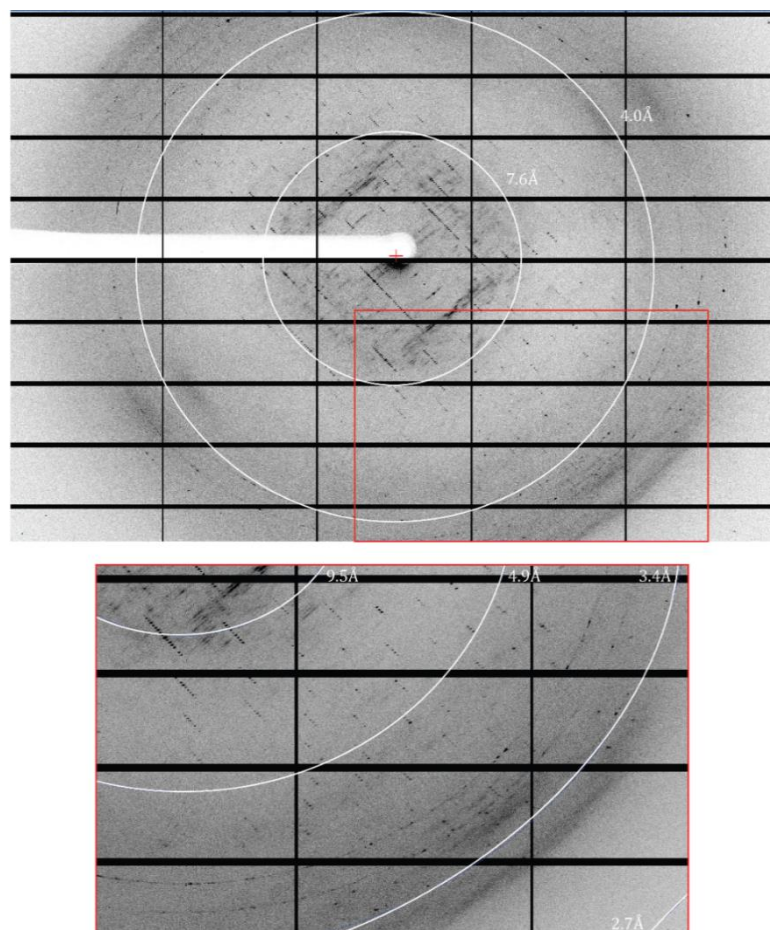
Figure 23: Diffraction image obtained from AhR/ARNT complex crystals on a Pilatus 6M detector. Oscillation range 1°, maximum resolution 3.5 Å. Close up (red box) shows the anisotropic diffraction up to 3.5 Å.

4.6.3   Molecular replacement

Using the collected native dataset, solving the structure via molecular replacement (MR) was attempted. Three differently scaled data sets were used as starting points.

- Data set scaled to 4.2Å (completeness 99.5%, R_{meas} 66.4%)
- Data set scaled to 3.5Å (completeness 82.7%, R_{meas} 97.8%)
- Data set ellipsoidally scaled to a= 3.5 Å ; b= 4.3 Å ; c= 4.5 Å

Furthermore, the hexagonal space groups with the six-fold screw axis $P6_1$, $P6_5$, $P6_122$, $P6_522$ were tested. Several available crystal structures were identified as potential starting models for molecular replacement. The crystal structures of the homologous transcription factor mouse CLOCK/BMAL1 (PDB: 4F3L), as well as of the PAS A domain of mouse AhR (PDB: 4M4X) and of the bHLH domains of human CLOCK and BMAL1 bound to a dsDNA (PDB: 4H10) were used in that regard (see **Figure 24**). The sequence

similarity of mouse ARNT and its homolog mouse BMAL1 within the bHLH and PAS A domains is 92% and the sequence similarity between the bHLH of human AhR and its homolog human CLOCK is 28%. Together with the crystal structure of the human AhR PAS A domain, these structures are suitable for MR in terms of sequence similarity and homology. They were used as starting models in various combinations, in truncated and in poly-alanine versions for extensive molecular replacement attempts using Phenix MR [136], MolRep [137], MrBump [138] and Balbes [139]. The different programs calculate TFZ scores or Z scores and LLG values in order to rank the probability of whether the MR approach led to a correct solution. The thresholds, variable for each program, which indicate that the MR solution could probably be correct, could not be reached via the used MR attempts. Additionally, the clash score of the proposed solutions was highly increased indicating that the predicted solutions were mistakenly overlapping within the crystal lattice. Subsequent rigid body refinement of the most promising MR solutions did not reveal any extra density or difference density, which could improve the initial model. Altogether, solving the crystal structure of the AhR/ARNT complex bound to 12mer dsDNA via MR with the mentioned crystal structures as starting model was not successful.
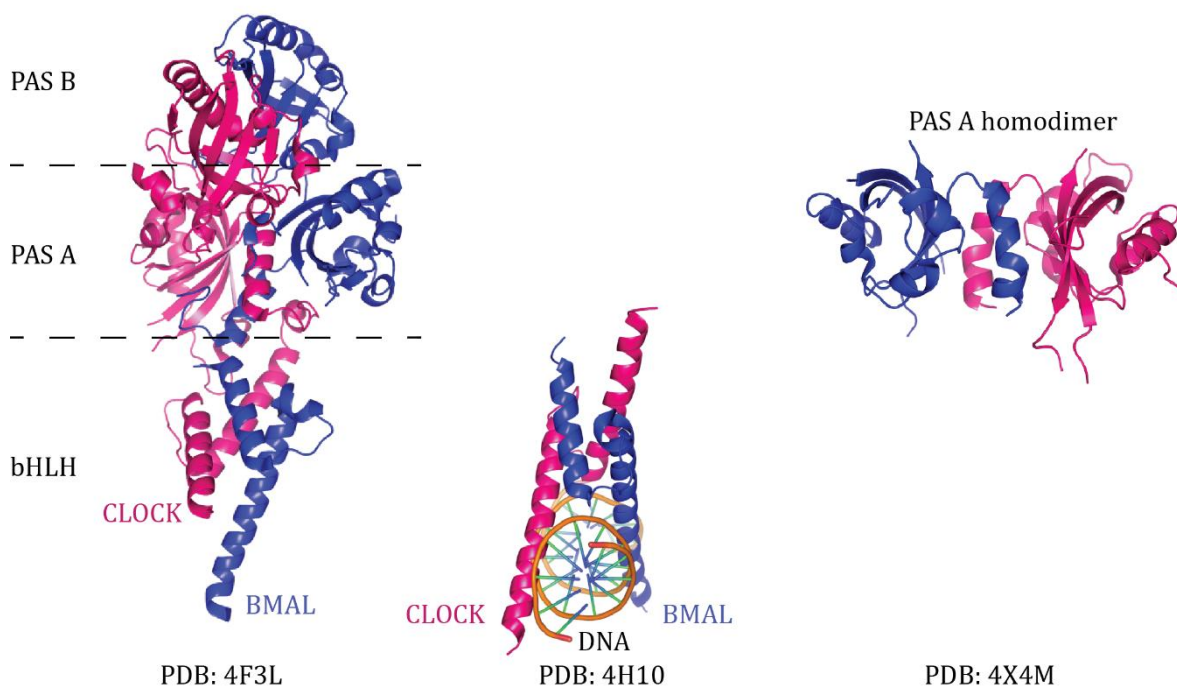
Figure 24: Crystal structures used as initial models for molecular replacement. Heterodimeric transcription factor mouse CLOCK/BMAL (PDB: 4F3L), bHLH of the heterodimeric transcription factor mouse CLOCK/BMAL bound to 14mer dsDNA (PDB: 4H10), PAS A domain homodimer of human AhR (PDB: 4X4M).

One possible reason for this could be that the crystals do not comprise the expected AhR/ARNT complex. In order to analyze this, some crystals were dissolved and loaded on an SDS-gel to check if the AhR/ARNT construct is really part of the crystals and whether degradation occurred. The SDS-gel showed unambiguously that the crystals contained the AhR/ARNT complex without any degradation occuring.

4.6.4 Experimental phasing

Since phasing via MR was not successful, experimental phasing was attempted in order to solve the crystal structure of the AhR/ARNT complex bound to 12mer dsDNA. In the following, the production of crystals of selenomethionine substituted AhR/ARNT complex bound to 12mer dsDNA (see 4.6.4.1) as well as heavy metal soaking of native crystals of AhR/ARNT complex bound to 12mer dsDNA (see 4.6.4.2) will be described. The data obtained from both approaches were combined in the computational analysis (see 4.6.4.3.).

4.6.4.1 Selenomethionine substituted Ahr/ARNT complex

The selenomethionine (SeMet) substituted AhR/ARNT complex was expressed as described in 3.3.3 and purified as in 3.3.4. The yield was decreased to 0.8 mg SeMet substituted protein per liter expression culture. Binding to the 12mer dsDNA, as described in 4.6.2, was successful (see **Figure 25**).
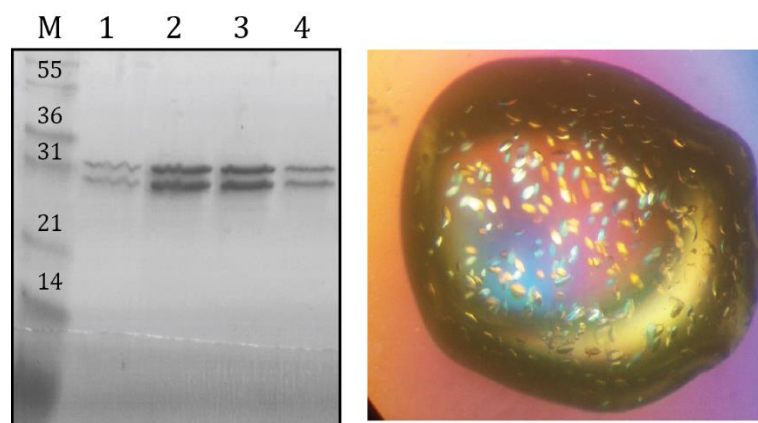
Figure 25: Preparation of SeMet substituted crystals. Left: SDS-Gel from preparative gel filtration S75 (10/300) peak fractions (lane 1-4) from SeMet substituted AhR/ARNT complex bound to 12mer dsDNA. Right: Crystals from SeMet substituted AhR/ARNT complex bound to 12mer dsDNA revealed by strike seeding with a cat whisker in 18% PEG 3350, 230 mM ammonium formate, pH 6.6. for explanation see text.

For re-screening the crystallization conditions, the SeMet substituted AhR/ARNT complex bound to the 12mer dsDNA was subjected to initial crystallization trials without success. Crystallization in the crystallization condition optimized for the native protein crystals (see 4.6.2) failed as well.

SeMet substituted crystals of the AhR/ARNT complex bound to the 12mer dsDNA could only be produced via streak seeding with a cat whisker using the native crystals of the AhR/ARNT complex bound to 12mer dsDNA as seeds (see **Figure 25**). With this approach, SeMet substituted crystals appeared after 6 days in 18% PEG 3350, 230 mM ammonium formate, pH 6.6 in a 24 well hand screen format at 4°C. Crystallization drops were set up by mixing 2 µL reservoir solution with 1.5 µL SeMet substituted protein solution (8.5 mg/mL) and subsequent streak seeding with native crystal seeds.

Cryo-conditions were screened and finally used as in 4.6.2 and slight dehydration (see 4.6.2) of the crystals was deployed in order to increase the diffraction properties of the SeMet substituted crystals of the AhR/ARNT complex bound to 12mer dsDNA.

	SeMet data set
Beamline	BESSY II BL 14.1
Wavelength [Å]	0.976
Space group	$P6_522/P6_122$
Cell dimensions	
a, b, c [Å]	91.86
	91.86
	462.18
α, β, γ [°]	90.00, 90.00, 120.00
Resolution [Å][a]	50.00 (3.91)- 3.69
R_{meas} (%)	34.1 (214.5)
I/σ	6.31 (1.27)
Completeness (%)	95.0 (78.7)
Redundancy	12.8 (7.8)

Table 5: Data collection statistics of the best datasets of the crystals of the SeMet substituted AhR/ARNT complex bound to 12mer dsDNA.
[a] Numbers in parentheses apply to the highest resolution shell.

Data collection was carried out at BESSY II 14.1 and the used wavelength was 0.976 Å in order to detect the anomalous signal of selenium. 3600 images were collected with an

oscillation range of 0.1°, an exposure time of 1s and a detector distance of 650 mm. The best crystal diffracted anisotropically up to 3.7 Å. The space group was $P6_122/P6_522$ with the cell dimensions a=b= 91.86 Å, c= 462.18 Å. The data set was processed anomalously and the anomalous signal could be detected up to 6.8 Å. Further data statistics are summarized in **Table 5**.

4.6.4.2 Heavy atom labelling of the crystals of the AhR/ARNT complex bound to 12mer dsDNA

In a parallel approach, heavy atom labelling of native protein crystals was tested. One strategy was, to bind the AhR/ARNT complex to bromoylated DNA. The 12mer ssDNA was modified to 5'-GUUCACGCAAAC-3' incorporating bromine labelled uracil (U; 5-Bromo-2'-deoxyuridine-5'-triphosphate; see **Figure 26**). The reverse complement ssDNA was designed comparably (3'-CAAGTGCGTUUG-5'). Although binding of the bromoylated 12mer dsDNA to the AhR/ARNT complex (cohsAhR 23-273; mmARNT 85-360, Δ274-303, C256S) could be detected, initial crystallization of the complex failed.
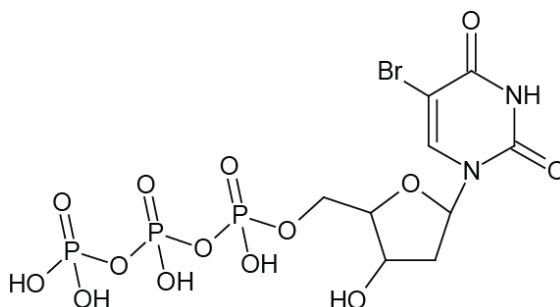
Figure 26: Structural formula of 5-Bromo-2'-deoxyuridine 5'-triphosphate, a bromine substituted uracil, which was incorporated in the ssDNA sequences 5'-GUUCACGCAAAC-3' and 3'-CAAGTGCGTUUG-5'. For explanation see text.

For heavy metal soaking, various derivatives were added to the cryo solution with a concentration ranging from 1-10 mM. Crystals were either incubated in the cryo solution during dehydration for 1.5 hrs at 4°C (see 4.6.2) or soaked for 5-30 s into the cryo solution plus derivative after dehydration. Heavy metal soaking with the following listed compounds either destroyed the crystals during soaking or drastically diminished the diffraction resolution.

- Methylmercury (II) acetate
- Methylmercury(II)chloride
- Ammoniumtetrachloroplatinate

- Potassium iodide
- Hexatantalum tetradecabromide

In the following, the preparation and data collection of the successful soaking approach with 5-amino-2.4.5-triiodoisophtalic acid (I3C) will be described. Native crystals of the AhR/ARNT complex bound to 12mer dsDNA were grown and slight dehydration was carried out as described in 4.6.2. The cryo-condition (see 4.6.2) was supplemented with 20-50 mM I3C. Crystals were dehydrated for 1.5 hrs at 4°C in the supplemented cryo condition and subsequently mounted into cryo-loops. Up to 500 crystals had to be tested in order to find two properly diffracting crystal from which a data set could be collected.

	iodine I data set	iodine II data set
Beamline	BESSY II BL 14.1	
Wavelength [Å]	1.54	
Space group	$P6_522/P6_122$	
Cell dimensions		
a, b, c [Å]	91.22	92.39
	91.22	92.39
	463.58	462.09
α, β, γ [°]	90.00, 90.00, 120.00	
Resolution [Å][a]	50.00 (4.07)- 3.84	50.00 (4.12)- 3.88
R_{meas} (%)	12.1 (176.5)	17.7 (408.8)
I/σ	15.54 (1.92)	9.67 (1.75)
Completeness (%)	99.5 (98.1)	99.4 (97.9)
Redundancy	20.5 (20.1)	18.4 (18.1)

Table 6: Data collection statistics of the best datasets of the crystals of the iodine soaked AhR/ARNT complex bound to 12mer dsDNA. [a] Numbers in parentheses apply to the highest resolution shell.

Data collection of two crystals was carried out at BESSY II BL 14.1 and the used wavelength was 1.54 Å in order to detect the anomalous signal of iodine. 1800 Images were collected with an oscillation rage of 0.2 °, an exposure time of 2 s and a detector distance of 403 mm. The crystals diffracted anisotropically up to 3.8/3.9 Å in the space group $P6_122/P6_522$ with the cell dimensions a=b= 91.22/92.39 Å, c= 462.18/462.09 Å.

The data sets were processed anomalously, and the anomalous signal could be detected up to 6.2 Å. Further information for the data statistics are depicted in **Table 6**.

4.6.4.3   **Computational approaches to solve the phase problem**

Isomorphous replacement or anomalous dispersion and a combination of both can be used for the computational approaches in order to solve the phase problem. SAD (single anomalous dispersion) requires only one dataset collected at the absorption edge of the derivative in order to get initial phases. This approach could be applied for the selenomethionine and for both collected iodine data sets separately.

Due to the fact that the collected native and derivative data sets are approximately isomorphous, anomalous dispersion could be combined with isomorphous replacement. The SIRAS (single isomorphous replacement anomalous dispersion) approach was applied combining one of the anomalous datasets with the collected native datasets. This can be further extended such that anomalous datasets from different derivatives (Selenomethionine and iodine) are combined with a native dataset, which is named MIRAS (multiple isomorphous replacement anomalous dispersion).

The programs shelX [140] , sharp [141] and Phenix.Autosol [136] were used for the different phasing attempts.

For SIRAS/MIRAS phasing, the anomalous iodine data sets I/II diffracting to 3.8/3.9 Å and/or the anomalous Selenomethionine data set diffracting to 3.7 Å were combined with the native data set diffracting to 3.5 Å. Although these phasing attempts combined the most information, phasing did not lead to a clear solution. From the calculated initial experimental phases, no conclusions could be drawn to the crystal structure of the AhR/ARNT complex bound to the 12mer dsDNA.

Furthermore, SAD phasing methods using the anomalous iodine I/II data set diffracting to 3.8/3.9 Å or the anomalous Selenomethionine data set diffracting to 3.7 Å were applied. Only SAD phasing using the anomalous iodine (I) data set with usage of PhenixAutosol combined with density modification led to a promising electron density in space group $P6_522$ that was of sufficient quality to allow fitting of two PAS A domains and one 12mer dsDNA molecule in the asymmetric unit (see **Figure 27**).

Subsequent rigid body refinement with Phenix.refine [136] was carried out. As initial model, the arrangement of the two fitted PAS A domains and the fitted 12mer dsDNA (see **Figure 27**) was used and refined against the native data set (see 4.6.2).
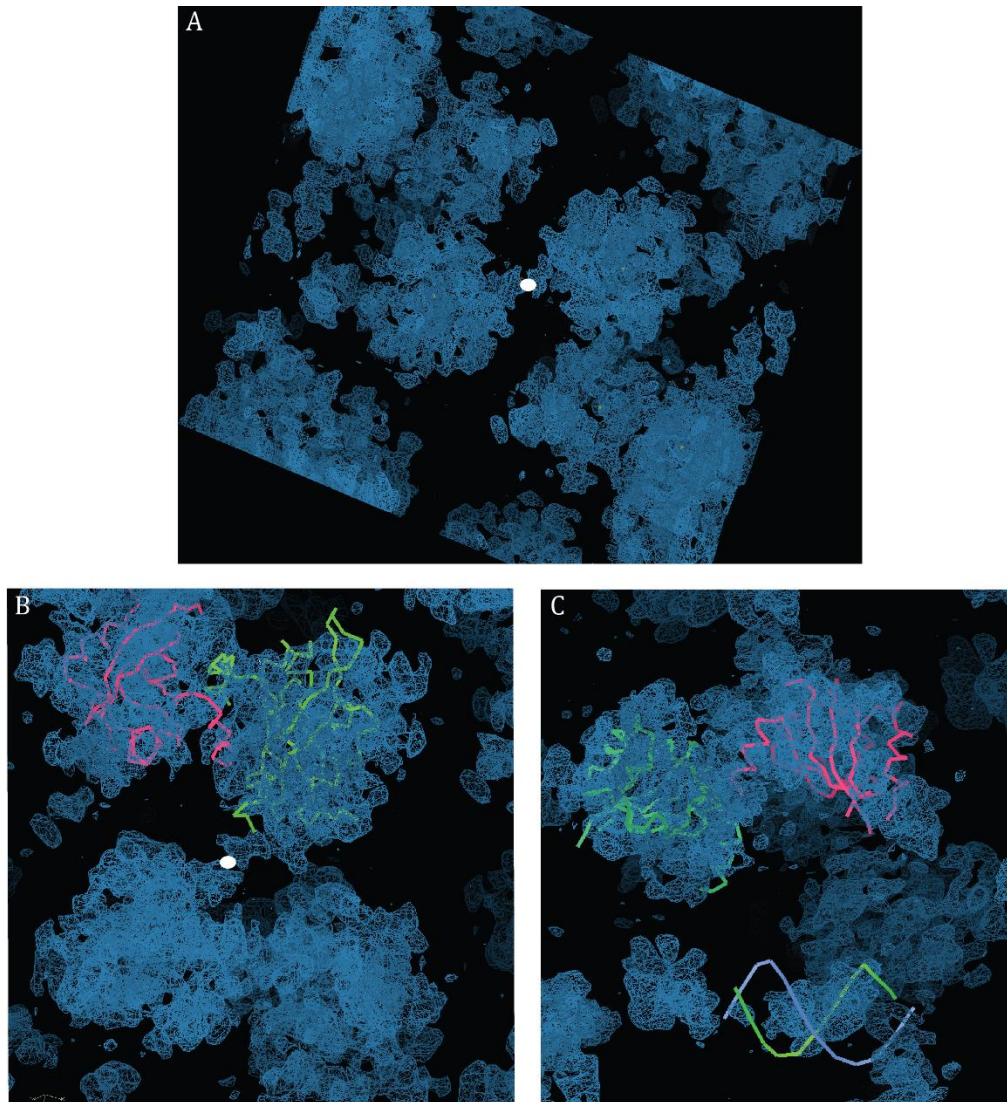
Figure 27: Preliminary electron density. A: Preliminary electron density (2Fo-Fc map; blue, contoured at 1.5 σ) obtained by SAD phasing using the Iodine (I)-SAD data and AutoSol. White dot represents a 2 fold axis. B: Two PAS A domains (green and pink) fitted into the preliminary electron density. Same view as in A. white dot represents the 2 fold axis. C: One DNA molecule could be fitted into additional density in distinct distance to the PAS domains. For explanation see text.

This rigid body refinement revealed some extra electron density suitable to fit two more PAS A domains into the asymmetric unit (see **Figure 28**, red circles).

Subsequent rigid body refinement was applied after fitting the two extra PAS A domains into the extra electron density. The updated model comprised now in total four PAS A domains and one 12mer dsDNA molecule within the asymmetric unit.

However, rigid body refinement of the native data against the extended model revealed no extra electron density or difference density. This would have been necessary to further extend the model by the bHLH domains and one further 12mer dsDNA molecule. Furthermore, the electron density describing the putative PAS A domains was ambiguous. Thus, it was impossible to orient the PAS A domains accurately into the electron density.

One reason for this is the globular PAS domain fold mainly comprising β-strands, which in turn forms a β-sheet. These are often poorly defined in low resolution electron density. Despite extensive crystallographic approaches the crystal structure of the AhR/ARNT complex bound to the 12mer dsDNA could not be solved via experimental phasing.

90 °
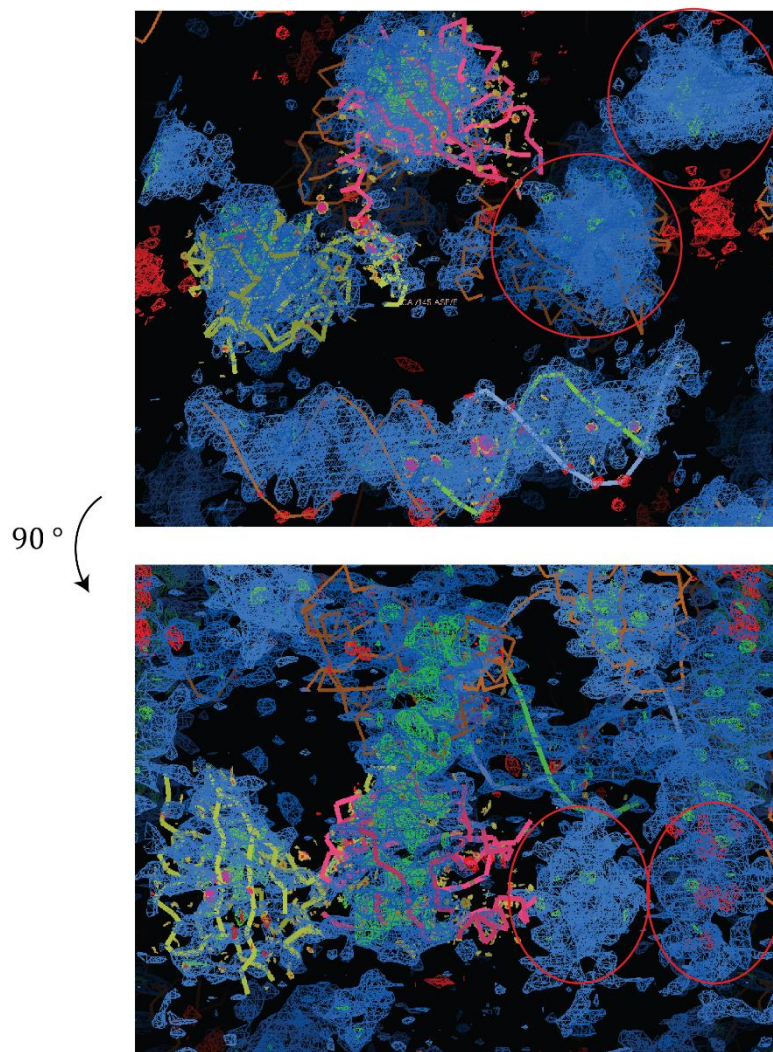
Figure 28: Electron density after rigid body refinement of the native dataset with the arrangement of two PAS A domains (yellow and pink) and one DNA molecule (green/blue). Brownish molecules are the symmetry mates of the initial model. Red circles highlight the extra electron density which appeared after refinement. $2F_0$-F_c map contoured at 1 σ (blue). F_0-F_c map contoured at 2.5 σ (red: negative; green: positive). For explanation see text.

4.6.5   Optimization of the AhR construct

Optimization of the crystals of the AhR/ARNT complex, comprising cohsAhR 23-273 and mmARNT 85-360(Δ274-303, C256S) did not lead to structure solution. Thus, further

construct optimization was done in order to optimize crystal packing and in turn to optimize diffraction properties.
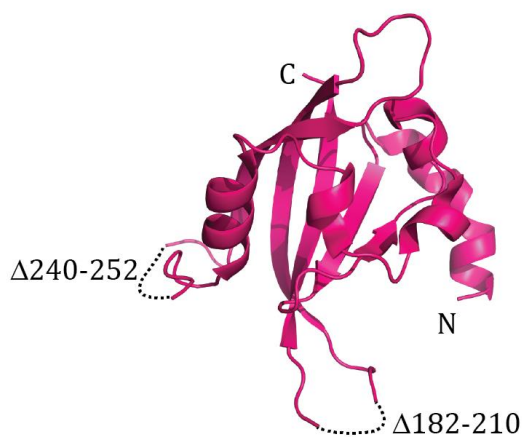
Figure 29: Ribbon representation of the PAS A domain of AhR (PDB: 4M4X; pink). N- and C-terminal end of the construct and the two identified missing loops are labelled.

The recently published crystal structure of the AhR PAS A domain (PDB: 4X4M) was analyzed in detail. It turned out that in this structure two loops (Δ182-210; Δ240-252) within the PAS A domain are not resolved, which indicates that these are unstructured and not necessarily important for PAS A domain folding and crystal formation. The loop Δ182-210 was deleted by site-directed mutagenesis. Subsequent co-expression and co-purification of cohsAhR 23-273 (Δ182-210) in complex with mmARNT 85-360 (Δ274-303, C256S) revealed soluble AhR/ARNT complex. However, initial crystallization trials of the apo complex were unsuccessful.

To this end, the optimized AhR/ARNT complex was incubated with a set of double stranded DNA constructs (18mer – 10mer) with blunt ends (see **Figure 17**) and prepared for crystallization as in 4.6.2. Although binding of the dsDNA constructs (18mer – 10mer) to the optimized AhR/ARNT complex was successful, initial crystallization did not reveal crystal hits.

Altogether, the optimization of the protein construct has been investigated without success regarding the crystallization attempts.

4.6.6   Optimization of the dsDNA constructs

So far, only blunt end dsDNA constructs (18mer-12mer) were used for the crystallization attempts. As previous results within this study showed already that the dsDNA length is

critical for crystallization of the AhR/ARNT complex (see 4.6.2), further screening of different dsDNA constructs was performed.

The sticky end dsDNA constructs were designed based on the 12mer sequence dsDNA 5'-GGTTGCGTGACC-3', which was used for successful crystallization trials (see 4.4). In order to improve possible crystallization contacts via the dsDNA, these were designed with one or two bases overhang. Furthermore, the 3' end of the blunt end 12mer dsDNA constructs used for successful crystallization trials were varied (see **Figure 30** and compare **Figure 17**) such that the 5'-GGTTGCGTGACC-3' was exchanged by a 5'- GGTTGCGTGAGG-3'. Corresponding to that, the complementary DNA strand was varied at the 5' end. Finally, a 24mer dsDNA was designed, which comprises twice the 12mer dsDNA sequence 5'-GGTTGCGTGACC-3'.

Figure 30: Double strand DNA constructs with blunt ends or sticky ends with one or two bases overhang. Construct design based on 12mer dsDNA blunt end, which was used for previous crystallization trials. Red: DRE binding motif. For explanation see text.

The dsDNA constructs were prepared as described in 3.3.8 and added to the AhR/ARNT complex (cohsAhR 23-273; mmARNT 85-360, Δ274-303, C256S) in a slight molar excess. Subsequent preparative SEC was carried out. The elution profile shows the AhR/ARNT complex bound to the 12mer blunt end dsDNA eluting at a retention volume of 9.2 mL and the excess of dsDNA eluting at a retention volume of 11.7 mL (see **Figure 30**). In the first peak, the ratio of the absorption of UV light at a wavelength of 260 and 280 nm was ~1.5. Previous experiments showed, that successful binding of dsDNA to the AhR/ARNT complex was obtained by exactly this UV_{260}/UV_{280} ratio.
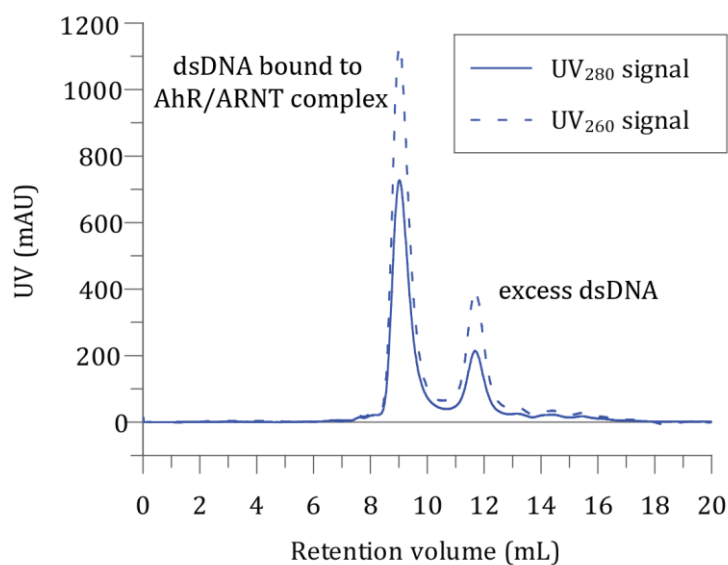
Figure 31: SEC elution profile of the AhR/ARNT complex bound to the 12mer blunt end dsDNA as described in **Figure 30**. Superdex S75 10/300 was used. For explanation see text.

Thus, the SEC experiments served as evaluation of the binding of the designed dsDNA constructs to the AhR/ARNT complex and, secondly, as purification step prior to the crystallization trials. Except for the 10mer sticky end dsDNA, binding could be observed for all designed dsDNA constructs listed in **Figure 31**. After the SEC, the AhR/ARNT complex bound to the dsDNA constructs was concentrated to 9 mg/mL and subjected to initial crystallization trials. No crystal hits could be obtained, which underlines the hypothesis that the dsDNA seems to play a role in forming crystallization contacts and thus is critical for AhR/ARNT complex crystallization.

5   **Discussion and Outlook**

This study describes the successful over-expression and purification of the bHLH and PAS A domain of AhR, achieved by co-expression with its most important interaction partner ARNT, comprising also the bHLH and PAS A domain.

The purified AhR/ARNT complex (cohsAhR 23-273; mmARNT 85-360, Δ274-303, C256S) was shown to be properly folded and to bind dsDNA comprising the DRE motif with very high affinity. Other groups published that the basic region of AhR (residue 1-26) is crucial for DNA binding. With cell-based assays, it was shown, that mutations of residue R14 and R15 reduce the affinity to DNA 10-fold[142] and that residue Y9 and especially its phosphorylation is necessary for DNA binding[143]. Here, the *in vitro* analyses shows, that this residues 1-22 (not included in construct) of cohsAhR are not necessary for DNA binding. Thus, the basic region of AhR is maybe necessary for recruiting further interaction partners in order to enhance the probability of DNA binding within the cells.

Additionally, I could show by ITC experiments that the additional base pair flanking each site of the 12mer dsDNA (compared to the 10mer dsDNA; see **Figure 17**) seem to strengthen the DNA affinity. The affinity to the 10mer dsDNA towards the AhR/ARNT complex drops 8-fold compared to the affinity of the 12mer DNA to the AhR/ARNT complex (see 4.4). In order to analyze whether the base pair on the 5' or 3' end mediates the strengthening of affinity, further ITC experiments could be carried out with distinct 11mer dsDNA constructs.

Whether the PAS B domain of AhR and ARNT can further increase or decrease the binding affinity to the DRE motif remains elusive. However, it could be shown via the ITC experiments that the PAS B domain (not included in the constructs of the AhR/ARNT complex) is not necessary for DNA binding.

Furthermore, this study shows that the purified AhR/ARNT complex bound to the 12mer dsDNA can be crystallized. After extensive optimization of the crystallization condition, the cryo condition and the dehydration condition, the crystals diffract anisotropically to 3.5 Å. Molecular replacement attempts to uncover the structure of the complex were unsuccessful, although the sequence similarity of the structures used for molecular replacement (see 4.6.3) compared to the AhR/ARNT complex were ≥ 85%. This allows the assumption that molecular replacement was not successful because of the borderline quality of the measured datasets in regard to the achieved resolution. Secondly, it is

conceivable that the crystal structures used as initial models for molecular replacement did not reflect the crystalized AhR/ARNT complex correctly.
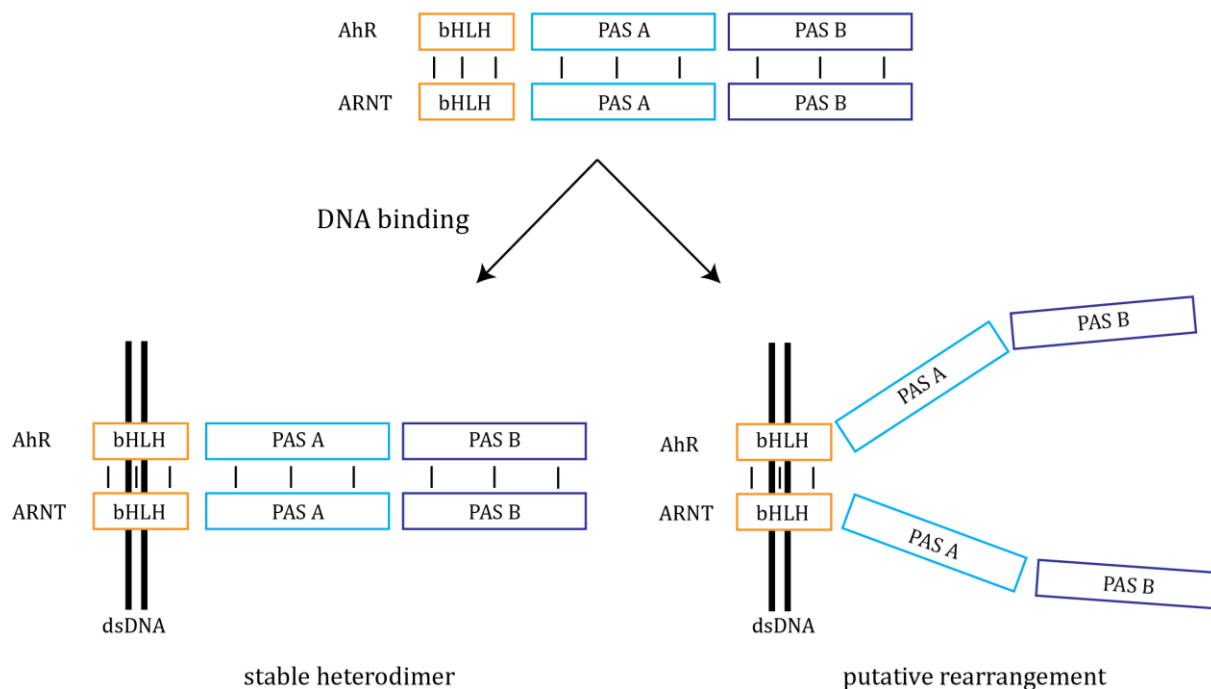
Figure 32: Model of rearrangement of the AhR/ARNT complex upon DNA binding. Top: The AhR/ARNT complex without DNA. The AhR/ARNT complex could stay stable upon dsDNA binding (bottom left) or the dimerization between the PAS A and PAS B domain breaks (bottom right) in order to facilitate the heterodimer to fall apart and thus to terminate transcriptional activation. Black lines indicate the interaction between the distinct domains of AhR and ARNT. For explanation see text.

Upon 12mer dsDNA binding some rearrangements within the AhR/ARNT complex could occur that are not predicted in the models used for molecular replacement. The structure of the homologous transcription factor complex of mouse CLOCK/BMAL1 (PDB: 4F3L) showed a heterodimerization which spans over the bHLH, PAS A and PAS B domain (see **Figure 24**). However, this heterodimer has to fall apart if the transcriptional activation is terminated. Up to now, no exact mechanism of AhR/ARNT complex disruption was proposed. Nevertheless, it is conceivable that upon DNA binding, structural rearrangements are introduced, which cause the interactions between the PAS folds of both proteins to break. This, in turn, would facilitate binding of other proteins, e.g. the repressor AhRR, and support the disruption of the complex resulting in termination of transcription.

In order to clarify this assumption, experimental phasing was used to uncover the structure of the AhR/ARNT complex bound to the 12mer dsDNA. Although several phasing attempts were carried out, only native crystals of the AhR/ARNT complex bound

to 12mer dsDNA soaked with 5-amino-2.4.5-triiodoisophthalic and crystals of the selenomethionine substituted AhR/ARNT complex bound to 12mer dsDNA provided anomalous data. Here, the anomalous signal was detected to a resolution of 6.2 Å, which is reasonable considering the maximum diffraction of the crystals, but borderline for experimental phasing in general. However, the structure could not be solved via the various experimental phasing approaches. Further screening for other anomalous scatterer should be attempted to succeed in the experimental phasing.

Several approaches other than crystallography can be used to obtain structural information of the AhR/ARNT complex and to answer the question of rearrangements within the complex upon DNA binding. Initial cross-linking experiments were already carried out in order to detect the dimerization interface of the AhR/ARNT complex and to reveal structural rearrangements upon DNA binding. Results of preliminary cross-linking experiments showed that cross-links are established all over the heterodimer within distances of up to 40 Å, calculated on a homology model of the AhR/ARNT complex based on the mouse CLOCK/BMAL1 heterodimer (PDB: 4F3L), although the used linker was only 12 Å in length (see **Figure 33**). These results could indicate that either the AhR/ARNT complex unfolds upon cross-linking or that cross-links between different protein complexes were observed or that the structure of the AhR/ARNT complex is completely different as predicted by the homology model.

However, further optimizations of this experimental set-up have to be performed as such that, on one hand, the AhR/ARNT complex stays stable under crosslinking conditions and that, on the other hand, only intramolecular cross-links are established.

NMR-studies could be a second approach to determine the dimerization interface of the AhR/ARNT complex and possible changes within this interface upon DNA binding. First expression trials of the AhR/ARNT complex in minimal medium supplemented with $^{15}NH_4Cl$ were already carried out, but protein yields have to be improved in order to produce the amount of isotope labelled AhR/ARNT complex needed for initial NMR studies.

