universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

Investigation of PRI as novel approach in structural determination of globular proteins

verfasst von / submitted by

Andreas Prattinger, Bsc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2017 / Vienna 2017

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 834
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Molekulare Biologie
Betreut von / Supervisor:	Univ.-Prof. Dr. Robert Konrat
Mitbetreut von / Co-Supervisor:	

Acknowledgements

I want to thank Univ. Prof. Dr. Robert Konrat for presenting the opportunity to learn and work under his guidance. I further want to thank the whole group for providing me with their support and insight, without which I could not have achieved this thesis. Hereby I want to especially express my gratitude towards Sven Brüschweiler who supervised me and had the patience to help me develop the presented work as well as Karin Ledolter, Leonhard Geist, Borja Mateos, Tomas Sara, Thomas Schwarz, Mate Somlyay, Nicolas Coudevylle and Andrea Flamm, who never hesitated to offer their knowledge and a helping hand.

I also want to thank my mother, father and grandmother for the support and love they lent me and my sister for critically reading this work. At last I want to thank Katharina, who not only offered her unconditional support but also her scientific expertise and helped me in every step I had to take to finish this thesis.

Glossary

ATM	Ataxia Telangiectasia Mutated
ATR	Ataxia telangiectasia and Rad3-related protein
b/HLH/Z	basic helix-loop-helix leucine zipper
Bard1	BRCA1-associated RING domain protein 1
BAX	BCL2 associated X
Brca1	BReast CAncer 1
BRCT	BRCA1 C Terminus
C/EBPb	CCAAT/enhancer-binding protein beta
CSR	curie spin relaxation
CtBP	C-terminal Binding Protein
CtIP	CtBP-interacting protein
DTT	Dithiothreitol
E-box	enhancer box
E.coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
EMT	epithelial-mesenchymal transition
ER	estrogen receptor
ERK1/2	extracellular signal-regulated kinase ½
FID	free induction decay
GB1	streptococcal protein G / domain B
HDAC	histone deacetylase
her2	Receptor tyrosine-protein kinase erbB-2
HLH	Helix-loop-helix
Hsqc	heteronuclear single quantum coherenc
hTERT	human Telomerase reverse transcriptase
IgG	Immunoglobulin G
IL-17	Interleukin-17
IL-1β	Interleukin-1-beta
INEPT	Insensitive nuclei enhanced by polarization transfer
INR	initiator element
IPTG	Isopropyl-β-D-thiogalactopyranosid

LB	lysogeny broth
LCN2	Lipocalin 2
MAD	Mothers against decapentaplegic
MAX	MYC associated factor X
MES	2-(N-morpholino)ethanesulfonic acid
Mga	MAX gene-associated protein
Miz-1	Myc-interacting zinc-finger protein 1
Mlx	Max-like protein X
MMP-9	matrix metallo proteinase gelatinase-B
Mnt	MAX network transcriptional repressor
MTSL	S-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonotl
MXI	MAX-interacting protein
Myc	MYC proto-oncogene, bHLH transcription factor
NGAL	Neutrophil gelatinase-associated lipocalin
Nmi	N-Myc-interacting protein)
NMR	nuclear magnetic resonance
NOE	nuclear Overhauser effect
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PCS	pseudo contact shifts
PolII	polymerase II
PR	progesterone receptor
PRE	paramagnetic relaxation enhancement
PRI	paramagnetic relaxation interference
RAC1	Ras-related C3 botulinum toxin substrate
Ras	Rat sarcoma
RDC	residual dipolar coupling
RING	Really interesting new gene
RMSD	Root-mean-square deviation
SCR	structural conserved region
SDS	sodium dodecyl sulfate
TEV	Tobacco Etch Virus nuclear-inclusion-a endopeptidase
TNF-α	tumor necrosis factor
Tris	tris(hydroxymethyl)aminomethane

1. Abstract

NMR-spectroscopy represents a well established and powerful tool in structural biology, which is not limited to only one method, but far more the basis for a wide array of ever developing experiments, some of which rely on the exploit of paramagnetic relaxation enhancement (PRE). This effect, paramagnetic entities have on nuclear spins, was already employed in NMR to refine structural analysis and also found applicability in the investigation of low populated transient states. PRE is a valuable addition to a classic NOE-based approach of structure solution, as PRE enables the measurement of significantly longer distances. Recently this technique has further evolved through the introduction of multiple sites of paramagnetism. Hereby it has been observed that the effect on one spin in range of multiple paramagnetic entities is not equal to the sum of the separate effects. This difference is due to dipole-dipole cross-correlation, of the paramagnetic electrons and the labelled proton. Paramagnetic relaxation interference (PRI) was already successfully used to study structural correlation within intrinsically disordered proteins. The aim of this study is to translate this approach to globular proteins and to evaluate the additional information gained through PRI. To assess accuracy and validity of this approach, different biological systems were recruited. The applicability was tested on the small globular protein GB1, NGAL, also known as LCN2 and Brca1 bound to a MAX homo-dimer.