This study could not offer insights regarding the ligand binding properties of the AhR PAS B domain, because the latter turned out to be unstable and unfolded to a large extent. A variety of attempts to overcome this stability problems failed (see 4.1). Based on the current knowledge of this study, over-expression of the AhR/ARNT complex (cohsAhR 23-395; mmARNT 85-465, Δ274-303, C256S), which was extended by the PAS B domain, was observed but I was not able to purify the complex with or without ligand.
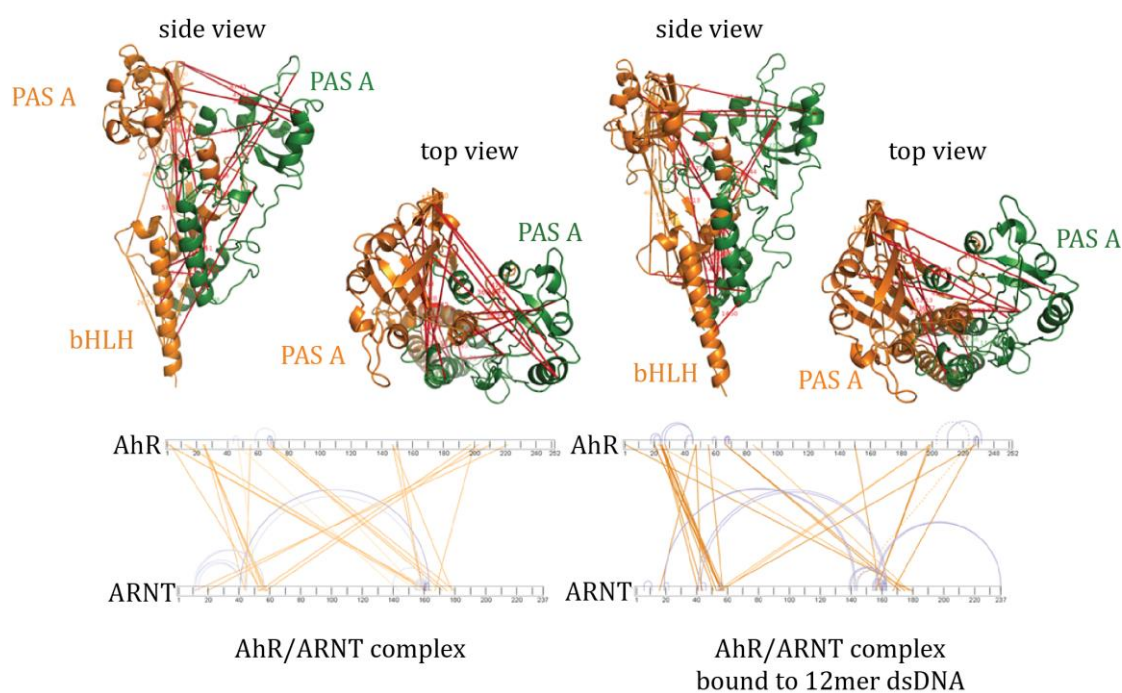
Figure 33: Preliminary cross-linking experiments with the AhR/ARNT complex (cohsAhR 23-273; mmARNT 85-360, Δ274-303, C256S) show cross-linking all over the heterodimer (homology model of the AhR/ARNT complex based on the mouse CLOCK/BMAL1 heterodimer (PDB: 4F3L)). Top: model of the AhR/ARNT complex based on the crystal structure of mouse CLOCK/BMAL1 (PDB: 4F3L). Red lines: Putative intramolecular crosslinks, yellow lines: putative intermolecular crosslinks. Bottom: Schematic view of the AhR/ARNT complex. Observed intramolecular crosslinks are shown as yellow lines. Putative intermolecular crosslinks are shown as grey lines.

In further studies, eukaryotic expression or *in vitro* expression of the AhR/ARNT complex including the AhR and ARNT PAS B domain could help to purify this complex. This could gain insights into the ligand binding to the AhR PAS B domain, the PAS B domain interaction within the complex as well as the influence of the PAS B domain on the DNA binding properties and on potential rearrangements of the AhR/ARNT complex.

Altogether, to my knowledge, the current study describes the first successful approach to purify the AhR/ARNT complex in large amounts for biochemical and structural characterization. The crystallization of the AhR/ARNT complex bound to the 12mer dsDNA was successful and diffraction could be improved after extensive optimizations. Despite all efforts, the structure could not be solved. Further experiments outlined here and initialized already during the study will help to understand the heterodimerization and DNA binding of the AhR/ARNT complex from a structural point of view.

Part II

6   Introduction

6.1   Guanine nucleotide binding and hydrolyzing proteins

Nucleotide triphosphate binding and hydrolyzing proteins (NTPs) are present in almost all living organisms and have in common the mononucleotide-binding fold (P-loop, NTPase fold), which represents 10-18% of all gene products in the cellular organism.[144] One of the seven major monophyletic lineages within the P loop NTPase fold is the Guanosinetriphosphatase (GTPase, GNBP, G protein) superclass.[145] Although involved in a broad variety of cellular functions like signal transduction, nuclear transport, vesicle trafficking, cytoskeleton regulation, endocytosis, translocation of proteins and cell division, the mediating molecular mechanism of guanine nucleotide binding and hydrolyzing is highly conserved.[145,146]

6.1.1   The G domain switch

Most GTPases act as regulatory switches. The interconversion between the ON and the OFF state is determined by nucleotide binding and hydrolysis. When GDP-bound, the G domain of the GTPase is in the OFF-state. Once GDP is exchanged for GTP, the G domain switches to the ON-state. In this state, the GTPase has the highest affinity towards different downstream effector molecules, which it can activate to initiate downstream signaling. The exchange from GDP to GTP is facilitated by auxiliary proteins named Guanine nucleotide exchange factors (GEF). The return to the OFF-state is regulated by GTP hydrolysis (see **Figure 34**). Due to an intrinsically slow GTP hydrolysis, auxiliary proteins named GTPase activating proteins (GAPs) accelerate the GTP hydrolysis reaction.[146-150]

In addition to the small GTPases using this mechanism, a growing number of GTPases have been identified, which are often of higher molecular mass, have a low nucleotide binding affinity and do not require GEFs for activation. Furthermore, they often have a built-in GAP function that is activated upon complex formation.[151-157]

The mechanism of the G domain switch on the molecular level is presented here using the three-dimensional structure of H-Ras p21 (see **Figure 35**). The latter raised great attention due to its mutation dependent promotion of malignant tumor growth and is therefore one of the best studied small GTPases.[158] The G domain of Ras contains 166 amino acids and folds into a central six-stranded β-sheet, surrounded by five α-helices.
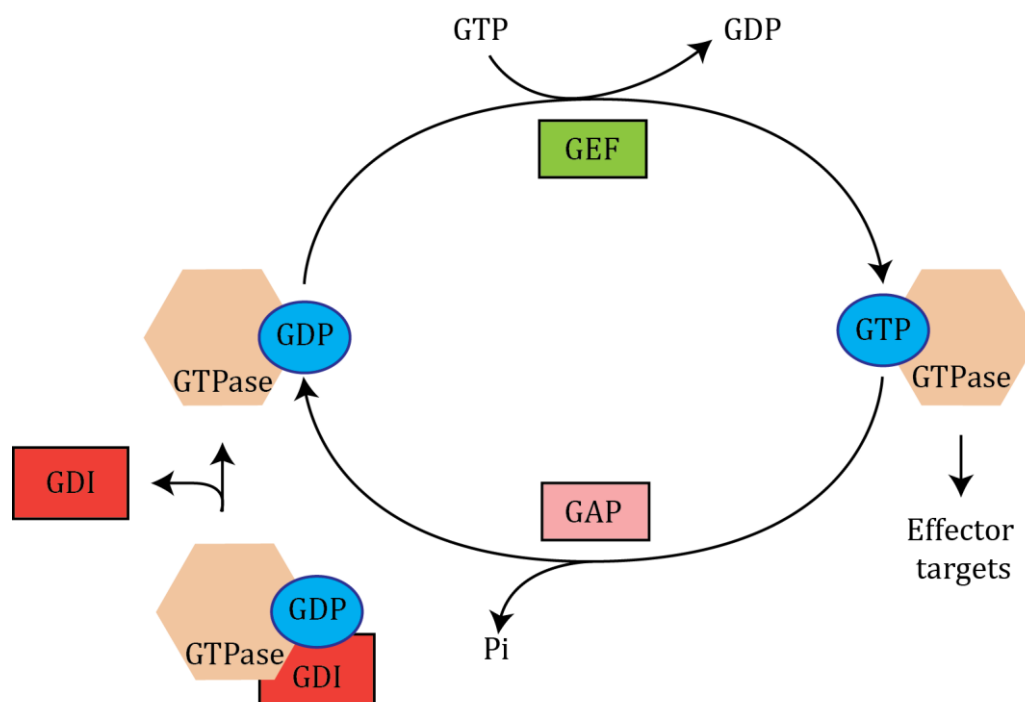
Figure 34: The GTPase cycle. GTPases can cycle between their active GTP-bound state and their inactive GDP-bound state. The active GTP-bound form can interact with effector targets. Regulation of both states is largely accomplished through the auxiliary proteins named GAPs, GEFs and GDIs (guanine nucleotide dissociation inhibitors). Modified from [147].

The G domain provides an environment for nucleotide binding formed by an appropriate spatial arrangement of five loops. These loops comprise the highly conserved sequence motifs G1-G5. The loops between α1 and β2 and between β3 and α2 are particularly important for the conformational rearrangement upon GTP hydrolysis and are called switch I and II. Switch I contains the G2 motif, whereas switch II comprise the G3 motif. Together with the G1 motif, they are involved in contacting the phosphate moiety of the nucleotide.

The G1 motif, also called phosphate-binding motif or Walker A motif, with the consensus sequence GxxxxGK(S/T) (Ras: ^{10}GAGGVGKS^{17}) interacts with the β- and γ-phosphate of the bound nucleotide. Whereas G13 and G15 form hydrogen bonds to the β-phosphate, K16 interacts electrostatically with the β- and γ-phosphate and S17 is involved in magnesium ion coordination (see **Figure 35**).

The G2-motif within the switch I region is also conserved. In H-Ras it is composed of the residues ^{29}VDEYDPT^{35}. The conserved residue T35 completes the octahedral magnesium coordination together with S17, the oxygens of the β- and γ-phosphates and two water molecules.
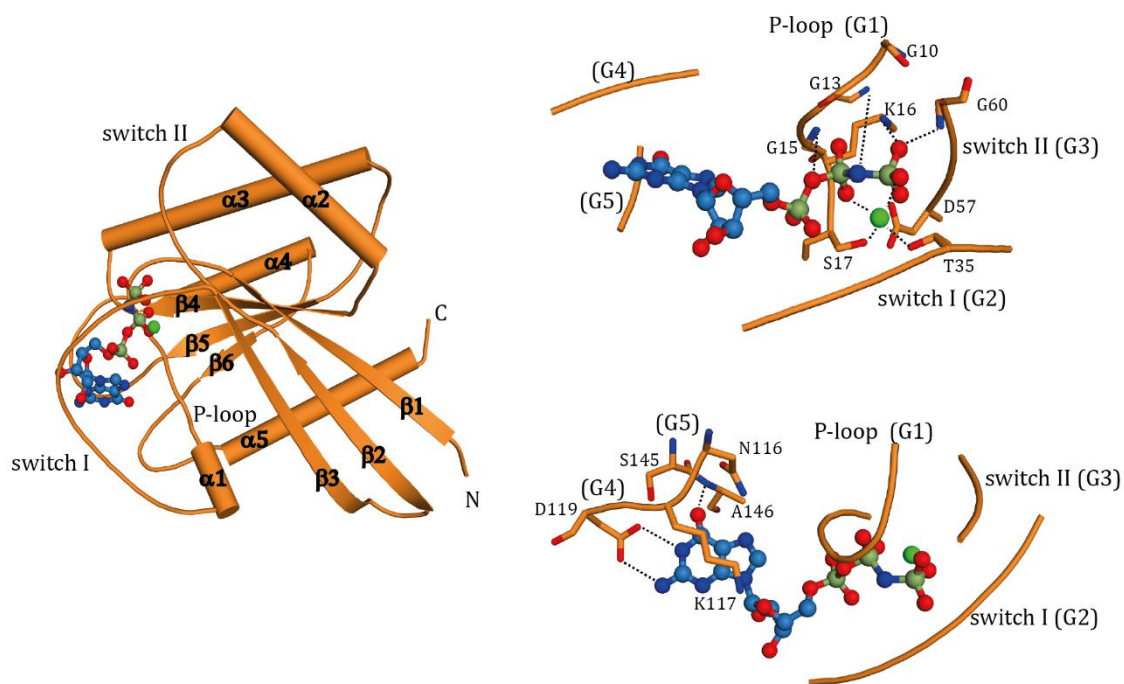
Figure 35: Ribbon representation of the structure of the H-Ras p21 G domain (PDB: 5P21) bound to the non-hydrolysable GTP analogue GMPPNP determined via x-ray crystallography. β-sheets are shown in flat arrows, α-helices in cylinders. GMPPNP is shown in sphere-and-stick representation. Carbon atoms are coloured light blue, nitrogen atoms are dark blue, oxygen atoms red and phosphate atoms dark green. Magnesium ion is coloured in light green. Right: Detailed view of H-Ras p21 nucleotide binding site. Loops involved in binding are shown in orange and selected amino acids are shown in stick-representation. Nitrogens are coloured in blue, oxygens in red. Water molecules are not shown for clarity. Dotted lines highlight interaction between GMPPNP and protein and are further described in the text.

The G3 motif, also known as walker B motif, with the consensus sequence DxxG (Ras: [57]DTAG[60]) is indirectly involved in coordination of the magnesium ion via positioning of the water molecule by the aspartate D57. Additionally, the conserved glycine G60 forms a hydrogen bond to the γ-phosphate.

The G4 motif with the consensus sequence (N/T)KxD (Ras: [116]NKCD[119]) is important for the specificity of the G domain towards the guanine nucleobase. The aspartate D119 is thereby playing a key role, as it forms a bifurcated hydrogen bond to the exocyclic amino group as well as to the endocyclic nitrogen of the guanine base. The G5 motif (Ras: [145]SAK[147]) further stabilizes the guanine nucleobase via hydrogen bonds between the exocyclic oxygen and the residue S145 (not shown) as well as the main chain nitrogen of A146.[146,149,159]

To undergo the GTPase molecular switch, the above-mentioned conserved threonine T35 in switch I and the glycine G60 in switch II interact with the γ-phosphate, stabilizing the switch regions in a rigid conformation. This enables the recruitment of effector proteins that in most cases interact with the switch region.

Upon GTP hydrolysis, the mentioned interactions are broken and the switch regions undergo a conformational change leading to the GDP-bound OFF-state and to dissociation of the effector proteins.

The generally accepted model is that a water molecule acts as nucleophile and attacks the γ-phosphate in-line.[160-163] However, the transition state in the protein environment remains still elusive.

6.1.2   GAPs and GEFs

Within the family of small GTPases the GTP hydrolysis rate is intrinsically very slow. As studied in the Ras-RasGAP system the GAP-stimulated GTP hydrolysis rate is increased by several orders of magnitude.[164] The RasGAP occludes the active site of Ras and thereby shielding the reaction volume from solvent molecules. Additionally, a positively charged arginine of RasGAP (arginine finger) is positioned into the phosphate-binding site of Ras and stabilizes the transition state by neutralizing the developing negative charge at the γ-phosphate. Further, the arginine finger of Ras-GAP forms main chain hydrogen bonds that stabilize the position of the catalytic glutamine N61 of Ras, which in turn coordinates the attacking water (see **Figure 36**B).
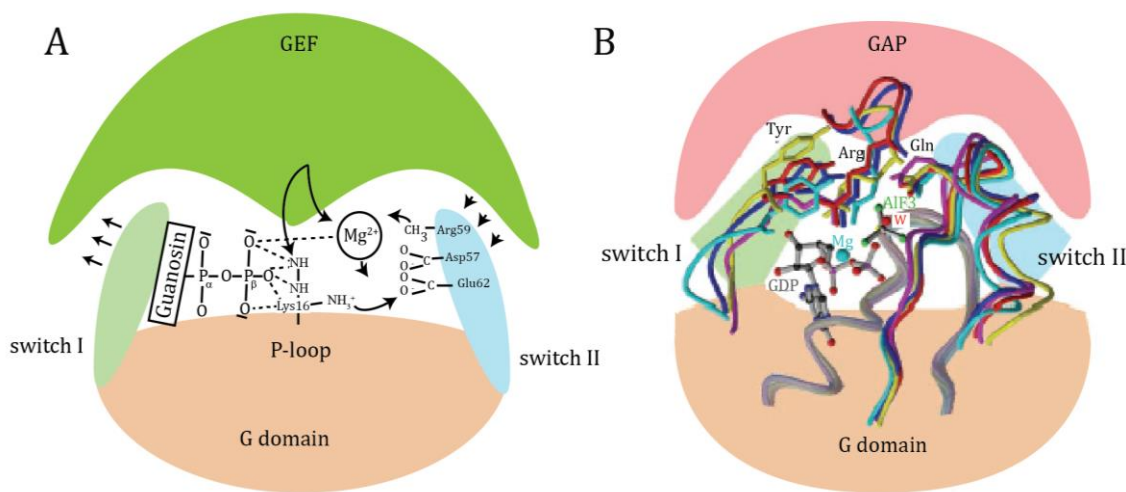
Figure 36: Mode of action of GEFs and GAPs. A: Schematic diagram of GEF action, showing the mechanistic principle of nucleotide release. The magnesium ion is pushed out of the active site via elements of the GEF or the switch II, which in turn is pulled towards the nucleotide-binding site. P loop conformation and GEF binding are disturbed leading to loss of the interaction with the phosphates. Figure taken from [149]. B: Common principle in the mechanism of GAP action on Ras, Rho and Gα proteins, as seen in the structures of Ras-RasGAP (yellow), Rho-Rho-GAP (red), RacExoS (magenta) complexes. All structures are solved in presence of GDP, aluminium fluoride and magnesium, which mimics the transition state during GTP hydrolysis. Catalytic water is labelled as "W". The arginine fingers and the catalytic glutamines are indicated as Arg and Gln. Figure modified from [149]. For explanation see text.

This arginine finger is essential for the described mechanism and is also conserved in several other G protein-GAP systems.

The release of the hydrolyzed GDP is also intrinsically very slow in small G proteins. The auxiliary GEFs can catalyze the nucleotide-release by modifying the active site. Upon GEF binding, structural changes within the P-loop and the switch regions lead to a release of the phosphate moiety and the magnesium ion from the active site. The affinity for the nucleotide is drastically reduced and the release of the nucleotide is possible. Due to a higher GTP concentration than the GDP concentration in the cell the rebinding results in a higher proportion of GTP-bound protein (see **Figure 36**A).[164,165]

6.1.3   GADs

The auxiliary proteins GAP and GEF are not necessary for the class of GAD proteins (G domain activation by dimerization). The nucleotide dependent dimerization of the G domains in GADs lead to the complementation of the active site by the neighboring protomer. Although the interaction takes place via the switch regions of the G domain, as for Ras-RasGAP, the mechanism of GTP hydrolysis differs. The dimerization leads to the stabilization of the flexible switch regions and thus to the stable formation of the transition state. One can distinguish between three different types of GADs (see **Figure 37**). (i) The pseudo-homodimerization takes place between G proteins within the same protein family (e.g. SRP-SR or Toc).[166,167] The dimerization mechanism could be clarified due to a crystal structure of the bacterial SRP-SR homolog Ffh and FtsY.[160,168] The G domains are dimerizing in a fashion that the nucleotides are oriented in a *head-to-tail* orientation whereas the 3'-OH group of the ribose interacts with the γ-phosphate of the opposing nucleotide. The aspartate (D135 in Ffh and D139 in FtsY) side chains, positioned by hydrogen bonds to a magnesium-ion-coordinated water, in turn position the nucleophilic water 3.3 Å from the γ-phosphate. A glutamine side chain (Q144 and Q148) is directed into the active site and forms hydrogen bonds to one of the coordinating waters of the magnesium ion and to a α-phosphate oxygen. The arginine residues (R138 and R142) are located between the γ-phosphates of GMPPCP and could play a role as arginine fingers, possibly in both *cis* and *trans* configurations relative to the two nucleotides.
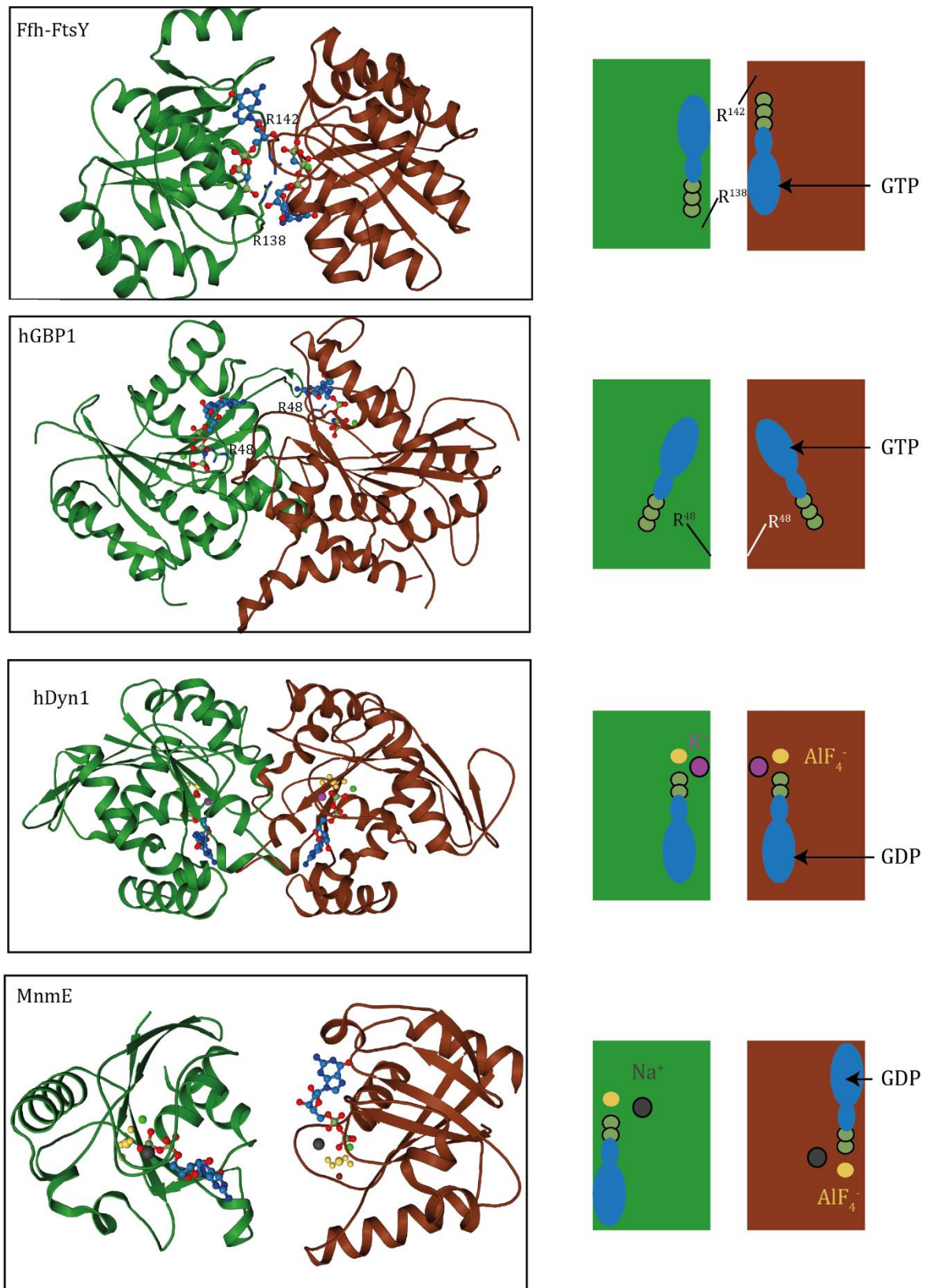
Figure 37: Structures of different GAD proteins. Left: Ribbon representation of G domains of different G-proteins coloured in green and brown. Ffh-FtsY (PDB: 1RJ9), human GBP1 (PDB: 2BC9), human dynamin 1 (PDB: 2X2E), MnmE (PDB: 2GJ8). Nucleotides shown in sphere-and-stick representation. Carbon light-blue, nitrogen dark-blue, oxygen red, phosphorous dark-green, AlF_4- yellow, potassium ion magenta, sodium ion grey, magnesium ion in light green. Arginine residues are shown in stick representation. Waters are not shown due to clarity. Right: Cartoon shows orientation of nucleotides in dimerizing G domains (green/brown) as well as the ions/arginines involved in transition state stabilization. Figure adapted from [153].

The *trans*-homodimerization is the formation of two identical G domains (e.g. GBP1 and dynamin).[169,170] Human GBP1 (Guanylate-binding protein 1) belongs to the dynamin superfamily and has the unique characteristic to hydrolyze GTP to GDP and further to GMP.[171] Here, upon G domain dimerization, the nucleotides are oriented in a *head-to-head* conformation and the positioning of the nucleophilic water is coordinated by S73 in the switch I. R48 in the P loop functions here as the arginine-finger. The hydrolysis to GMP is catalyzed by the same residues.

The crystal structure of the human dynamin G domain dimer also reveal a *head-to-head* orientation of the nucleotides. Instead of an arginine finger, additional electron density was identified in the active site and interpreted as a sodium ion. This sodium ion could undertake the function of the charge compensator.[170]

In *cis*-homodimerization, additional domains are involved in dimerization (e.g. MnmE).[154] MnmE forms dimers via the G domain, the N-terminal domain and the helical domain, whereas the GTPase activity is stimulated via the G domain dimerization. Here, a potassium-ion mimics the positive charge of the arginine-finger in the transition state and neutralizes the negatively charged γ-phosphate.[153,154]

6.2 The dynamin superfamily of G proteins

The dynamin superfamily members are large G proteins involved in budding of transport vesicles, division of organelles, cytokines and pathogen resistance. They have in common the low micro-molar affinity to nucleotides and a relatively high basal GTPase activity, which is enhanced in presence of lipid-templates. Furthermore, they share a lipid-template dependent oligomerization, the propensity to self-assemble and the ability to dimerize across the conserved G domain interface.

Within the dynamin superfamily, the G domain is the most conserved part and has a canonical G domain fold, as found in Ras (see 6.1.1) with two additional α-helices and two additional β-strands. These extensions are located between the G2 and the G3 motif and downstream the G5 motif and are supposed to fulfil additional regulatory functions.[172,173] Besides the G domain, additional helical domains, named bundle signaling element (BSE) and stalk are present. They can mediate self-assembly or transduce conformational changes elicited by nucleotide hydrolysis. These domains vary in length within the family members and in some cases the stalk is an extended single domain including the BSE. Besides the classical dynamin, dynamin-like proteins (DMNL), optic atrophy protein 1

(OPA1) , Myxovirus resistance proteins (Mx), mitofusin proteins (MNF), Eps15 homology domain-containing proteins (EHD2) and guanylate-binding proteins (GBP)/atlastins belong to this superfamily. In the following, the current knowledge of selected proteins of this superfamily will be summarized regarding the specific molecular architectures, the functions and the G domain dimerization mode.

6.2.1   Dynamins

Dynamin is the founding member of the dynamin superfamily of large G proteins.[174] There are three dynamin homologs in mammals (and in almost all eukaryotes) that are involved in several membrane deforming functions as in budding of caveolae, phagocytosis, podosome formation and most notably in budding of clathrin coated vesicles (CCV) during clathrin mediated endocytosis. Dynamin 2 is ubiquitously expressed, whereas Dynamin 1 and 3 are mainly found in brain, lung, heart and testis.[156,175]

Dynamin oligomerizes in helical structures in the presence of appropriate membrane template thereby deforming them to membrane tubules *in vitro*.[176,177] Once assembled the GTPase activity increases 100-fold which in turn leads to a conformational change and results in vesicle scission.[178-182]
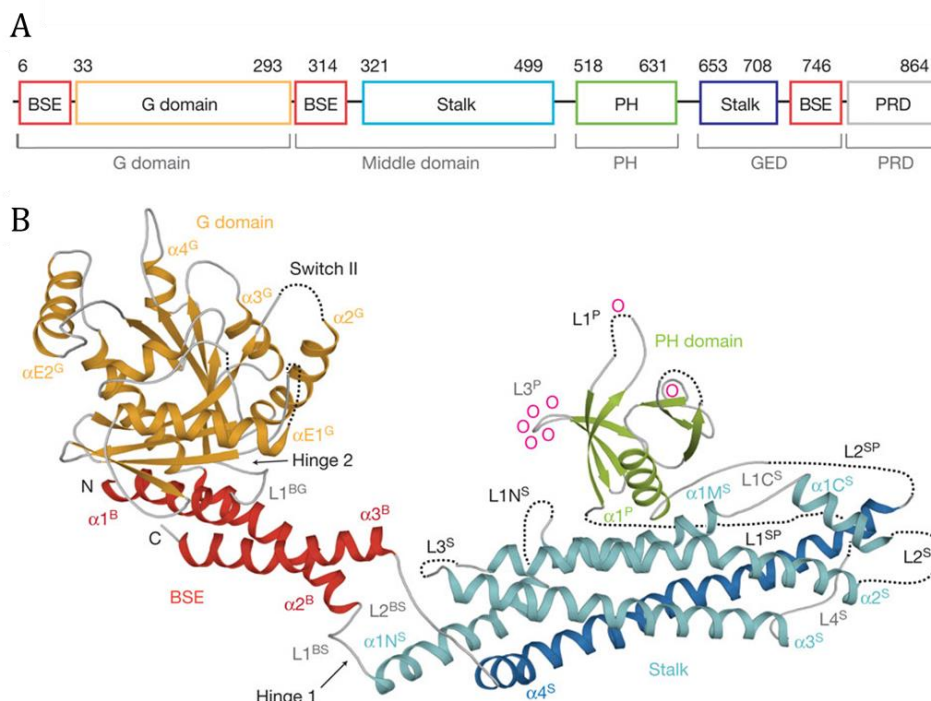
Figure 38: Human dynamin 1. A: Domain diagram shows the structural domains (coloured) and the old sequence derived domain assignments (grey). Depicted numbers correspond to the first/last amino acid of the respective domain. B: Crystal structure of human dynamin 1 ΔPRD (PDB: 3SNH). Figure taken from[183].

Dynamin has a molecular weight of 100 kDa and consists of the G domain, a BSE, a stalk, a pleckstrin homology (PH) domain and a proline rich domain (PRD).

The BSE consists of three helices, spread over the whole protein sequence (see **Figure 38**A), and is connecting the G domain with the stalk region. It has a central role in dynamin function, as mutations in this interface disrupt assembly-stimulated GTPase activity and impair dynamin catalyzed membrane fission.[170,184,185]

The PH domain, first identified in the pleckstrin protein, consists of one α-helix, seven β-strands and three lipid binding loops pointing towards the solvent. Due to its specificity to phosphatidylinositol-4.5-biphosphate (PIP2), which is found mainly in the plasma membrane, dynamin is thought to exclusively act there and not at internal membranes. The PH domain is embedded in the stalk between $\alpha3^S$ and $\alpha4^S$ (see **Figure 38**).[186,187]
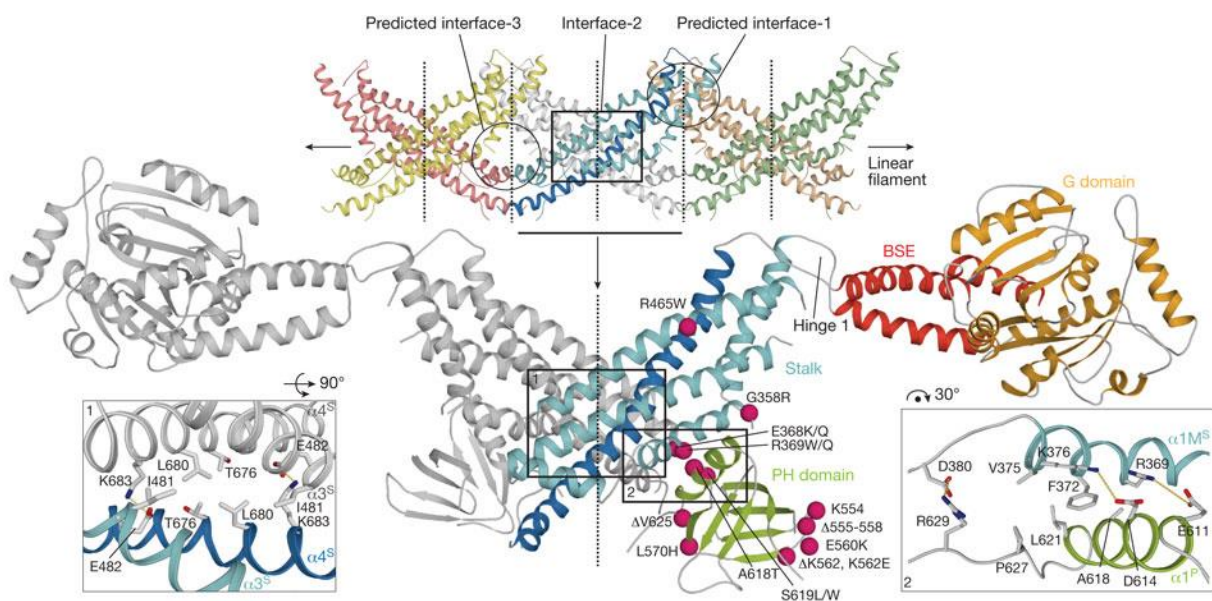
Figure 39: Crystal structure of the human dynamin 1 ΔPRD (PDB: 3SNH) dimer. For clarity, the upper panel shows only the stalks, forming the X-shaped minimal building block via Interface 2 as well as the filament formation within the crystal structure via further oligomerization through interfaces 1 and 3. The lower panel shows a dynamin 1 ΔPRD dimer. One monomer is color coded as in **Figure 38** and the protomer in grey. Disease-related dynamin 2 mutations are indicated as pink spheres. Figure taken from [183].

The stalk and thus the central assembly hub could be described as the structure of dynamin 1 ΔPRD was solved.[183,188] Assembly via the highly conserved stalk interface 2 leads to the formation of an x-shaped dynamin dimer, which constitutes the minimal building block in dynamin oligomerization. The conserved Interfaces 1 and 3 mediate further oligomerization of these dimers into linear filaments.[183] Furthermore, the stalk interacts with the PH domain via another conserved interface involving α1 of the stalk

and the PH domain (**Figure 39**). A salt-bridge between stalk (R440) and BSE (D744) of the neighboring, parallel dynamin dimer was proposed to have a regulatory function in oligomerization, as mutations in this interface lead to increased dynamin-mediated endocytosis *in vivo* (not shown).[183]

Two crystal structures containing the G domain and BSE of human dynamin 1 in the presence of GDP and aluminium tetrafluoride (GDP·AlF$_4^-$) (PDB: 2X2E) or GMPPCP (PBD: 3ZYC) revealed the G domain dimerization mode. The dimerization can be divided in three interfaces and is described in the following in the GDP·AlF$_4^-$–bound form (PDB: 2X2E).
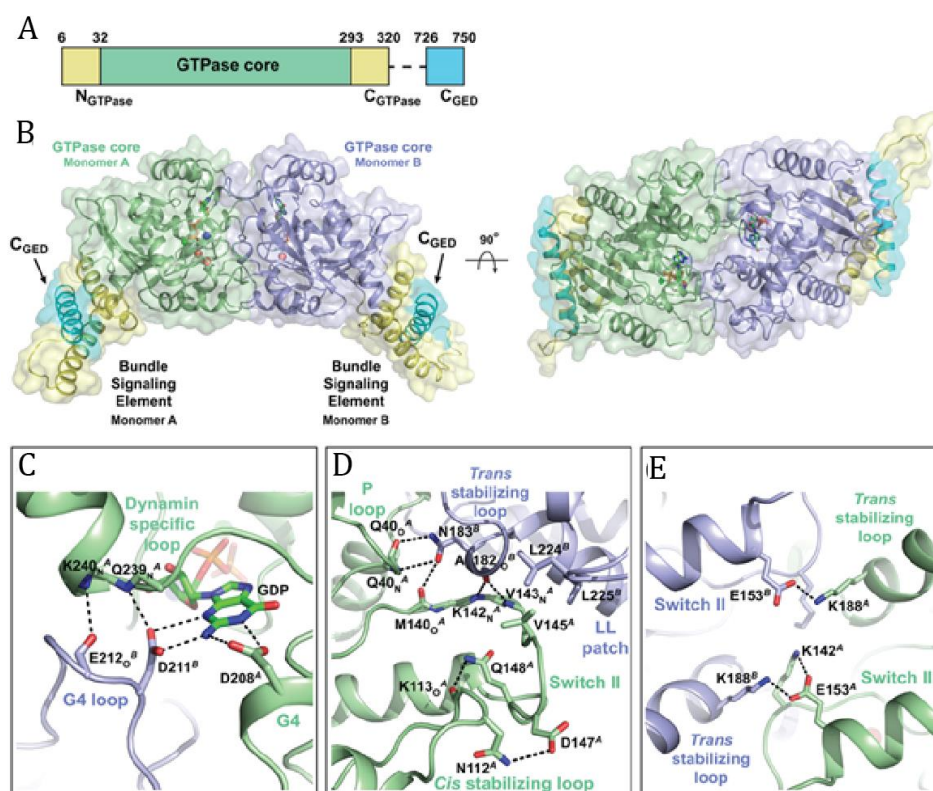
Figure 40: Structure of the G domain dimer of human dynamin 1 (PDB: 2X2E). A: Domain arrangement of GTPase-GED (GG) fusion constructed from human dynamin 1. Each monomer contains a GTPase core (green) and three helical segments: N$_{GTPase}$, C$_{GTPase}$ (yellow) and C$_{GED}$ (cyan), known as BSE. Dashed line denotes the flexible linker that tethers the C$_{GED}$ helix. B: Structure of GDP.AlF$_4^-$-stabilized GG dimer from the "long axis" crystal form shown from the side (left) and top (right). The GTPase cores of individual monomers are coloured in green and blue. The helical segments in each monomer that constitute the bundle signaling element (BSE) are coloured yellow with the C$_{GED}$ helix highlighted in cyan. C, D, E: Structural interactions that stabilize the GG dimer. Coloring is the same as B. Key residues are labelled with a superscript (A or B) to indicate from which monomer they originate. The additional subscript (O or N) signifies the interaction of a main chain carbonyl or nitrogen. Dashed lines indicate hydrogen-bonding interactions. Figure taken from[170].

Interface I (**Figure 40**C) is formed via dual *cis*/*trans* coordination of the guanine base by the G4 motif (^{205}TKLD^{208}). In *cis*, D208 interacts with the guanine base whereas upon dimerization D211 in *trans* forms two additional hydrogen bonds to the exo- and

endocyclic nitrogens at position N2 and N3. Concomitantly, D211 and E212 interact with Q239 and K240 of the adjacent dimer, partially stabilizing the dynamin specific loop (DSL; residue 236-246).

Interface II (**Figure 40**D) is formed between switch II (residue 136-153) and the *trans*-stabilizing loop (residue 176-188), primarily by main chain interactions. This interface is bolstered by interactions between V145 of switch II and the leucine residues L224 and L225 and hydrogen bonding between N183 and the main chain atoms of Q40 (P-loop).

Interface III (**Figure 40**E) is build up by a pair of symmetric salt bridges between E153 and K188, anchoring the base of the dimer in *trans*. Furthermore K142 forms a hydrogen bond with E153 in *cis*.[170]
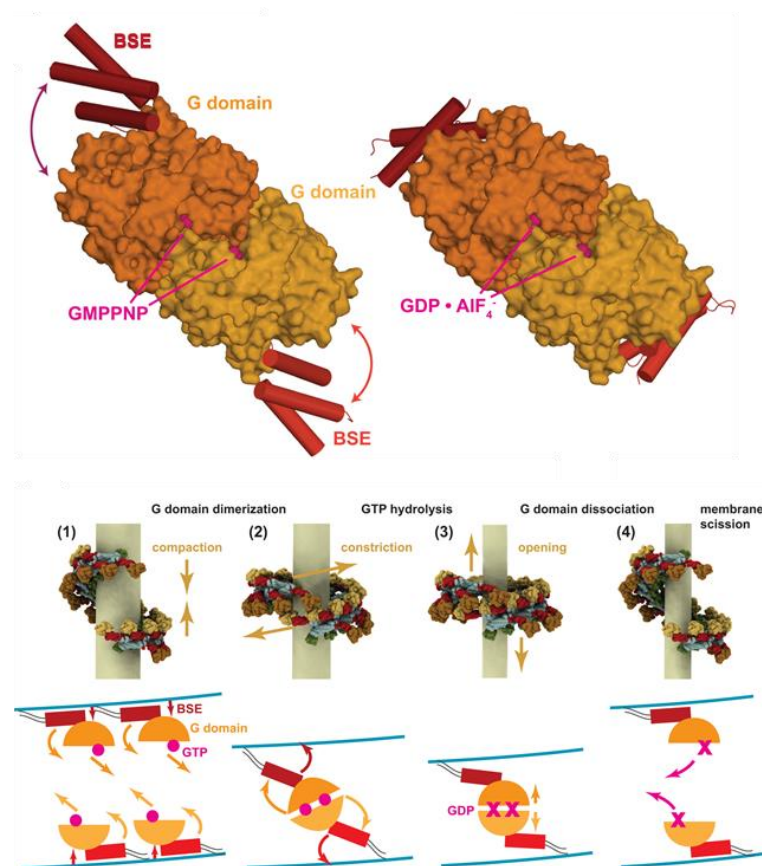
Figure 41: The power stroke of dynamin. Top: Comparism of the GMPPNP (PDB: 3ZYC) and the GDP-AlF$_4^-$-bound structures (PDB: 2X2E) of the G domain-minimal BSE construct. Bottom: Model for dynamin action during membrane scission: (1) when the growing dynamin filament has embraced the membrane templates, GTP-loaded G domains of neighboring turns oppose each other. (2) G domains or neighboring filaments dimerize leading to a compaction of the dynamin helix. (3) Nucleotide-hydrolysis induce conformational changes in the position of the BSE with respect to the G domain might induce a power stroke actively pulling filaments versus each other. (4) GDP-bound G domains separate leading to dissociation of neighboring filaments. Figure taken from [189].