Zusammenfassung

NMR-Spektroskopie ist ein gut etabliertes und wichtiges Werkzeug der Strukturbiologie. Hierbei ist diese nicht beschränkt auf eine Methode, sondern ist viel mehr die Basis für eine Vielzahl an sich stets weiterentwickelnden Experimenten. Einige dieser Experimente nutzen PRE (paramagnetic relaxation enhancement). PRE ist ein Effekt welcher paramagnetischen Molekülen entspringt und hierbei das Relaxationsverhalten von umliegenden Spins beeinflusst. Dies wurde bereits in der NMR-spektroskopie zur Strukturbestimmung wie auch zur Untersuchung von gering besiedelten Zuständen von Proteinen eingesetzt. PRE stellt eine bedeutende Ergänzung zu auf NOE-basierenden Methoden der Strukturbestimmung dar, da Information von signifikant längeren Distanzen ermittelt werden kann. In jüngster Zeit entwickelte sich diese Methode durch die Einführung mehrerer paramagnetischer Zentren innerhalb des zu messenden Moleküls weiter. Es konnte aufgezeigt werden, dass ein Spin innerhalb des Wirkungsradius mehrerer paramagnetischer Entitäten einem Effekt ausgesetzt ist welcher nicht gleich der Summe der Einzeleffekte ist. Diese Differenz entsteht aufgrund von Dipol-Dipol Cross-Korrelation der paramagnetischen Labels und des betroffenen Protons.

PRI (paramagnetic relaxation interference) wurde bereits erfolgreich verwendet um strukturelle Korrelation in IDPs (intrinsically disordered proteins) zu untersuchen. Das Ziel dieser Arbeit ist nun diese Methode an globulären Proteinen zu testen und die daraus gewonnen Daten zu evaluieren. Um die Genauigkeit wie auch die Gültigkeit von PRIs zu determinieren wurden unterschiedliche biologische Systeme verwendet. Die Anwendbarkeit wurde getestet mittels GB1, NGAL und Brca1 gebunden an ein MAX Homodimer.

Table of Contents

2. Introduction

2.1 NGAL

First discovered in human neutrophils[1], the 25kDa large protein NGAL was purified associated to MMP-9, hence the term Neutrophil gelatinase-associated lipocalin was established to describe this finding. This interaction is comprised of a disulfide linkage at NGAL Cys 87, which further inhibits the proteolytic function of MMP-9 through sequestration[2]. NGAL, also mentioned as LCN2 or 24p3, was found in homo-, dimer- and trimeric oligomeric state. The most prominent function of NGAL is the binding of small hydrophobic molecules. This includes the siderophores enterobactin and bacillobactin, which are utilised by gram negative bacteria in the uptake of ferric iron. Through this interaction, and further the endocytosis of both by the host, a bacteriostatic function arises. However, the biological relevance of NGAL is not just limited to this role, but this protein is present in multiple physiological and pathological pathways.

2.1.1 Structure

NGAL, as member of the Lipocalin family, exhibits the structural features characteristic for this group. The Lipocalin family is a group of secreted proteins characterised through a specific fold. This conserved three dimensional structure was shown through crystallographic studies[3]. On a functional level Lipocalins are known for the binding of cell-surface receptors, the formation of macromolecular complexes and the binding of small hydrophobic molecules. Throughout the group there is only a low pairwise sequence identity, often below 20%. The family of Lipocalins is further separated within, according to the presence of structural conserved regions (SCRs). Whereas the “kernel” Lipocalins, NGAL being one of them, exhibit all three of these SCRs, the “outliers” only displaying two.

The Lipocalin fold is characterised by eight antiparallel beta-sheets forming a hydrogen bonded barrel structure. In case of NGAL, the structure additionally possesses a short N – terminal 3_{10} helix and a C – terminal alpha-helix followed by another short beta-sheet. Further crystal structures revealed the presence of 3_{10} helices within proximity of beta-strand B2 and B5. This barrel structure encloses the ligand binding site. The specificity of the Lipocalin is defined by the sequence of amino acids located in this recess and the loops between the beta sheets, as well as the overall conformation and size of the calyx. Regarding NGAL, this specificity is created through an arrangement in three subpockets. Within these recesses, the ligand binding specific amino acids Lys 125, Lys 134 and Arg 81 are located[4]. Further, Tyr 106 was shown to stabilize this interaction via hydrogen bonds. The overall structure of NGAL is also maintained through an intramolecular cysteine bond (Cys 76 - Cys 176). As mentioned above, a third cysteine is present. This residue enables interaction with other proteins but also serves homodimerisation.

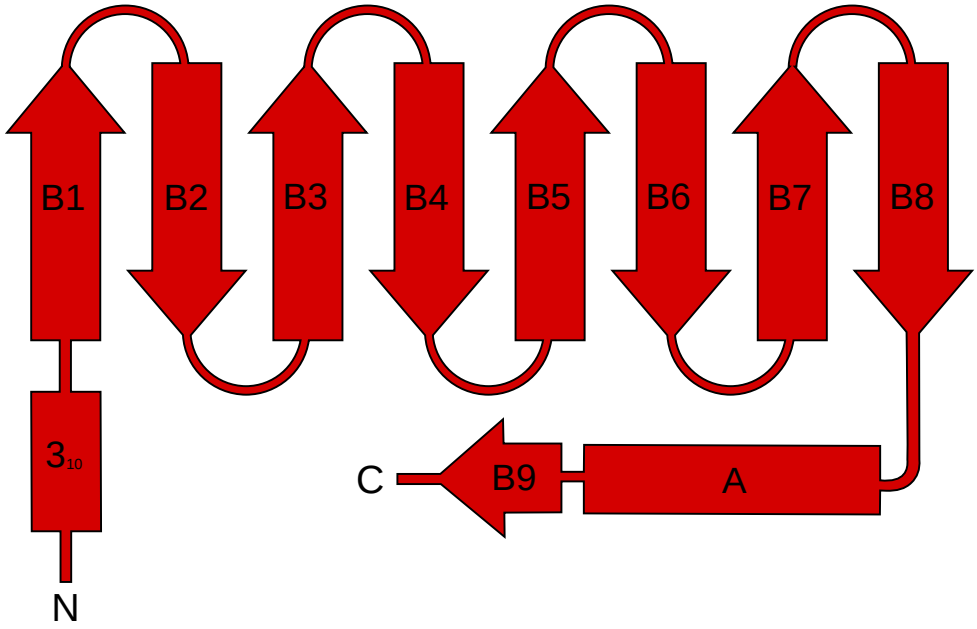

Figure 2.1: Representation of the Lipocalin fold, whereas B1 to B9 signify the beta sheets and A represents the alpha helix.

2.1.2 Physiology

Despite observing low common sequence identity within the group of Lipocalins, functional homology could be demonstrated. This functionality is maintained through high conservation in the SCRs, in which hydrophobic residues involved in ligand binding reside. Studying the expression levels, it became apparent, that though first located in neutrophil granules, evidence suggests that several tissue types express NGAL, although at a relatively low level. Further NGAL is predominantly secreted, exhibiting a 200-fold higher concentration extracellular. In contrast to healthy tissue, a broad range of diseases induces a significant upregulation of NGAL expression. These stimuli are not restricted to one type of condition, but vary from cancer to renal diseases or neurological diseases. Further instances of increased levels have been found in adipose tissue and in the context of diabetes[4].

The best described function of NGAL is its bacteriostatic effect through the binding of siderophores, therefore assigning it to the group of siderocalins. Siderocalins have been found in several species. This emphasizes its biological relevance as a general strategy of higher eucaryotes to fend off bacterial infection[5]. Due to the low concentration of free iron in the human body, it is essential for bacteria to efficiently harvest this scarce resource. Siderophores hereby represent an approach to outcompete host iron-binding proteins in their association strength to iron. This influx of iron, mediated through siderophores, into bacterial cells is essential for their proliferation. NGAL does not directly compete with the bacterial iron uptake, but instead binds the siderophore with high affinity (Enterobactin: KD = 0.4 and Bacillobactin: K_D = 0.5 nM) and transports the complex of NGAL-Ent-Fe into the host cell through endocytosis. It is assumed, that there is a counterpart to siderophores in mammalians, called catechols, which are also proficient in NGAL binding[4]. The role of NGAL in innate immune response can be shown through injection with E.coli, which resulted in a 22-fold increased serum level. It is hypothesised that there is another function of NGAL participating in inflammation, namely to operate as a chemoattractant for neutrophils[2].