The most striking difference in the two observed G domain dimer structures is the movement of the BSE relative to the G domain between the ground-state (GMPPCP-

bound) and the transition-state (GDP·AlF$_4^-$-bound). The rearrangement is a rigid body rotation of 69° which brings the BSE close to the G domain (see **Figure 40**B and **Figure 41**). This confirms that the BSE is indeed a flexible domain (hinge region) transducing a "power stroke" upon nucleotide-binding and hydrolysis which is necessary for dynamin's membrane-remodeling activity.[190]

A wealth of structures illuminates the function of dynamin dependent membrane fission, which is proposed to proceed as follows: Dynamin is recruited to the bud neck in the GTP-bound state. Intra- and intermolecular domain interactions are released upon membrane binding (e.g. interactions between stalk and BSE, PH domain respectively), which in turn enables recruiting of more dynamin subunits. After one helical turn is assembled around the neck in a right-handed helical orientation, the G domains of the neighboring filaments are proposed to dimerize via the G interface. G domain dimerization leads to a flexible BSE-G domain interaction. Following GTP hydrolysis, a conformational change moves the BSE back to the G domain. Repeating this cycle of GTP hydrolysis in a nut runner-like manner might end up in constricting the underlying membrane tubules (see **Figure 41**).[189]

6.2.2   Bacterial dynamin-like proteins (BDLPs)

Many bacteria have genes encoding for hypothetical dynamin-like proteins (BDLPs), which have a low sequence identity of around 20% relative to eukaryotic dynamins. In case of BDLP from the filamentous cyanobacterium *Nostoc punctiforme* (npBDLP), it could be shown that npBDLP is predominantly localized at the cell membrane. It self-assembles in the presence of GMPPNP, binds to bacterial whole-cell lipid extract in a regular pattern and tubulates liposomes in a fashion similar to dynamin and MxA.[157]

The crystal structures of npBDLP in the nucleotide-free and in the GDP-associated state reveal a domain organization slightly different to dynamin 1. The protein is composed of a globular G domain, located at the N-terminus followed by a neck-, a trunk- and a tip-domain. Neck and trunk both form a four-helical bundle and are reminiscent of the BSE (neck) and the stalk (trunk) domains of dynamin 1 and MxA. The neck is, as the BSE in dynamin and MxA, also not a discrete entity but the four helices are rather spread over the whole protein sequence. Both domains are bend towards each other and are connected via a flexible loop termed hinge 1. The connection between the G domain and the neck is partially conserved and termed hinge 2 as it can reorient upon oligomerization

and lipid binding. The trunk also consists of four helices and the paddle and tip are embedded into the trunk domain similar to the PH domain of dynamin 1 (see **Figure 42**A). Both paddle helices are predicted to be transmembrane helices and its insertion into the lipid membrane could introduce lipid curvature promoting membrane remodeling.[157,191] The npBDLP crystal structures in the nucleotide-free and in the GDP-bound form reveal a dimerization interface via the G domains (see **Figure 42**B). Helix 9 thereby crosses the interface and E245 at its C-terminal tip forms a hydrogen bond to the nucleotide base in *trans*, as it was seen already for the residue D211 in dynamin (see **Figure 40**C). K79 in the P-loop is oriented in *cis* for efficient nucleotide catalysis. Nevertheless, in the GDP-bound structure the most striking conformational changes occur within the tip region, thereby providing a potential mechanism to shear the GDP dimer apart.
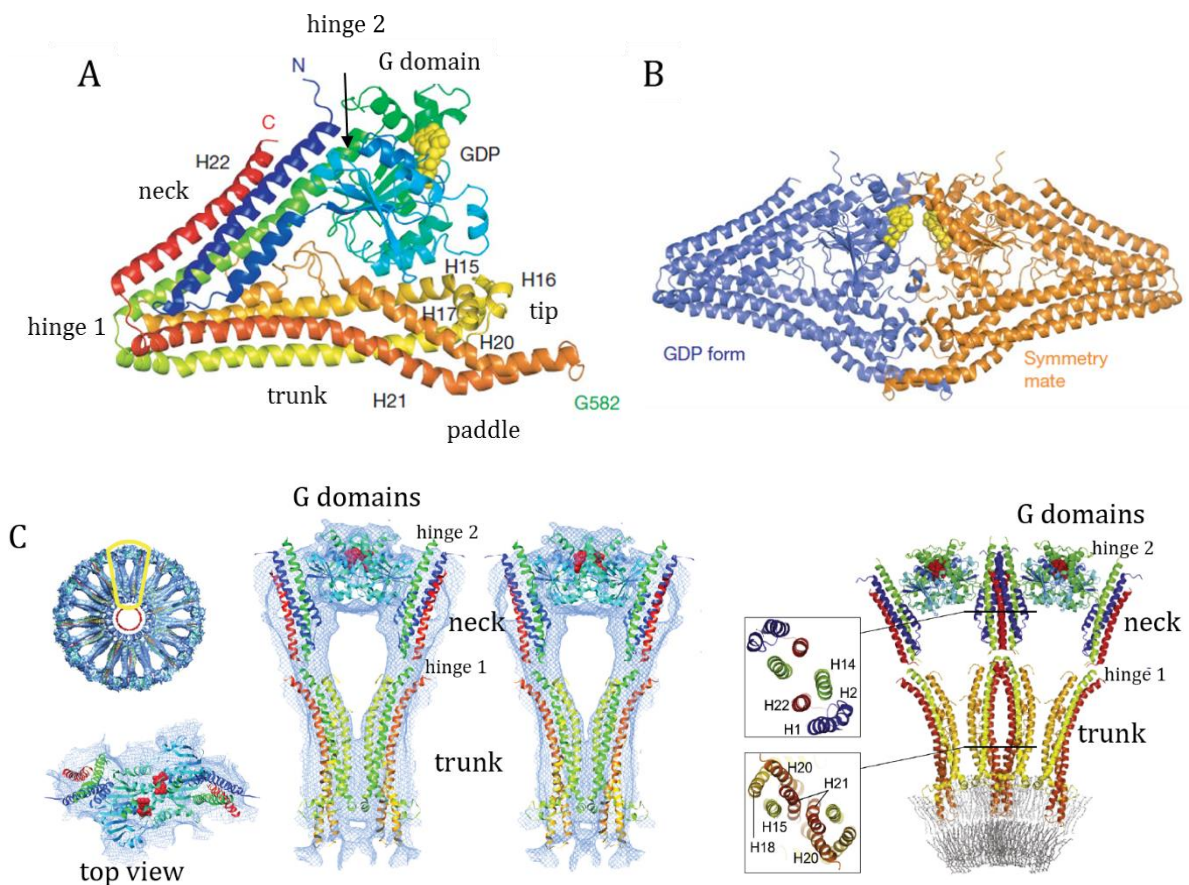
Figure 42: Structures of npBDLP. A: The BDLP-GDP monomer in ribbon representation rainbow-coloured from N (blue) to C (red) terminus (PDB: 2J86). B: GDP-bound BDLP and its symmetry mate forming a homodimer in the crystal packing. C: Cryo-EM reconstruction. Model of the helical BDLP filament in cross-section to the helix axis showing a fitted lipid bilayer. Only the inner ring of the lipid head groups (and hence lipids) is clearly observed in the 3D density. Shown close ups are protein-protein contacts between a pair of asymmetric units. The focus is on interaction between the central neighboring neck and trunk helices. Figure modified from [157,191]. For explanation see text.

A cryo electron-microscopy reconstruction of 50 nm BDLP decorated tubes in presence of GMPPNP revealed a T-shaped repeating assembly unit most likely composed of a BDLP dimer (see **Figure 42**C). Compared to the more compact crystal structures of BDLP, neck and trunk have to undergo a drastic conformational change relative to each other mediated by the hinge 1 and 2 regions. This arrangement is more linear with the hydrophobic paddle facing the lipid tubule.

However, although BDLP tubulates liposomes similar to dynamin, the observed domain rearrangement driven by GTP binding and/or hydrolysis are tremendous.

6.2.3   Atlastins

The atlastin GTPases (Atl) are key players involved in the maintenance of normal ER morphology. It was shown that down regulation of atlastin function lead to shorter, unconnected ER tubules in *Drosophila* motor neurons whereas in yeast and human, long unbranched ER tubules are the result of interfering the atlastin function.[192,193] In humans, three atlastin isoforms exist. Whereas isoform 2 and 3 are expressed ubiquitously, atlastin 1 (Atl1) is found in neurons exclusively. Atlastins contain an N-terminal G domain, a three helical bundle domain termed middle domain (MD) followed by two predicted transmembrane helices and a short C terminal cytosolic part.

Four crystal structures of Atl1 (not including the transmembrane helices and the C-terminal cytosolic part) bound to GDP (crystal form 1 + 2), GDP·AlF$_4^-$ (crystal form 3) and GMPPNP (also crystal form 3) shed light on how atlastin could catalyze membrane fusion in a GTP binding and hydrolysis-dependent fashion (biochemical data see [193]).[194,195] Strikingly, the orientation of the MD relative to the G domain differs dramatically in the crystal structures (see **Figure 43**). This movement is mediated around the linker region between both domains with a conserved proline residue recognized as a hinge region. Crystal form 2 (PDB: 3Q5E; see **Figure 43**C) reveals an extended Atl1 dimer, bound to GDP, with the MD domains pointing to opposing sides whereas in crystal form 1 (PDB: 3Q5D; see **Figure 43**B) Atl1, also bound to GDP, features a cross-over of the linker and the MD without forming an additional interface between the MDs. This crystal form 2 is termed a relaxed-parallel dimer. The recently observed crystal form 3 (PDB: 4IDO, 4IDN; see **Figure 43**D) from Atl1 bound to GDP·AlF$_4^-$ or GMPPNP is termed a tight-parallel dimer. Independently from the nucleotide loading state, it shows that the interfacial

surface extends to the first and third helix of the middle domains, in addition to the interface formed by the linker cross-over (crystal form 1).

While the G domains in the GDP bound form (crystal form 1+2) adopt a somewhat open conformation with regard to the nucleotide-binding pocket, in the GMPPNP or GDP·AlF$_4^-$ loaded state (crystal form 3) switch I and II undergo a conformational rearrangement folding over the phosphate and AlF$_4^-$ moieties. This positions the conserved R77 to interact with the α- and γ-phosphate and establishes its role as a catalytic arginine finger. In the GDP bound structures (crystal form 1+2), this residue supports the G domain dimerization by forming a salt bridge to a neighboring glutamate residue (see **Figure 43**E).
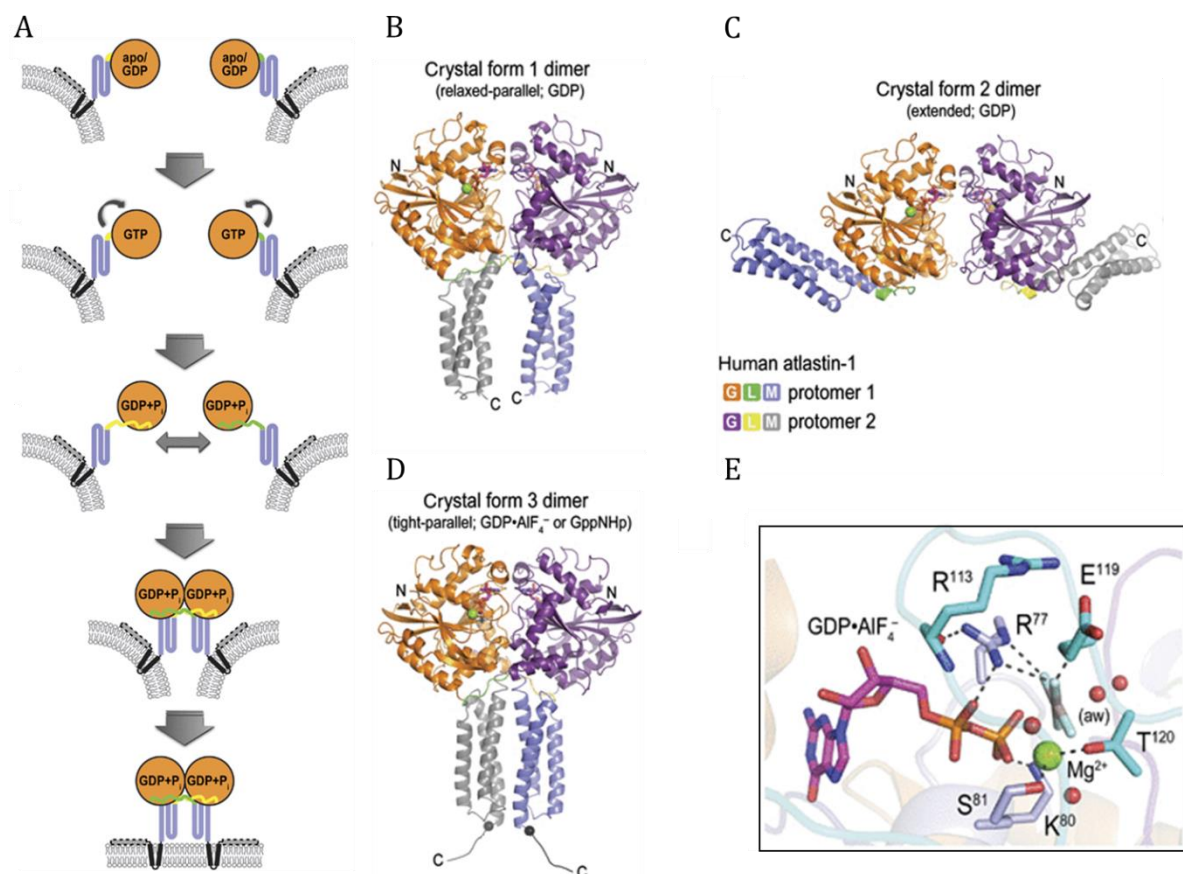
Figure 43: Crystal structures and mode of action of atlastins. A: Model for atlastin-mediated membrane fusion. The G domain (orange) is connected to the middle domain (blue) by a short linker (green/yellow). The two transmembrane α helices (black) and a C-terminal amphipathic helix (grey) are shown. B, C, D: The three obtained crystal forms of atlastin 1, bound to GDP or GMPPNP/ GDP·AlF$_4^-$. E: Residues involved in nucleotide hydrolysis. The nucleotide binding pocket of crystal form 3 is shown. Residues making direct or indirect contacts with the phosphate groups of the nucleotide analogue are shown as sticks. Mg^{2+} are shown as a green sphere. The attacking water molecule (labelled 'aw') and Mg^{2+}-coordinating water molecules are shown as red spheres. Figure modified from [194]. For explanation see text.

The proposed model suggests that membrane fusion follows a tethering step, in which both the G domains and the MDs, dimerize upon GTP binding and hydrolysis. This tight pairwise interaction would bring opposing membranes into close proximity and the increased membrane curvature could facilitate membrane fusion (see **Figure 43**A). Once the phosphate moiety is released, relaxation and subsequent disassembly of the dimer could occur. Within the model, the middle domain plays a dual role as it is required for efficient GTP binding and for homotypic dimerization of atlastin.[194]

6.3 GTPases in immunity

6.3.1 Resistance and immunity

For several centuries, multicellular organisms have evolved mechanisms to fight invading pathogens. Under tremendous selective pressure by these aggressors, the immune system developed the ability to recognize and eliminate them. The immune system can be divided in the innate and the adaptive branch. The fundamental principle of the innate immune system is to distinguish between non-infectious self and infectious non-self. Whereas components of the innate immune system can be found in all multicellular organisms, the adaptive immune system is mostly known in vertebrates.[196,197]

The innate immune system recognizes molecular patterns of microorganisms which are conserved and essential for their pathogenicity (pathogen-associated molecular patterns (PAMPs)). Lipopolysaccharides - shared by gram-negative and gram-positive bacteria, un-methylated CpG motifs within the DNA - also discriminatory for bacteria, double stranded RNA - a signature for RNA viruses, as well as mannans - components of the yeast cell wall, all belong to the PAMPs and are recognized by the host organism via different sets of pattern-recognition-receptors (PRRs). PRRs (e.g. Toll-like receptors (TLRs) or macrophage scavenger) are strategically expressed in cells which are the first to encounter pathogens during infection such as the epithelia surface or all types of effector cells of the innate immune system. The recognition of PAMPs by PRRs results in immediate induction of apoptosis and activation of the complement system leading to activation of phagocytosis, apoptosis or secretion of pro-inflammatory cytokines and chemokines.[198-200]

In contrast, the adaptive immune system requires days to establish an effective response after contact with a new pathogen due to clonal expression and differentiation to effector lymphocytes. This antigen-specific immune response is mediated by B- and T-

lymphocytes and results from somatic recombination and hyper mutation for B- and T-cell receptor encoding gene elements.

Thus, the innate and the adaptive immune system do not comprise two independent units but rather the innate immune response precedes and is necessary for the adaptive response.[198,201]

6.3.2   Interferons

Cytokines and chemokines help to orchestrate the immune response. They mediate the intercellular communication and the spatial localization during inflammation.[201] Interferons (INF) describe one class of cytokines and are important signal mediators of the innate and adaptive immune system. INFs can be further subdivided in three classes.
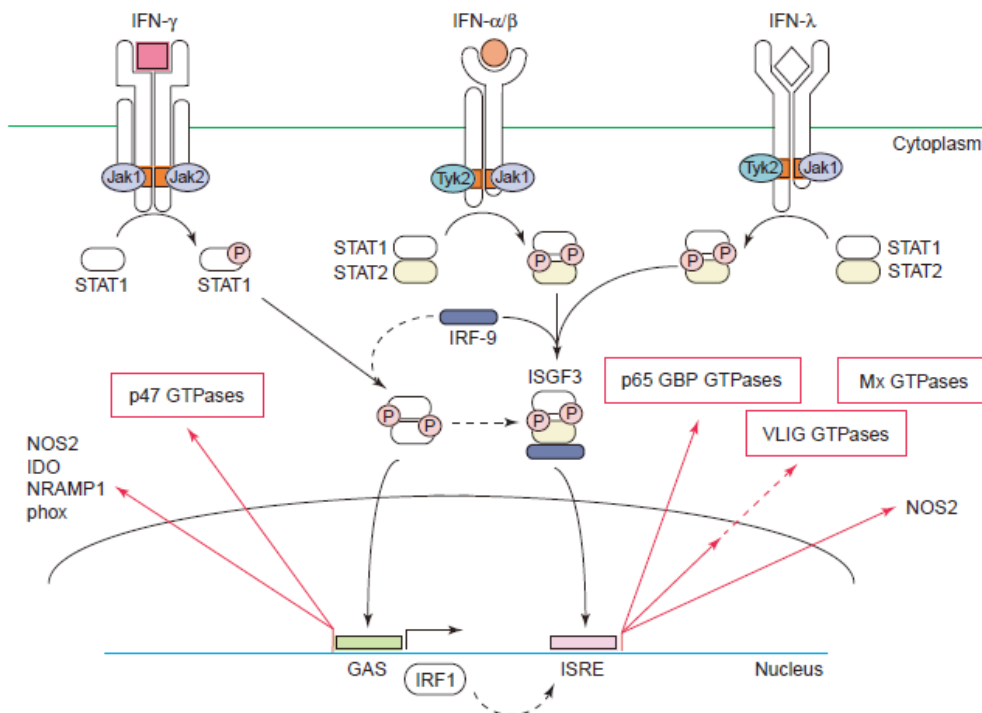
Figure 44: Response of the IFN-inducible GTPase families and other well characterized antimicrobial pathways to type I (IFN-α/β, IFN-λ) and type II (IFN-γ) IFN signaling. Binding of soluble IFN-α/β, IFN-λ or IFN-γ to their cognate receptors primarily enlists the Jak–STAT signaling cascade. Signaling specificity arises from the combination of certain Jaks with a particular STAT pairing. For IFN-γ, receptor engagement allows Jak1/2 activation, docking and phosphorylation of STAT1 and rapid translocation of STAT1 dimers into the nucleus. Abbreviations: GAS, IFN-γ activation site; GBP, guanylate-binding protein; GTPase, guanosine triphosphatase; IFN, interferon; IFN regulatory factor-1; ISGF, interferon-stimulated gene factor; ISRE, interferon-stimulated response element; Jak, Janus kinase; IDO, indoleamine-2,3-dioxygenase; IRF-1, IFN regulatory factor-1; NOS2, nitric oxide synthase 2; NRAMP1, natural-resistance associated macrophage protein-1; Phox, phagocyte oxidase; STAT, signal transducer and activator of transcription; Tyk, Jak protein tyrosine kinase; VLIG, very large inducible GTPases. Figure taken from [202]. For explanation see text.

Type I INFs (α, β and more) are secreted by virus-infected cells and induce the antiviral programs of the neighboring host cells. Type II INFs (γ) are secreted by thymus derived (T) cells and natural killer (NK) cells and a few others. They can activate a large number of genes involved in the defense against the pathogenic invaders.[203]

Type I and II Interferons bind to their respective INF receptors followed by activation of the JAK-STAT signal transduction pathway which stimulates the gene transcription of several effectors (e.g. p47 GTPases, p65 GTPases, Mx GTPases and others) controlled by IFN-stimulated response element (ISRE) and INFγ-activated sites (see **Figure 44**).[204]

6.3.3 Cell-autonomous immunity

Antimicrobial defense is not only shaped by a set of specialized immune cells but rather a more ubiquitous interplay of cytokines and lymphocytes. This interplay is named cell-autonomous immunity and is spread over all three kingdoms of life.

In archaea and bacteria, clustered regularly interspace short palindromic repeats (CRISPR) contribute to the cell autonomous immunity as exemplified here.[205]

In vertebrates, Mx proteins mediate inhibition of viral replication [206,207], Tryptophan depletion is induced by the indole 2,3-dioxegenase (IDO) [208,209], depletion of arginine and production of nitric oxide (NO) is catalyzed by nitric oxide synthase (iNOS) [210] and immunity-related GTPases (IRGs) are active against intracellular bacteria and protozoa [155,211].

6.3.4 Toxoplasma gondii

Toxoplasma gondii (*T. gondii*) is a ubiquitous protozoan parasite estimated to infect one third of the world's population. It can infect nearly any warm blooded animal and invades the host cell via its characteristic apical secretory organelles. Micronemes, rhoptries and dense granules are specialized for invasion and formation of a non-fusiogenic parasitophorous vacuole (PV) within the host. In this compartment, the parasite can acquire nutrients from and proliferate within the host cell.[212,213] The parasitophorous vacuole membrane is modified extensively by the parasite and contains multiple proteins that interact with the host cell organelles including host cell mitochondria and the endoplasmatic reticulum. *T. gondii* possesses a novel non-photosynthetic chloroplast-like organelle called apicoplast which is essential for lipid biosynthesis, heme synthesis and

the isoprenoid pathway. Antibiotics like clindamycin, tetracyclines, chloramphenicol and quinolones target apicoplast processes.[214]

T. gondii can differentiate into the rapidly multiplying tachyzoite. Tachyzoites differentiate into latent bradyzoites, which are surrounded by the cyst wall within the parasitophorous vacuole. The tissue cyst can persist indefinitely in the host.[213,214]

T.gondii isolated from Europe and North America can be grouped into three major genotypes (I, II and III). Type I strain is highly virulent in mice and seems to be more frequently associated with ocular toxoplasmosis. Type II and III are significantly less virulent to mice and readily form cysts in vitro.[214-216]

Mouse immunity-related GTPases (IRGs) mediate early resistance to *T. gondii* infection in mice. IRG proteins are able to accumulate on the parasitophorous vacuole of invading tachyzoites, resulting in disruption of the vacuolar membrane. Interestingly, the virulent strains of *T. gondii* are able to overcome the IRG resistance thus suggesting that different virulence types interact differently with the IRG system.[217-224]

6.3.5   p47 GTPases

p47 GTPases are immunity-related GTPases (IRG) which are expressed in several species, including humans, mice, rats, cows, pigs, dogs, chickens and jawed fish.[225,226] Although the mouse is often a good experimental model system for human, some immune mechanisms have evolved rather differently. Whereas in human, just one p47 GTPase (Irgc) is present, bioinformatics studies suggest that 23 paralogues exist within the mouse genome. In mice, the interferon-inducible p47 GTPases constitute one of the most powerful cell-autonomous resistance systems against several intracellular pathogens.[202,227,228] IRGs are found to be mainly induced by IFNγ and to a lesser extent also by IFNα/β in all cell types analyzed.[155,190,203,225,229-234] However, only some of them have been functionally characterized so far. These members localize to intracellular membranes in interferon-induced cells with some favoring the endoplasmatic reticulum [235,236] and others the Golgi membrane [224,236]. However, infection or phagocytosis initiates redistribution of several examined members to the phagocytic vacuole (see **Figure 45**B).[224,225,236] Thus, resistance to Gram-positive and Gram-negative bacteria, mycobacteria and protozoal pathogens (e.g.: *Mycobacterium tuberculosis* [225], *Mycobacterium avium* [237], *Listeria monocytogenes* [238], *Tripanosoma cruzi* [239], *Toxoplasma gondii* [220,221,223,224,238]) is mediated by p47 GTPases. The six founding members, Irgm1 (LRG47)[232], Irgm2 (GTPI)[226], Irgm3 (IGTP)[235], Irgb6

(TGTP)[240,241], Irgd (IRG47)[231] and Irga6 (IIGP1)[226] were investigated in the context of immunity in some more detail. Three of these founding members have a radical substitution of methionine for lysine in the G1 motif and are entitled GMS GTPases (IRGM, Irgm1-3)(see **Figure 45**).[226,229]
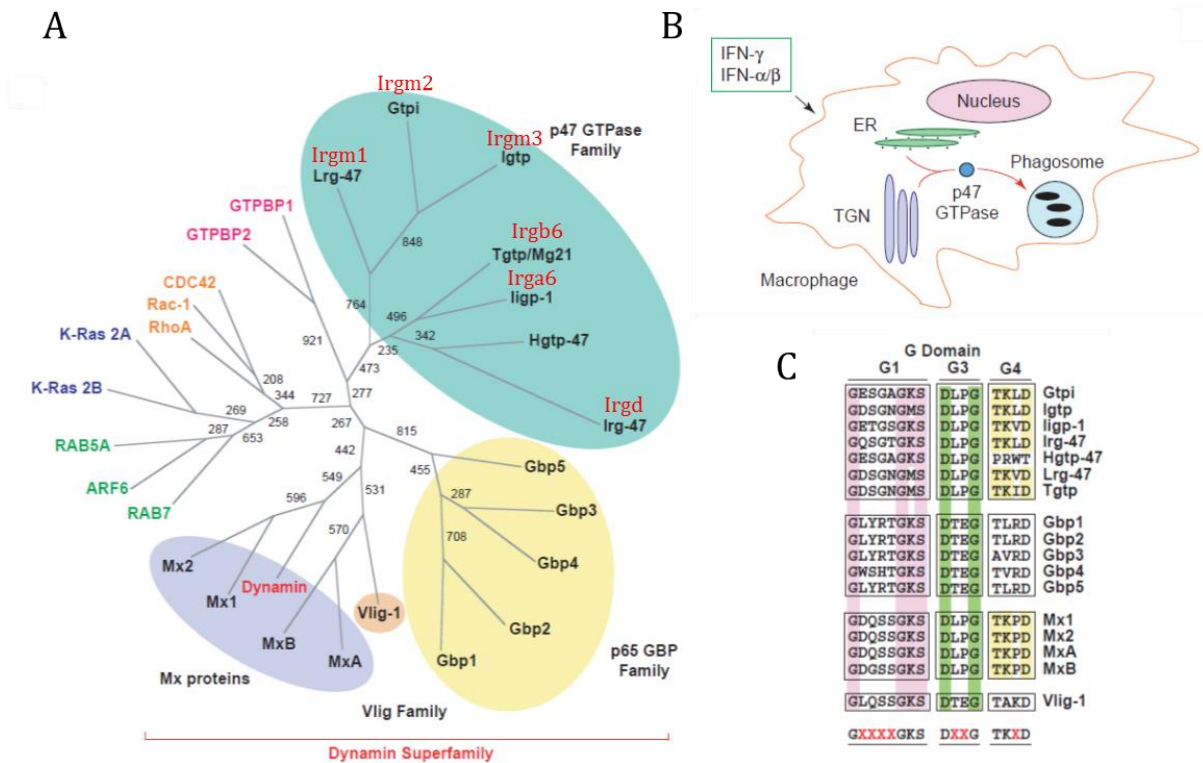
Figure 45: Evolutionary tree of the IFN-inducible GTPase families. A: Neighbor-joining dendogram depict similarities between the p47, Gbp, Vlig and Mx proteins with selected small 20kDa GTPases and dynamin. ClustalW-aligned sequences spanning the entire G-domain were examined as unrooted trees using Phylip 3.6b Prodist distance analysis. B: The predicted mechanism for the p47 family shows trafficking of p47 GTPase from the trans-Golgi-network or the endoplasmatic reticulum to remodel the pathogen phagosome within infected macrophages. C: Amino acid alignment of the conserved G1, G3 and G4 consensus regions among IFN-inducible GTPase families. Abbreviations: ARF, ADP-ribosylation factor; GP-1, G protein-1; GTPBP1, GTP-binding protein 1; VLIG, very large inducible GTPases. Figure modified from [202].

IRG proteins have in common their ability to accumulate on pathogen containing vacuoles in INFγ stimulated cells. Irgm2-3, Igra6, Irgb6, Irgd and Irgb10 were found to re-localize from their resting compartments to the PVM (parasitophorous vacuole membrane) of avirulent *T. gondii* but only very inefficiently to the PVM of virulent *T.gondii*.[218,221,224,242,243] The PVM of avirulent *T. gondii* was shown to disrupt in IFNγ stimulated cells and vesiculation and sequestration of the membranes was reported. Irga6 was shown to localize to vesicular structures next to disrupted sites of the PVM, as it was reported for

the *T. gondii* granuole protein GRA7 (see **Figure 46**). Although Irga6 seems to be crucial for pathogenic defense, its molecular mechanism of action remains elusive.[217,224,244,245] Opposed to this, IRGMs are mainly localized at the cellular membrane in the GDP-bound state. It was shown that Irgm3 can interact with Irga6 only if both proteins are GDP bound. Furthermore, Irga6 accumulates in the GTP-bound state in INFγ-non-induced cells, when the IRGMs are knocked out.[242] Thus, it was speculated that IRGMs in general attenuate the activation process of Irga6 at intracellular membranes by retaining Irga6 in the GDP-bound state. It was further shown that the myristoylation site of Irga6 at residue G2 is necessary for the recruitment to the PVM but not for the membrane binding itself.[236,243] Additionally, the myristoylation site as well as the "hinge" region between amino acid 7 and 12 are just accessible in the GTP bound state of Irga6 and not in the GDP bound state.[242,243] This could be explained by large structural changes within Irga6 upon nucleotide exchange and hydrolysis.
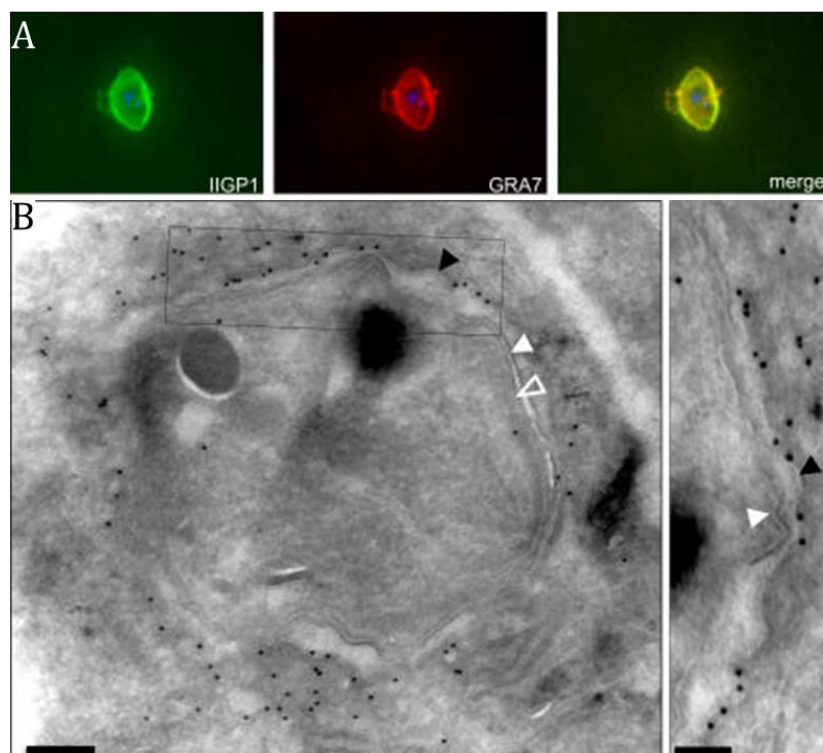
Figure 46: Localization of Irga6 within the cell. A: IFN-γ-induced astrocytes were infected with *T. gondii* for 2 h and stained for Irga6 (green) and GRA7 (red). Nuclei were stained with DAPI. B: IFN-γ-induced astrocytes were infected with *T. gondii* for 6 h, fixed, and subjected to ultra-thin cryo sectioning. Sections were labelled for Irga6 using the 165 antiserum and 10 nm gold particles coupled to protein A. Right: Enlarged view of the boxed region showing that the Irga6 label was found in close proximity to the PVM (black arrowhead: PVM; white arrowhead: *T. gondii* plasma membrane; open arrowhead: *T. gondii* inner membrane complex; bars 200 nm and 100 nm [inset]). Figure modified from [224]. For explanation see text.

Recent studies shed light on the mechanism how *T.gondii* proteins are involved in the process of IRG accumulation on the PVM.[246] ROP18 is a kinase injected into the host cell from the *T.gondii* rhoptries upon infection and is found in a complex with other pseudo-kinases (i.e. ROP8/2) and the dense granule protein GRA7. It could be shown that the parasite secretory kinase rhoptry protein 18 (ROP18) phosphorylates Irga6 and prevents recruitment to the vacuole surrounding virulent parasites. GRA7 binds to oligomers of Irga6, promoting initially polymerization but ultimately leading to rapid disassembly. The activity of GRA7 may prevent proper IRG loading onto the parasite-containing vacuole and may make substrates available for ROP18, thus explaining its contribution to virulence.[246]

These data delineate a lifecycle of Irga6 starting in the GDP-bound state at the resting compartment (golgi/ER) where they interact with the IRGMs. Upon INFγ induction, Irga6 binds to GTP and gets recruited to the PVM, which is directed by the myristoylation of Irga6. At the PVM, Irga6 interacts with the membrane and with GRA7. Whether the binding to GRA7 is required for Irga6 function remains elusive. Irga6 then disrupt the PVM in GTP hydrolysis dependent manner.
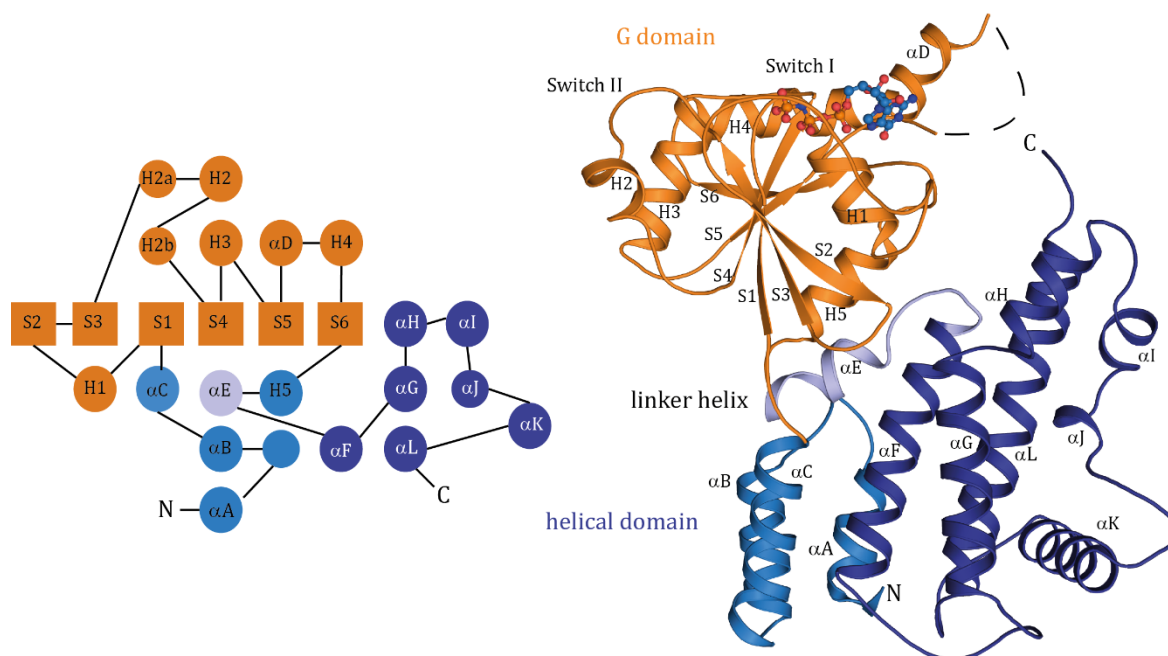
Figure 47: Crystal structure of Irga6-GMPPNP (PDB: 1TQ2) in ribbon representation. G domain is coloured in orange, N- and C-terminal helical regions coloured in marine blue (αA-C) and dark blue (αF-L). The linker helix αE connecting the G domain and the C-terminal helical region is shown in light blue. GMPPNP is shown in sphere-and-stick representation. Carbon atoms are coloured light blue, nitrogen atoms are dark blue, oxygen atoms red and phosphate atoms orange. The topology is shown schematically using the same color code. Figure adapted to [247]. For explanation see text.

In the following, the structural and enzymatic features of Irga6, one of the founding member of the IRG protein family, will be discussed in more detail.

Several crystal structures of Irga6 with and without GDP or GMPPNP bound were determined.[247] Irga6 is composed of a central G domain flanked by an N-terminal and a C-terminal part of the helical domain (see **Figure 47**). The G domain topology is related to H-Ras p21, with additional loop insertions and deletions.[248] It consists of a six stranded β-sheet (S1-S6) of mixed topology, surrounded by six α-helices (H1-H5, αd).

backside dimer
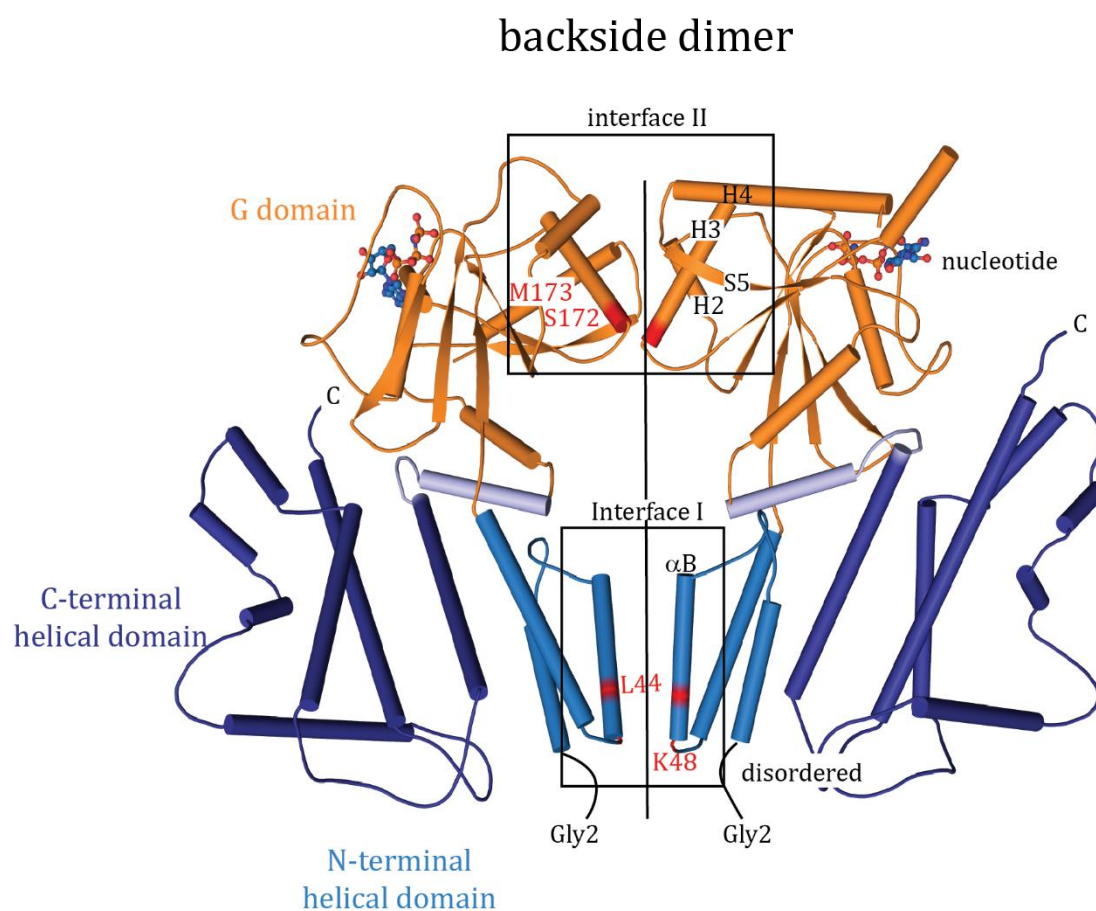
Figure 48: Structure of the GMPPNP-bound Irga6 dimer (PDB: 1TQ2) in ribbon representation. Domains are colored as in **Figure 47**. Secondary structure elements involved in the dimer interface I and II are labelled. The 2-fold non-crystallographic symmetry axis is shown as black line. Residues involved in the interface formation are shown and labelled in red. GMPPNP is shown in sphere-and-stick representation. Carbon atoms are coloured light blue, nitrogen atoms are dark blue, oxygen atoms red and phosphate atoms orange. Residues 1-13 are not resolved and shown as black line with Gly2. Interface I is mainly formed by αB in the N-terminal helical region. Interface II is formed by H2, H3, S5 and H4 of the G domain. For explanation see text.

The helical domain comprises ten helices (αA-L). Whereas αA-C form the N-terminal part preceding the G domain, αF-L constitute the C-terminal region following the G domain. The linker helix αE connects the G domain with the C-terminal helical domain. The helices αA-C and αF form an antiparallel four-helical bundle with a characteristic up-down

topology and establish the interaction between the N- and C-terminus of the helical domain.

All three wild type crystal structures of nucleotide-free, GDP- and GMPPNP-bound Irga6 show dimers, which are related by a 2-fold non-crystallographic symmetry. Thereby, two dimerization interfaces at the backside of the G-interface are formed. Interface I describes the interaction between the N-terminal helical regions of the opposing molecules. The αB helix and Ser18 (αA) are involved in this interface and form a knobs-into-holes side chain interaction. Interface II is formed by H2, H3, S5 and H4 of the G domains, not including the nucleotide-binding pocket (see **Figure 48**). It was proposed that dimerization may therefore stabilize membrane localization by bridging the two myristoylation sites at residue G2 into proximity.[247] Based on structure analysis, mutations within the dimer interface (K48A, L44R, S172R and M173A; see **Figure 48**) were introduced.
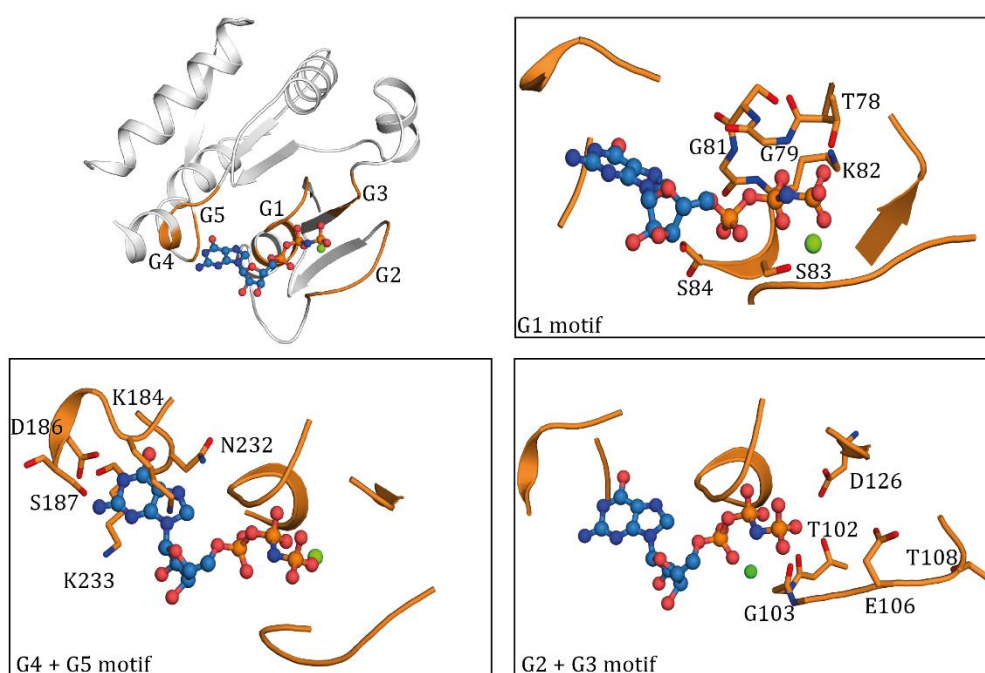
Figure 49: Ribbon-representation of the G domain of GMPPNP-bound Irga6 (PDB: 1TQ2). Upper left: G domain is coloured in white, nucleotide binding motifs G1-G5 are labelled and coloured in orange. Nucleotide is shown in sphere-and-stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. The close-ups show the residues of the nucleotide binding motif G1+G2G3+G4/G5 which are directly involved in nucleotide interaction in stick representation. For explanation see text.

For all mutants, the nucleotide association and dissociation rates are similar to those of the wild-type Irga6 but the mutants failed to display cooperativity in assembly and nucleotide hydrolysis. The activity of the mutants is comparable to the basal activity of

wild-type Irga6. Additionally, the mutants (K48A-GDP, M173A-GMPPNP) crystallized as monomers with minor changes in the G domain but larger movements in the helical domain due to loss of dimerization interface I.[247] Altogether, this indicates, that the backside dimer interface of Irga6 is necessary for oligomerization and GTP hydrolysis.

In the following, the nucleotide binding pocket of wt Irga6 bound to GMPPNP (PDB: 1TQ2) will be described in more detail (see **Figure 49** and **Figure 50**). The α- and β-phosphates form a set of hydrogen bonds to the main chains of G79, G81, K82, S84 (G1 motif/P-loop) and G103 (G2 motif/switch I) as well as to the side chains of K82, S83, S84 (G1 motif/P-loop) and the Mg²⁺ ion. The γ-phosphate contacts the side chains of T78, K82 (G1 motif/P-loop) and the Mg²⁺ ion. The switch I region of Irga6 contains two threonine residues T108 and T102, which could be equivalent to the γ-phosphate sensor T35 in Ras. However, neither is involved in Mg²⁺ or γ-phosphate coordination.
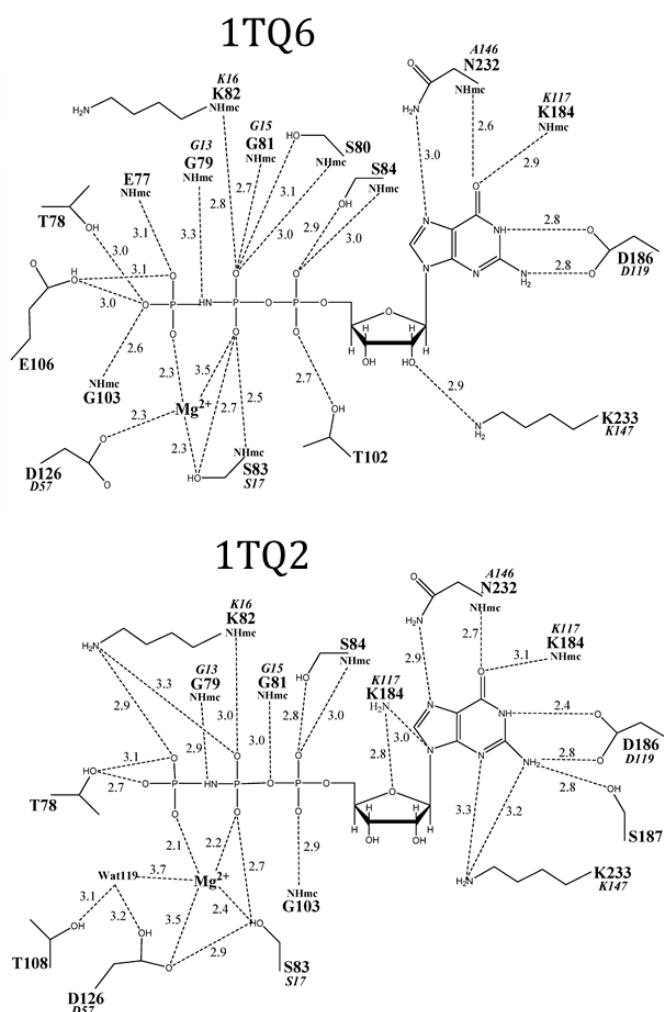
Figure 50: Schematic drawing of the GMPPNP interactions with wtIrga6 (PDB: 1TQ2) and M173A-Irga6 (PEB: 1TQ6). Nucleotide-interacting residues (bold) are labeled. Dashed lines indicate hydrogen bonds. The respective interatomic distances (Å) are annotated. Figure taken from [247]. For explanation see text.

Instead, T78 (G1 motif/P-loop), and in M173A-GMPPNP (PDB: 1TQ6) additionally G103 and E106 (G2 motif, switch I; see **Figure 50**), coordinate the γ-phosphate. The Mg^{2+} ion appears to be loosely bound by D126 (G3 motif/switch II) and S83 (G1 motif/P-loop) and by the β-, γ-phosphates. In contrast to Ras, no direct contact between the γ-phosphate and the G3 motif (^{126}DLPG^{129}) could be observed.[247]

The guanine base forms hydrogen bonds between N1, N2 and O6 and the side chain of D186, the main chain of K184 (^{183}TKVD^{186}/G4 motif) as well as between O6 and N7 and N232. Thus, ^{231}SNK^{233} was identified as G5 motif. Furthermore, the guanine base is sandwiched between K184 and K233. The latter forms flexible side chain contacts to the base or an exocyclic ribose oxygen, whereas K184 contacts the endocyclic ribose oxygen.[247]
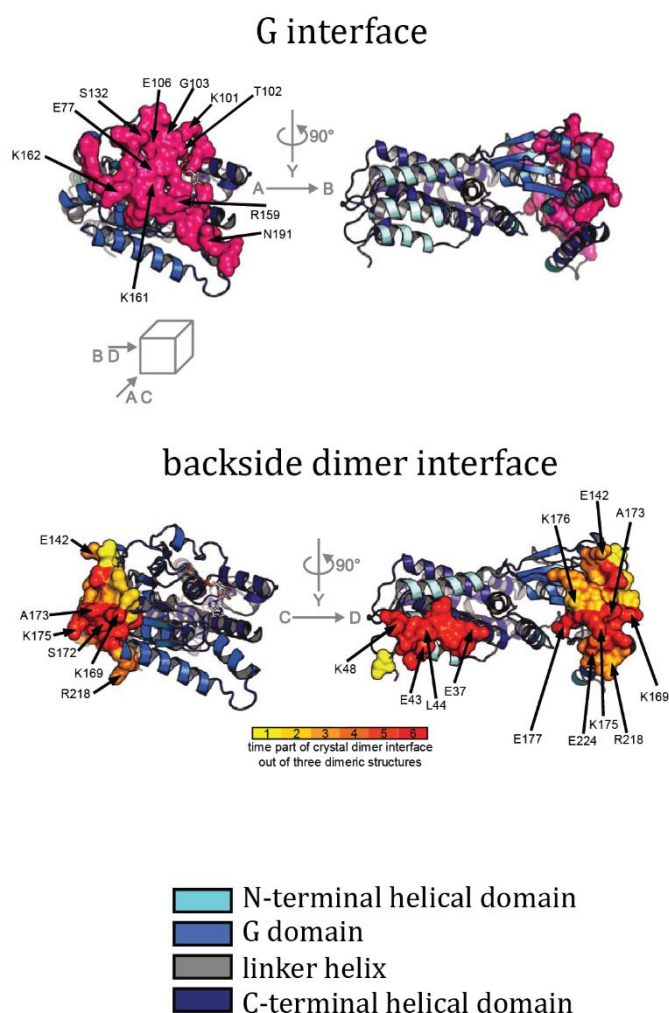
Figure 51: Relative position of the G dimer interface and the backside dimer interface. The structure of Irga6-M173A (PDB: 1TQ6) is shown in ribbon representation. Residues found to be involved in the G interface are shown in surface representation and colored in pink. Three dimeric crystal structures of Irga6 are available (PDB: 1TPZ, 1TQ2; 1TQD) therefore each residue can be maximum six time involved in the backside dimer interface. Residues highly relevant for the backside dimer interface are shown in red, less relevant in yellow. Figure taken from [249].

In contrast to other large GTPases, Irga6 exhibits a rather low enzymatic activity (2/min under saturating GTP conditions) and has a 15-fold higher affinity to GDP (K_D= 1μM) compared to GTP due to higher dissociation rates for GTP (21 s^{-1}) compared to GDP (3.6 s^{-1}). Furthermore, it has been shown biochemically that Irga6 forms GTP-dependent oligomeric complexes *in vitro* and *in vivo* and that GTP hydrolysis is cooperatively activated.[243,250] Besides the shown backside dimer interface, extensive mutational studies revealed an additional interface across the nucleotide binding site equivalent to the G dimer interface in dynamin GTPases, required for accelerated GTP hydrolysis and oligomerization, including the G1, G2 and G3 motif and the nucleotide itself.[249] This G dimer interface is located opposite the backside dimer interface of Irga6 seen in the crystal structures (see **Figure 51**) and was not obtained in the solved crystal structures. The reason for that is, that the GMPPNP-bound form of Irga6 was obtained by soaking GMPPNP into the nucleotide free crystals. This in turn did not allow conformational rearrangements upon nucleotide binding and thus hinders G domain dimer formation. Overall, these findings imply that the biochemically analyzed catalytic interface of Irga6, is a key component for oligomerization and cooperative GTP-hydrolysis and therefore for recruiting to and disrupting of the PVM. However, the structural information about the G domain dimerization of Irga6 has remained elusive so far.

7   Scope of this work

Immunity-related GTPases (IRGs) constitute one of the most powerful cell-autonomous resistance systems against several intracellular pathogens in mice. IRGs accumulate on the *T. gondii* parasitophorous vacuole membrane (PVM) in interferon γ (INFγ) induced cells. Subsequently IRGs participate in the destruction of the PVM via vesiculation and sequestration.

Biochemical studies of Irga6 revealed an interface crucial for cooperative GTP hydrolysis and oligomerization. This interface is established via the G domains and spans over the nucleotide binding pocket of the G domains and involves the nucleotide itself. Furthermore, the residue E106 within the switch I was shown to be essential for catalysis. The available crystal structures of Irga6 revealed a dimeric arrangement via the backside dimer, but could not show the G dimerization. The structures also did not elucidate, whether residue E106 in switch I points towards the nucleotide and works as γ-phosphate sensor and Mg^{2+} coordinator or not.

The main objective of this study was to uncover the G domain dimerization interface of Irga6. Furthermore, I intended to clarify, whether residue E106 is involved in catalysis or whether it has an additional role in G domain dimerization.

Thus, the G domain dimer of Irga6 has to be stabilized prior crystallization. Four mutations were introduced (K31E, R32E, K176E, K246E) into Irga6, which were shown to increase the half-life of the nucleotide dependent complex and, in turn, were proposed to stabilize the G domain dimer. This should allow Irga6 to form a stable G domain dimer interface upon nucleotide binding. In addition to that, the Irga6 mutant was incubated with the GTP-analog GMPPNP prior crystallization in order to allow putative rearrangements upon nucleotide binding and G domain dimerization.

The G domain dimerization mode of GIMAPs, dynamin, Toc and septin GTPases is highly similar. The G domains forming dimers in a *head-to-head* orientation. Contrary to that, based on biochemical analysis Irga6 dimerization in a *head-to-tail* orientation was proposed. The crystal structure shall clarify which orientation Irga6 has in G dimerization. In analogy to the knowledge about oligomerization of dynamin superfamily members, I predicted a nucleotide dependent higher order assembly of Irga6. In case of BDLP, a large rearrangement between the GDP- (kincked) and the GMPPNP-bound (linear) form was observed. Based on these findings in combination with the published biochemical data an oligomerization mode of Irga6 was predicted.

8   Materials and Methods

8.1   Materials

A detailed list of instruments and chemicals can be found in the appendices.

8.2   Protein expression and purification

Full length *mus musculus* Irga6 R31E, K32E, K176E, K246E was cloned in expression vector pGEX-4T-2.[250] Protein was expressed as described here[127] and purified by Nico Pawlowski as described here.[249] Protein was concentrated to 118 mg/mL and stored in aliquots at –80 °C in 50 mM Tris-HCl pH 7.4, 5 mM $MgCl_2$, 2 mM DTT.

8.3   Protein crystallization

The protein solution was gently thawed and diluted to a final concentration of 10 mg/mL with buffer containing 20 mM Tris-HCl, pH 7.5, 8 mM $MgCl_2$, 3 mM DTT. The non-hydrolysable GTP analog GMPPNP was added to a final concentration of 2 mM. Crystallization was carried out in a 96 well format using the sitting drop vapor diffusion method. 70 μL of the premixed reservoir solution were automatically pipetted to the 96 well Greiner Crystal Quick crystallization plates. The sitting drop consisted of 200 nL protein solution and 200 nL reservoir solution using an Art Robbins Gryphon system.

8.4   Cryo-protection of crystals

Crystals were treated with suitable cryo-protectant prior data collection to prevent or alleviate radiation damage. Cryo-solutions were composed of equal volumes of protein buffer and reservoir solution from the corresponding crystallization condition plus varying concentrations of PEG 4000 or 3350. The final cryo solution contained 33% PEG4000, 3% Isopropanol, 50 mM HEPES pH 7.0, 4mM $MgCl_2$, 2 mM DTT and 2 mM GMPPNP. Crystals were soaked into 4 μL cryo-solution at 4°C for at least 5 s before flash freezing in liquid nitrogen. Crystals were screened at beamline BL 14.1 at BESSY II, Berlin.

8.5   Data collection

All data were recorded at beamline P11 at PETRA III, DESY Hamburg using a PILATUS 6M detector. For structure determination, three data sets were collected with a φ increment of 0.05/0.1° at a temperature of 100 K. The used wavelength was 0.979574 Å and the detector distance varied between 1300 and 598.5 mm. Images were recorded with an exposure time of 1 s. Calculation of an optimal data collection strategy was done with the Mosflm software.[131] The datasets were processed using XDS program suite.[132]

8.6   Protein structure solution

General principles of protein structure solution are described in 3.4.4. Structure solution was done by molecular replacement with Phaser [251] employing the structure of the Irga6 protein without nucleotide (PDB: 1TQ2) as search model.[247]

8.7   Atomic model building and refinement

Atomic model building was done by COOT.[252] Iterative refinement was done using Phenix.refine at a maximum resolution of 3.2 Å.[136,253,254] For the refinement strategy a seven-fold non-crystallographic symmetry as well as one molecule Irga6 (PDB: 1TQ4)[247] as high resolution reference structure were chosen. 5% of the measured X-ray intensities were set aside from the refinement as cross-validation.[255]

8.8   Structure analysis and figure preparation

Protein superposition was done with lsqkab [256] and PyMol [257] and figures were prepared using PyMol.

8.9   Protein structure validation and deposition

Evaluation of all-atom contacts and geometry of the atomic model was done by the Molprobity server.[258]

9   Results

9.1   Structural analysis of Irga6

9.1.1   Crystallization and structure determination

To obtain insights into the GTP dependent G-domain dimerization interface of Irga6, full length protein containing the mutations R31E, K32E, K176E, K246E, which have significantly decreased nucleotide binding affinities and show impaired oligomerization,[259] were subjected to crystallization trials (see 8.3).
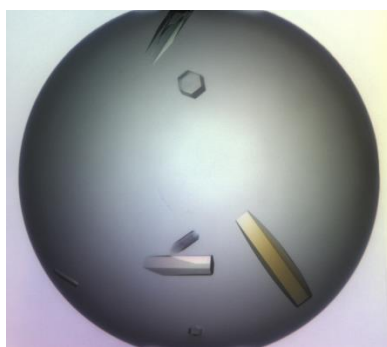
Figure 52: Hexagonal, rod-shaped crystals from Irga6 R31E, K32E, K176E, K246E at day 4 after crystallization set up. Drop size is 400 nL, crystal size is up to 400 μm. Crystallization condition is 100mM HEPES pH 7.0, 9% PEG4000, 6% Isopropanol.

Using the sitting drop method, hexagonal-shaped crystals were grown in the presence of the non-hydrolysable GTP analog GMPPNP using 6-12% PEG3350 or PEG4000, 1-6% Isopropanol as precipitant and 100 mM HEPES pH 7.0-7.5 as buffer system. Crystals appeared after two days at 4°C and grew for two more days.

The crystals with a maximum size of 400 μm x 50 μm x 50 μm were obtained in the optimized crystallization condition containing 100mM HEPES pH 7.0, 9% PEG4000, 6% Isopropanol. To reduce radiation damage during X-ray exposure, crystals were transferred into a cryo-solution containing 33% PEG4000, 3% Isopropanol, 50 mM HEPES pH 7.0, 4mM $MgCl_2$, 2 mM DTT and 2 mM GMPPNP.

X-ray data sets were collected as described (see 8.5). The crystals diffracted X-rays up to 3.0 Å and belong to the hexagonal space group $P6_122$ (178) with the cell dimensions a=b= 98.51 Å, c= 1289.40 Å, α=β= 90.00°, c= 120.00° (see **Table 7**).

Data collection	data set 1	data set 2	data set 3	merged
Beamline	PETRA III P11			
Wavelength [Å]	0.972	0.972	0.979	
Space group	P6$_1$22			
Cell dimensions				
a, b, c [Å]	98.51	98.51	98.96	98.51
	98.51	98.51	98.96	98.51
	1289.40	1289.40	1.286.21	1289.40
α, β, γ [°]	90.00, 90.00, 120.00			
Resolution [Å]^a	50.00 (5.80)-5.48	50.00 (3.16) - 2.99	50.00 (3.06) - 2.90	50.00 (3.20) - 3.00
R$_{meas}$ (%)	10.0 (19.1)	19.5 (97.0)	11.9 (142.7)	22.2 (71.0)
I/σ	24.00 (12.36)	12.03 (1.14)	12.19 (0.8)	10.9 (3.2)
Completeness (%)	99.1 (95.2)	78.1 (37.8)	99.6 (97.5)	100.0 (100.0)
Redundancy	18.3 (13.9)	9.5 (4.2)	7.6 (4.4)	18.4 (6.2)

Table 7: Data collection statistics of three relevant anomalous datasets of Selenomethionine substituted Irga6 R31E, K32E, K176E, K246E. ^{a} Numbers in parentheses apply to the highest resolution shell.

The c-axis is of extraordinary cell length. Longer cell axes have only been observed in five structures currently available in the PDB (24.4.2015), namely the crystal structure of the 70S ribosome of *E.coli* and four viral particle structures. These circumstances required a judicious strategy and set up for structure determination, because the collection of a high resolution data set resulted in the overlap of the detected spots in the high resolution shells (see **Figure 53**). In this particular case, it was not possible to index the collected high resolution data set. To overcome this problem, a low resolution data set was initially collected by using a large detector distance (see **Table 1**, data set 1). Additionally, the exact information of the detector center and detector distance to the crystal was essential for successful indexing of the low/high resolution data sets. This was achieved by measuring well diffracting protein crystals of Thaumatin.
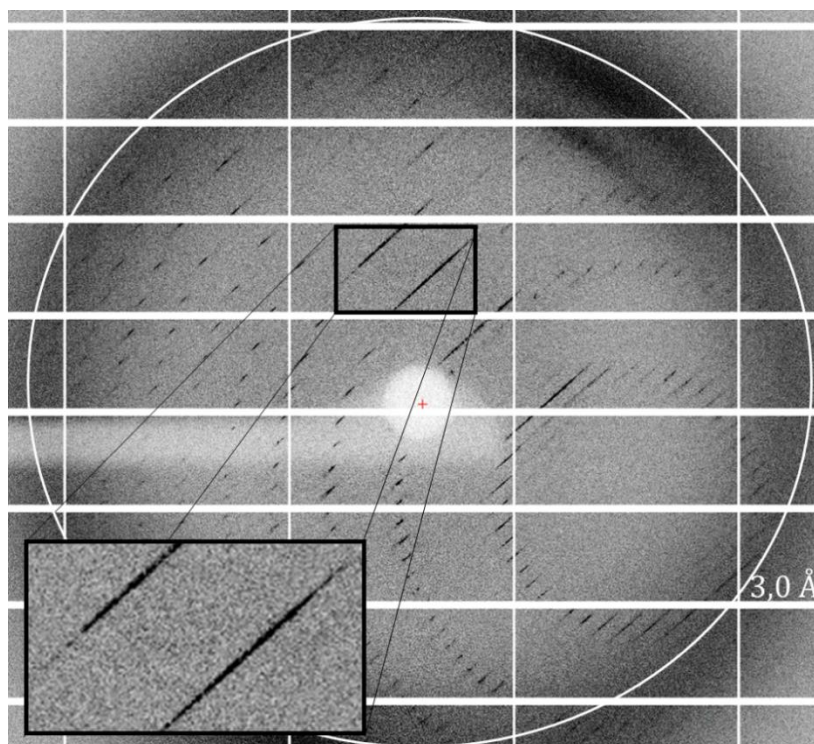
Figure 53: Diffraction image of Irga6 crystals collected by a Pilatus 6M detector. Oscillation range 1°, maximum resolution 2.9 Å. Close up shows the intertwining of the spots in the high resolution range.

The high precision values determined by this method were fixed during processing with XDS. The described combination of different approaches was indispensable to unambiguously separate the spots and successfully index the low resolution data set. An additional high resolution data set was merged to the indexed low resolution data set in order to succeed high-resolution structure solution. For this project, a beamline set up equipped with a state-of-the-art detector with a highly increased sensitivity, pixel number and overall size as the Pilatus 6M was required. Secondly, there was need of a detector that can be moved far away from the mounted crystal due to collection of low resolution data. For structure determination, the three listed isomorphous datasets were merged (see **Table 7**).

The merged datasets were used for molecular replacement as described here (8.6). One Irga6 molecule (PDB: 1TQ2) without nucleotide served as initial search model. In total, seven molecules were found within the asymmetric unit. Due to the use of 7-fold non-crystallographic symmetry and a high resolution reference model (PDB: 1TQ4, chain A 14-413) during refinement, the resulting map was clearly improved compared to the map obtained by molecular replacement. Positive difference density within the nucleotide binding pocket appeared, which allowed fitting of GMPPNP (except in molecule F).
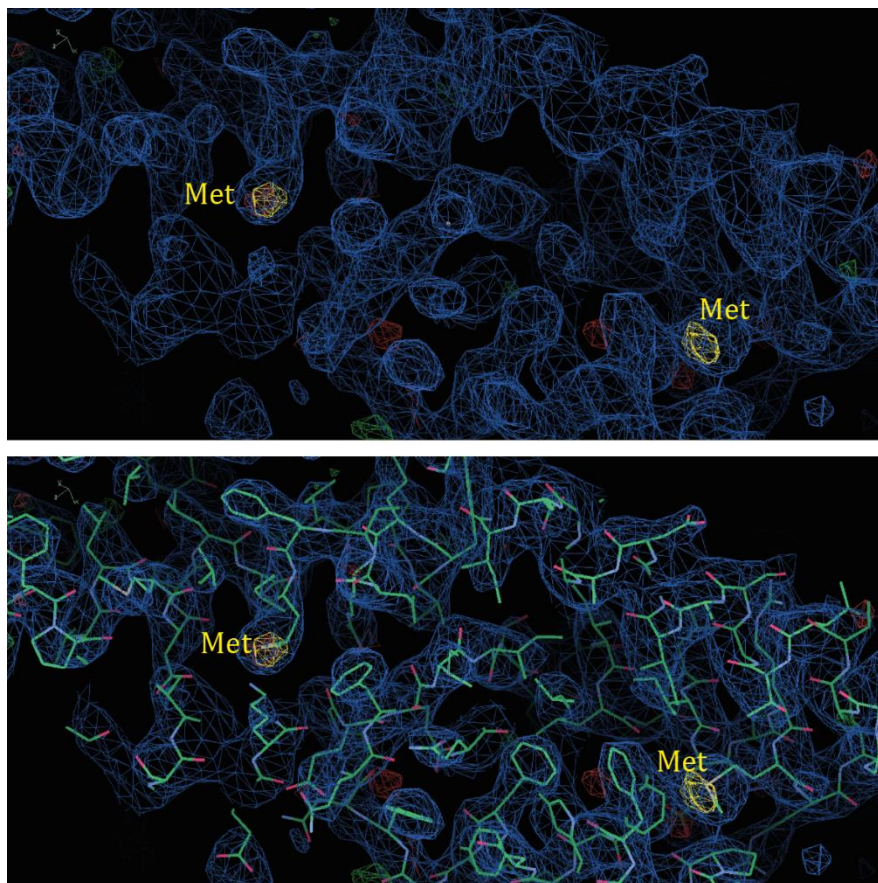
Figure 54: Representative electron density of the Irga6 helical domain. The $2F_0$-F_c map is shown in blue and contoured at 1.12σ. the F_0-F_c map is shown in green (positive) and red (negative) mesh at 3σ, anomalous signal is shown in yellow mesh and contoured at 4σ. Bottom: Model of the helical domain of Irga6 in stick representation fitted into the electron density. Green: carbon atoms, red: oxygen atoms, blue: nitrogen atoms, grey selenium moieties.

The anomalous map was considered during refinement for cross-validation, e.g. to confirm the methionine sites. Thus, the anomalous map was calculated using the anomalous diffraction pattern of the selenomethionine sites. These selenomethionine sites were found using the phases, obtained from the refinement.

The final model consisted of seven molecules named chain A-G. The following residues were included:

chain A	15-199, 203-356, 358-413
chain B	15-355, 358-412
chain C	15-194, 200-332, 335-354, 358-413
chain D	14-196, 201-413
chain E	14-196, 200-355, 357-413
chain F	13-104, 107-195, 199-297, 299-355, 358-413
chain G	14-412

In the final model, 95,6% of all residues are in the favored regions of the Ramachandran plot and 0,2% are outliers, as analyzed by MolProbity.[258] The structure was refined to R_{work}/R_{free} of 29.09% /31.24%. For further refinement statistics see **Table 8**.

Refinement	merged
Resolution [Å]	50.00 - 3.2
No. Reflections	73.219
$R_{work}{}^{a}/R_{free}{}^{b}$	29.1 / 31.2
mol/asu	7
No. protein atoms	22 352
B-factor protein [Å²]	73.52
R.m.s. deviations	
Bond length [Å]	0.002
Angles [°]	0.79

Table 8: Refinement statistics of three merged datasets. [a] $R_{work} = \frac{\sum h,k,l||F_{obs}(h,k,l)|-|F_{calc}(h,k,l)||}{\sum h,k,l|F_{obs}(h,k,l)|}$. [b] R_{free} was calculated with 5% of reflections excluded from refinement.

9.1.2   The overall structure of Igra6, R31E, K32E, K176E, K246E

The monomeric structure of Irga6 R31E, K32E, K176E, K246E is very similar to the GMPPNP-bound structure solved previously (PDB: 1TQ2). Irga6 consists of a central G-domain and a helical domain. The helical domain of Irga6 comprises ten helices. αA-C (residues 14-65) forming the N-terminal part followed by the G-domain. αE (residue 252-265) connects the G-domain to the C-terminal helical domain formed by αF-L (residue 265-413) wherein αF forms a typical four-helical bundle with αA-C (see **Figure 55**).[247]

The G-domain topology (residues 67-252) is highly similar to those found in the Ras family.[248] It contains a six-stranded β-sheets (S1-S6) and six α-helices (H1-H5, αD). The H1-S2 loop corresponding to the switch I (G2) region is followed by the switch II (G3) region (helix H2).[247]

The non-hydrolysable GTP analog GMPPNP binds into the nucleotide binding pocket as known for GTPases and as described already for Irga6 (PDB: 1TQ2). All residues mentioned in the following are oriented highly similar and therefor fulfil the same function except E106 within switch I. The α- and β-phosphates from GMPPNP form hydrogen bonds to the main chain of the P loop (G79-S84) as well as to the side chains of S80, S83, S84 (G1) and T102 (G2). The γ-phosphate of GMPPNP interacts with the side

chain of T78 (G1) and the main chain of G79 (G1) via hydrogen bonding as well as with the magnesium ion. The guanine base forms hydrogen bonds to the side chain of D186 (G4) via the nitrogen moieties N1 and N2 as well as to the main chain of residues K184 (G4) and N232 (G5) via the endocyclic O6 oxygen. Additionally, the side chain of N232 (G5) is involved in hydrogen bonding to the N7 moiety of the guanine base.
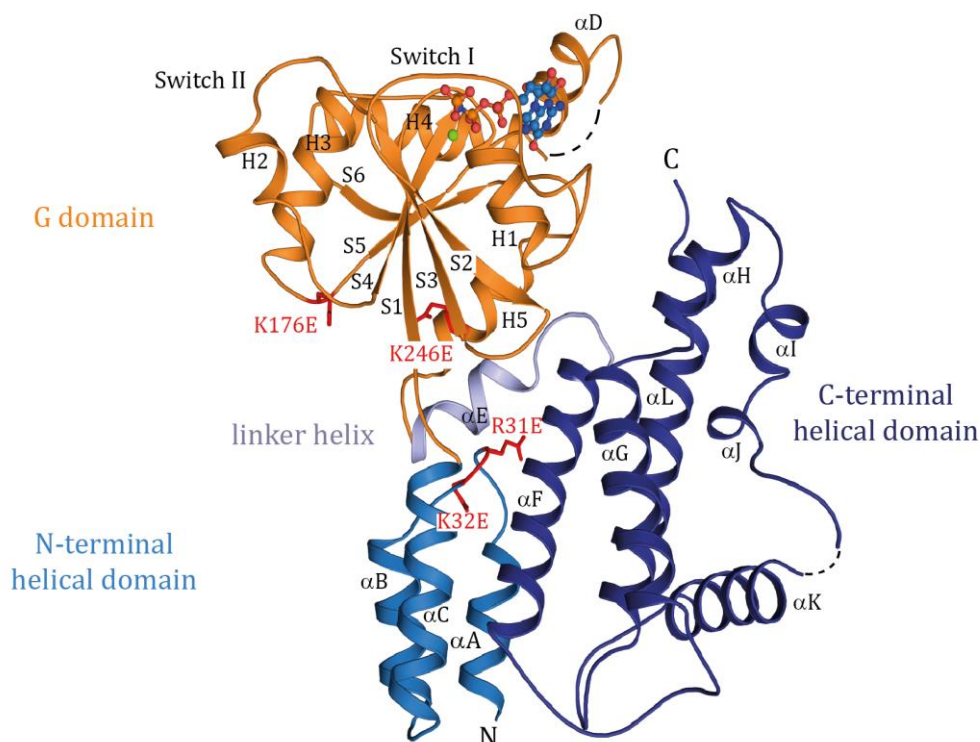
Figure 55: The structure of Irga6 R31E, K32E, K176E, K246E (chain A). A: Domain architecture of Irga6. The first and last residues of each domain are indicated. Mutations are depicted at relative positions in the protein. B: Ribbon presentation of Irga6 R31E, K32E, K176E, K246E. Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphates orange. Mg^{2+} is coloured in green. Domains are color-coded as in A. Regions which are not resolved are indicated as dotted lines. N- and C-termini are labelled. Mutations are shown in stick-representation and are coloured and labelled in red. For explanation see text.

The ribose holds its place due to side chain hydrogen boning via the exocyclic ribose 2'-oxygen and K233 (G5) and via the endocyclic ribose oxygen and K184 (G4) (see **Figure 56** and compare with **Figure 50**). The side chain E106 in the structure solved here positions the magnesium ion whereas in the structure 1TQ2 the side chain positions a water residue and in the structure 1TQ6 it interacts directly with the γ-phosphate of the bound GMPPNP.
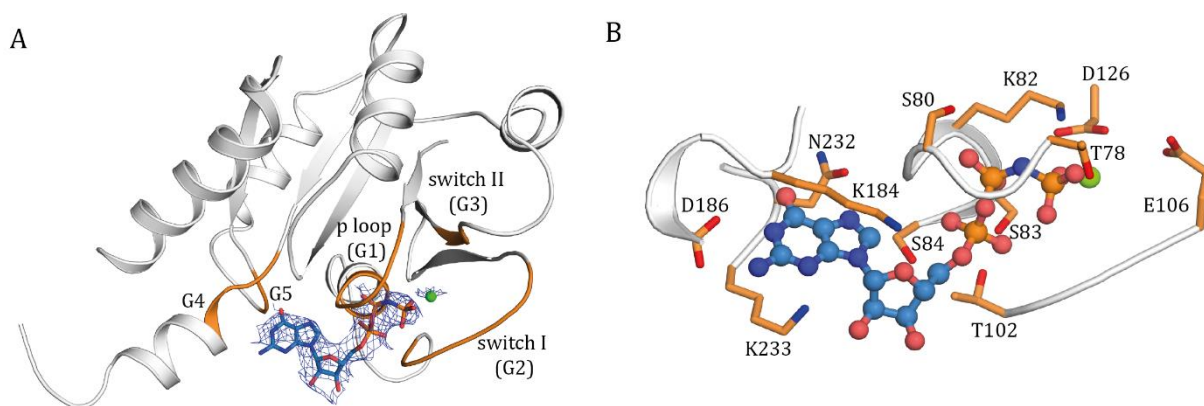
Figure 56: Ribbon representation of the G domain of Irga6 R31E, K32E, K176E, K246E bound to GMPPNP. A: GMPPNP binding to the nucleotide binding pocket of Irga6 G domain mediated by the binding motifs G1-G5 (highlighted in orange). Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphates orange. Mg^{2+} is coloured in green. 2Fo-Fc map for GMPPNP is shown as blue mesh and contoured at 1.1 σ. B: Detailed view of nucleotide binding by the motifs G1-G5 of the G domain. Residues forming hydrogen bonds via side chains are labelled and shown in stick representation. For explanation see text.

9.1.3 Intramolecular interactions within the crystal lattice

Within the crystal lattice, the Irga6 molecules are forming dimers which are organized in a continuous helical arrangement. The unit cell consists of seven Irga6 molecules. Six of them form dimers within the unit cell and the remaining seventh molecule (molecule G) forms a dimer with its symmetry mate of the neighboring unit cell (molecule G') related by a 2-fold crystallographic symmetry (see **Figure 57**).

	chain A	chain B	chain C	chain D	chain E	chain F	chain G	chain G'
chain A	0.00	0.36	0.45	0.39	0.34	0.43	0.32	
1TQ6	1.11	1.06	1.13	1.00	1.09	1.18	1.00	
1TQ2	3.68		3.44		3.46		3.60	

Table 9: Calculated r.m.s.d values [Å] from superposition of 1TQ6, 1TQ2 and chain A to the structure of Irga6 R31E, K32E, K176E, K246E (chain A-G). For explanation see text

The protein superposition of the molecules within the unit cell and to the already known structures 1TQ6 and 1TQ2 was done as described in 8.8. The r.m.s.d. calculation was based on the peptide backbone only.

All seven molecules within the unit cell reveal a highly conserved overall fold with an r.m.s.d below 0.50 Å, as evaluated by superposition of chain A to chain B-G.
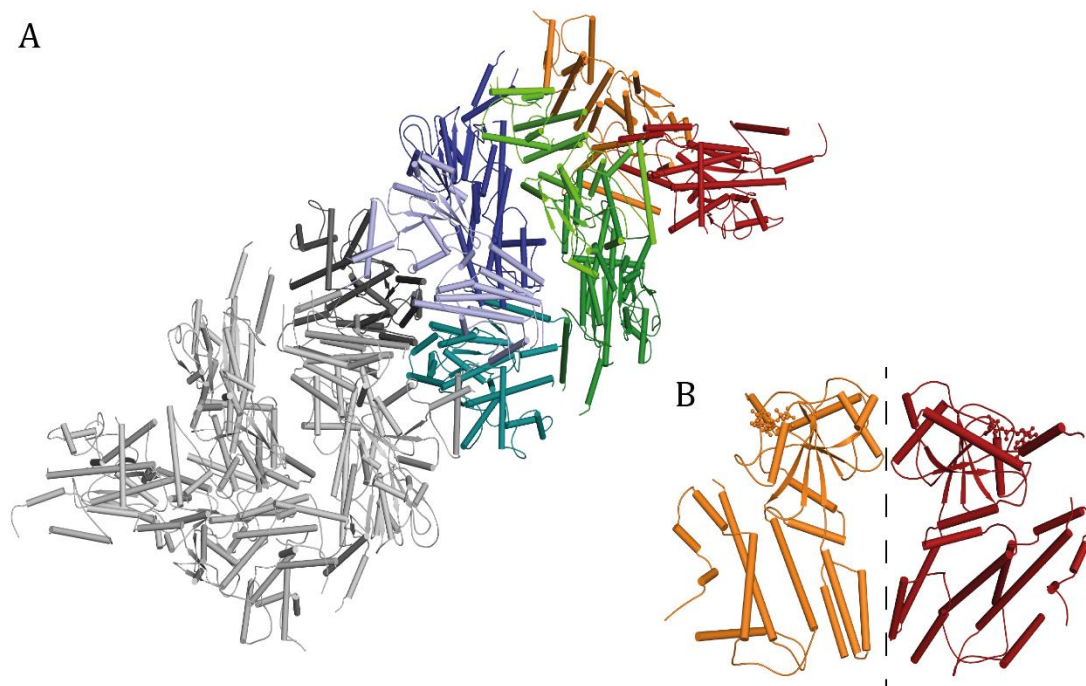
Figure 57: Overall organization of the crystal packing. A: Ribbon representation of seven Irga6 molecules in one unit cell (coloured) connected to the neighboring unit cell (grey). (chain A: red, chain B: orange, chain C: green, chain D: dark green, chain E: dark blue, chain F: light blue, chain G: cyan, chain G': dark grey). Dimer formation between chain A and B, chain C and D, chain E and F, chain G and G'. B: Ribbon representation of Irga6 dimer formation (chain A and B) via the so called backside interface. Nucleotides are represented in sphere and stick-representation in the same color as the molecules. For explanation see text.

The superposition of the monomeric Irga6 M173A structure loaded with GMPPNP (PDB: 1TQ6) to each chain of the Irga6 structure solved in this study reveals almost no conformational changes as indicated by an r.m.s.d not higher than 1.20 Å (see **Table 9** and **Figure 58**).

Superposing the dimeric Irga6 structure loaded with GMPPNP (PDB: 1TQ2) to each individual dimer (chain A + B, chain C + D, chain E + F, chain G + G') appearing in the crystal lattice revealed some differences as discernible in an increased r.m.s.d of approximately 3.50 Å (see **Table 9** and **Figure 58**). Whereas the helical domains superpose very well, the G domains (**Figure 58**, lower right) are shifted against each other along the z axis by 15 Å overall.

This rearrangement of the G domains towards each other cause some changes in the interface formation of the backside dimer as it will be discussed in the following section.
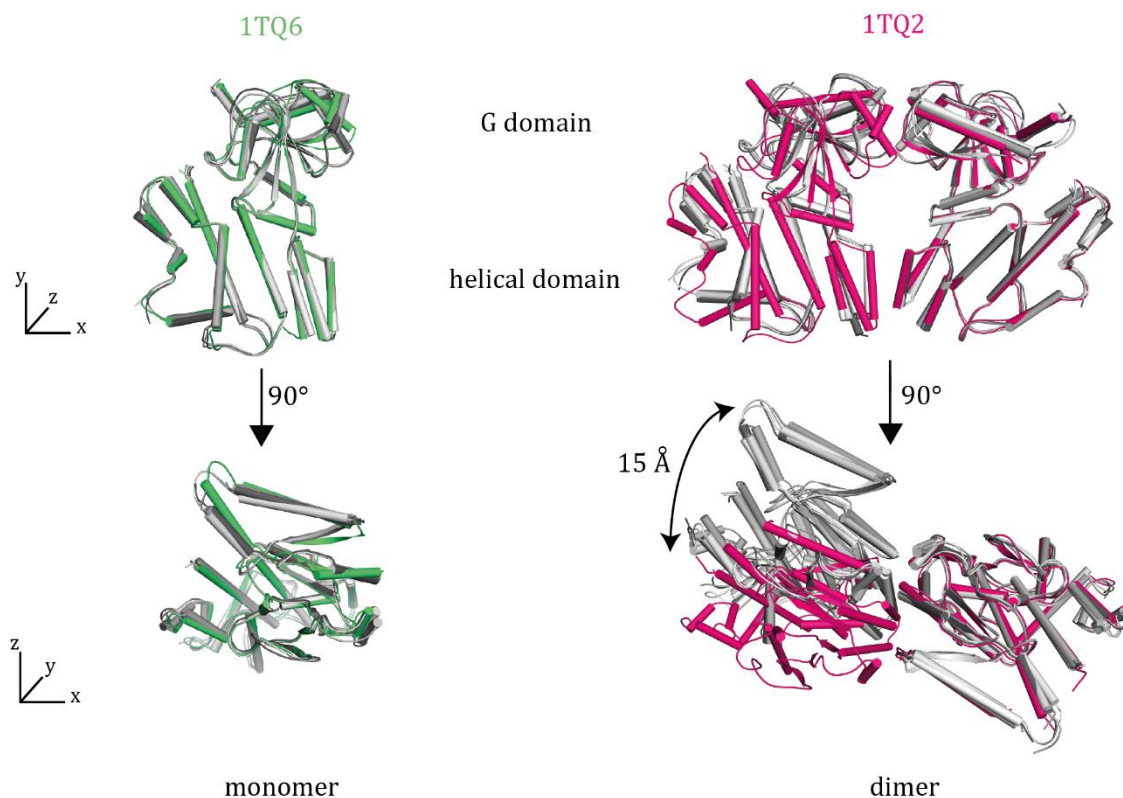
Figure 58: Ribbon representation of superposition of Irga6 R32E, K32E, K176E, K246E monomers (chain A+C+E+G)) (coloured in different grey scales) to1TQ6 (green, left) or Irga6 R32E, K32E, K176E, K246E dimers (chain A+ B, chain C + D, chain E + F, chain G + G') (coloured in different grey scales) to 1TQ2 (pink, right).