Apart from its role as bacteriostatic agent, NGAL has been identified to be upregulated in several types of non-microbially-associated cancers, whereas the source for this impulse is the sterile inflammation in the tumor microenvironment. Through the cytokines IL-1β, TNF-α and IL-17, NF-κB (IL-1β, TNF-α) and C/EBPb (IL-17) are affected, which further increases expression of NGAL. Despite inconclusiveness regarding the role of NGAL as tumorigenic factor, two primary functions have been elucidated:

I. Promotion of metastasis
II. Promotion of cell survival

I. Essential for the initiation of metastasis is the ability of cells to invade tissue, this is enabled through EMT. NGAL affects this pathway in a MMP-9-dependent and -independent way. If MMP-9 is bound to NGAL, it is protected from degradation and can restructure the extracellular matrix. Independent to MMP-9 NGAL can upregulate ERK1/2, hereby increasing cell migration. This development is further affected by the induced decrease in E-cadherin and increased expression of vimentin and fibronectin. In colon cancer it was shown that NGAL can stimulate EMT additionally through RAC1[6]. It has to be mentioned that, regarding the invasiveness of breast and colon cancer also contrary results were found[7].

II. The high proliferation of cancer cells leads to an increased demand for iron. Therefore, increased iron uptake may lead to increased survivability of cells. It was proposed that additionally to transferrin, NGAL can participate in iron harvest, by binding iron-bound catechols in a similar fashion to the bacteriostatic effect of NGAL.

By means of systemic injection of LPS (lipopolysaccharide) in mice, a neuroinflammatory state was mimicked, which lead to a dramatically increased expression of NGAL. These studies exposed NGAL as mediator of neuroinflammatory signals but also as neuroinflammatogen[8]. The neurotoxic effect is limited to neurons and does not affect astrocytes, microglia or oligodendrocytes. The secretion of NGAL is performed by reactive astrocytes and microglia. In neuronal cells the expression levels of NGAL, measured through RNA abundance, are dismissible, the corresponding receptor 24p3R however is strongly expressed. The

high levels of the protein NGAL found within neuronal cells can therefore be explained through the active uptake of extracellular protein. This mechanism further leads to accumulation of protein, which supports the role of NGAL as neurotoxin. Another consequence of NGAL secretion is stimulation of neighbouring astrocytes for reactivation[9]. This neurotoxic effect of reactive astrocytes through increased secretion of NGAL is also present in neurodegenerative diseases, such as ALS or Alzheimer's disease. Due to this correlation, NGAL was also proposed as possible biomarker[10].

The best known and also the most understood pathological involvement is the role of NGAL in kidney diseases. In both acute kidney injury and chronic kidney diseases an elevated level of iron in blood and urine are common. It was shown that during septic as well as aseptic conditions in kidney injury, NGAL is released by the damaged nephron. This further induces NGAL expression in nephrons throughout the ischemic domain. Regarding kidney injury, these increased expression levels act in response to infection, but were also found in aseptic conditions. This suggests a function to prevent iron-catalyzed damage through catechol exploiting scavenging of free iron. The secreted NGAL was found in oligomeric states, differing between chronic and acute renal defect. Whereas acute kidney injury only caused monomeric oligomeric state, chronic disease produced dimers and hetero-multimers[11]. Due to the fast response and significantly elevated levels of expression, NGAL was proposed as possible biomarker for acute kidney injury, proposing an alternative to the observation of serum creatinine levels, which can be observed only after 24-72 hours[12].

2.2 GB1

GB1 is the isolated domain B of the streptococcal protein G, which was first observed in Streptococcus group C and G. This domain is repeated twice or thrice within the sequence, depending on the specific strain. It possesses the ability to bind immunoglobulin, specifically several members of the IgG family. The streptococcal protein G is located in the cell membrane of streptococcal bacteria, therefore allowing interaction with host IgG[13]. Owing to this interaction, protein G is used to specifically purify and label antibodies[14]. A structurally separate domain also located on the streptococcal protein G interacts with human serum albumin[15]. GB1 is a well structured protein of 55aa, that serves frequently as model to investigate methods of structural biology. GB1 earned this role due to its already well defined structure and small size, but also due to the straightforward expressibility and handling. Due to its mechanical stability GB1 was used as a system to study the mechanical response of proteins. This was examined through e.g. protein unfolding studies, applying forces across different axes throughout the protein[16].