9.1.4 The backside dimer of Igra6 R31E, K32E, K176E, K246E

The backside dimer of Irga6 R31E, K32E, K176E, K246E is formed via two interfaces (see **Figure 59**). Interface I is built by the αB helices of the N-terminal helical domain and is identical to the interface described by Ghosh et al (see 6.3.5).[247]

The interaction takes place via a knobs-into-holes arrangement of side chains packing against each other.

The hydrophobic interactions are mainly mediated by N40, L41, L44, R47 and K48 of helix αB and the opposing helix αB'. The residues R31E and K32E mutated in this study are not involved in the interface formation.

The previously mentioned reorientation of the G domains within the backside dimer leads to changes in the interaction interface II compared to Ghosh et al.[247] This interface is mainly characterized by hydrophobic interactions of residues in helix H2 (135-150) and helix H3 (162-177).
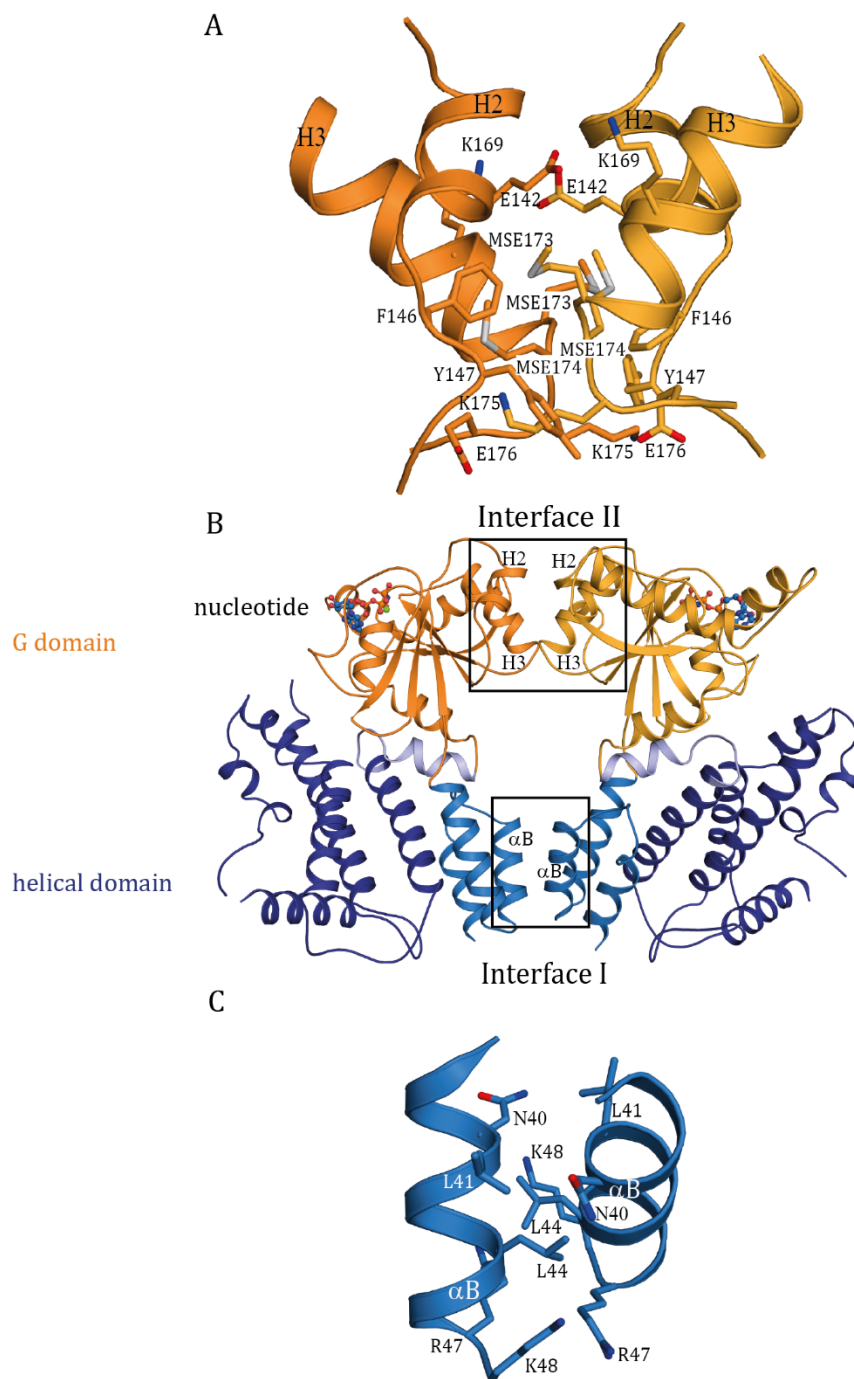
Figure 59: The Irga6 R31E, K32E, K176E, K246E backside dimer interface shown in ribbon representation. A: Interface II formed by H2 and H3 of the G domain. Residues which are involved in the interface are labelled and shown in stick representation. MSE= Selenomethionine. B: Overview of dimerization. Domains are coloured as in **Figure 55**. Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphates orange. Mg^{2+} is coloured in green. Helices involved in backside dimer interaction are labelled and the interfaces are highlighted and labelled. C: Interface I formed by helix αB in the N-terminal helical region. Residues involved in the interface are labelled and shown in stick representation.

The network of hydrophobic interactions is formed by helix H3 and helices H3' and H2' of the opposing molecule. The side chains of the following residues in H3 are involved in this

hydrophobic network: MSE173 → F146', K175 → Y147', MSE174 → MSE173' and vice versa.

Four hydrogen bonds are further stabilizing this interface. The sidechains of K175 and the mutation K176E of helix H3 are hydrogen bonding to K176E' and K175' of helix H3'. The sidechain of residue E142, located in helix H2, also forms a hydrogen bond to K169 (H3') and vice versa. Due to G domain reorientation, H4 and S5 (see **Figure 48**) are not involved in the formation of interface II in contrast to the structure described by Gosh et al.[247] The mutation K246E is not involved in the backside dimer interface.

9.1.5   The G domain dimer

Apart from the backside interaction, one backside dimer (I: molecule A + B) forms an additional G dimer interface with its symmetry mate (II: molecule A' + B') from the adjacent unit cell related by a 2-fold crystallographic symmetry (**Figure 60**).
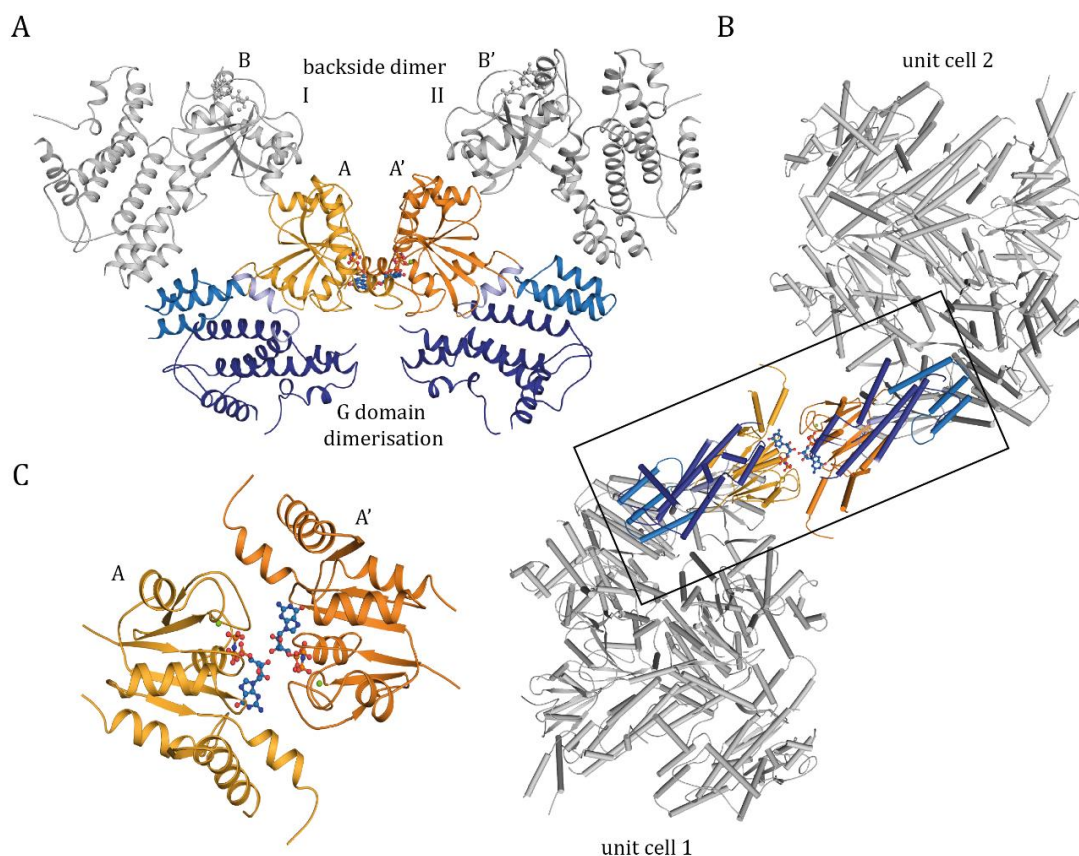
Figure 60: G domain dimer formation in ribbon representation. Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. Domains are color-coded as in **Figure 55**. A: Close up of the backside dimer I (molecule A + B) and II (molecule A' + B') involved in G domain dimerization. B: Adjacent unit cells 1 and 2 connected via the G domain dimerization of the backside dimers I + II. C: Close up of G domains of molecule A+A' responsible for G domain dimerization. For explanation see text.

This G dimer is formed by just one monomer of each backside dimer (molecule A + A') (**Figure 60**). Directly involved in this G domain interaction are the switch I region, the G4 loop as well as the *trans* stabilizing loop, the switch II loop directly following the switch II motif and the nucleotide itself. G domain dimerization is mediated mainly by hydrogen bonds and hydrophobic interactions and can be subdivided into three interfaces (see **Figure 61**).
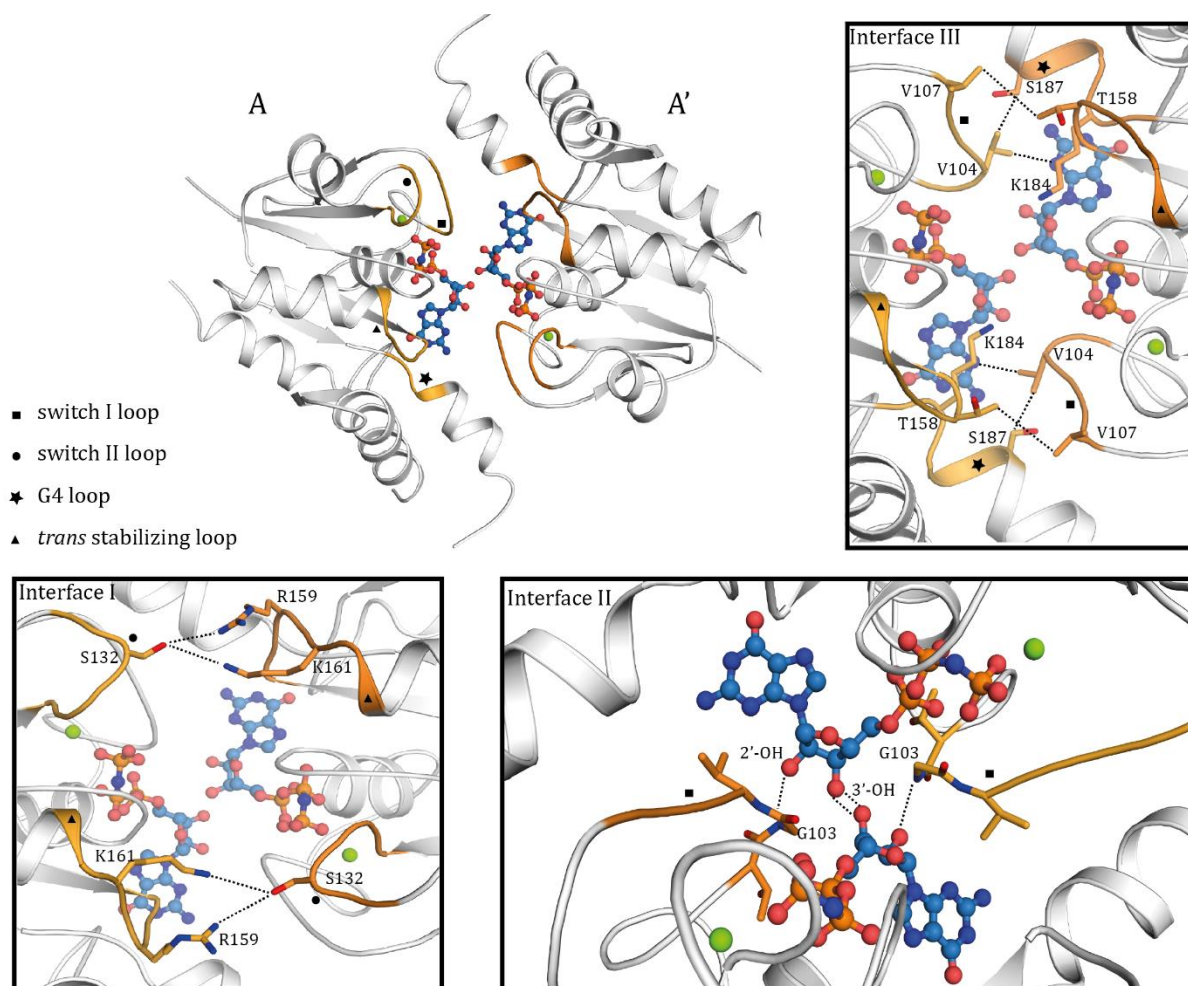
Figure 61: The G dimerization mode. G domains are shown in ribbon representation. Upper left: Overview of G domain dimerization. G domains of molecule A and A' are shown in ribbon-representation and are coloured in white. Switch I loop, switch II loop, G4 loop and *trans* stabilizing loop directly involved in G dimer formation are coloured in yellow (molecule A) and orange (molecule A'). Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. Interface I-III: Residues involved in G domain formation are shown in stick representation. Interactions between residues are shown as dotted lines. Interface I: Hydrogen bonding between *trans* stabilizing loop and S132. Interface II: Hydrogen bonding between nucleotide and G103 in *trans* and between the opposing nucleotides via the 3'-OH group of the ribose. Interface III: Hydrophobic interactions between switch I and G4 and the *trans* stabilizing loop.

Interface I is formed by hydrogen bonds between R159 and K161, located in the *trans* stabilizing loop, and S132 belonging to the switch II loop of the opposing G domain and vice versa.

Interface III is a hydrophobic patch formed by the switch I of one G domain with the G4 loop as well as with the *trans* stabilizing loop from the opposing G domain. V104 (switch I) forms hydrophobic contacts to S187 and K184 (G4 loop) while V107 (switch I) interacts with T158 (*trans* stabilizing loop).

Interface II has a crucial role in G domain dimerization and is mediated by the ribose of GMPPNP. The exocyclic 2'-OH group of the ribose interacts in *trans* with the main chain of G103 (switch I) via hydrogen bonding whereas the exocyclic 3'-OH group of the ribose forms a hydrogen bond with the 3'-OH group of the opposing nucleotide.

Overall, this newly found G domain interface confirms the biochemical results [249,259] in large parts and gives additional structural insights into the G domain interaction of Irga6. Although, it remains elusive how the oligomerization mechanism of Irga6 takes place in detail, an important key interface is now identified and allows some tentative conclusions which will be discussed in the next section.

10   Discussion

10.1   Mechanistic insights into Irga6 oligomerization

The involvement of Irga6 G domain dimerization was investigated by Nico Pawlowski via biochemical methods.[249,259] The suggested model for the G domain dimer interface was based on the crystal structure of the protein dimer Ffh-FtsY (see 6.1.2 and **Figure 62**).
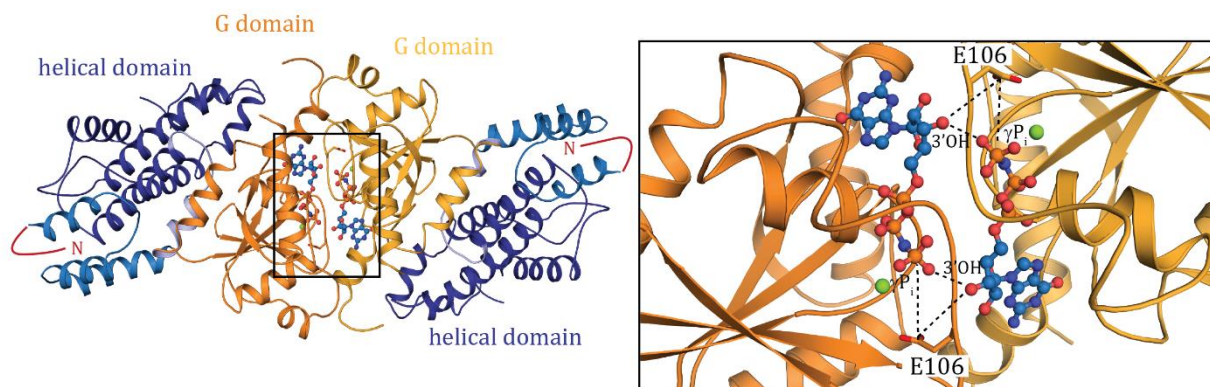
Figure 62: Model of Irga6 G domain dimerization based on the known structure of Ffh-FtsY (PDB: 1RJ9). Left: Ribbon-representation of the G domain dimerization via the *head*-to-*tail* oriented GTP-analog GMPPNP. Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. N-terminus is not resolved. Residues 1-14 are shown as red line. Right: Close up of the G domain dimerization interface of Irga6, which shows the predicted influence of E106 (shown in stick representation) in *cis* and *trans* if loaded with GMPPNP. For explanation see text.

Although usually discussed in the context of the dynamin superfamily members, Irga6 and Ffh-FtsY share some similarities such as the low nucleotide-binding affinities, the wide-open binding pocket and the key idea that the nucleotide is directly involved in the G domain dimerization. The model shows a G domain interface with nucleotides in a *head*-to-*tail* orientation and the 3'OH group of the ribose interacting with the γ-phosphate in *trans*. Furthermore the model suggests that the residue E106 is interacting in *cis* with the γ-phosphate and in *trans* with the 3'OH of the ribose promoting the G domain interface.
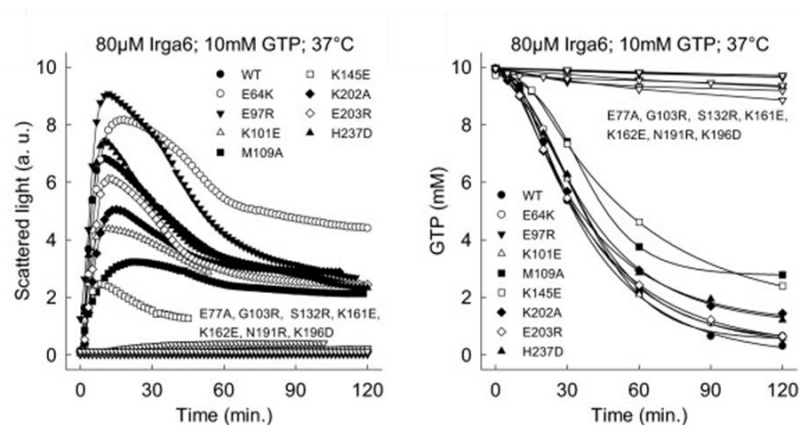
Figure 63: The G domain involved in Irga6 oligomerization. Left: Oligomerization of wt or mutant Irga6 in presence of GTP detected by light scattering. Right: GTP hydrolysis measurement of Irga6 wt and mutants. Samples assayed by TLC and autoradiography. Figure taken from[249]. For explanation see text.

In this context, the solved crystal structure now clarifies the G domain dimerization mode, which is in large parts in agreement with the biochemical data[249,259] and strongly opposes the presented model. The biochemical studies showed that mutation of the residues E77, G103, E106, S132, R159, K161, K162, D164, N191, K196 prevent GTP hydrolysis and oligomerization completely (see **Figure 63** and **Figure 64**).
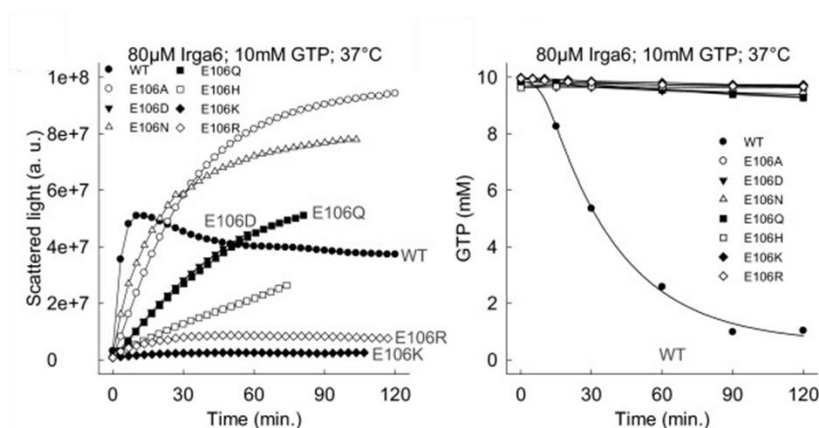
Figure 64: The Glu106 is essential for the activation of GTP hydrolysis. Left: Oligomerization of wt. or Glu106 mutant Irga6 monitored by light scattering. Right: GTP hydrolysis measurements of Irga6 wt or mutants. Samples assayed by TLC and autoradiography. Figure taken from [249]. For explanation see text.

10.1.1   Catalysis

Residue E106 is part of the flexible switch I region which undergoes nucleotide-dependent conformational changes. In the structure solved in this work, this residue seems to play a role in positioning the Mg^{2+} ion in *cis*, similar to the related T65 in dynamin, and is not involved in a *cis* and *trans* interaction network with the 3'OH of the

ribose and the γ-phosphate of the GMPPNP as predicted in the previous model (see **Figure 62**). Mutating this residue led to a complete inhibition of GTP hydrolysis (see **Figure 64**) which is indicative of its important function and underlines that E106 could act as key residue in the nucleophilic attack.[249] Furthermore, in the previously solved Irga6 structure bound to GDP (PDB: 1TPZ) residue E106 pointed into the opposite direction and thus is surface exposed (see **Figure 65** right).
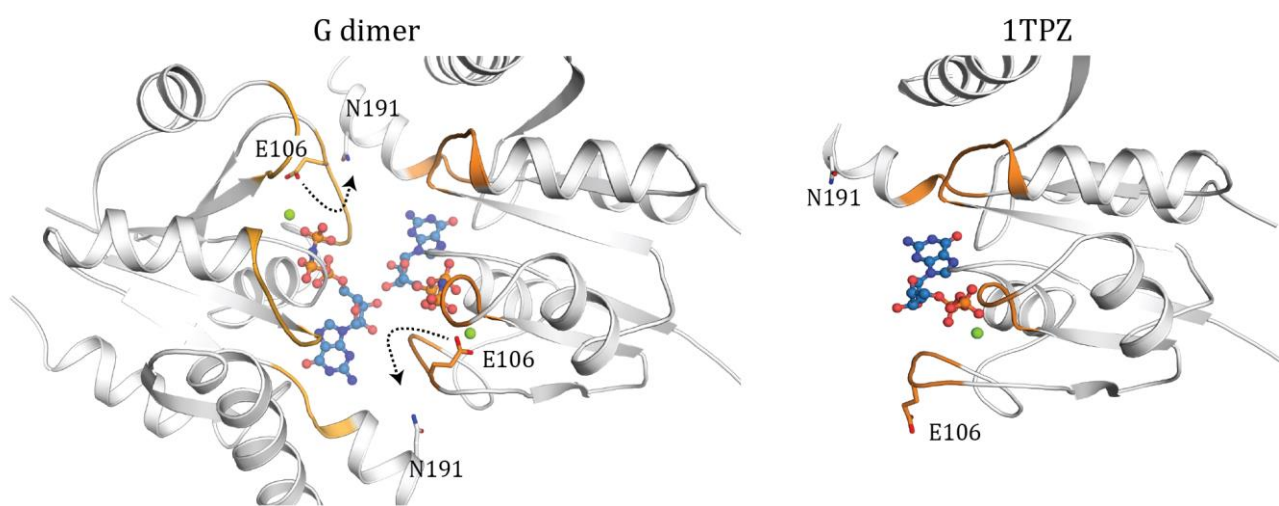
Figure 65: Ribbon representation of the G domain of the G dimer (left) and Irga6 bound to GDP (PDB: 1TPZ). Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. Residues E106 and N191 are shown in sphere- and stick-representation.

If one would confer the orientation of E106 from the letter mentioned structure (PDB: 1TPZ; see **Figure 65**) to the structure solved in this work, the residue E106 would point into the dimerization interface in very close proximity to residue N191 and could therefore establish a hydrogen bond (see **Figure 65**, arrows). This theory is quite interesting due to the fact that N191 was also biochemically identified to completely inhibit GTP hydrolysis and oligomerization (see **Figure 62**). This would substantiate the accepted theory that E106 works as a switch. In *trans* it could first support the formation of the dimer interface whereas subsequently it has its role as coordinator of the Mg^{2+} in *cis* due to GTP hydrolysis.

10.1.2   The G domain interface

The inhibition of GTP hydrolysis and oligomerization by residue G103 within the switch I can be explained by the G dimer structure. The backbone nitrogen stabilizes in this

structure stabilizes neither the α-phosphate (PDB: 1TQ2) nor the γ-phosphate (PDB: 1TQ6) in *cis* (see **Figure 50**) but rather the position of the nucleotide in *trans* via hydrogen bonding to the 2'OH of the ribose and therefore promotes the orientation of both 3'OH groups of the ribose relative to each other (see **Figure 61**).

The 3'OH of the ribose was suggested to interact in *trans* with the γ-phosphate (see **Figure 62**). Biochemical data showed that the 3'OH group is mandatory for GTP hydrolysis and oligomerization (see **Figure 66**). This can only be explained by a cooperative involvement, because in the Irga6 monomer the 3'OH group is in a surface exposed orientation in the wide-opened nucleotide-binding pocket. Indeed, the solved crystal structure of the G dimer provides a convincing explanation for this finding, as the 3'OH group forms a hydrogen bond to the 3'OH group of the opposing nucleotide and is therefore directly involved in G domain dimerization.
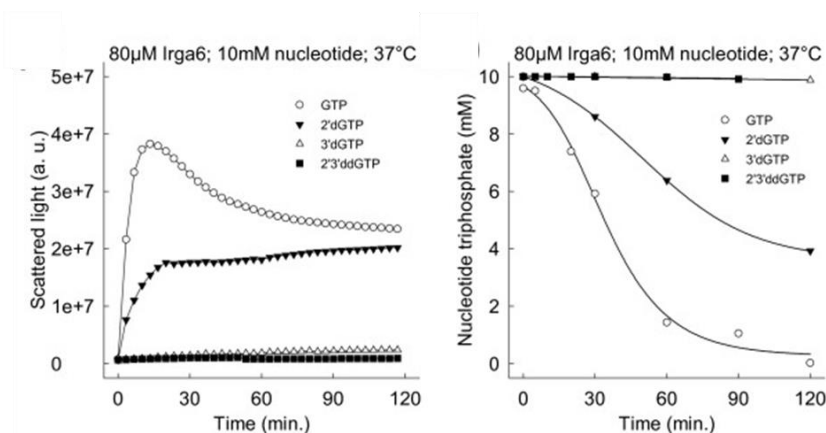
Figure 66: The GTP ribose 3'OH is essential for the activation of oligomerization and GTP hydrolysis. Left: Oligomerization of wtIrga6 in presence of GTP, 2'dGTP, 3'dGTP, 2',3'ddGTP monitored by light scattering. Right: Hydrolysis of GTP, 2'dGTP, 3'dGTP, 2',3'ddGTP in presence of wt Irga6. Samples assayed by TLC and autoradiography. Figure taken from [249]. For explanation see text.

Finally, residue R159 together with residue K161 forms a hydrogen bonding network to residue S132 of the opposing molecule. This is one of the most important interfaces in the G domain dimerization, as the biochemical data show abolished G domain dimerization and oligomerization for each of these residues (see **Figure 63** and **Figure 61**).

10.1.3   Other residues

Contrary, the residues K162 and D164 are influencing GTP hydrolysis and oligomerization but they are not part of the G domain dimer interface of the crystal structure solved in this work. Instead they form a hydrogen bonding network together

with residue D166 which in turn could help to orient the *trans* stabilization loop in the correct position and to establish the G domain dimerization interface of Irga6.

In case of residue K196, the biochemical results cannot be explained by the structure solved in this work as the residue is quite far away from the dimerization interface. It could be conceived that this residue is involved in another oligomerization interface in order to form Irga6 oligomers.

Also the residue E77 shows no interaction with the opposing molecule in order to stabilize the G domain dimer interface, although it is important for GTP hydrolysis and oligomerization. However, it could be envisaged to stabilize the residue K161 and therefor contribute to the assembly of the *trans* stabilization loop.
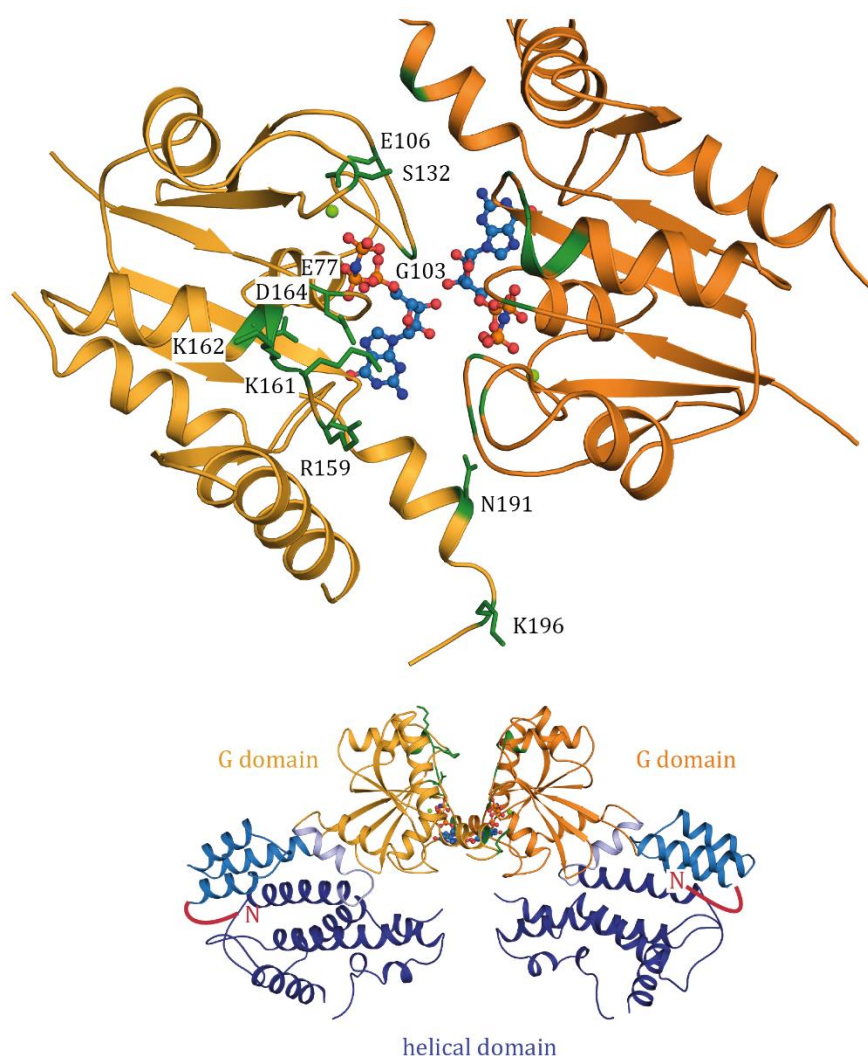
Figure 67: The Irga6 G domain dimer. Ribbon representation of Irga6 R31E, K32E, K176E, K246E. Residues which were biochemically shown to interfere G domain dimerization and oligomerization are shown in stick representation and labelled and colored in green. Nucleotide and Mg^{2+} ion are shown in sphere and stick representation. Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. Domains are color-coded as in **Figure 55**. Red lines represent the not resolved N-terminus (residue 1-13) with the myristoylation site at the N-terminus. For explanation see text.

Overall, there are marked differences between the previously suggested model of the G domain dimer and the crystal structure solved in this work (compare **Figure 67** and **Figure 62**). The solved structure of the G domain dimer resembles a butterfly with the G domain pointing to the top and the helical domain pointing to the bottom. On the contrary, the antiparallel model (see **Figure 62**) shows a more elongated Irga6 dimer with the G domains in the core of the dimer and the helical domains flanking this core in opposing directions. It also shows a buried nucleotide binding pocket where the nucleotide is oriented in a *head*-to-*tail* orientation interacting via the 3'OH group of the ribose and the γ-phosphate in *trans*. In comparism to that, the G dimer structure solved here reveals a G domain dimer in a more open conformation (see **Figure 67**) with a nucleotide interaction only via the 3'OH group. This hydroxyl group can be seen as the core of a cross formed by the two nucleotides which are rotated 90° to each other.

10.1.4   The backside dimer

As discussed, the crystal structure of Irga6 presented in this work shows the G domain dimerization of Irga6 in detail. However, the extent to which the backside dimer contributes to the oligomerization of Irga6 remains to be seen. The backside dimer is built out of two interfaces. Interface II in our crystal structure differs to some extent from interface II seen in the already obtained crystal structure of Irga6 (see 9.1.4). Nevertheless, biochemical investigations showed that mutations of the residues L44 and K48 in interface I and residues S172 and M173 in interface II reduce the cooperativity of GTP hydrolysis significantly as well as the oligomerization (see **Figure 48**).[247] Except for S172, these residues are also involved in backside dimer formation in the structure solved here (see **Figure 59**). Interestingly, the backside dimer is only slightly rearranged upon GTP binding and G dimer formation (interface II), but overall forms a stable minimal building block. Additionally, mutations of the residues R31, K32, K176 and K246 were shown elsewhere to diminish GTP hydrolysis and oligomerization.[259] Interestingly, theses residues are close to but not part of the backside dimer interface in any of the various crystal structures, except for residue K176 (see **Figure 55**). This could give a hint to an additional oligomerization interface not seen yet in the crystal structures or implies that this residues form a rearranged backside dimer interface together with the already known backside dimer residues.

Overall the solved crystal structure of Irga6 R31E, K32E, K176E, K246E is in agreement with the biochemical data and reveals the mode of G domain dimerization. Nevertheless, both the model and the solved crystal structure, give no explanation how the helical domain containing the myristoylation site G2 at the N-terminus (see red lines in **Figure 67** and **Figure 62**) can assemble around a convex parasitophorous vacuole membrane barring some large rearrangements.

10.2   Comparism of Irga6 with Atlastin

Human atlastin 1 (Atl1) is the member of the dynamin superfamily, which has together with GBP the closest phylogenetic distance to Irga6. Due to this, Irga6 will be compared in detail with Atl1 as a representative for the dynamin superfamily.
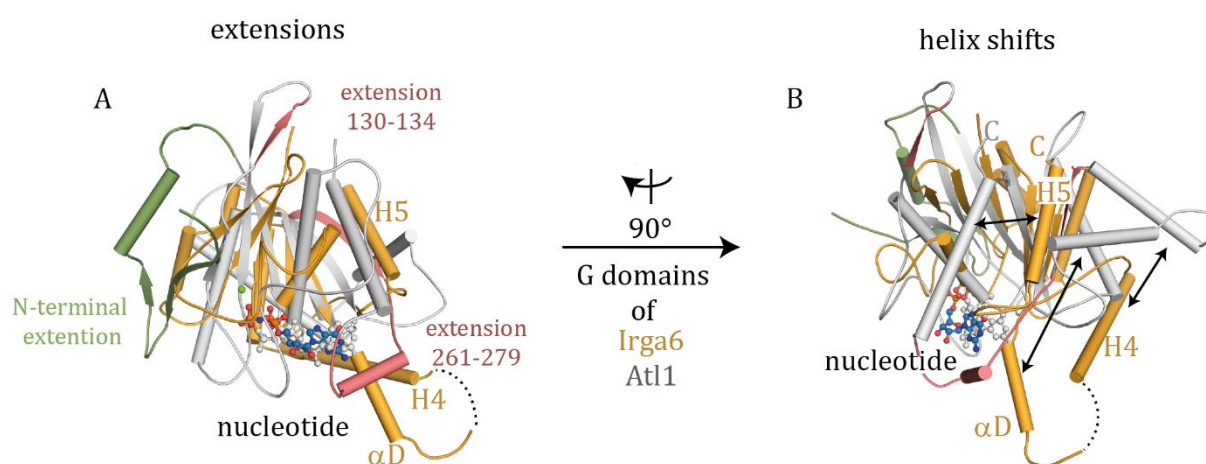
Figure 68: Superposition of the G domains of Irga6 and Atl1 (PDB: 4IDN) shown in ribbon representation. Yellow: Irga6 bound to GMPPNP (shown in sphere and stick representation). Nucleotide oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. Grey: Atl1 bound to GMPPNP (shown in sphere and stick representation). Salmon: extensions in Atl1 within the G domain. Green: N-terminal extension of the G domain of Atl1. A: Overlay of both G domains. B: Overlay of both G domains rotated around 90° compared to A. Shifts of the helices are labelled with arrows. For explanation see text.

As introduced in chapter 6.2.3, Atl1 was crystallized in 3 different isoforms which revealed no changes in the G domain dimerization interface. Comparing the G domain organization of Atl1 and Irga6, some differences can be identified. Atl1 belongs to the dynamin superfamily and contains several insertions besides the typical Ras-fold (see 6.2.1). Atl1 contains an N-terminal extension of the G domain (see **Figure 68**, coloured in green). Additional to that, Atl1 possesses an extra sheet (residue 130-134) between the G2 and G3 motif and, more strikingly, a larger insertion between the G3 and G4 motif compared to Irga6. The helix of Atl1 comparable to helix H4 of Irga6 is extended by 60

residues that form a sheet-helix arrangement (residues 261-279), which is missing in Irga6. The G4 motif in Irga6 faces only the nucleobase and the binding pocket is wide open. In Atl1, the G4 motif is located in a loop (residues 271-293), which folds over the complete nucleotide and closes the nucleotide binding pocket as a lid. Altogether, this insertion in Atl1 lead to a rearrangement of the helices comparable to αD, H4 and H5 in Irga6 (see **Figure 68**).

Consequently, the G domain dimerization interface differs as well. In case of Atl1, the G domain dimerization interface encompasses a much larger surface compared to Irga6. Irga6 only forms a hydrophobic interface via switch I and the G4 loop, hydrogen bonds between switch I and the *trans* stabilizing loop, a hydrogen bonding network between the switch II loop and the *trans* stabilizing loop and hydrogen bonding via the nucleotide itself. On the other hand Atl1 forms a large hydrophobic and hydrogen bonding network involving the switch I loop, switch II loop, G3 loop and the G4 loop. It could be shown that further residues not located in these known nucleotide binding loops are additionally involved in G domain dimerization.[195] Residue R77 (switch I) was proposed to act as a molecular switch upon nucleotide exchange (see 6.2.3). Possibly, E106 has the same role in Irga6. Interestingly, the nucleotide is not involved in the Atl1 dimerization at all in contrast to Irga6 dimerization.

For both proteins, a linker region follows C-terminally from the G domain connecting the helical domain (Irga6) and middle domain (Atl1), respectively.

The striking difference is the distinct localizations of the middle domain of Atl1 relative to the G domain. The middle domain has to rotate by 90° to undergo the rearrangement which were seen between the crystal structures in isoform 1+2 (see **Figure 69**A). The flexibility for this large rearrangement is given by the linker segment of Atl1 (residues 337-346), which is in case of isoform 2 a short helix connecting the G domain and the middle domain. This linker helix comprises the so called proline cap (P344 + P342) that is also found in dynamin and other members of this family. These prolines are a pivot point during nucleotide dependent transition. In the crystal structure of isoform 3 of Atl1, the linker region loses its secondary structure and mediates the inter domain swap of the middle domain. This leads to the formation of a more elongated Atl1 dimer that includes an additional interface between the middle domains.
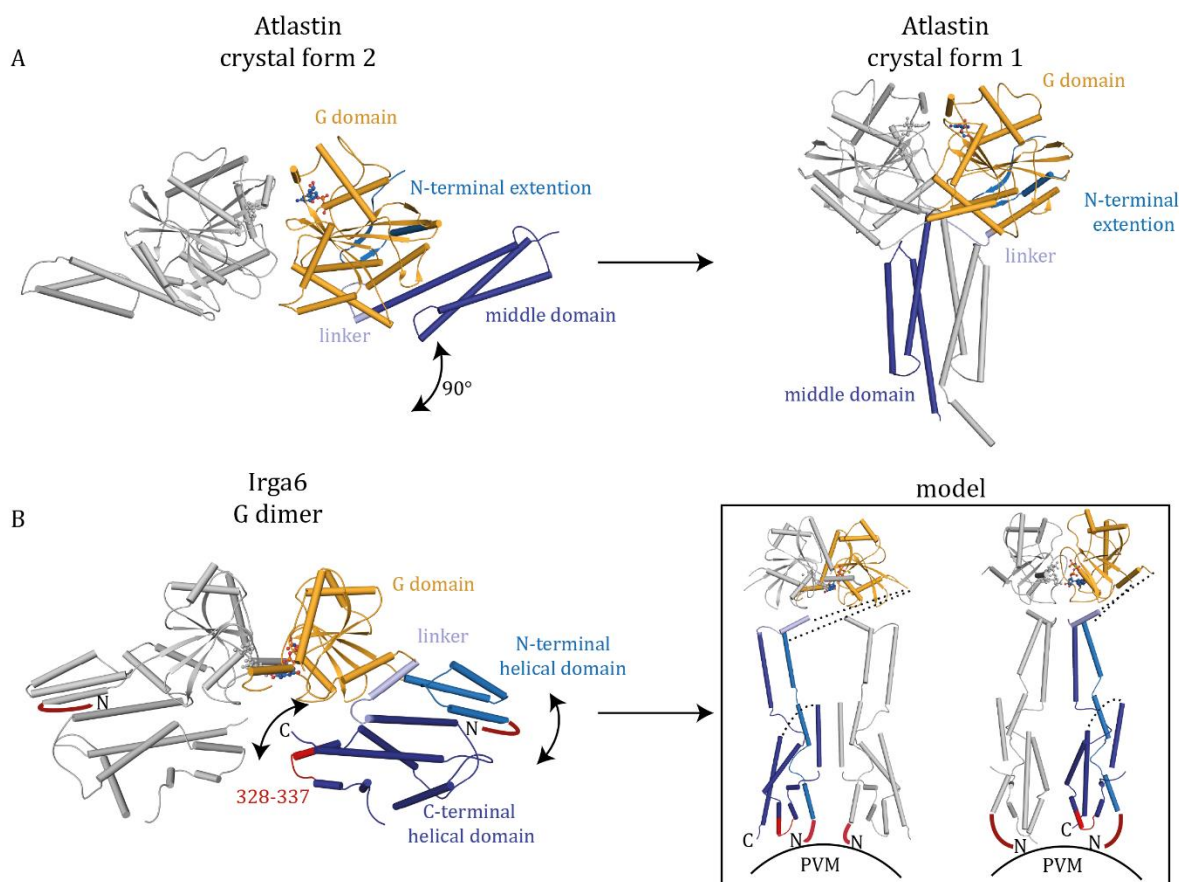
Figure 69: Comparison of the overall fold and predicted assembly mode of Irga6 with Atl1. Ribbon representations of Atl1 and Irga6. One monomer coloured in grey. Domains of the second monomer are color-coded as in **Figure 55**. Domains of Atl1 are coloured according to the domain architecture of Irga6. Nucleotides are shown in sphere and stick representation and coloured in grey or oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. A: left: ribbon representation of isoform 2 of Atl1 bound to GDP (PDB: 3Q5E), right: ribbon representation of isoform 3 of Atl1 bound to GMPPNP (PDB: 4IDN). B: left: ribbon representation of Irga6 G domain dimer. Red lines represent the N-terminus containing the myristoylation site. Predicted membrane accoziative parts are labelled and colored in red. Right: predicted model of an elongated Irga6 dimer binding to the PVM with (left) or without (right) an inter domain swap as seen for Atl1. For explanation see text.

A linker helix can also be found in the liker region of Irga6 (helix αE, residues 252-265), which is flanked by two prolines (P251 + P264). It is conceivable that the linker helix of Irga6 also rearranges upon GTP hydrolysis. This would allow the C-terminal helical domains to swing out, possibly also including an inter domain swap. The N-terminal helical domain could rearrange similarly, forming a more elongated helical bundle. The proposed buried region in Irga6 (residue 7-12)[243] could release the N-terminus and thus bring the myristoylated G2 to the tip in order to guide Irga6 to the parasitophorous vacuole membrane (see **Figure 69**B)

As it was shown that the myristoylation of Irga6 at G2 is crucial for its guidance to the PVM, but not exclusively for the binding [236,243], an additional membrane binding site

should be present. Due to the fact that Atl1 is membrane-associated via a transmembrane-region, it does not lend itself to comparisons in this regard. Nevertheless, the two regions which could be located at the very tip of the proposed elongated Irga6 dimer (see **Figure 69** model) harbor positively charged patches interspersed with some hydrophobic amino acids, as it is common for membrane associating parts in proteins. On the one hand, the C-terminus 407-413 (KE$_{ICL}$RN) and, on the other hand, the loop region 328-337 (QR$_{LA}$RD$_W$E$_I$E) were identified as potential membrane binding regions.

So far, this model would explain how Irga6 dimerizes in an elongated fashion consistent with membrane association. The rearrangement of the N-terminal helical domain of Irga6 would bring the N-terminus with the myristoylation site close to the PVM. The same is true for the C-terminal helical domain, as the identified membrane associative regions would be located near the PVM upon rearrangement. Whether the rearrangement of Irga6 includes also an interdomain swap (see **Figure 69** left model), as it was seen for Atl1, remains unclear. Nevertheless, the presented models both with and without interdomain swap appear reasonable upon membrane binding.

In case of atlastin, only dimerization is proposed to be required during fusion of opposing membranes (see 6.2.3). However, the rearrangement of Irga6 would not only be necessary for dimerization, PVM guidance and membrane binding but also for oligomerization around the PVM in order to disrupt the PVM. How oligomerization of Irga6 could take place will be discussed in the following section with the help of the already known oligomerization modes of other members of the dynamin superfamily.

10.3   Oligomerization of Irga6 on the PVM based on the oligomerization mode of dynamin and BDLP

Comparing dimerization and oligomerization of dynamin (see 6.2.1) with Irga6, two major differences become clear. Dynamin dimerize via the stalk in a criss-cross manner whereas Irga6 dimerizes via the backside interface without any crossing over. Oligomerization in dynamin occurs also via the stalk, and only the connections to the adjacent rungs in dynamin helices are mediated by the G domains (see 6.2.1). There is no evidence that Irga6 forms helical arrangements around the PVM as described for dynamin. Thus, only a filamentous oligomerization mode of Irga6 will be taken into account for the model, which assumes that G domain dimerization of Irga6 takes place in the same filament.
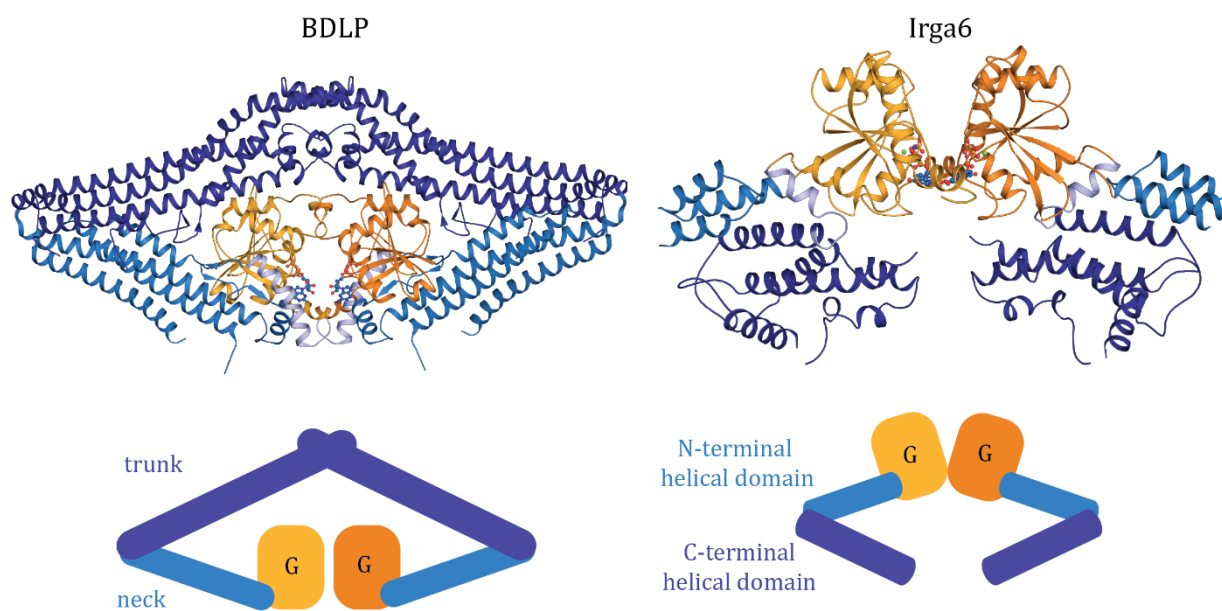
Figure 70: Ribbon representation of BDLP bound to GDP (PDB: 2J68) and of Irga6 K31E, R32E, K176E, K246E bound to GMPPNP. Domains are color-coded as in **Figure 55**. Domains of BDLP are coloured according to the domain architecture of Irga6. Nucleotides are shown in sphere and stick representation and coloured in grey or oxygens are red, nitrogens dark-blue, carbons light-blue and phosphate moieties orange. Mg^{2+} is coloured in green. Cartoon visualizes the same overall architecture of both GTPases. Nucleotides are not shown for clarity. For explanation see text.

The crystal structure of BDLP was solved in the nucleotide-free and in the GDP-bound form and reveals a kinked conformation of BDLP as seen for Irga6 (see **Figure 70**). A cryo-EM reconstruction of a tubular fragment of GMPPNP bound BDLP led to molecular insights into the conformational changes upon nucleotide exchange and membrane binding (see 6.2.2 and **Figure 42**). A large rearrangement mediated by the hinge 1 and hinge 2 regions induced the elongated linear structure with the paddle in close proximity to the lipid tubule and the G domain at the outer side. Besides the G domain dimerization interface, the cryo-EM reconstruction shows a second interface including the neck and the trunk.

Comparison of the crystal structure of Irga6 bound to GDP with that of Irga6 R31E, K32E, K176E, K246E bound to GMPPNP reveals no such large rearrangement in the helical domains. Nevertheless, the overall fold of BDLP bound to GDP and Irga6 is comparable (see **Figure 70**). The G domains are facing each other in a similar orientation and the C-terminal and N-terminal helical domains introduce a kink comparable to the trunk and the neck in BDLP. This underlines the idea that the arrangement of the membrane-associating parts of the helical domain towards the PVM could take place similarly to the reorientation of the trunk and neck of BDLP towards the membrane.

Taken together, Irga6 oligomerization based on the mode of action proposed for BDLP is conceivable. The oligomerization could take place in a parallel, linear organization of the helical regions as it was predicted for BDLP in cryo-EM reconstructions. The G domain dimerization would serve as the second interface necessary for the formation of a filamentous oligomer. How Irga6 could fulfill these criteria in order to wrap around the PVM will be discussed in the following chapter.

10.4   Model of Irga6 oligomerization at the PVM

The newly identified G domain interface of Irga6 combined with the information given by the crystal structures and cryo-EM reconstructions of atlastin and BDLP lead to a model visualizing how Irga6 could oligomerize and wrap around the PVM. This model is based on the assumption that the helical domain of Irga6 has to undergo a large rearrangement upon GTP binding and hydrolysis, whereas the newly identified G domain interface remains unchanged in the model.

In order to predict the second interface necessary for oligomerization, the amino acids R31, K32, L44, K48, S172, M173, K176 and K246 were taken into account. These were biochemically identified to impair GTP hydrolysis and oligomerization [247,249,259] although they are not located near the G domain dimerization interface.

Based on the parallel elongated Irga6 dimer model (see **Figure 69**), an oligomerization mode is proposed similar to that of BDLP as seen in the cryo-EM reconstruction (see **Figure 71** left). Irga6 oligomerize in a linear fashion with the elongated helical domain pointing towards the PVM. As described in 10.3, the putative membrane associating parts within the C-terminal helical domain (residues 328-337 and 407-413) and the N-terminus with the myristoylation site (residue 1-13) are in close proximity to the PVM. The G domain is located at the outer ring. Here, the residues S172, M173, K176 and K246 are located at the backside of the G domain and are in close proximity to the backside of the G domain of the next protomer in order to form a stable interface to promote further oligomerization. The patch of amino acids R31, K32, L44 and K48, located in or near the αB helix of the helical domain, would not be involved in the oligomerization interface in this model but they could be involved in stabilizing the elongated form of the helical domain and therefore ensure the possibility of oligomerization.
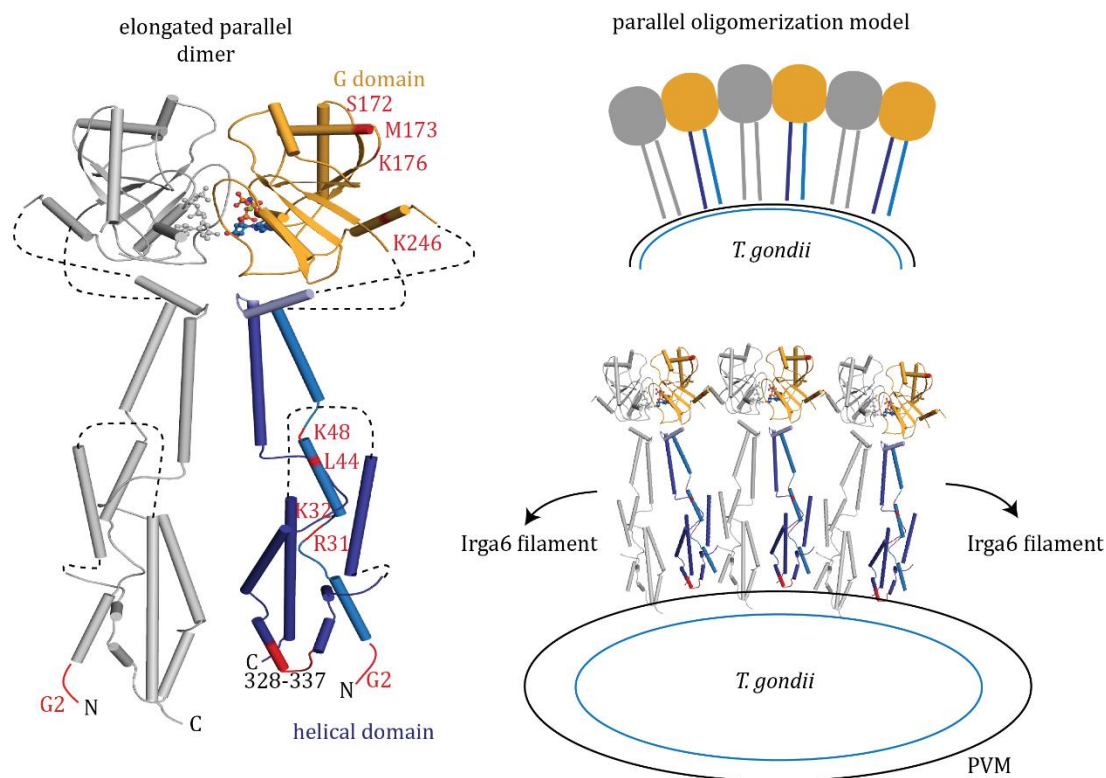
Figure 71: Proposed oligomerization mode of the elongated Irga6 model on the PVM. Left: Irga6 elongated dimer model. Residues for putative interface are colored and labelled in red. Putative membrane associative parts (G2 and residue 328-337) are labelled and colored in red. Dashed lines show missing connections between the secondary structure elements. N- and C-terminus are labelled. Right: Irga6 oligomers are oriented as linear dimers wrapping around the PVM. Minimal building block is the Irga6 elongated dimer shown on the left. Model is based on the cryo-EM reconstruction of BDLP. For explanation see text.

Altogether, the oligomerization model reveals a possible second interface necessary for dimerization involving residues which are known to impair GTP hydrolysis and oligomerization. Furthermore, the model would introduce a slight curvature due to oligomerization ideal to wrap around the convex PVM.

Nevertheless, more experiments are necessary to clarify the oligomerization mode of Irga6 around the PVM and to reveal whether PVM disruption is mediated by a constriction mode as suggested for dynamin.

10.5   Phylogenetic aspects of the G domain dimerization

The G domain dimerization mode of GIMAPs, dynamin, Toc and septin GTPases is highly similar. The G domains forming dimers in a *head-to-head* orientation and the dimer is build up via the switch I and the *trans* stabilizing loop (dynamin 1; equivalent to the residues 154-163 in mouse septin 2), and the G4 loop, which coordinates the nucleotide

base in *trans*. Based on this information, the evolutionary relationships of these membrane-associated GTPases was analyze in detail (see **Figure 72**).[260] The analysis revealed that dynamin and septin superfamily GTPases indeed share a common ancestor which likely dimerized on membranes via the G interface leading to GTPase stimulation (see **Figure 72**). Included in this analysis was also Irga6 (PDB: 1TPZ), which in this analysis was predicted to dimerize in the same fashion. Contrary models were derived based on biochemical analysis indicating that Irga6 forms a dimer in a *head-to-tail* orientation, similar to the very distantly related SRPs.[249]
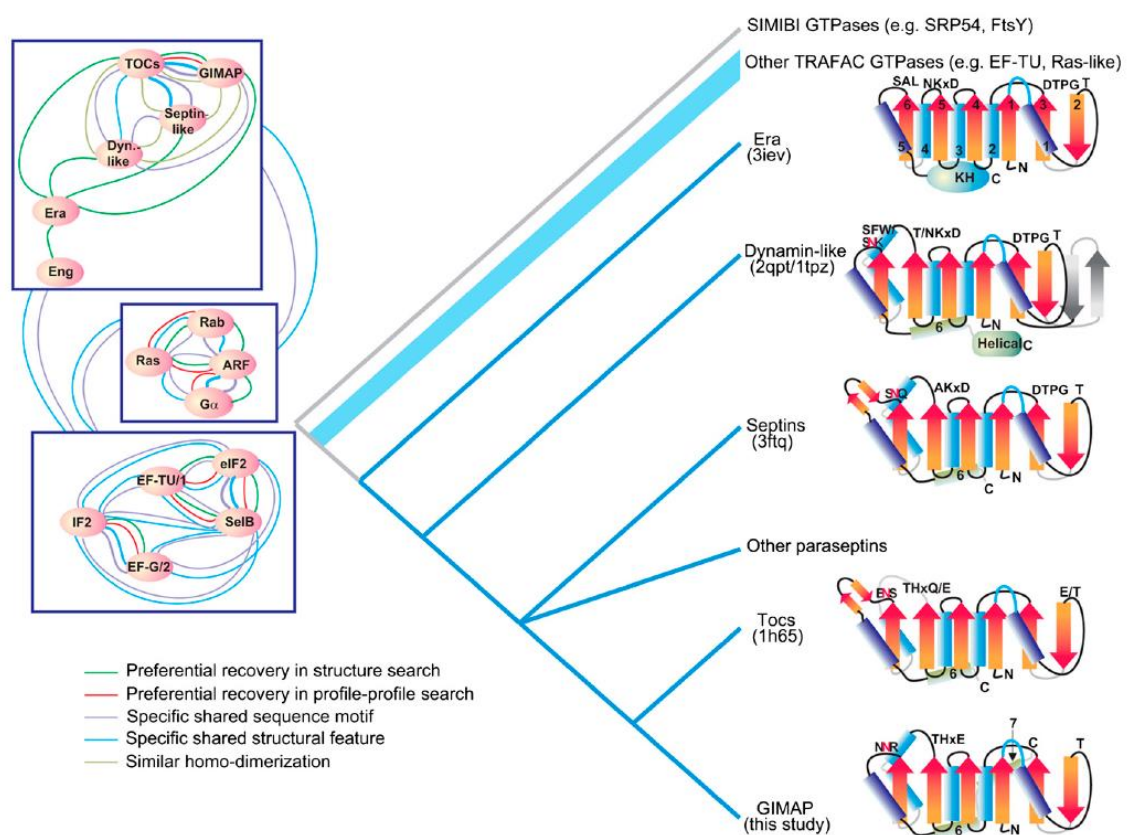
Figure 72: Higher-order relationship analysis of the GIMAP family. (Left) The relationship network constructed based on different lines of evidence. The relative thickness of the edges between the nodes in the case of the specific sequence and structure features reflects the relative number of such features that are shared. At the higher level the major groups are boxed for simplicity of illustration, and the edges between them (i.e., those that define the TRAFAC class of GTPases) are shown. (Right) The inferred higher-order relationships as a cladogram, along with the simplified topologies of representative members of each clade for which structures are available (see PDB ID codes). For the dynamins a reconstructed structure of the ancestral version is depicted based on the two derived versions of the clade typified by EHD2 (PDB ID code 2QPT) and IIGP (PDB ID code 1TPZ). Also indicated on the topology are the typical forms of the different conserved GTPase motifs in a given group. The P loop is coloured blue. The strands and helices are numbered in the Era structure, with the additional helices indicated in the structures in which they occur. Figure taken from [260]. For explanation see text.

The solved crystal structure of the G dimer of Irga6 shows a *head-to-head* G domain dimer that also employs switch I and the *trans* stabilizing loop as seen for dynamins. In contrast

to dynamins, the G4 loop participates in alternative contacts, and the nucleotide itself is involved in the dimer interface. Thus, our structural data clearly indicate that Irga6 is indeed a bona-fide member of the dynamin superfamily. Variations in the exact G domain dimerization mode may confer the specific oligomerization mode of IRG GTPases and, consequently, may be crucial for its function as anti-microbial agents.

10.6   Outlook

Irga6 as a member of the cell autonomous resistance system of mice can disrupt PVM harboring *T.gondii*. The structure of Irga6 bound to GMPPNP presented here reveals a new mode of G domain dimerization with direct involvement of the nucleotide, which indicates that binding to GTP is necessary to form this interface. Furthermore, it has been shown that Irga6 oligomerizes in a nucleotide-dependent manner *in vitro* and localizes to the PVM harboring *T.gondii* if bound to GTP and activated by INFγ.

In order to disrupt the PVM, Irga6 may undergo a mechanochemical conformational change upon GTP hydrolysis. The G domain could serve to initiate a power stroke, the linker helix could mediate this reaction and the helical domain undertakes a mechanical movement similar to the constriction mode predicted for dynamin. This could increase the mechanical force on the PVM and finally break the membranes of the parasitophorous vacuole. However, the cooperativity of GTP hydrolysis and the oligomerization of Irga6 in the presence of membranes or liposomes has not been investigated so far. Possibly, the GTP hydrolysis and oligomerization is increased if Irga6 is bound to lipids.

A dye leakage assay could give hints whether Irga6 itself is able to disrupt membranes upon GTP hydrolysis mediating a mechanochemical reaction or whether further interaction partners are necessary. Thus, large unilamelar vesicles filled with fluorescent dye would serve as template and potential membrane leakage is analyzed by detecting fluorescence. Both experiments could be done with mutants of the G domain dimer interface and the putative second oligomerization interface in order to analyze the ability of GTP hydrolysis, oligomerization and mechanochemical reaction in the presence of lipids.

In the determined crystal structure of Irga6 bound to GMPPNP, no conformational changes of the helical domain was observed compared to the known structures of Irga6 bound to GDP. In case of BDLP, there were also only minor changes observed in the crystal structures of the nucleotide free and the GDP-loaded state. Nevertheless, cryo-EM

reconstructions of BDLP in the presence of lipids showed a large rearrangement from its kinked organization to an elongated parallel oligomer wrapping around the lipid tubules. Cryo-EM reconstructions on Irga6 in the presence of GMPPNP or GTP and liposomes could possibly reveal such a mechanochemical movement as seen for BDLP in a more elongated form or a criss-cross organization, as it was seen for dynamin. Further electron microscopy studies could address the question whether Irga6 can form filaments around liposomes or even tubulate them and, if they do so, whether the adjacent filaments are connected via a third interface, as it was seen for dynamin and other members of the dynamin superfamily.

Except that, the region which mediates the lipid binding of Irga6 has to be identified. Cell biological studies, similar as they were done in order to analyze the importance of the myristoylation site of Irga6,[243] could give an answer whether the proposed regions 328-337 and 407-413 serve as lipid binding patches by mutating or deleting them. Irga6 could be fused to a fluorescent protein and its re-localization upon *T. gondii* infection and INFγ activation could be recorded. Altering the latter mentioned patches could give a hint to which extent these are involved in PVM binding.

In summary, these experiments would give a more detailed answer on how Irga6 interacts with the PVM. Nevertheless, to understand the whole orchestration of IRGs *in vivo* further investigations have to be done.

Bibliography

1   Taylor, B. L. & Zhulin, I. B. PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and molecular biology reviews : MMBR* **63**, 479-506 (1999).

2   Wang, G. L., Jiang, B. H., Rue, E. A. & Semenza, G. L. Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 5510-5514 (1995).

3   Hahn, M. E. The aryl hydrocarbon receptor: a comparative perspective. *Comparative biochemistry and physiology. Part C, Pharmacology, toxicology & endocrinology* **121**, 23-53 (1998).

4   Gekakis, N. *et al.* Role of the CLOCK protein in the mammalian circadian mechanism. *Science* **280**, 1564-1569 (1998).

5   Bersten, D. C., Sullivan, A. E., Peet, D. J. & Whitelaw, M. L. bHLH-PAS proteins in cancer. *Nature reviews. Cancer* **13**, 827-841, doi:10.1038/nrc3621 (2013).

6   Hahn, M. E., Karchner, S. I., Shapiro, M. A. & Perera, S. A. Molecular evolution of two vertebrate aryl hydrocarbon (dioxin) receptors (AHR1 and AHR2) and the PAS family. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 13743-13748 (1997).

7   Poland, A., Palen, D. & Glover, E. Tumour promotion by TCDD in skin of HRS/J hairless mice. *Nature* **300**, 271-273 (1982).

8   Fukunaga, B. N., Probst, M. R., Reisz-Porszasz, S. & Hankinson, O. Identification of functional domains of the aryl hydrocarbon receptor. *The Journal of biological chemistry* **270**, 29270-29278 (1995).

9   Ikuta, T., Eguchi, H., Tachibana, T., Yoneda, Y. & Kawajiri, K. Nuclear localization and export signals of the human aryl hydrocarbon receptor. *The Journal of biological chemistry* **273**, 2895-2904 (1998).

10 Pongratz, I., Mason, G. G. & Poellinger, L. Dual roles of the 90-kDa heat shock protein hsp90 in modulating functional activities of the dioxin receptor. Evidence that the dioxin receptor functionally belongs to a subclass of nuclear receptors which require hsp90 both for ligand binding activity and repression of intrinsic DNA binding activity. *The Journal of biological chemistry* **267**, 13728-13734 (1992).

11 Jain, S., Dolwick, K. M., Schmidt, J. V. & Bradfield, C. A. Potent transactivation domains of the Ah receptor and the Ah receptor nuclear translocator map to their carboxyl termini. *The Journal of biological chemistry* **269**, 31518-31524 (1994).

12 Rowlands, J. C., McEwan, I. J. & Gustafsson, J. A. Trans-activation by the human aryl hydrocarbon receptor and aryl hydrocarbon receptor nuclear translocator proteins: direct interactions with basal transcription factors. *Molecular pharmacology* **50**, 538-548 (1996).

13 Watt, K., Jess, T. J., Kelly, S. M., Price, N. C. & McEwan, I. J. Induced alpha-helix structure in the aryl hydrocarbon receptor transactivation domain modulates protein-protein interactions. *Biochemistry* **44**, 734-743, doi:10.1021/bi0487701 (2005).

14 Kumar, M. B. & Perdew, G. H. Nuclear receptor coactivator SRC-1 interacts with the Q-rich subdomain of the AhR and modulates its transactivation potential. *Gene expression* **8**, 273-286 (1999).

15 Jones, L. C. & Whitlock, J. P., Jr. Dioxin-inducible transactivation in a chromosomal setting. Analysis of the acidic domain of the Ah receptor. *The Journal of biological chemistry* **276**, 25037-25042, doi:10.1074/jbc.M102910200 (2001).

16 Bradfield, C. A., Kende, A. S. & Poland, A. Kinetic and equilibrium studies of Ah receptor-ligand binding: use of [125I]2-iodo-7,8-dibromodibenzo-p-dioxin. *Molecular pharmacology* **34**, 229-237 (1988).

17 Ponting, C. P. & Aravind, L. PAS: a multifunctional domain family comes to light. *Current biology : CB* **7**, R674-677 (1997).

18   Borgstahl, G. E., Williams, D. R. & Getzoff, E. D. 1.4 A structure of photoactive yellow protein, a cytosolic photoreceptor: unusual fold, active site, and chromophore. *Biochemistry* **34**, 6278-6287 (1995).

19   Zhulin, I. B., Taylor, B. L. & Dixon, R. PAS domain S-boxes in Archaea, Bacteria and sensors for oxygen and redox. *Trends in biochemical sciences* **22**, 331-333 (1997).

20   Ma, Q., Baldwin, K. T., Renzelli, A. J., McDaniel, A. & Dong, L. TCDD-inducible poly(ADP-ribose) polymerase: a novel response to 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Biochemical and biophysical research communications* **289**, 499-506, doi:10.1006/bbrc.2001.5987 (2001).

21   Boverhof, D. R. *et al.* Temporal and dose-dependent hepatic gene expression patterns in mice provide new insights into TCDD-Mediated hepatotoxicity. *Toxicological sciences : an official journal of the Society of Toxicology* **85**, 1048-1063, doi:10.1093/toxsci/kfi162 (2005).

22   Hahn, M. E. Aryl hydrocarbon receptors: diversity and evolution. *Chemico-biological interactions* **141**, 131-160 (2002).

23   Denison, M. S. & Nagy, S. R. Activation of the aryl hydrocarbon receptor by structurally diverse exogenous and endogenous chemicals. *Annual review of pharmacology and toxicology* **43**, 309-334, doi:10.1146/annurev.pharmtox.43.100901.135828 (2003).

24   Gillner, M. *et al.* Interactions of indolo[3,2-b]carbazoles and related polycyclic aromatic hydrocarbons with specific binding sites for 2,3,7,8-tetrachlorodibenzo-p-dioxin in rat liver. *Molecular pharmacology* **44**, 336-345 (1993).

25   Jellinck, P. H. *et al.* Ah receptor binding properties of indole carbinols and induction of hepatic estradiol hydroxylation. *Biochemical pharmacology* **45**, 1129-1136 (1993).

26   Schecter, A., Birnbaum, L., Ryan, J. J. & Constable, J. D. Dioxins: an overview. *Environmental research* **101**, 419-428, doi:10.1016/j.envres.2005.12.003 (2006).

27   Schecter, A. & Olson, J. R. Cancer risk assessment using blood dioxin levels and daily dietary TEQ intake in general populations of industrial and non-industrial countries. *Chemosphere* **34**, 1569-1577 (1997).

28   Bandiera, S., Safe, S. & Okey, A. B. Binding of polychlorinated biphenyls classified as either phenobarbitone-, 3-methylcholanthrene- or mixed-type inducers to cytosolic Ah receptor. *Chemico-biological interactions* **39**, 259-277 (1982).

29   Piskorska-Pliszczynska, J., Keys, B., Safe, S. & Newman, M. S. The cytosolic receptor binding affinities and AHH induction potencies of 29 polynuclear aromatic hydrocarbons. *Toxicology letters* **34**, 67-74 (1986).

30   Oberg, M., Bergander, L., Hakansson, H., Rannug, U. & Rannug, A. Identification of the tryptophan photoproduct 6-formylindolo[3,2-b]carbazole, in cell culture medium, as a factor that controls the background aryl hydrocarbon receptor activity. *Toxicological sciences : an official journal of the Society of Toxicology* **85**, 935-943, doi:10.1093/toxsci/kfi154 (2005).

31   Rannug, A. The tryptophan photoproduct 6-formylindolo[3,2-b]carbazole helps genes jump. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 18239-18240, doi:10.1073/pnas.1013653107 (2010).

32   Savouret, J. F. *et al.* 7-ketocholesterol is an endogenous modulator for the arylhydrocarbon receptor. *The Journal of biological chemistry* **276**, 3054-3059, doi:10.1074/jbc.M005988200 (2001).

33   Schaldach, C. M., Riby, J. & Bjeldanes, L. F. Lipoxin A4: a new class of ligand for the Ah receptor. *Biochemistry* **38**, 7594-7600, doi:10.1021/bi982861e (1999).

34   Opitz, C. A. *et al.* An endogenous tumour-promoting ligand of the human aryl hydrocarbon receptor. *Nature* **478**, 197-203, doi:10.1038/nature10491 (2011).

35   Kazlauskas, A., Poellinger, L. & Pongratz, I. Evidence that the co-chaperone p23 regulates ligand responsiveness of the dioxin (Aryl hydrocarbon) receptor. *The Journal of biological chemistry* **274**, 13519-13524 (1999).

36   Petrulis, J. R., Kusnadi, A., Ramadoss, P., Hollingshead, B. & Perdew, G. H. The hsp90 Co-chaperone XAP2 alters importin beta recognition of the bipartite nuclear localization

signal of the Ah receptor and represses transcriptional activity. *The Journal of biological chemistry* **278**, 2677-2685, doi:10.1074/jbc.M209331200 (2003).

37   Petrulis, J. R. & Perdew, G. H. The role of chaperone proteins in the aryl hydrocarbon receptor core complex. *Chemico-biological interactions* **141**, 25-40 (2002).

38   Perdew, G. H. Association of the Ah receptor with the 90-kDa heat shock protein. *The Journal of biological chemistry* **263**, 13802-13805 (1988).

39   Denis, M., Cuthill, S., Wikstrom, A. C., Poellinger, L. & Gustafsson, J. A. Association of the dioxin receptor with the Mr 90,000 heat shock protein: a structural kinship with the glucocorticoid receptor. *Biochemical and biophysical research communications* **155**, 801-807 (1988).

40   Antonsson, C., Whitelaw, M. L., McGuire, J., Gustafsson, J. A. & Poellinger, L. Distinct roles of the molecular chaperone hsp90 in modulating dioxin receptor function via the basic helix-loop-helix and PAS domains. *Molecular and cellular biology* **15**, 756-765 (1995).

41   Perdew, G. H. & Bradfield, C. A. Mapping the 90 kDa heat shock protein binding region of the Ah receptor. *Biochemistry and molecular biology international* **39**, 589-593 (1996).

42   Meyer, B. K., Pray-Grant, M. G., Vanden Heuvel, J. P. & Perdew, G. H. Hepatitis B virus X-associated protein 2 is a subunit of the unliganded aryl hydrocarbon receptor core complex and exhibits transcriptional enhancer activity. *Molecular and cellular biology* **18**, 978-988 (1998).

43   Meyer, B. K. & Perdew, G. H. Characterization of the AhR-hsp90-XAP2 core complex and the role of the immunophilin-related protein XAP2 in AhR stabilization. *Biochemistry* **38**, 8907-8917, doi:10.1021/bi982223w (1999).

44   Grenert, J. P. *et al.* The amino-terminal domain of heat shock protein 90 (hsp90) that binds geldanamycin is an ATP/ADP switch domain that regulates hsp90 conformation. *The Journal of biological chemistry* **272**, 23843-23850 (1997).

45   Nair, S. C. *et al.* A pathway of multi-chaperone interactions common to diverse regulatory proteins: estrogen receptor, Fes tyrosine kinase, heat shock transcription factor Hsf1, and the aryl hydrocarbon receptor. *Cell stress & chaperones* **1**, 237-250 (1996).

46   Stockinger, B., Di Meglio, P., Gialitakis, M. & Duarte, J. H. The aryl hydrocarbon receptor: multitasking in the immune system. *Annual review of immunology* **32**, 403-432, doi:10.1146/annurev-immunol-032713-120245 (2014).

47   McGuire, J., Whitelaw, M. L., Pongratz, I., Gustafsson, J. A. & Poellinger, L. A cellular factor stimulates ligand-dependent release of hsp90 from the basic helix-loop-helix dioxin receptor. *Molecular and cellular biology* **14**, 2438-2446 (1994).

48   Reyes, H., Reisz-Porszasz, S. & Hankinson, O. Identification of the Ah receptor nuclear translocator protein (Arnt) as a component of the DNA binding form of the Ah receptor. *Science* **256**, 1193-1195 (1992).

49   Yao, G., Craven, M., Drinkwater, N. & Bradfield, C. A. Interaction networks in yeast define and enumerate the signaling steps of the vertebrate aryl hydrocarbon receptor. *PLoS biology* **2**, E65, doi:10.1371/journal.pbio.0020065 (2004).

50   Probst, M. R., Reisz-Porszasz, S., Agbunag, R. V., Ong, M. S. & Hankinson, O. Role of the aryl hydrocarbon receptor nuclear translocator protein in aryl hydrocarbon (dioxin) receptor action. *Molecular pharmacology* **44**, 511-518 (1993).

51   Bacsi, S. G. & Hankinson, O. Functional characterization of DNA-binding domains of the subunits of the heterodimeric aryl hydrocarbon receptor complex imputing novel and canonical basic helix-loop-helix protein-DNA interactions. *The Journal of biological chemistry* **271**, 8843-8850 (1996).

52   Shen, E. S. & Whitlock, J. P., Jr. Protein-DNA interactions at a dioxin-responsive enhancer. Mutational analysis of the DNA-binding site for the liganded Ah receptor. *The Journal of biological chemistry* **267**, 6815-6819 (1992).

53   Lusska, A., Shen, E. & Whitlock, J. P., Jr. Protein-DNA interactions at a dioxin-responsive enhancer. Analysis of six bona fide DNA-binding sites for the liganded Ah receptor. *The Journal of biological chemistry* **268**, 6575-6580 (1993).

54   Swanson, H. I., Chan, W. K. & Bradfield, C. A. DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *The Journal of biological chemistry* **270**, 26292-26302 (1995).

55   Swanson, H. I. & Yang, J. H. The aryl hydrocarbon receptor interacts with transcription factor IIB. *Molecular pharmacology* **54**, 671-677 (1998).

56   Tian, Y., Ke, S., Chen, M. & Sheng, T. Interactions between the aryl hydrocarbon receptor and P-TEFb. Sequential recruitment of transcription factors and differential phosphorylation of C-terminal domain of RNA polymerase II at cyp1a1 promoter. *The Journal of biological chemistry* **278**, 44041-44048, doi:10.1074/jbc.M306443200 (2003).

57   Wang, S., Ge, K., Roeder, R. G. & Hankinson, O. Role of mediator in transcriptional activation by the aryl hydrocarbon receptor. *The Journal of biological chemistry* **279**, 13593-13600, doi:10.1074/jbc.M312274200 (2004).

58   Okino, S. T. & Whitlock, J. P., Jr. Dioxin induces localized, graded changes in chromatin structure: implications for Cyp1A1 gene transcription. *Molecular and cellular biology* **15**, 3714-3721 (1995).

59   Wang, S. & Hankinson, O. Functional involvement of the Brahma/SWI2-related gene 1 protein in cytochrome P4501A1 transcription mediated by the aryl hydrocarbon receptor complex. *The Journal of biological chemistry* **277**, 11821-11827, doi:10.1074/jbc.M110122200 (2002).

60   Schnekenburger, M., Peng, L. & Puga, A. HDAC1 bound to the Cyp1a1 promoter blocks histone acetylation associated with Ah receptor-mediated trans-activation. *Biochimica et biophysica acta* **1769**, 569-578, doi:10.1016/j.bbaexp.2007.07.002 (2007).

61   Beischlag, T. V. *et al.* Recruitment of the NCoA/SRC-1/p160 family of transcriptional coactivators by the aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator complex. *Molecular and cellular biology* **22**, 4319-4333 (2002).

62   DiNatale, B. C., Schroeder, J. C., Francey, L. J., Kusnadi, A. & Perdew, G. H. Mechanistic insights into the events that lead to synergistic induction of interleukin 6 transcription upon activation of the aryl hydrocarbon receptor and inflammatory signaling. *The Journal of biological chemistry* **285**, 24388-24397, doi:10.1074/jbc.M110.118570 (2010).

63   Nguyen, T. A., Hoivik, D., Lee, J. E. & Safe, S. Interactions of nuclear receptor coactivator/corepressor proteins with the aryl hydrocarbon receptor complex. *Archives of biochemistry and biophysics* **367**, 250-257, doi:10.1006/abbi.1999.1282 (1999).

64   Klinge, C. M. *et al.* Short heterodimer partner (SHP) orphan nuclear receptor inhibits the transcriptional activity of aryl hydrocarbon receptor (AHR)/AHR nuclear translocator (ARNT). *Archives of biochemistry and biophysics* **390**, 64-70, doi:10.1006/abbi.2001.2366 (2001).

65   Nebert, D. W., Dalton, T. P., Okey, A. B. & Gonzalez, F. J. Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. *The Journal of biological chemistry* **279**, 23847-23850, doi:10.1074/jbc.R400004200 (2004).

66   Ma, Q. & Baldwin, K. T. A cycloheximide-sensitive factor regulates TCDD-induced degradation of the aryl hydrocarbon receptor. *Chemosphere* **46**, 1491-1500 (2002).

67   Pollenz, R. S. Specific blockage of ligand-induced degradation of the Ah receptor by proteasome but not calpain inhibitors in cell culture lines from different species. *Biochemical pharmacology* **74**, 131-143, doi:10.1016/j.bcp.2007.03.021 (2007).

68   Pollenz, R. S. & Buggy, C. Ligand-dependent and -independent degradation of the human aryl hydrocarbon receptor (hAHR) in cell culture models. *Chemico-biological interactions* **164**, 49-59, doi:10.1016/j.cbi.2006.08.014 (2006).

69   Mimura, J., Ema, M., Sogawa, K. & Fujii-Kuriyama, Y. Identification of a novel mechanism of regulation of Ah (dioxin) receptor function. *Genes & development* **13**, 20-25 (1999).

70   Pansoy, A., Ahmed, S., Valen, E., Sandelin, A. & Matthews, J. 3-methylcholanthrene induces differential recruitment of aryl hydrocarbon receptor to human promoters. *Toxicological sciences : an official journal of the Society of Toxicology* **117**, 90-100, doi:10.1093/toxsci/kfq096 (2010).

71   Evans, B. R. *et al.* Repression of aryl hydrocarbon receptor (AHR) signaling by AHR repressor: role of DNA binding and competition for AHR nuclear translocator. *Molecular pharmacology* **73**, 387-398, doi:10.1124/mol.107.040204 (2008).

72   McMillan, B. J. & Bradfield, C. A. The aryl hydrocarbon receptor sans xenobiotics: endogenous function in genetic model systems. *Molecular pharmacology* **72**, 487-498, doi:10.1124/mol.107.037259 (2007).

73   Tomita, S. *et al.* T cell-specific disruption of arylhydrocarbon receptor nuclear translocator (Arnt) gene causes resistance to 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced thymic involution. *Journal of immunology* **171**, 4113-4120 (2003).

74   Pitot, H. C., Goldsworthy, T., Campbell, H. A. & Poland, A. Quantitative evaluation of the promotion by 2,3,7,8-tetrachlorodibenzo-p-dioxin of hepatocarcinogenesis from diethylnitrosamine. *Cancer research* **40**, 3616-3620 (1980).

75   Holsapple, M. P., Morris, D. L., Wood, S. C. & Snyder, N. K. 2,3,7,8-tetrachlorodibenzo-p-dioxin-induced changes in immunocompetence: possible mechanisms. *Annual review of pharmacology and toxicology* **31**, 73-100, doi:10.1146/annurev.pa.31.040191.000445 (1991).

76   Schmidt, J. V., Su, G. H., Reddy, J. K., Simon, M. C. & Bradfield, C. A. Characterization of a murine Ahr null allele: involvement of the Ah receptor in hepatic growth and development. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6731-6736 (1996).

77   Mimura, J. *et al.* Loss of teratogenic response to 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) in mice lacking the Ah (dioxin) receptor. *Genes to cells : devoted to molecular & cellular mechanisms* **2**, 645-654 (1997).

78   Fernandez-Salguero, P. *et al.* Immune system impairment and hepatic fibrosis in mice lacking the dioxin-binding Ah receptor. *Science* **268**, 722-726 (1995).