2.2.1 Structure

GB1 exhibits a distinct three-dimensional structure, consisting of an alpha-helix packed against a four-stranded beta sheet. The alpha-helix hereby connects strand two and three, granting the protein a succession of features in a fashion of (N-B1-B2-A-B3-B4-C). This is further represented in *figure 2.2*. The β-strands themselves are aligned to each other antiparallel with exception of the orientation of B1 to B4, which is parallel[17].

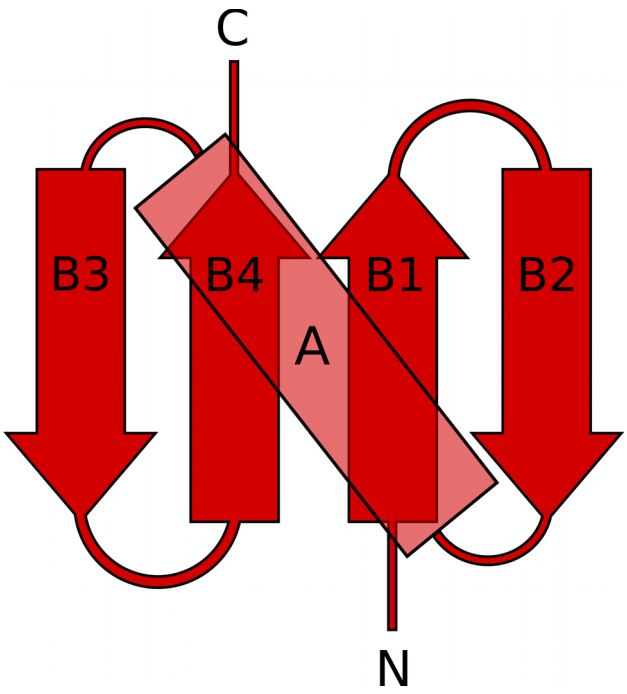

Figure 2.2: Representation of the secondary structure arrangment of GB1

Partly owing to its small size, it was possible to apply molecular dynamics to simulate the folding pathway of GB1. It was concluded that folding is initialised at the second hairpin, followed by stabilisation of the central region of the alpha helix. Regarding the interaction between GB1 and immunoglobulin, results demonstrating binding of the Fc, as well as the Fab fragment of IgG, have been reported. The interaction with the Fc fragment was characterised to occur in a “knobs-to-holes” fashion. This binding motif can be separated into two features, the first being comprised of the GB1 B1 residues E27, held in position by K31, interacting with the opening formed by I253 and S254 of the Fc fragment. The other built up by N434/Fc interacting with B1 of GB1 (N35, D36, D40, E42, W43)[18]. The interaction between Fab and GB1 was characterised to emerge from the last beta-strand of the constant CH1 domain. From this point, several hydrogen bonds reach out to GB1/B2. Additionally T16, T21 and T22 of GB1 secure the interaction with IgG[19].

2.3 Brca1 – Peptide

As final experimental system to measure intermolecular PREs and further PRI, the Max homodimer bound to a Brca1 peptide was chosen. Although the interaction between the two full-length proteins is not published and was argued against[20], NMR results clearly show strong interaction between MAX and different Brca1 peptides. The sequence chosen for this experimental set-up is a 58 amino acid long peptide located in the disordered region of Brca1. The BRCA1 gene is located on chromosome 17q21 and encoding for a 1863 amino acid long protein, best known for its role in hereditary breast cancer. Though the function of its natural form is as a tumor suppressor, repairing double strand breaks of DNA, a large variety of different mutations lead to a significant increase of breast and ovarian cancer. These mutations are frequently found in the N-terminal Ring domain, the BRCT domain or in exons 11-13[21].