79   Lahvis, G. P. *et al.* Portosystemic shunting and persistent fetal vascular structures in aryl hydrocarbon receptor-deficient mice. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 10442-10447, doi:10.1073/pnas.190256997 (2000).

80   Lund, A. K., Goens, M. B., Kanagy, N. L. & Walker, M. K. Cardiac hypertrophy in aryl hydrocarbon receptor null mice is correlated with elevated angiotensin II, endothelin-1, and mean arterial blood pressure. *Toxicology and applied pharmacology* **193**, 177-187 (2003).

81   Lahvis, G. P. *et al.* The aryl hydrocarbon receptor is required for developmental closure of the ductus venosus in the neonatal mouse. *Molecular pharmacology* **67**, 714-720, doi:10.1124/mol.104.008888 (2005).

82   Kiserud, T. Physiology of the fetal circulation. *Seminars in fetal & neonatal medicine* **10**, 493-503, doi:10.1016/j.siny.2005.08.007 (2005).

83   Marlowe, J. L. & Puga, A. Aryl hydrocarbon receptor, cell cycle regulation, toxicity, and tumorigenesis. *Journal of cellular biochemistry* **96**, 1174-1184, doi:10.1002/jcb.20656 (2005).

84   Coqueret, O. Linking cyclins to transcriptional control. *Gene* **299**, 35-55 (2002).

85   Ge, N. L. & Elferink, C. J. A direct interaction between the aryl hydrocarbon receptor and retinoblastoma protein. Linking dioxin signaling to the cell cycle. *The Journal of biological chemistry* **273**, 22708-22713 (1998).

86   Barhoover, M. A., Hall, J. M., Greenlee, W. F. & Thomas, R. S. Aryl hydrocarbon receptor regulates cell cycle progression in human breast cancer cells via a functional interaction with cyclin-dependent kinase 4. *Molecular pharmacology* **77**, 195-201, doi:10.1124/mol.109.059675 (2010).

87   Elizondo, G. *et al.* Altered cell cycle control at the G(2)/M phases in aryl hydrocarbon receptor-null embryo fibroblast. *Molecular pharmacology* **57**, 1056-1063 (2000).

88   Gonzalez, F. J. & Fernandez-Salguero, P. The aryl hydrocarbon receptor: studies using the AHR-null mice. *Drug metabolism and disposition: the biological fate of chemicals* **26**, 1194-1198 (1998).

89   Enan, E. & Matsumura, F. Significance of TCDD-induced changes in protein phosphorylation in the adipocyte of male guinea pigs. *Journal of biochemical toxicology* **9**, 159-170 (1994).

90   Worner, W. & Schrenk, D. Influence of liver tumor promoters on apoptosis in rat hepatocytes induced by 2-acetylaminofluorene, ultraviolet light, or transforming growth factor beta 1. *Cancer research* **56**, 1272-1278 (1996).

91   McGregor, D. B., Partensky, C., Wilbourn, J. & Rice, J. M. An IARC evaluation of polychlorinated dibenzo-p-dioxins and polychlorinated dibenzofurans as risk factors in human carcinogenesis. *Environmental health perspectives* **106 Suppl 2**, 755-760 (1998).

92   Kohle, C. *et al.* 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD)-mediated membrane translocation of c-Src protein kinase in liver WB-F344 cells. *Archives of toxicology* **73**, 152-158 (1999).

93   Gaido, K. W., Maness, S. C., Leonard, L. S. & Greenlee, W. F. 2,3,7,8-Tetrachlorodibenzo-p-dioxin-dependent regulation of transforming growth factors-alpha and -beta 2 expression in a human keratinocyte cell line involves both transcriptional and post-transcriptional control. *The Journal of biological chemistry* **267**, 24591-24595 (1992).

94   Hoffer, A., Chang, C. Y. & Puga, A. Dioxin induces transcription of fos and jun genes by Ah receptor-dependent and -independent pathways. *Toxicology and applied pharmacology* **141**, 238-247, doi:10.1006/taap.1996.0280 (1996).

95   Kolluri, S. K., Weiss, C., Koff, A. & Gottlicher, M. p27(Kip1) induction and inhibition of proliferation by the intracellular Ah receptor in developing thymus and hepatoma cells. *Genes & development* **13**, 1742-1753 (1999).

96   Paajarvi, G., Viluksela, M., Pohjanvirta, R., Stenius, U. & Hogberg, J. TCDD activates Mdm2 and attenuates the p53 response to DNA damaging agents. *Carcinogenesis* **26**, 201-208, doi:10.1093/carcin/bgh289 (2005).

97   Marshall, N. B., Vorachek, W. R., Steppan, L. B., Mourich, D. V. & Kerkvliet, N. I. Functional characterization and gene expression analysis of CD4+ CD25+ regulatory T cells generated in mice treated with 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Journal of immunology* **181**, 2382-2391 (2008).

98   Quintana, F. J. *et al.* Control of T(reg) and T(H)17 cell differentiation by the aryl hydrocarbon receptor. *Nature* **453**, 65-71, doi:10.1038/nature06880 (2008).

99   Nakahama, T. *et al.* Aryl hydrocarbon receptor deficiency in T cells suppresses the development of collagen-induced arthritis. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 14222-14227, doi:10.1073/pnas.1111786108 (2011).

100   Monteleone, I. *et al.* Aryl hydrocarbon receptor-induced signals up-regulate IL-22 production and inhibit inflammation in the gastrointestinal tract. *Gastroenterology* **141**, 237-248, 248 e231, doi:10.1053/j.gastro.2011.04.007 (2011).

101   Lee, J. A. *et al.* 2,3,7,8-Tetrachlorodibenzo-p-dioxin modulates functional differentiation of mouse bone marrow-derived dendritic cells Downregulation of RelB by 2,3,7,8-tetrachlorodibenzo-p-dioxin. *Toxicology letters* **173**, 31-40, doi:10.1016/j.toxlet.2007.06.012 (2007).

102   Kimura, A. *et al.* Aryl hydrocarbon receptor in combination with Stat1 regulates LPS-induced inflammatory responses. *The Journal of experimental medicine* **206**, 2027-2035, doi:10.1084/jem.20090560 (2009).

103   Sanchez, Y. *et al.* The unexpected role for the aryl hydrocarbon receptor on susceptibility to experimental toxoplasmosis. *Journal of biomedicine & biotechnology* **2010**, 505694, doi:10.1155/2010/505694 (2010).

104   Elizondo, G., Rodriguez-Sosa, M., Estrada-Muniz, E., Gonzalez, F. J. & Vega, L. Deletion of the aryl hydrocarbon receptor enhances the inflammatory response to Leishmania major infection. *International journal of biological sciences* **7**, 1220-1229 (2011).

105   Neff-LaFord, H., Teske, S., Bushnell, T. P. & Lawrence, B. P. Aryl hydrocarbon receptor activation during influenza virus infection unveils a novel pathway of IFN-gamma production by phagocytic cells. *Journal of immunology* **179**, 247-255 (2007).

106   Hao, N. & Whitelaw, M. L. The emerging roles of AhR in physiology and immunity. *Biochemical pharmacology* **86**, 561-570, doi:10.1016/j.bcp.2013.07.004 (2013).

107   Safe, S., Lee, S. O. & Jin, U. H. Role of the aryl hydrocarbon receptor in carcinogenesis and potential as a drug target. *Toxicological sciences : an official journal of the Society of Toxicology* **135**, 1-16, doi:10.1093/toxsci/kft128 (2013).

108   Saito, R. *et al.* Aryl hydrocarbon receptor in breast cancer-a newly defined prognostic marker. *Hormones & cancer* **5**, 11-21, doi:10.1007/s12672-013-0160-z (2014).

109   Richmond, O. *et al.* The aryl hydrocarbon receptor is constitutively active in advanced prostate cancer cells. *PloS one* **9**, e95058, doi:10.1371/journal.pone.0095058 (2014).

110   Yin, X. F., Chen, J., Mao, W., Wang, Y. H. & Chen, M. H. Downregulation of aryl hydrocarbon receptor expression decreases gastric cancer cell growth and invasion. *Oncology reports* **30**, 364-370, doi:10.3892/or.2013.2410 (2013).

111   Su, J. M., Lin, P. & Chang, H. Prognostic value of nuclear translocation of aryl hydrocarbon receptor for non-small cell lung cancer. *Anticancer research* **33**, 3953-3961 (2013).

112   Liu, Z. *et al.* AhR expression is increased in hepatocellular carcinoma. *Journal of molecular histology* **44**, 455-461, doi:10.1007/s10735-013-9495-6 (2013).

113   Mezrich, J. D. *et al.* An interaction between kynurenine and the aryl hydrocarbon receptor can generate regulatory T cells. *Journal of immunology* **185**, 3190-3198, doi:10.4049/jimmunol.0903670 (2010).

114   Stone, T. W., Stoy, N. & Darlington, L. G. An expanding range of targets for kynurenine metabolites of tryptophan. *Trends in pharmacological sciences* **34**, 136-143, doi:10.1016/j.tips.2012.09.006 (2013).

115   Huang, N. *et al.* Crystal structure of the heterodimeric CLOCK:BMAL1 transcriptional activator complex. *Science* **337**, 189-194, doi:10.1126/science.1222804 (2012).

116   Wang, Z., Wu, Y., Li, L. & Su, X. D. Intermolecular recognition revealed by the complex structure of human CLOCK-BMAL1 basic helix-loop-helix domains with E-box DNA. *Cell research* **23**, 213-224, doi:10.1038/cr.2012.170 (2013).

117   Wu, D., Potluri, N., Kim, Y. & Rastinejad, F. Structure and dimerization properties of the aryl hydrocarbon receptor PAS-A domain. *Molecular and cellular biology* **33**, 4346-4356, doi:10.1128/MCB.00698-13 (2013).

118   Scheuermann, T. H. *et al.* Artificial ligand binding within the HIF2alpha PAS-B domain of the HIF2 transcription factor. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 450-455, doi:10.1073/pnas.0808092106 (2009).

119   Evans, M. R., Card, P. B. & Gardner, K. H. ARNT PAS-B has a fragile native state structure with an alternative beta-sheet register nearby in sequence space. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 2617-2622, doi:10.1073/pnas.0808270106 (2009).

120   Bisson, W. H. *et al.* Modeling of the aryl hydrocarbon receptor (AhR) ligand binding domain and its utility in virtual ligand screening to predict new AhR ligands. *Journal of medicinal chemistry* **52**, 5635-5641, doi:10.1021/jm900199u (2009).

121   Sambrook, J., MacCallum P., and Russel, D. . A Laboratory Manual. *Molecular Cloning* (2001).

122   Aslanidis, C. & de Jong, P. J. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic acids research* **18**, 6069-6074 (1990).

123   Chung, C. T., Niemela, S. L. & Miller, R. H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 2172-2175 (1989).

124   Chung, C. T., Niemela, S. L., and Miller, R. H. *Proc. NAtl. Acad. Sci. U.S.A* **86**, 2172-2175 (1989).

125   Moos, M., Jr., Nguyen, N. Y., and Liu, T. Y. . *J. Biol. Chem.* **263**, 6005-6008 (1988).

126   Laemmli, U. K. *Nature* **227**, 680-685 (1970).

127   Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *Journal of molecular biology* **229**, 105-124, doi:10.1006/jmbi.1993.1012 (1993).

128   Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. . *Methods Mol. Biol.* **112**, 531-552 (1999).

129   Bohm, G., Muhr, R. & Jaenicke, R. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. *Protein engineering* **5**, 191-195 (1992).

130   Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochimica et biophysica acta* **1751**, 119-139, doi:10.1016/j.bbapap.2005.06.005 (2005).

131   Leslie, A. G. *Acta Crystallogr. D. Biol. Crystallogr.* **62**, 48-57 (2006).

132   Kabsch, W. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125-132 (2010).

133   Hahn, T., et al.,. International tables of cystallography. *Kluwer academic Publisher Group* (1984).

134   Taylor, G. The phase problem. *Acta crystallographica. Section D, Biological crystallography* **59**, 1881-1890 (2003).

135   Swanson, H. I. & Yang, J. H. Specificity of DNA binding of the c-Myc/Max and ARNT/ARNT dimers at the CACGTG recognition site. *Nucleic acids research* **27**, 3205-3212 (1999).

136   Adams PD, A. P., Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung, LW, K. G., Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, & Richardson DC, R. J., Terwilliger TC, Zwart PH. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst* **D66**, 213-221 (2010).

137   Vagin, A. T., A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst* **30**, 1022-1025 (1997).

138   Keegan, R. M. & Winn, M. D. MrBUMP: an automated pipeline for molecular replacement. *Acta crystallographica. Section D, Biological crystallography* **64**, 119-124, doi:10.1107/S0907444907037195 (2008).

139   Long, F., Vagin, A. A., Young, P. & Murshudov, G. N. BALBES: a molecular-replacement pipeline. *Acta crystallographica. Section D, Biological crystallography* **64**, 125-132, doi:10.1107/S0907444907050172 (2008).

140   Sheldrick, G. M. A short history of SHELX. *Acta crystallographica. Section A, Foundations of crystallography* **64**, 112-122, doi:10.1107/S0108767307043930 (2008).

141   Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta crystallographica. Section D, Biological crystallography* **59**, 2023-2030 (2003).

142   Wache, S. C., Hoagland, E. M., Zeigler, G. & Swanson, H. I. Role of arginine residues 14 and 15 in dictating DNA binding stability and transactivation of the aryl hydrocarbon receptor/aryl hydrocarbon receptor nuclear translocator heterodimer. *Gene expression* **12**, 231-243 (2005).

143   Minsavage, G. D., Park, S. K. & Gasiewicz, T. A. The aryl hydrocarbon receptor (AhR) tyrosine 9, a residue that is essential for AhR DNA binding activity, is not a phosphoresidue but augments AhR phosphorylation. *The Journal of biological chemistry* **279**, 20582-20593, doi:10.1074/jbc.M312977200 (2004).

144   Koonin, E. V., Wolf, Y. I. & Aravind, L. Protein fold recognition using sequence profiles and its application in structural genomics. *Advances in protein chemistry* **54**, 245-275 (2000).

145   Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. Classification and evolution of P-loop GTPases and related ATPases. *Journal of molecular biology* **317**, 41-72, doi:10.1006/jmbi.2001.5378 (2002).

146   Bourne, H. R., Sanders, D. A. & McCormick, F. The GTPase superfamily: a conserved switch for diverse cell functions. *Nature* **348**, 125-132, doi:10.1038/348125a0 (1990).

147   Bernards, A. & Settleman, J. GAP control: regulating the regulators of small GTPases. *Trends in cell biology* **14**, 377-385, doi:10.1016/j.tcb.2004.05.003 (2004).

148   Geyer, M. & Wittinghofer, A. GEFs, GAPs, GDIs and effectors: taking a closer (3D) look at the regulation of Ras-related GTP-binding proteins. *Current opinion in structural biology* **7**, 786-792 (1997).

149   Vetter, I. R. & Wittinghofer, A. The guanine nucleotide-binding switch in three dimensions. *Science* **294**, 1299-1304, doi:10.1126/science.1062023 (2001).

150   Wittinghofer, A. & Nassar, N. How Ras-related proteins talk to their effectors. *Trends in biochemical sciences* **21**, 488-491 (1996).

151   Arozarena, I. *et al.* Activation of H-Ras in the endoplasmic reticulum by the RasGRF family guanine nucleotide exchange factors. *Molecular and cellular biology* **24**, 1516-1530 (2004).

152   Gasper, R., Scrima, A. & Wittinghofer, A. Structural insights into HypB, a GTP-binding protein that regulates metal binding. *The Journal of biological chemistry* **281**, 27492-27502, doi:10.1074/jbc.M600809200 (2006).

153   Gasper, R., Meyer, S., Gotthardt, K., Sirajuddin, M. & Wittinghofer, A. It takes two to tango: regulation of G proteins by dimerization. *Nature reviews. Molecular cell biology* **10**, 423-429, doi:10.1038/nrm2689 (2009).

154   Scrima, A. & Wittinghofer, A. Dimerisation-dependent GTPase reaction of MnmE: how potassium acts as GTPase-activating element. *The EMBO journal* **25**, 2940-2951, doi:10.1038/sj.emboj.7601171 (2006).

155   Martens, S. & Howard, J. The interferon-inducible GTPases. *Annual review of cell and developmental biology* **22**, 559-589, doi:10.1146/annurev.cellbio.22.010305.104619 (2006).

156   Praefcke, G. J. & McMahon, H. T. The dynamin superfamily: universal membrane tubulation and fission molecules? *Nature reviews. Molecular cell biology* **5**, 133-147, doi:10.1038/nrm1313 (2004).

157   Low, H. H. & Lowe, J. A bacterial dynamin-like protein. *Nature* **444**, 766-769, doi:10.1038/nature05312 (2006).

158   Wittinghofer, A., Waldmann, H. Ras - a molecular switch involved in tumor formation. *Angewandte Chemie* **39**, 4192-4214 (2000).

159   Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *The EMBO journal* **1**, 945-951 (1982).

160   Feuerstein, J., Goody, R. S. & Webb, M. R. The mechanism of guanosine nucleotide hydrolysis by p21 c-Ha-ras. The stereochemical course of the GTPase reaction. *The Journal of biological chemistry* **264**, 6188-6190 (1989).

161   Blackburn, G. M., Williams, N. H., Gamblin, S. J. & Smerdon, S. J. Comment on "The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction". *Science* **301**, 1184; author reply 1184, doi:10.1126/science.1085796 (2003).

162   Lahiri, S. D., Zhang, G., Dunaway-Mariano, D. & Allen, K. N. The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction. *Science* **299**, 2067-2071, doi:10.1126/science.1082710 (2003).

163   Rudack, T., Xia, F., Schlitter, J., Kotting, C. & Gerwert, K. Ras and GTPase-activating protein (GAP) drive GTP into a precatalytic state as revealed by combining FTIR and biomolecular simulations. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 15295-15300, doi:10.1073/pnas.1204333109 (2012).

164   Bos, J. L., Rehmann, H. & Wittinghofer, A. GEFs and GAPs: critical elements in the control of small G proteins. *Cell* **129**, 865-877, doi:10.1016/j.cell.2007.05.018 (2007).

165   Kleineke, J., Duls, C. & Soling, H. D. Subcellular compartmentation of guanine nucleotides and functional relationships between the adenine and guanine nucleotide systems in isolated hepatocytes. *FEBS letters* **107**, 198-202 (1979).

166   Meyer, D. I., Krause, E. & Dobberstein, B. Secretory protein translocation across membranes-the role of the "docking protein'. *Nature* **297**, 647-650 (1982).

167   Gilmore, R., Blobel, G. & Walter, P. Protein translocation across the endoplasmic reticulum. I. Detection in the microsomal membrane of a receptor for the signal recognition particle. *The Journal of cell biology* **95**, 463-469 (1982).

168   Focia, P. J., Shepotinovskaya, I. V., Seidler, J. A. & Freymann, D. M. Heterodimeric GTPase core of the SRP targeting complex. *Science* **303**, 373-377, doi:10.1126/science.1090827 (2004).

169   Ghosh, A., Praefcke, G. J., Renault, L., Wittinghofer, A. & Herrmann, C. How guanylate-binding proteins achieve assembly-stimulated processive cleavage of GTP to GMP. *Nature* **440**, 101-104, doi:10.1038/nature04510 (2006).

170   Chappie, J. S., Acharya, S., Leonard, M., Schmid, S. L. & Dyda, F. G domain dimerization controls dynamin's assembly-stimulated GTPase activity. *Nature* **465**, 435-440, doi:10.1038/nature09032 (2010).

171   Schwemmle, M. & Staeheli, P. The interferon-induced 67-kDa guanylate-binding protein (hGBP1) is a GTPase that converts GTP to GMP. *The Journal of biological chemistry* **269**, 11299-11305 (1994).

172   Mears, J. A., Ray, P. & Hinshaw, J. E. A corkscrew model for dynamin constriction. *Structure* **15**, 1190-1202, doi:10.1016/j.str.2007.08.012 (2007).

173   Reubold, T. F. *et al.* Crystal structure of the GTPase domain of rat dynamin 1. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13093-13098, doi:10.1073/pnas.0506491102 (2005).

174   Grigliatti, T. A., Hall, L., Rosenbluth, R. & Suzuki, D. T. Temperature-sensitive mutations in Drosophila melanogaster. XIV. A selection of immobile adults. *Molecular & general genetics : MGG* **120**, 107-114 (1973).

175   Urrutia, R., Henley, J. R., Cook, T. & McNiven, M. A. The dynamins: redundant or distinct functions for an expanding family of related GTPases? *Proceedings of the National Academy of Sciences of the United States of America* **94**, 377-384 (1997).

176   Muhlberg, A. B., Warnock, D. E. & Schmid, S. L. Domain structure and intramolecular regulation of dynamin GTPase. *The EMBO journal* **16**, 6676-6683, doi:10.1093/emboj/16.22.6676 (1997).

177   Hinshaw, J. E. & Schmid, S. L. Dynamin self-assembles into rings suggesting a mechanism for coated vesicle budding. *Nature* **374**, 190-192, doi:10.1038/374190a0 (1995).

178   Tuma, P. L., Stachniak, M. C. & Collins, C. A. Activation of dynamin GTPase by acidic phospholipids and endogenous rat brain vesicles. *The Journal of biological chemistry* **268**, 17240-17246 (1993).

179   Stowell, M. H., Marks, B., Wigge, P. & McMahon, H. T. Nucleotide-dependent conformational changes in dynamin: evidence for a mechanochemical molecular spring. *Nature cell biology* **1**, 27-32, doi:10.1038/8997 (1999).

180   Sweitzer, S. M. & Hinshaw, J. E. Dynamin undergoes a GTP-dependent conformational change causing vesiculation. *Cell* **93**, 1021-1029 (1998).

181   Marks, B. *et al.* GTPase activity of dynamin and resulting conformation change are essential for endocytosis. *Nature* **410**, 231-235, doi:10.1038/35065645 (2001).

182   Roux, A., Uyhazi, K., Frost, A. & De Camilli, P. GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission. *Nature* **441**, 528-531, doi:10.1038/nature04718 (2006).

183   Faelber, K. *et al.* Crystal structure of nucleotide-free dynamin. *Nature* **477**, 556-560, doi:10.1038/nature10369 (2011).

184   Sever, S., Muhlberg, A. B. & Schmid, S. L. Impairment of dynamin's GAP domain stimulates receptor-mediated endocytosis. *Nature* **398**, 481-486, doi:10.1038/19024 (1999).

185   Smirnova, E., Shurland, D. L., Newman-Smith, E. D., Pishvaee, B. & van der Bliek, A. M. A model for dynamin self-assembly based on binding between three different protein domains. *The Journal of biological chemistry* **274**, 14942-14947 (1999).

186   Haslam, R. J., Koide, H. B. & Hemmings, B. A. Pleckstrin domain homology. *Nature* **363**, 309-310, doi:10.1038/363309b0 (1993).

187   Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. Crystal structure at 2.2 A resolution of the pleckstrin homology domain from human dynamin. *Cell* **79**, 199-209 (1994).

188   Ford, M. G., Jenni, S. & Nunnari, J. The crystal structure of dynamin. *Nature* **477**, 561-566, doi:10.1038/nature10441 (2011).

189   Faelber, K. *et al.* Structural insights into dynamin-mediated membrane fission. *Structure* **20**, 1621-1628, doi:10.1016/j.str.2012.08.028 (2012).

190   Chappie, J. S. *et al.* A pseudoatomic model of the dynamin polymer identifies a hydrolysis-dependent powerstroke. *Cell* **147**, 209-222, doi:10.1016/j.cell.2011.09.003 (2011).

191   Low, H. H., Sachse, C., Amos, L. A. & Lowe, J. Structure of a bacterial dynamin-like protein lipid tube provides a mechanism for assembly and membrane curving. *Cell* **139**, 1342-1352, doi:10.1016/j.cell.2009.11.003 (2009).

192   Hu, J. *et al.* A class of dynamin-like GTPases involved in the generation of the tubular ER network. *Cell* **138**, 549-561, doi:10.1016/j.cell.2009.05.025 (2009).

193   Orso, G. *et al.* Homotypic fusion of ER membranes requires the dynamin-like GTPase atlastin. *Nature* **460**, 978-983, doi:10.1038/nature08280 (2009).

194   Byrnes, L. J. *et al.* Structural basis for conformational switching and GTP loading of the large G protein atlastin. *The EMBO journal* **32**, 369-384, doi:10.1038/emboj.2012.353 (2013).

195   Byrnes, L. J. & Sondermann, H. Structural basis for the nucleotide-dependent dimerization of the large G protein atlastin-1/SPG3A. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2216-2221, doi:10.1073/pnas.1012792108 (2011).

196   Janeway, C. A., Jr. & Medzhitov, R. Innate immune recognition. *Annual review of immunology* **20**, 197-216, doi:10.1146/annurev.immunol.20.083001.084359 (2002).

197   Janeway, C. A., Jr. How the immune system works to protect the host from infection: a personal view. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 7461-7468, doi:10.1073/pnas.131202998 (2001).

198   Medzhitov, R. & Janeway, C. A., Jr. Innate immunity: impact on the adaptive immune response. *Current opinion in immunology* **9**, 4-9 (1997).

199   Medzhitov, R. & Janeway, C., Jr. Innate immunity. *The New England journal of medicine* **343**, 338-344, doi:10.1056/NEJM200008033430506 (2000).

200   Medzhitov, R. Toll-like receptors and innate immunity. *Nature reviews. Immunology* **1**, 135-145, doi:10.1038/35100529 (2001).

201   Janeway Jr, C. Immunobiology. *Garland Publishing* (2001).

202   MacMicking, J. D. IFN-inducible GTPases and immunity to intracellular pathogens. *Trends in immunology* **25**, 601-609, doi:10.1016/j.it.2004.08.010 (2004).

203   Boehm, U., Klamp, T., Groot, M. & Howard, J. C. Cellular responses to interferon-gamma. *Annual review of immunology* **15**, 749-795, doi:10.1146/annurev.immunol.15.1.749 (1997).

204   Decker, T., Stockinger, S., Karaghiosoff, M., Muller, M. & Kovarik, P. IFNs and STATs in innate immunity to microorganisms. *The Journal of clinical investigation* **109**, 1271-1277, doi:10.1172/JCI15770 (2002).

205   Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167-170, doi:10.1126/science.1179555 (2010).

206   Haller, O., Staeheli, P. & Kochs, G. Interferon-induced Mx proteins in antiviral host defense. *Biochimie* **89**, 812-818, doi:10.1016/j.biochi.2007.04.015 (2007).

207   Staeheli, P. *et al.* Interferon-regulated influenza virus resistance gene Mx is localized on mouse chromosome 16. *Journal of virology* **58**, 967-969 (1986).

208   Murray, H. W. *et al.* Role of tryptophan degradation in respiratory burst-independent antimicrobial activity of gamma interferon-stimulated human macrophages. *Infection and immunity* **57**, 845-849 (1989).

209   Pfefferkorn, E. R., Rebhun, S. & Eckel, M. Characterization of an indoleamine 2,3-dioxygenase induced by gamma-interferon in cultured human fibroblasts. *Journal of interferon research* **6**, 267-279 (1986).

210   Adams, L. B., Hibbs, J. B., Jr., Taintor, R. R. & Krahenbuhl, J. L. Microbiostatic effect of murine-activated macrophages for Toxoplasma gondii. Role for synthesis of inorganic nitrogen oxides from L-arginine. *Journal of immunology* **144**, 2725-2729 (1990).

211   Taylor, G. A., Feng, C. G. & Sher, A. Control of IFN-gamma-mediated host resistance to intracellular pathogens by immunity-related GTPases (p47 GTPases). *Microbes and infection / Institut Pasteur* **9**, 1644-1651, doi:10.1016/j.micinf.2007.09.004 (2007).

212   Plattner, F. & Soldati-Favre, D. Hijacking of host cellular functions by the Apicomplexa. *Annual review of microbiology* **62**, 471-487, doi:10.1146/annurev.micro.62.081307.162802 (2008).

213   Dubey, J. P. The history of Toxoplasma gondii--the first 100 years. *The Journal of eukaryotic microbiology* **55**, 467-475, doi:10.1111/j.1550-7408.2008.00345.x (2008).

214   Kim, K. & Weiss, L. M. Toxoplasma: the next 100years. *Microbes and infection / Institut Pasteur* **10**, 978-984, doi:10.1016/j.micinf.2008.07.015 (2008).

215   Konen-Waisman, S. & Howard, J. C. Cell-autonomous immunity to Toxoplasma gondii in mouse and man. *Microbes and infection / Institut Pasteur* **9**, 1652-1661, doi:10.1016/j.micinf.2007.09.005 (2007).

216   Howe, D. K. & Sibley, L. D. Toxoplasma gondii comprises three clonal lineages: correlation of parasite genotype with human disease. *The Journal of infectious diseases* **172**, 1561-1566 (1995).

217   Zhao, Y. O., Khaminets, A., Hunn, J. P. & Howard, J. C. Disruption of the Toxoplasma gondii parasitophorous vacuole by IFNgamma-inducible immunity-related GTPases (IRG proteins) triggers necrotic cell death. *PLoS pathogens* **5**, e1000288, doi:10.1371/journal.ppat.1000288 (2009).

218   Zhao, Y. O. *et al.* Toxoplasma gondii and the Immunity-Related GTPase (IRG) resistance system in mice: a review. *Memorias do Instituto Oswaldo Cruz* **104**, 234-240 (2009).

219   Collazo, C. M. *et al.* The function of gamma interferon-inducible GTP-binding protein IGTP in host resistance to Toxoplasma gondii is Stat1 dependent and requires expression in both hematopoietic and nonhematopoietic cellular compartments. *Infection and immunity* **70**, 6933-6939 (2002).

220   Butcher, B. A. *et al.* p47 GTPases regulate Toxoplasma gondii survival in activated macrophages. *Infection and immunity* **73**, 3278-3286, doi:10.1128/IAI.73.6.3278-3286.2005 (2005).

221   Henry, S. C. *et al.* Balance of Irgm protein activities determines IFN-gamma-induced host defense. *Journal of leukocyte biology* **85**, 877-885, doi:10.1189/jlb.1008599 (2009).

222   Halonen, S. K., Taylor, G. A. & Weiss, L. M. Gamma interferon-induced inhibition of Toxoplasma gondii in astrocytes is mediated by IGTP. *Infection and immunity* **69**, 5573-5576 (2001).

223   Taylor, G. A. *et al.* Pathogen-specific loss of host resistance in mice lacking the IFN-gamma-inducible gene IGTP. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 751-755 (2000).

224   Martens, S. *et al.* Disruption of Toxoplasma gondii parasitophorous vacuoles by the mouse p47-resistance GTPases. *PLoS pathogens* **1**, e24, doi:10.1371/journal.ppat.0010024 (2005).

225   MacMicking, J. D., Taylor, G. A. & McKinney, J. D. Immune control of tuberculosis by IFN-gamma-inducible LRG-47. *Science* **302**, 654-659, doi:10.1126/science.1088063 (2003).

226   Boehm, U. *et al.* Two families of GTPases dominate the complex cellular response to IFN-gamma. *Journal of immunology* **161**, 6715-6723 (1998).

227   Taylor, G. A., Feng, C. G. & Sher, A. p47 GTPases: regulators of immunity to intracellular pathogens. *Nature reviews. Immunology* **4**, 100-109, doi:10.1038/nri1270 (2004).

228   MacMicking, J. D. Immune control of phagosomal bacteria by p47 GTPases. *Current opinion in microbiology* **8**, 74-82, doi:10.1016/j.mib.2004.12.012 (2005).

229   Bekpen, C. *et al.* The interferon-inducible p47 (IRG) GTPases in vertebrates: loss of the cell autonomous resistance mechanism in the human lineage. *Genome biology* **6**, R92, doi:10.1186/gb-2005-6-11-r92 (2005).

230   Gavrilescu, L. C., Butcher, B. A., Del Rio, L., Taylor, G. A. & Denkers, E. Y. STAT1 is essential for antimicrobial effector function but dispensable for gamma interferon production during Toxoplasma gondii infection. *Infection and immunity* **72**, 1257-1264 (2004).

231   Gilly, M. & Wall, R. The IRG-47 gene is IFN-gamma induced in B cells and encodes a protein with GTP-binding motifs. *Journal of immunology* **148**, 3275-3281 (1992).

232   Sorace, J. M., Johnson, R. J., Howard, D. L. & Drysdale, B. E. Identification of an endotoxin and IFN-inducible cDNA: possible identification of a novel protein family. *Journal of leukocyte biology* **58**, 477-484 (1995).

233   Zerrahn, J., Schaible, U. E., Brinkmann, V., Guhlich, U. & Kaufmann, S. H. The IFN-inducible Golgi- and endoplasmic reticulum- associated 47-kDa GTPase IIGP is transiently expressed during listeriosis. *Journal of immunology* **168**, 3428-3436 (2002).

234   Taylor, G. A. *et al.* Identification of a novel GTPase, the inducibly expressed GTPase, that accumulates in response to interferon gamma. *The Journal of biological chemistry* **271**, 20399-20405 (1996).

235   Taylor, G. A. *et al.* The inducibly expressed GTPase localizes to the endoplasmic reticulum, independently of GTP binding. *The Journal of biological chemistry* **272**, 10639-10645 (1997).

236   Martens, S. *et al.* Mechanisms regulating the positioning of mouse p47 resistance GTPases LRG-47 and IIGP1 on cellular membranes: retargeting to plasma membrane induced by phagocytosis. *Journal of immunology* **173**, 2594-2606 (2004).

237   Feng, C. G. *et al.* Mice deficient in LRG-47 display increased susceptibility to mycobacterial infection associated with the induction of lymphopenia. *Journal of immunology* **172**, 1163-1168 (2004).

238   Collazo, C. M. *et al.* Inactivation of LRG-47 and IRG-47 reveals a family of interferon gamma-inducible genes with essential, pathogen-specific roles in resistance to infection. *The Journal of experimental medicine* **194**, 181-188 (2001).

239   Santiago, H. C. *et al.* Mice deficient in LRG-47 display enhanced susceptibility to Trypanosoma cruzi infection associated with defective hemopoiesis and intracellular control of parasite growth. *Journal of immunology* **175**, 8165-8172 (2005).

240   Carlow, D. A., Marth, J., Clark-Lewis, I. & Teh, H. S. Isolation of a gene encoding a developmentally regulated T cell-specific protein with a guanine nucleotide triphosphate-binding motif. *Journal of immunology* **154**, 1724-1734 (1995).

241   Lafuse, W. P., Brown, D., Castle, L. & Zwilling, B. S. Cloning and characterization of a novel cDNA that is IFN-gamma-induced in mouse peritoneal macrophages and encodes a putative GTP-binding protein. *Journal of leukocyte biology* **57**, 477-483 (1995).

242   Hunn, J. P. *et al.* Regulatory interactions between IRG resistance GTPases in the cellular response to Toxoplasma gondii. *The EMBO journal* **27**, 2495-2509, doi:10.1038/emboj.2008.176 (2008).

243   Papic, N., Hunn, J. P., Pawlowski, N., Zerrahn, J. & Howard, J. C. Inactive and active states of the interferon-inducible resistance GTPase, Irga6, in vivo. *The Journal of biological chemistry* **283**, 32143-32151, doi:10.1074/jbc.M804846200 (2008).

244   Melzer, T., Duffy, A., Weiss, L. M. & Halonen, S. K. The gamma interferon (IFN-gamma)-inducible GTP-binding protein IGTP is necessary for toxoplasma vacuolar disruption and induces parasite egression in IFN-gamma-stimulated astrocytes. *Infection and immunity* **76**, 4883-4894, doi:10.1128/IAI.01288-07 (2008).

245   Ling, Y. M. *et al.* Vacuolar and plasma membrane stripping and autophagic elimination of Toxoplasma gondii in primed effector macrophages. *The Journal of experimental medicine* **203**, 2063-2071, doi:10.1084/jem.20061318 (2006).

246   Alaganan, A., Fentress, S. J., Tang, K., Wang, Q. & Sibley, L. D. Toxoplasma GRA7 effector increases turnover of immunity-related GTPases and contributes to acute virulence in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 1126-1131, doi:10.1073/pnas.1313501111 (2014).

247   Ghosh, A., Uthaiah, R., Howard, J., Herrmann, C. & Wolf, E. Crystal structure of IIGP1: a paradigm for interferon-inducible p47 resistance GTPases. *Molecular cell* **15**, 727-739, doi:10.1016/j.molcel.2004.07.017 (2004).

248   Wittinghofer, A. & Pai, E. F. The structure of Ras protein: a model for a universal molecular switch. *Trends in biochemical sciences* **16**, 382-387 (1991).

249   Pawlowski, N. *et al.* The activation mechanism of Irga6, an interferon-inducible GTPase contributing to mouse resistance against Toxoplasma gondii. *BMC biology* **9**, 7, doi:10.1186/1741-7007-9-7 (2011).

250 Uthaiah, R. C., Praefcke, G. J., Howard, J. C. & Herrmann, C. IIGP1, an interferon-gamma-inducible 47-kDa GTPase of the mouse, showing cooperative enzymatic activity and GTP-dependent multimerization. *The Journal of biological chemistry* **278**, 29336-29343, doi:10.1074/jbc.M211973200 (2003).

251 McCoy AJ, G.-K. R., Adams PD, Winn MD, Storoni LC, Read RJ. . *J Appl Cryst* **40**, 658-674 (2007).

252 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501, doi:10.1107/S0907444910007493 (2010).

253 hen VB, A. W., Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, & Richardson JS, R. D. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst* **D66**, 16-21 (2010).

254 onine PV, G.-K. R., Echols N, Headd JJ, Moriarty NW, Mustyakimov & M, T. T., Urzhumtsev A, Zwart PH, Adams PD. Towards automated crystallographic structure refinement with phenix.refine. *Acta crystallographica. Section D, Biological crystallography* **68**, 352-367 (2012).

255 Brunger, A. T. Free R value: cross-validation in crystallography. *Methods in enzymology* **277**, 366-396 (1997).

256 W., K. *Acta. Cryst.* **A32**, 922-923 (1976).

257 Schrödinger, L. The PyMOL Molecular Graphics System Version 1.7.4.

258 al., C. e. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica* **D66**, 12-21 (2010).

259 Pawlowski, N. Complex Formation and the GTP Hydrolysis Mechanism of the Immunity-Related GTPase Irga6 (2009).

260 Schwefel, D. *et al.* Structural basis of oligomerization in septin-like GTPase of immunity-associated protein 2 (GIMAP2). *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20299-20304, doi:10.1073/pnas.1010322107 (2010).

261 Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* **22**, 4673-4680 (1994).

Abstract

In my PhD thesis, I addressed two different topics. The first part dealt with the structural and biochemical characterization on the aryl hydrocarbon receptor (AhR). This protein belongs to the bHLH-PAS family of transcription factors that sense environmental signals. As all members of this family, AhR comprise a bHLH domain, necessary for DNA binding, a PAS A domain, necessary for dimerization, a PAS B domain, which in case of AhR is able to bind ligands and an unstructured TAD domain, involved in transactivation. Upon ligand binding, AhR heterodimerizes with the aryl hydrocarbon receptor nuclear translocator (ARNT), which also belongs to the bHLH-PAS family of transcription factors. Through this dimerization, the fully active heterodimeric transcription factor is formed. Depending on the ligand, transcriptional activation orchestrates the expression of genes required for detoxification, pro-tumor and anti-tumor pathway or immune response.

In this study, AhR was functionally and structurally characterized to better understand the process of heterodimerization and DNA binding on a molecular level. This study showed for the first time successful co-expression and co-purification of the AhR/ARNT heterodimeric transcription factor complex comprising the bHLH and PAS A domains. The AhR/ARNT complex appeared properly folded and heterodimerized in a 1:1 stoichiometry. The affinity of the AhR/ARNT complex towards dsDNA containing the DRE was in the lower nano-molar range. This interaction was formed by one molecule AhR/ARNT complex and one molecule double stranded (ds)DNA. The AhR/ARNT complex bound to 12mer dsDNA was crystallized and crystals were extensively optimized. The optimized crystals diffracted anisotropically up to 3.5 Å and various phasing attempts were carried out in order to solve the structure of the AhR/ARNT complex bound to dsDNA. Although structure determination was not achieved within this study, the success of high yield purification of AhR/ARNT complex and biochemical characterization *in vitro* is a mile stone for further structural investigations.

The second part of this study dealt with the structural characterization of the immunity-related GTPase Irga6. Immunity-related GTPases (IRGs) constitute a powerful cell-autonomous resistance systems against several intracellular pathogens in mice. In mice, 23 paralogs of IRG proteins are identified and Irga6 is one of the best studied members of the family. Irga6 is composed of a G domain flanked by a C-terminal and an N-terminal helical domain. Biochemical studies analyzed an interface, which spans over the nucleotide binding pocket of the G domains and is crucial for nucleotide-dependent

oligomerization and cooperative GTP-hydrolysis. However, the already known crystal structures of Irga6 show a dimerization mode via the backside dimer but do not reveal this G domain interface.

In this part of the study, the structure of a non-oligomerizing Irga6 mutant bound to GMPPNP was solved. This new structure reveals the G domain interface and thus a new mode of G domain dimerization for the IRG family. Beside the nucleotide, two additional interfaces participate in G domain dimerization. A hydrogen bonding network between the switch II and the *trans* stabilizing loop as well as hydrophobic interactions between the switch I loop and the *trans* activation loop and the G4 loop establish the G domain dimer interface. These findings are in excellent agreement with the previously published biochemical data. Based on the solved G dimer structure in combination with the published biochemical data and the structural information available for BDLP I proposed an oligomerization model of Irga6 with the G dimer as a nucleotide-dependent interaction interface.

Zusammenfassung

Innerhalb dieser Arbeit wurden zwei Themen behandelt. Zunächst befasst sich die Arbeit mit der strukturellen und biochemischen Charakterisierung des Aryl-hydrocarbon-rezeptors (AhR). Dieser gehört zu der bHLH-PAS-Domänen Familie der Transkriptionsfaktoren und reagiert auf unterschiedlichste Umwelteinflüsse. AhR besteht, wie alle Mitglieder dieser Proteinfamilie, aus einer bHLH Domäne, welche die DNA Bindung ermöglicht, aus einer PAS A Domäne, notwendig für die Heterodimerisierung, einer PAS B Domäne, welche im Falle von AhR Liganden binden kann, sowie einer unstrukturierten TAD Domäne. Durch Ligandenbindung wird AhR aktiviert und heterodimerisiert mit dem Aryl-hydrocarbon-rezeptor-nuclear-translocator (ARNT), ebenso Mitglied der oben genannten Proteinfamilie, zu einem aktiven, heterodimeren Transkriptionsfaktor. Dieser Komplex kann bestimmte DNA Motive binden und dadurch die Transkription aktivieren. Abhängig von der Art des Liganden wird die Transkription von unterschiedlichen Genen aktiviert, welche wiederum die Detoxification, das Tumorwachstum oder aber die Immunantwort beeinflussen können.