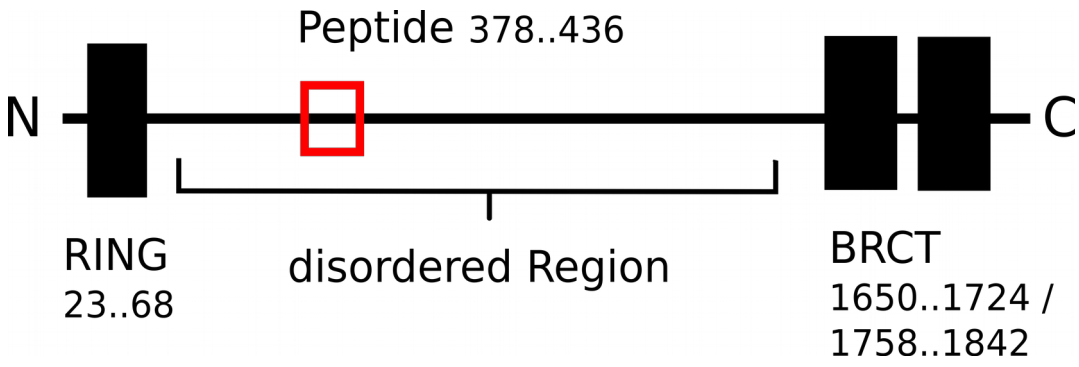

Figure 2.3.1: The figure shows a simplified representation of the Brca1 protein. We hereby highlight the N-terminal Ring domain and the BRCT domains, as well as their respective position within the protein sequence. The red square symbolises the position of the peptide which was chosen as construct. All depicted numbers indicate the specific position of the respective feature, given in number of amino acids.

2.3.1 Structure

Brca1 is classically divided into three domains. First of which is the highly conserved RING (Really Interesting New Gene) domain (aa 23-68). The RING – finger motif, first only characterised as a cysteine rich region, is a specific class of zinc finger proteins. In case of Brca1, the RING domain is comprised of one zinc finger, which is build up by a three-stranded BETA-sheet and a central helix surrounded by two alpha helices[21]. The RING domain exhibits E3-ubiquitin ligase activity and serves as dimerisation interface. Hereby noteworthy is the Brca1 – Bard1 heterodimer[22]. The other structurally well defined domain is the C-terminal domain of Brca1 (BRCT), which is present as a tandem repeat[23]. This domain acts as binding partner for proteins phosphorylated by ATM and ATR but also enables interaction with DNA[24], specifically with double strand breaks[25]. Structurally, the BRCT motif is comprised of a four stranded beta sheet surrounded by three alpha helices, whereas the two modules are orientated in a head-to-tail fashion. Despite these well structured sites the majority of the Brca1 sequence is predominantly disordered. The peptide used for our studies, as well as the exons 11-13, which are known loci for cancer-inducing mutations, are both located within this region. Nevertheless, this does not negate functionality, since several binding sites are located within this disordered region.

2.3.2 Physiology

Brca1 is an intensively studied protein, mostly focusing on the interplay with factors leading to tumorigenesis. Mutations of Brca1 are most commonly found in breast and ovarian cancer. The protein engages in transcriptional regulation, stress response, cell cycle regulation, apoptosis and DNA repair. Further, an effect on mammary gland stem cell differentiation has been proposed[26]. Expression of Brca1 is highly tissue specific, exhibiting increased expression levels in testis, thymus, spleen, ovaries and mammary glands, whereby the expression increases during puberty[27]. Further

expression is also dependent on cell-cycle phase, whereby the highest protein concentration was observed before entry into S phase. Although hereditary breast cancer only accounts for 5-10% of breast cancer cases[28], within this fraction, mutations in Brca1 and Brca2 were found in nearly half of patients[29]. Even though these mutations don't directly lead to tumour formation, the presence of additional oncogenic factors leads to cancer development. This profound involvement in cancerogenisis is partially due to the participation in the Tricomplex of Nmi-*wt*Brca1-c-Myc and the consequential regulation of the promoter activity of hTERT[30]. Regarding its contribution to molecular signalling, Brca1 further controls ERK activation through inhibition of cyclines released in response to estrogen receptor activation as well as signals from EGFR[31].

The role of Brca1 in DNA repair was first suggested due to the discovery of hyperphosphorylation by the serine protein kinase ATM, in response to DNA damage. This is accompanied by transfer to the site of replication[32].. Not only is Brca1 involved in the repair of DNA damage, but is also involved in active repression through interaction with HDACs or the CtIP-CtBP corepressor complex[27]. Brca1 was found to interact with p53, which plays a pivotal role in cell cycle control and DNA damage repair. Here it functions as coactivator of p53, regulating the transcription of P21 and BAX[29]. In addition, Brca1 is part of the RNA polymerase II holoenzyme and is integrated through an interaction between RNA helicase A and its C – terminus[32]. Cancer arising due to Brca1 mutations belongs to the group of basal-like carcinoma, closely resembling cancers caused by down-regulated Brca1 expression, as caused by the “triple negative type” (PR-/ER-/Her2-)[26].

2.4 Max

Max, the MYC-associated factor X, is as the name suggests is best known for its regulatory interaction with MYC. MYC acts as a transcription factor and is involved in the control of cell proliferation, apoptosis and differentiation. This profound systemic involvement leads to a high risk of MYC turning into an oncogene through mutation, therefore Max possesses significant importance as interaction partner, and competitor for binding sites, if dimerised with other proteins apart from MYC. Max is a member of the family of b/HLH/Z proteins, exhibiting a leucine-zipper and helix-loop-helix motif, which enable dimerisation. This dimerisation is a prerequisite for interaction between DNA and the basic domain of MAX. Max is able to form homodimers as well as heterodimers with several different transcription factors sharing the same family and hence structural features.

2.4.1 Structure

As part of the b/HLH/Z transcription factor family, Max exhibits distinct structural regions. Each Max chain hereby forms two alpha helices interrupted by a short loop. This also positions Max in the family of helix-loop-helix proteins, and due to its structural features into the subclass IV, together with MAD and MXI. The family of HLH transcription factors is involved in a vast variety of biological processes including heart development, myogenesis, neurogenesis and sex determination, e.g. in drosophila[33]. Whereas the N-terminal helix contains the basic region(b-H1), the C-terminal helix is divided into H2 helix and the leucine zipper region(H2-Z). Crystal structures of the Max homodimer and the heterodimers of Myc-Max and Mad-Max showed that the pairs of helices arrange as a parallel left-handed four-helix bundle. Within H2, conserved hydrophobic residues enable stabilisation of the packed dimer structure[34]. The binding of these dimers induces a bend in the DNA strands. The extent is however specific to the composition of the dimer. Whereas the Max homodimer causes a flexure of 53°, the heterodimer comprised of Myc and Max leads to a bend of 80[35].

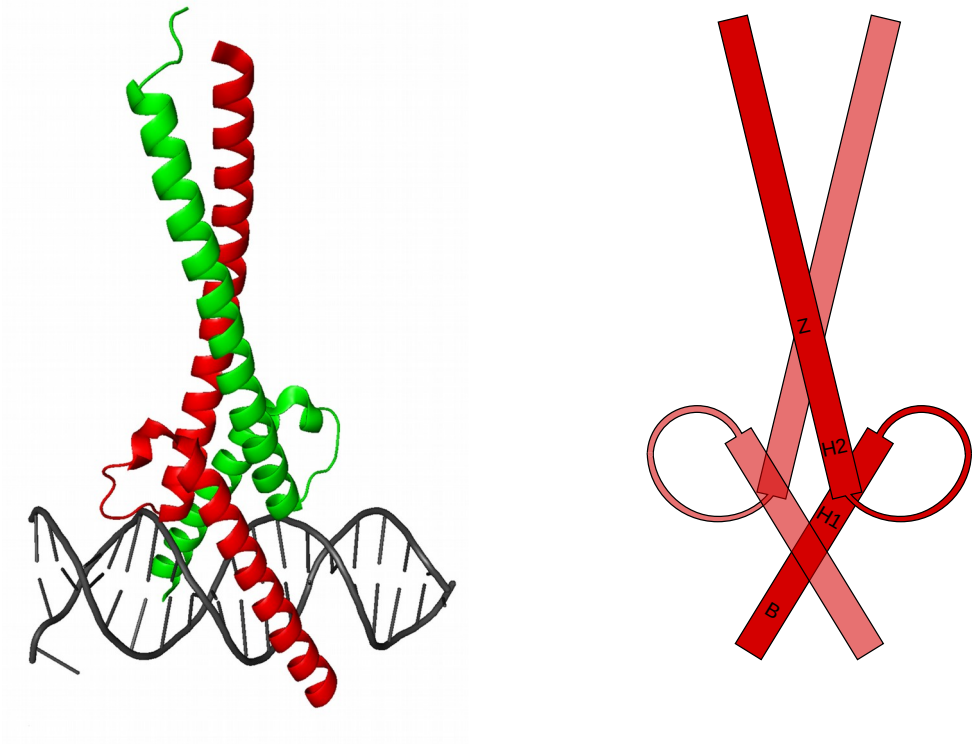

Figure 2.4.1: X-ray structure of Myc-Max dimer interacting with DNA (left)[34]. Depiction of the b/HLH/z fold as dimer, whereas the helices are named as H1 and H2 (right).

2.4.1.1 Leucine Zipper

Dimerisation of MAX is stabilised through a leucine zipper motif, which is characterised through a coiled-coil structure, stabilised through distinct hydrophobic residues. The positioning of leucines along the sequence in intervals of seven, donates an amphiphatic property to the alpha helix. Imperative for this motif is the mirrored positioning of leucines at the dimerisation partner[36].

The C-terminus of the leucine zipper possesses protein specific polar residues, which form an intermolecular interface comprised of four amino acids. In case of the Max homodimer this interface is built up by Gln 91-Asn 92-Gln 91*-Asn 92*. These two amino acids are substituted for Arg 423-Arg 424 in Myc and Glu 125-Gln 126 in Mad.