Um die Heterodimerisierung und DNA-Bindung auf molekularer Ebene zu verstehen, wurde in dieser Studie AhR funktionell und strukturell charakterisiert. Erstmalig konnte hier die Coaufreinigung und Coexpression der bHLH und PAS A Domänen des AhR/ARNT Komplexes gezeigt werden. Die weitere Charakterisierung ergab, dass der AhR/ARNT Komplex korrekt gefaltet ist und aus jeweils einem Molekül AhR und ARNT geformt wird. Die Affinität des AhR/ARNT Komplexes zu doppelsträngiger DNA, welche ein DRE Motiv beinhaltet, liegt im niedrigen nanomolaren Bereich. Jeweils ein Molekül AhR/ARNT Komplex und ein Molekül doppelsträngige DNA sind an der Interaktion beteiligt. Der AhR/ARNT Komplex, gebunden an ein Molekül doppelsträngiger DNA, wurde kristallisiert und die Kristalle wurden umfangreich optimiert. Durch die Optimierung konnte eine Diffraktion der Kristalle bis zu einer Auflösung von 3.5 Å erreicht werden. Zahlreiche Phasierungsmethoden wurden angewendet, um die Kristallstruktur des AhR/ARNT Komplexes, gebunden an doppelsträngige DNA, zu lösen. Obwohl die Strukturbestimmung in dieser Studie nicht gelang, ist der Erfolg der Aufreinigung von AhR im Komplex mit seinem Interaktionspartner ARNT und dessen biochemischen Charakterisierung *in vitro* als hervorragende Basis für weiterführende strukturelle Studien zu sehen.

Der zweite Teil dieser Arbeit behandelt die strukturelle Charakterisierung der immunity-related GTPase (IRG) Irga6. Diese stellen in Mäusen ein wirkungsvolles, zellautonomes Resistenzsystem gegen diverse interzelluläre Pathogene dar. In Mäusen sind insgesamt 23 Paraloge der IRG Proteine identifiziert, wobei Irga6 das wohl am besten studierte Mitglied der Proteinfamilie ist. Irga6 besteht aus einer N- und C-terminalen, helikalen Domäne sowie aus einer G Domäne, welche Nukleotid-abhängige Oligomerisation und kooperative GTP Hydrolyse bewirkt. Biochemische Studien zeigten, dass die für Oligomerisation und kooperative GTP Hydrolyse nötige Interaktionsfläche den Bereich der Nukleotidbindetasche der G Domäne von Irga6 mit einschließt. Die bisher bekannten Kristallstrukturen von Irga6 konnten jedoch diese G Domänen Dimerisierung nicht darstellen.

In diesem Teil der Studie wurde eine oligomerisations-defizite Irga6 Mutante, gebunden an GMPPNP, strukturell charakterisiert, um die G Domänen Interaktion zu verstehen und die Oligomerisation von Irga6 zu erklären. In der Tat konnte die Kristallstruktur der Irga6 Mutante, gebunden an GMPPNP, gelöst und ein neuer Mechanismus der G Domänen Dimerisierung beschrieben werden, welcher das Nukleotid involviert. Neben dem Nukleotid konnten zwei weitere Interaktionselemente identifiziert werden, die die G Domänen Dimerisierung stabilisieren. Ein Wasserstoffbrückennetzwerk zwischen der switch-II-Schleife und der *trans*-Stabilisierungs-Schleife sowie hydrophobe Wechselwirkungen zwischen der Switch-I-Schleife und der G4-Schleife bzw. der *trans*-Stabilisierungs-Schleife tragen hier zur Stabilisierung bei. Insgesamt bestätigt die hier präsentierte Kristallstruktur die bekannten biochemischen Daten. Basierend auf der hier ermittelten Kristallstruktur von Irga6, den strukturellen Informationen zu BDLP und den bekannten biochemischen Daten schlage ich ein Oligomerisationsmodell für Irga6 vor, welches die neu entdeckte nukleotidabhängige G Domänen Dimerisierung zur Grundlage hat.

Appendix A Instrument list

Instrument	Manufacturer
24-well crystallization plates	Hampton Research, Aliso Viejo,USA
45 Ti rotor	Beckman Coulter, Krefeld, D
96 well crystallization plates	Sigma-Aldrich, Steinheim, D
Agarose Gel Electrophoresis System	OLS, Bremen, D
Amicon centrifugal filter device	Millipore, Billerica, USA
Automated imaging and storing system Rock Imager	Formulatrix, Waltham, USA
Benchtop Centrifuge 5415 R	Eppendorf, Hamburg, D
Benchtop Centrifuge 5804 R	Eppendorf, Hamburg, D
Block Heater Rotilabo H250	Roth, Karlsruhe, D
Centrifuge Avanti J-26 XP	Beckman Coulter, Krefeld, D
Chromatography column HisTrap HP 1 ml	GE Healthcare, Piscataway, USA
Chromatography columns Superdex200 16/60, 26/60	GE Healthcare, Piscataway, USA
Chromatography columns Superdex75 16/60, 26/60	GE Healthcare, Piscataway, USA
Cryo-Fridge VIP Series -86°C	Sanyo, Moriguchi, J
CryoLoops, various sizes	Hampton Research, Aliso Viejo, USA
Fluidizer M-110 L Pneumatic	Microfluidics, Newtown, USA
FPLC Äkta Prime Plus / Purifier	GE Healthcare, Piscataway, USA
Fridge N3956 4°C/-20°C	Liebherr, Biberach an der Riss, D
Imaging system LAS4000 mini	FujiFilm, Düsseldorf, D
Isothermal Titration Calorimeter ITC200 (MicroCal)	GE Healthcare, Piscataway, USA
JLA 8.100 rotor	Beckman Coulter, Krefeld, D
Microscope FluoView FV1000	Olympus, Hamburg, D
NanoDrop 2000	Thermo Scientific, Wilmington, USA
Optima MAX-XP benchtop ultracentrifuge	Beckman Coulter, Krefeld, D
PCR thermocycler T-Gradient thermoblock	Biometra, Göttingen, D
Peristaltic pump ISM 827 B	Ismatec, Wertheim, D
pH-Meter	Mettler-Toledo, Gießen, D
Pipettes Eppendorf Research vario	Eppendorf, Hamburg, D
Pipetting robot Gryphon	Art Robins Instruments, Sunnyvale, USA
Precision scales	Mettler-Toledo, Gießen, D
RALS 270 dual detector	Malvern Instruments, Worcestershire, U.K.
Refractive index detector VE 3580	Malvern Instruments, Worcestershire, U.K.
SDS PAGE System Xcell Sure Lock	Life Techologies, Karlsruhe, D
Shaker Incubator Innova 44 R	New Brunswick Scientific, Enfield, USA
Thermoblock MKR13	HLC Biotech, Bovenden, D
Ultracentrifuge Optima L-100 K	Beckman Coulter, Krefeld, D
Vacuum pump	Vacuubrand, Wertheim, D
Vortex Genie 2	Bender+Hobien, Zurich, CH
Water quench Julabo TW20	Julabo, Seelbach, D

Appendix B Chemical list

Chemical	Cat.-No.	Manufacturer
2-Log DNA ladder	N3200S	NEB, Frankfurt a. M., D
Acetic Acid	3783.5	Roth, Karlsruhe, D
Acetone	9372.2	Roth, Karlsruhe, D
Acetonitrile	CN20.2	Roth, Karlsruhe, D
Additive Screen	HR2-428	Hampton Research, Aliso Viejo, USA
Agarose	2267.3	Roth, Karlsruhe, D
Ammonium acetate	09689	Sigma-Aldrich, Steinheim, D
Ammonium chloride	09700	Sigma-Aldrich, Steinheim, D
Ammonium citrate dibasic	09833	Sigma-Aldrich, Steinheim, D
Ammonium fluoride	09737	Sigma-Aldrich, Steinheim, D
Ammonium formate	09735	Sigma-Aldrich, Steinheim, D
Ammonium iodide	09874	Sigma-Aldrich, Steinheim, D
Ammonium nitrate	09889	Sigma-Aldrich, Steinheim, D
Ammonium phosphate monobasic	9709	Sigma-Aldrich, Steinheim, D
Ammonium sulfate	9212.2	Roth, Karlsruhe, D
Calcium acetate Hydrate	21056	Sigma-Aldrich, Steinheim, D
Calcium chloride	A119.1	Roth, Karlsruhe, D
Chloramphenicol	3886.3	Roth, Karlsruhe, D
Coomassie 108rilliant blue R 250 (C.I. 42660)	3862.2	Roth, Karlsruhe, D
di-Ammonium hydrogen phosphate	09839	Sigma-Aldrich, Steinheim, D
di-Potassium hydrogen phosphate anhydrous	P749.2	Roth, Karlsruhe, D
di-Sodium hydrogen phosphate anhydrous	P030.2	Roth, Karlsruhe, D
DTT	6908.2	Roth, Karlsruhe, D
EDTA	8040.2	Roth, Karlsruhe, D
Ethanol	5054.2	Roth, Karlsruhe, D
Ethidium bromide	2218.1	Roth, Karlsruhe, D
GeneAmp©dNTPs	N8080007	Roche Molecular, Branchburg, USA
Glutathione SepharoseTM 4B	27-4574-01	Amersham, Piscatawy, USA
Glycerol	3783.1	Roth, Karlsruhe, D
GSH reduced	3541	Calbiochem, Darmstadt, D
GTP analog GMPPNP	NU-1012-1G	Jena Bioscience, Jena, D
Guanidinehydrochloride	0037.1	Roth, Karlsruhe, D
HEPES	9105.4	Roth, Karlsruhe, D
Imidazole	3899.3	Roth, Karlsruhe, D
Isopropanol	9866.5	Roth, Karlsruhe, D
Kanamycinsulfate	T823.4	Roth, Karlsruhe, D
Lithium acetate dihydrate	62393	Sigma-Aldrich, Steinheim, D
Lithium citrate tribasic tetrahydrate	62484	Sigma-Aldrich, Steinheim, D
Lithium nitrate	62574	Sigma-Aldrich, Steinheim, D
Magnesium acetate tetrahydrate	63049	Sigma-Aldrich, Steinheim, D
Magnesium chloride hexahydrate	63065	Sigma-Aldrich, Steinheim, D

Appendix B Chemical list

Magnesium formate dihydrate	00793	Sigma-Aldrich, Steinheim, D
Magnesium nitrate	237175-100G	Sigma-Aldrich, Steinheim, D
Magnesium sulfate heptahydrate	63138	Sigma-Aldrich, Steinheim, D
Malonic acid	63290	Sigma-Aldrich, Steinheim, D
Methanol	4627.5	Roth, Karlsruhe, D
MPD	68340	Sigma-Aldrich, Steinheim, D
Ni Sepharose HP	71-5027-67 AD	GE Healthcare, München, D
NuPAGE©LDS Sample Buffer (4x)	NP0007	Life Techologies, Karlsruhe, D
NuPAGE©MES SDS Buffer Kit	NP0060	Life Techologies, Karlsruhe, D
NuPAGE©MOPS SDS Buffer Kit	NP0050	Life Techologies, Karlsruhe, D
NuPAGE©Novex 4-12% Bis-Tris Gel 1.5 mm, 10/15 well	NP0335BOX / NP0336BOX	Life Techologies, Karlsruhe, D
Pefabloc©SC-Protease inhibitor	A154.2	Roth, Karlsruhe, D
PEG 1000	81188	Sigma-Aldrich, Steinheim, D
PEG 2000MME	81321	Sigma-Aldrich, Steinheim, D
PEG 3350	88276	Sigma-Aldrich, Steinheim, D
PEG 400	91893	Sigma-Aldrich, Steinheim, D
PEG 4000	95904	Sigma-Aldrich, Steinheim, D
PEG 500MME	71578	Sigma-Aldrich, Steinheim, D
PEG 8000	89510	Sigma-Aldrich, Steinheim, D
Potassium acetate	60035	Sigma-Aldrich, Steinheim, D
Potassium chloride	6781.3	Roth, Karlsruhe, D
Potassium citrate tribasic monohydrate	25107	Sigma-Aldrich, Steinheim, D
Potassium dihydrogen phosphate	3904.1	Roth, Karlsruhe, D
Potassium fluoride	60239	Sigma-Aldrich, Steinheim, D
Potassium formate	60246	Sigma-Aldrich, Steinheim, D
Potassium iodide	60400	Sigma-Aldrich, Steinheim, D
Potassium nitrate	60414	Sigma-Aldrich, Steinheim, D
Potassium phosphate	3904.3	Roth, Karlsruhe, D
Potassium sulfate	60528	Sigma-Aldrich, Steinheim, D
Potassium thiocyanate	60517	Sigma-Aldrich, Steinheim, D
Sodium acetate trihydrate	71188	Sigma-Aldrich, Steinheim, D
Sodium chloride	9265.2	Roth, Karlsruhe, D
Sodium citrate tribasic dihydrate	71402	Sigma-Aldrich, Steinheim, D
Sodium di-hydrogen phosphate 2-hydrate	T879.1	Roth, Karlsruhe, D
Sodium fluoride	71519	Sigma-Aldrich, Steinheim, D
Sodium hydroxide	6771.1	Roth, Karlsruhe, D
Sodium nitrate	71755	Sigma-Aldrich, Steinheim, D
Sodium sulfate decahydrate	71969	Sigma-Aldrich, Steinheim, D
Sodium tartrate dibasic dihydrate	71994	Sigma-Aldrich, Steinheim, D
Sodium thiocyanate	71938	Sigma-Aldrich, Steinheim, D
Terrific-Broth medium	HP61.1	Roth, Karlsruhe, D
Tetrabutylammonium bromide	86860-500G	Sigma-Aldrich, Steinheim, D
Trichloromethane	6340.2	Roth, Karlsruhe, D
Triton X 100	X100-500ml	Sigma Aldrich, Steinheim, D
Tris-HCl	PHG0002	Sigma Aldrich, Steinhiem, D

Tryptone/peptone	8952.2	Roth, Karlsruhe, D
Tween20	P9416	Sigma Aldrich, Steinheim, D
Uranyl acetate dihydrate	73943	Sigma-Aldrich, Steinheim, D
Urea	U4883	Sigam Aldrich, Steinheim, D
Yeast extract	2363.2	Roth, Karlsruhe, D
Zinc acetate Dihydrate	96459	Sigma-Aldrich, Steinheim, D

Appendix C Alignments


```
mm_BMAL1   1 : ------MADQRMDISSTISDFMSPGPTDLLSGSLGTSGVDCNRKRKGSATLYQLDDFAFE :  54
mm_ARNT    1 : MAATTANPEMTSDVPSLGPTIASGNPGPGIQGGGAVVQRAIKRRSGLDFDLEVEVNTKFL :  60

                                              helix 1

mm_BMAL1  55 : ESMDTDKDLPHGRLEYAEHCG----RIKNAREAHSQIEKRRRDKMNSFIDELASLVPTCN : 110
mm_ARNT   61 : RCDDDQMCNDKERFARSDDEQSSADKERLARENHSEIERRRRNKMTAYITELSDMVPTCS : 120

                        helix 2              helix 3        sheet 1

mm_BMAL1 111 : AMSRKLDKLTVLRMAVQHMKTLRGATNPYTEANYKPTFLSDDELKHLILRAADGFLFVVG : 170
mm_ARNT  121 : AIARKPDKLTILRMAVSHMKSLRGTGNTSTDGSYKPSFLTDQELKHLILEAADGFLFIVS : 180

               sheet 2   helix 6              helix 7     not resolved

mm_BMAL1 171 : CDRGKILFVSESVFKILNYSQNDLIGQSLFDYLHPKDIAKVKEQLSSSDTAPRERLIDAK : 230
mm_ARNT  181 : CETGRVVYVSDSVTPVLNQPQSEWFGSTLYDQVHPDDVDKLREQLSTSENALTGRVLDLK : 240

                              sheet 3              not resolved

mm_BMAL1 231 : TGLPVKTDITPGPSRLGSGARRSFFCRMKCNRPSVKVEDKD----FASTCSK-----KKD : 281
mm_ARNT  241 : TGT-VKKEGQQSSMRMCMGSRRSFICRMRCGTSSVDPVSMNRLSFLRNRCRNGLGSVKEG : 299

                   sheet 4      not resolved        sheet 5

mm_BMAL1 282 : RKSFCTIHSTGYLKSWPETKMGLDEDNEPDNEGCNLSCLVAIGRLHSHMVFQPANGEIRV : 341
mm_ARNT  300 : EPHFVVVHCTGYIKAWPEAGVSLPDDPEAGQGS-KFCLVAIGRLQVTSSFNCTLMSNIC  : 358

                sheet 6   sheet 7  helix 8   helix 9   helix 10

mm_BMAL1 342 : KSMEYVSRHAIDGKFVFVLQRATAILAYLPQELLGTSCYEYFHQDLIGHLAECHRQVLQT : 401
mm_ARNT  359 : QPTEFISRHNIEGIFTFVLHRCVATVGYQPQELLGKNIVEFCHPEDQQLIRDSFQQVVKL : 418

                   sheet 8         sheet 9         sheet 10

mm_BMAL1 402 : REKITTNCYKFKIKDGSFITLRSRWFSFMNPWTKEVEYIVSTNTVVLAN----------- : 450
mm_ARNT  419 : KGQVLSVMFRFRSKTREWLWMRTSSFTFQNPYSDEIEYIICTNTNVKNSSQEPRPTLSNT : 478

mm_BMAL1 451 : ------------VLEGGDPTFP------------------QLTAPPHSMDSMIESGEGGP : 480
mm_ARNT  479 : IPRSQLGPTANLSLEMGTGQLPSRQQQQQHTELDMVPGRDGLASYNHSQVSVQEVASAGS : 538

mm_BMAL1 481 : KRTHPT------------------VEGIPGGTRAGAG---------------------- : 499
mm_ARNT  539 : EHSKPLEKSEGLFAQDRDPRFPEIYPSITADQSKGISSSTVPATQQLFSQGSSFPPNPRP : 598

mm_BMAL1 500 : --------------------KIGRMIAEEIMEIHRIRGSSPSSCGSS------------- : 526
mm_ARNT  599 : AENFRNSGLTPPVTIVQPSSSAGQILAQISRHSNPAQGSAPTWTSSSRPGFAAQQVPTQA : 658

mm_BMAL1 527 : ----------------------PLNITSTPPPDASSPGGKKILNGGTPDIPSTGLLPGQAC : 565
mm_ARNT  659 : TAKTRSSQFGVNNFQTSSSFSAMSLPGAPTASSGTAAYPALPNRGSNFPEETGQTTGCFC : 718

mm_BMAL1 566 : ETPGYPYSDSSSILGENPHIGIDMID--NDCGSSSPSNDEAAMAVIMSLIEADAGLGGPV : 623
mm_ARNT  719 : ARTAEGVGVWPQWQGQQPHHRSSSSEQHVQCTQAQAPSQPEVFQEMLSMIGDQSNTYNNE : 778

mm_BMAL1 624 : DFSDLPWPL---- : 632
mm_ARNT  779 : EFPDITMFPPFSE : 791
```

helix 1

```
mm_CLOCK    1 : MVFTVSCSKMSSIVDRDDSSIFDGLVEEDDKDKAKRVSRNKSEKKRRDQFNVLIKEIGSM : 60
hs_AhR      1 : --MNSSSANITYASRKRRK-----PVQKTVKPIPAEGIKSNPSKRHRDRINTEIDRIASL : 53
```

helix 2   helix 3

```
mm_CLOCK   61 : LPG---NARKMDKSTVLQKSIDFLRKHKETTAQSDASEIRQDWKPTFLSNEEFTQ----- : 112
hs_AhR     54 : LPFPQDVINKLDKLSVIRLSVSYLRAKSFFDVALKSSPTERNGGQDNCRAANFREGLNLQ : 113
```

helix 3   sheet 1   sheet 2   helix 4   helix 5   helix 6

```
mm_CLOCK  113 : ---LMLEALDGFFLAIMTDGSIIYVSESVTSLLEHLPSDLVDQSIFNFIPEGEHSEVYKI : 169
hs_AhR    114 : EGEFLLQALNGFVLVVTTDALVFYASSTIQDYLGFQQSDVIHQSVYELIHTEDRAEFQRQ : 173
```

helix 6   helix 7   sheet 3   sheet 4

```
mm_CLOCK  170 : LSTHILESDSLTP-EYIKSKNQIEFCCHMLRGTIDPKEPSTYEYVRFIGNFKSITSVSTS : 228
hs_AhR    174 : LHWAINPSQCTESGQGIEEATGLPQTVVCYNPDQIPPENSPLMERCFICRLRCILDNSSG : 233
```

not resolved   sheet 5   sheet 6

```
mm_CLOCK  229 : THNG-FEGTIQRTHRPSYEDR-----VCFVATVRLATPQFIKEMCTVEEPNEEFTSRHSL : 282
hs_AhR    234 : FLAMNFQGKLKYLHGQKKKGKDGSILPPQLALFAIATPLQPPSILEIRTKNFIFRTKHKL : 293
```

sheet 7   helix 8   helix 9   helix 10   sheet 8

```
mm_CLOCK  283 : EWKFIFLDHRAPPIIGYLPFEVLG-TSGYDYYHVDDIENIAKCHEHLMQYGKGKSCYYRF : 341
hs_AhR    294 : DFTPIGCDAKGRIVLGYTEAELCTRGSGYQFIHAADMLYCAESHIRMIKTGESGMIVFRL : 353
```

sheet 9   sheet 10

```
mm_CLOCK  342 : LTKGQQWIWLQTHYYITYHQWNSRPEFIVCTHTVVSYAEVRAERRRELGIEESLPETAAD : 401
hs_AhR    354 : LTKNNRWTWVQSNARLLYK--NGRPDYIIVTQRPLTDEEG-TEHLRKRNTKLPFMFTTGE : 410
```

```
mm_CLOCK  402 : KSQDSGSDNRINTVSLKEALERFDHSPTPSASSRSSRKSSHTAVSDPSSTPTKIPTDTST : 461
hs_AhR    411 : AVLYEATNPFPAIMDPLPLRTKNGTSGKDSATTSTLSKDSLNPSSLLAAMMQQDESIYLY : 470
```

```
mm_CLOCK  462 : PPRQHLPAHEKMTQRRSSFSSQSINSQSVGPSITQPAMSQAANIPIPCGMSQFQFSAQLG : 521
hs_AhR    471 : PASSTSSTAPFENNFFNESMNECRNWQDNTAPMGNDTILKHEQIDQPQDVNSFAG----G : 526
```

```
mm_CLOCK  522 : AMQHLKDQLEQRTRMIEANIHRQQEELRKIQEQLQMVHGQGLQMFLQQSNPGLNEGSVQL : 581
hs_AhR    527 : HPGLFQDSKNSDLYSIMKNLGIDFEDIRHMQNE--------KFFRNDFSGEVDFRDIDI : 577
```

```
mm_CLOCK  582 : SSGNSNIQQLTPVNMQGQVVPANQVCSGHISTGQHMICQCTIQSTSTQCSQCSVMSGHSC : 641
hs_AhR    578 : T--DEILTYVQDSLSKSPFIPSDYQCQQSIALNSSCMVQEHIHLEQQCHHCKQVVVEPC : 635
```

```
mm_CLOCK  642 : CTSLPSQTPSTLTAPLYNTMVISQPAAGSMVCIPSSMPQNSTQSATVTTFTQDRQIRFSQ : 701
hs_AhR    636 : CQLCQKMKHMCVNGMFENWNSNQFVPFNCPQQDPQQYNVFTDLHGISQEFPYKSEMDSMP : 695
```

```
mm_CLOCK  702 : GQCLVTKLVTAPVACGAVMVPSTMLMG-QVVTAYPTFATQCQQAQTLSVTCCCQQ-QQQC : 759
hs_AhR    696 : YTQNFISCNQPVLPQHSKCTELDYPMGSFEPSPYPTTSSLEDFVTCIQIPENCKHGLNPC : 755
```

```
mm_CLOCK  760 : PPQQQQCQQQSSQEQQLPSVQPAQAQLGQPPQQFIQTSRLLHGNPSTQLIISAAFPLCQ : 819
hs_AhR    756 : SAIITPCTCYAGAVSMYQCQPEPQHTHVGQMQYNPVLPGQQAFLNKFQNGVINETYEAEL : 815
```

```
mm_CLOCK  820 : STFPPSHHQQHQPQQQQQLPRHRTDSLTDPSKVQPQ : 855
hs_AhR    816 : NNINNTQTTTHLQPLHHPSEARPFPDLTSSGFL--- : 848
```

Amino acid sequences of mouse (mm) BMAL1 (Uniprot accession: Q9WTL8) and mouse (mm) ARNT (P53762) and the amino acid sequences of mouse (mm) CLOCK (O08785) and human (hs) AhR (P35869) were aligned using Clustal W [261]. Residues with a conservation greater than 80% are colored in red. mmBMAL1 and mmCLOCK (PDB: 4F3L) were used as a reference to determine the secondary structure. Alpha-helices are shown as cylinders, beta-sheets are shows as arrows. Not resolved regions within the structure (PDB: 4F3L) are shown as dotted lines. bHLH is colored in orange, PAS A is colored in blue, PAS B is colored in purple.

Appendix D Clone list

residues	protein	vector backbone	primer	mutations/deletions
90-465	mmARNT	pETDuet MSC1	C659 C660	
90-465	mmARNT	Precission-pETDuet MSC1	C910 C911	
85-465	mmARNT	Precission-pETDuet MSC1	C939 C40	
90-465	mmARNT	Precission-pETDuet MSC1	C941 C942	Δ166-189
90-465	mmARNT	Precission-pETDuet MSC1	C943 C944	polyA 182-189
85-465	mmARNT	Precission-pETDuet MSC1	C982 C983	Δ274-303
85-465	mmARNT	Precission-pETDuet MSC1	C984 C985	Δ277-300
85-465	mmARNT	Precission-pETDuet MSC1	D034 D035	Δ277-300, C256S
85-465	mmARNT	Precission-pETDuet MSC1	D036 D037	Δ277-300, CC366/367 SS
85-465	mmARNT	Precission-pETDuet MSC1	D038 D039	Δ277-300, Δ365-372
85-465	mmARNT	Precission-pETDuet MSC1	D034 D035	Δ274-303, C256S
85-465	mmARNT	Precission-pETDuet MSC1	D038 D039	Δ274-303, Δ365-372
85-465	mmARNT	Precission-pETDuet MSC1	D034 D035	Δ277-300, Δ365-372, C256S
85-465	mmARNT	Precission-pETDuet MSC1	D034 D035	Δ274-303, Δ365-372, C256S
85-360	mmARNT	pET30 EK/LIC	D079 D080	
85-360	mmARNT	pET30 EK/LIC	D079 D080	Δ274-303, C256S
85-465	mmARNT	pET30 EK/LIC	D079 D322	
23-487	cohsAhR	pET32 EK/LIC	D159b D158	
23-273	cohsAhR	pET32 EK/LIC	D158 D159	
23-487	cohsAhR	pET46 EK/LIC	D158 D159b	
23-373	cohsAhR	pET46 EK/LIC	D158 D159	
85-465	mmARNT	Precission-pETDuet MSC1	D237 D238	Δ274-303, Δ365-372, C256S, H94A, E98A
85-465	mmARNT	Precission-pETDuet MSC1	D239 D240	Δ274-303, Δ365-372, C256S, H94A, R102A
85-465	mmARNT	Precission-pETDuet MSC1	D245 D246	Δ274-303, Δ365-372, C256S, E98A
85-465	mmARNT	Precission-pETDuet MSC1	D235 D236	Δ274-303, Δ365-372, C256S, H94A, E98A, R102A
85-465	mmARNT	Precission-pETDuet MSC1	D241 D242	Δ274-303 Δ365-372, C256S, E98A, R102A
85-465	mmARNT	Precission-pETDuet MSC1	D247 D248	Δ274-303, Δ365-372, C256S, R102A
23-273	cohsAhR	pET32 EK/LIC	D446 D447	Δ182-210
85-465	mmARNT	pET30 EK LIC	D034 D035	Δ274-303 Δ365-372 C256S
23-394	cohsAhR	pET32 EK/LIC	D446 D447	Δ182-210
11-186	cohsAhR	pGEX-6P1 / pSKB2 LN2	C265 C267	
11-347	cohsAhR	pGEX-6P1 / pSKB2 LN2	C265C269	

11-411	cohsAhR	pGEX-6P1 / pSKB2 LN2	C265	C270
96-186	cohsAhR	pGEX-6P1 / pSKB2 LN2	C266	C267
96-347	cohsAhR	pGEX-6P1 / pSKB2 LN2	C266	C269
96-411	cohsAhR	pGEX-6P1 / pSKB2 LN2	C266	C270
256-347	cohsAhR	pGEX-6P1 / pSKB2 LN2	C268	C269
256-411	cohsAhR	pGEX-6P1 / pSKB2 LN2	C268	C270
91-216	cohsAhR	pGEX-6P1 / pSKB2 LN2	C281	C282
91-391	cohsAhR	pGEX-6P1 / pSKB2 LN2	C281	C284
91-470	cohsAhR	pGEX-6P1 / pSKB2 LN2	C281	C286
281-391	cohsAhR	pGEX-6P1 / pSKB2 LN2	C283	C284
281-470	cohsAhR	pGEX-6P1 / pSKB2 LN2	C283	C286
356-470	cohsAhR	pGEX-6P1 / pSKB2 LN2	C285	C286
103-239	mmARNT	pGEX-6P1 / pSKB2 LN2	C386	C388
103-421	mmARNT	pGEX-6P1 / pSKB2 LN2	C386	C340
161-239	mmARNT	pGEX-6P1 / pSKB2 LN2	C387	C388
161-421	mmARNT	pGEX-6P1 / pSKB2 LN2	C387	C340
347-421	mmARNT	pGEX-6P1 / pSKB2 LN2	C389	C340
11-391	cohsAhR	pMAL-C2X/P2X	C265	C344
11-216	cohsAhR	pMAL-C2X/P2X	C265	C343
96-216	cohsAhR	pMAL-C2X/P2X	C266	C343
96-391	cohsAhR	pMAL-C2X/P2X	C266	C344
256-391	cohsAhR	pMAL-C2X/P2X	C268	C344
91-216	cohsAhR	pMAL-C2X/P2X	C281	C343
91-391	cohsAhR	pMAL-C2X/P2X	C281	C344
281-391	cohsAhR	pMAL-C2X/P2X	C283	C344
47-482	cohsAhR	pGEX-6P1 / pSKB2 LN2	C596	C597
47-181	cohsAhR	pGEX-6P1 / pSKB2 LN2	C596	C599
47-191	cohsAhR	pGEX-6P1 / pSKB2 LN2	C596	C600
47-474	cohsAhR	pGEX-6P1 / pSKB2 LN2	C596	C601
47-391	cohsAhR	pGEX-6P1 / pSKB2 LN2	C596	C602
109-482	cohsAhR	pGEX-6P1 / pSKB2 LN2	C598	C597
109-181	cohsAhR	pGEX-6P1 / pSKB2 LN2	C598	C599
109-191	cohsAhR	pGEX-6P1 / pSKB2 LN2	C598	C600
109-474	cohsAhR	pGEX-6P1 / pSKB2 LN2	C598	C601
109-391	cohsAhR	pGEX-6P1 / pSKB2 LN2	C598	C602
41-192	cohsAhR	pGEX-6P1 / pSKB2 LN2	C427	C428
41-419	cohsAhR	pGEX-6P1 / pSKB2 LN2	C427	C430
181-419	cohsAhR	pGEX-6P1 / pSKB2 LN2	C429	C430
269-419	cohsAhR	pGEX-6P1 / pSKB2 LN2	C431	C430

Table 10: Designed constructs of mmARNT and cohsAhR. First and last residue of the construct are indicated in the left column. Mentioned primers are the forward and reverse primer of the mentioned construct. If a mutation/deletion is highlighted in red in the right column, respective primers are the mutation/deletion primers.
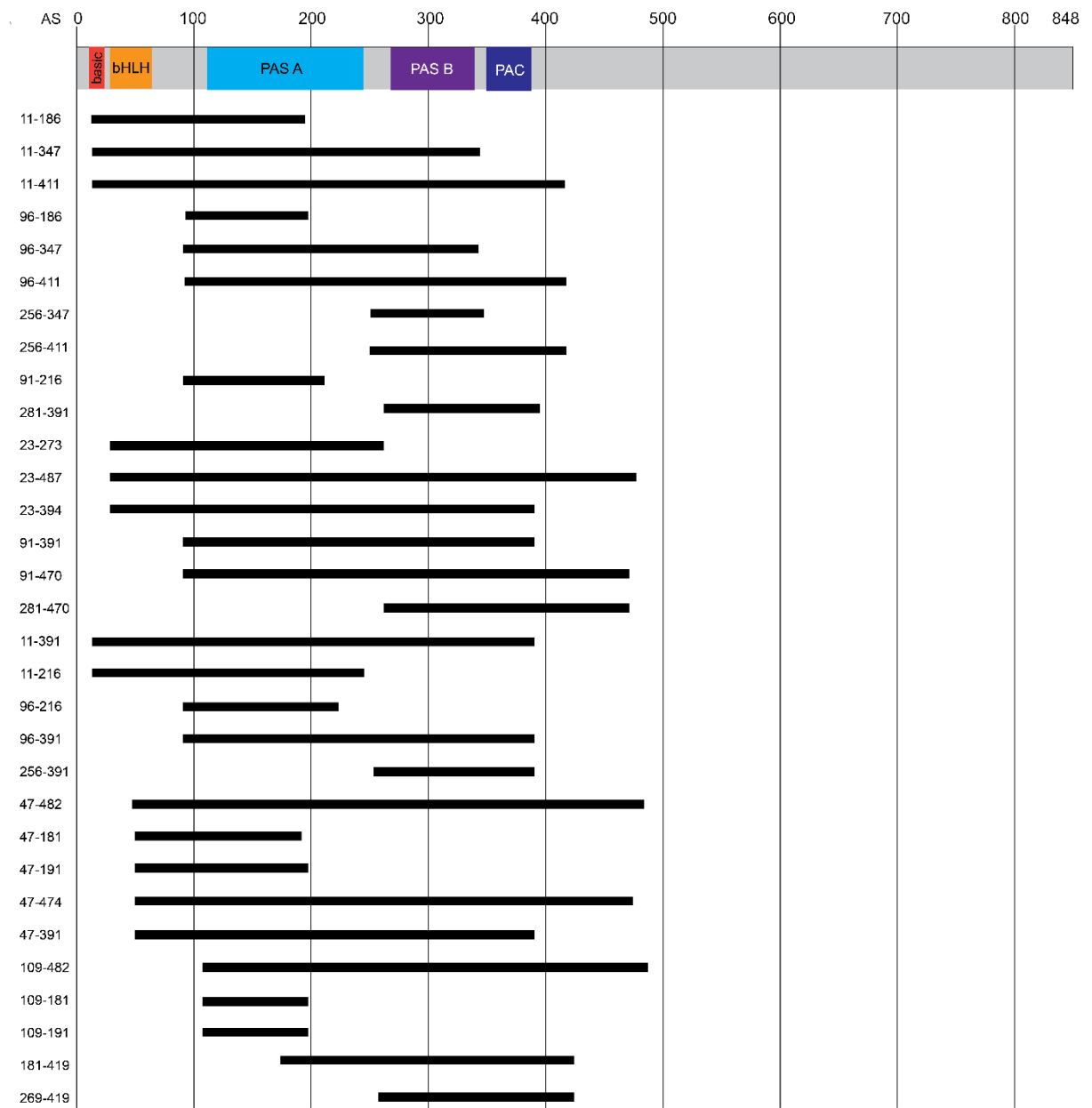
Figure 73: Schematic view of the Aryl hydrocarbon receptor. Designed constructs for cohsAhR are labelled with the first and the last amino acid and are shown as black lines in order to illustrate the domains each construct encompasses.
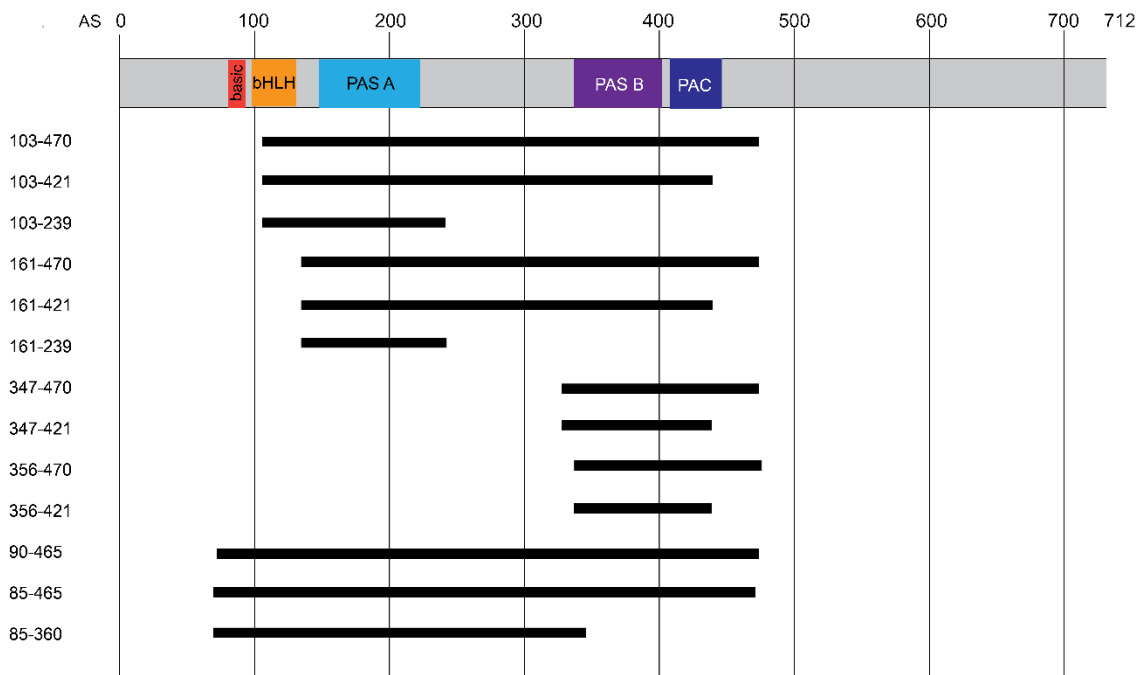
Appendix D Clone list

Figure 74: Schematic view of the Aryl hydrocarbon receptor nuclear translocator (ARNT). Designed constructs for mmARNT are labelled with the first and the last amino acid and are shown as black lines in order to illustrate the domains each construct encompasses.

Appendix E Abbreviations

AC	Affinity chromatography
AhR	Aryl hydrocarbon receptor
ARNT	Aryl hydrocarbon receptor nuclear translocator
Atl1	Atlastin 1
bHLH	basic helix-loop-helix
BMAL	Brain muscle ARNT like
cDNA	complementary deoxyribonucleic acid
cDNA	complementary deoxyribonucleic acid
CV	column volume
CLOCK	Circadian locomotor output cycles kaput
dd	double destilled
Ds	Double stranded
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
Dpn	Diplococcus pneumoniae
dsDNA	double stranded DNA
DTT	dithiothreitol
E	extinction
E. coli	Escherichia coli
e.g.	exempli gratia
EDTA	ethylenediaminetetraacetic acid
EM	electron microscopy
fw	forward
g	G-force
GED	GTPase effector domain
GMPPCP	guanosine-5'-[(β,γ)-methyleno]triphosphate
GMP-PNP	5'-guanylyl-β-γ-imidodiphosphate
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HP	high performance
hs	homo sapiens
Irga6	Immunity-related GTPase 6
L	litre
LB	Luria-Bertani
Log	logarithm
MES	2-(N-morpholino)ethanesulfonic acid
MOPS	3-(N-morpholino)propanesulfonic acid
MPD	2-methyl-1,3-propanediol
NMR	nuclear magnetic resonance
NTP	nucleotide triphosphate binding protein
O/N	overnight

151

Appendix E Abbreviations

OD	optical density
PAGE	polyacrylamide gel electrophoresis
PAS	period-arnt-sim
PEG	polyethylene glycol
R.m.s.	root mean square
rev	reverse
RT	room temperature
SDS	sodium dodecyl sulfate
SEC	size-exclusion chromatography
TB	Terrific-Broth
TBA(B)	tetrabutylammonium (bromide)
TBE	Tris/Borate/EDTA
Tris	tris(hydroxymethyl)aminomethane
TSS	transformation and storage solution
U	unit
w/o	without
w/v	weight per volume

Appendix F Amino acids

one letter code	three letter code	amino acid
A	Ala	alanine
C	Cys	cysteine
D	Asp	aspartate
E	Glu	glutamate
F	Phe	phenylalanine
G	Gly	glycine
H	His	histidine
I	Ile	isoleucine
K	Lys	lysine
L	Leu	leucine
M	Met	methionine
N	Asn	asparagine
P	Pro	proline
Q	Gln	glutamine
R	Arg	arginine
S	Ser	serine
T	Thr	threonine
V	Val	valine
W	Trp	tryptophane
Y	Tyr	tyrosine

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen – die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Oliver Daumke und Prof. Dr. Udo Heinemann betreut worden.

Berlin, Juni 2015

Kathrin Wiebke Schulte

Acknowledgement

I hereby would like to thank Prof. Dr. Oliver Daumke for his excellent supervision. His positive entity and enthusiasm were super-motivating, especially during lean periods.

Special thanks to Dr. Katja Fälber, who taught me a lot about crystallography. Without you, the Irga6 structure would not have been solved yet.

Many thanks to Dr. Eva Rosenbaum, who supported me a lot especially during writing my thesis.

I want to thank Sabine Werner and Jeanette Schlegel for being the good souls in our group and the whole AG Daumke for the great time I had.

Thanks to Dr. Yvette Roske for spending a lot of time with me at the beamline and teaching me everything related to protein crystals.

The MDC provides an excellent research environment, thanks for that.

I want to thank my family and friends for supporting and inspiring me on many different ways.

Finally, I want to thank Chris for being a bastion of calm and much more than that.