The electrostatic interaction at this site gives rise to binding partner specificity, which also prevents the formation of Myc homo-dimers. This argumentation is supported through successful dimerisation after targeted mutation of mentioned amino acids[34]. This specificity was further tested by binding studies between MAX and other b/HLH/z proteins, which showed that the presence of all structural domains is not enough to support interaction[37].

2.4.1.2 Basic Region

As well as the leucine zipper and helix loop helix motif the basic region is a common feature for Transcription factors. Contrary to the leucine zipper, the basic region is not essential for dimerisation. This N-terminal region of MAX enables sequence-specific DNA binding via the major groove. Max homodimer hereby discriminates for the class A E-box, 5'-C(1)A(2)C(3)G(4)T(5)G(6)-3'. Each half of this palindromic sequence is recognised by one participant of the dimer establishing the four contacts: His 28-Gua 3'; Glu 32-Cyt 3; Glu 32-Ade 2; Arg 36-Gua 1', whereas the prime indicates this base to be located on the opposite strand. Arg 36 specifies the central dinucleotide, distinguishing it from other class E-box binders[34]. This interaction is further stabilised through interactions between the basic region and phosphates of the DNA backbone[38].

2.4.2 Physiology

Max represents a key figure within the Myc-Max-Mad network. The ability to form heterodimers with other transcription factors, which are far more favourable than the corresponding homodimers, assigns Max the role of a center piece and mediator in a complex system which is crucial in regulating cellular behaviour. In contrast to Max, which exhibits a mainly constitutive expression, the interaction partners are tightly regulated and display a fast turnover rate[39]. This contrast is not just present on protein level, but also the mRNA is significantly short lived in comparison to MAX. Both factors are crucial to proper control and reactivity of the cellular system[40].

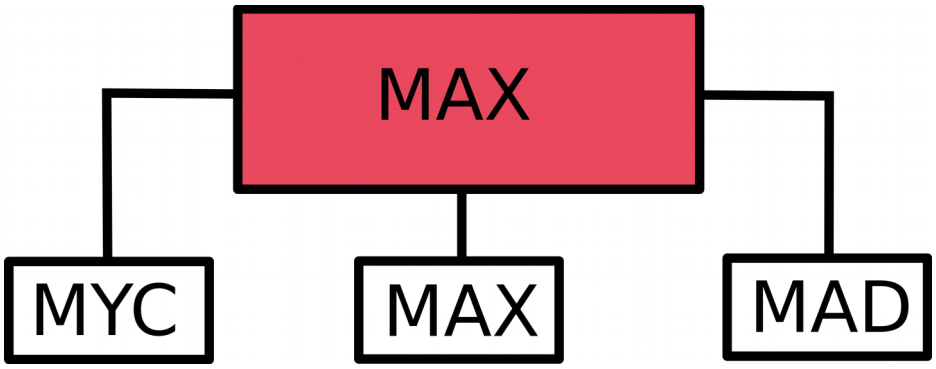

Figure 2.4.2: The figure above illustrates the major dimerisation partners of MAX

The transcription factor family MYC is well known for its role as protooncogene, often listed as trigger factor leading to cancer. The members c-Myc, N-Myc and L-Myc are all able to bind MAX. In addition to the canonical E-box sequence, the Myc-Max heterodimer is able to recognize and bind the signatures CATGTG, CATGCG, CACGCG, CACGAG and CAACGTG. A parameter also affecting binding of the recognition site is DNA methylation. Thus it was shown that Myc-Max and Mad-Max heterodimer are sensitive to methylated bases, not binding to the partially modified sequence CA^{m}CGTG[41]. In general, Myc associated to Max promotes proliferation and differentiation of cells but is also relevant to induce apoptosis. This is achieved through indirect control of the G1-S checkpoint. Due to this scope of functions it is apparent how alterations may affect tumorigenesis[39]. Further it was also shown that overexpression of Myc causes genomic destabilisation[40]. The dimer composed of Max and Myc, if bound to DNA, is able to recruit and interact with HAT complexes such as SAGA (Spt-Ada-Gcn5 acetyltransferase). The consequential histone modification ability promotes transcription[33]. To further affect transcription, Myc also possesses a transactivating domain located at the N-terminus. Opposed to this general positive transcriptory effect, Myc can function as repressor as well. This was shown through the blocking of the positive initiator Miz-1, which usually activates the transcription of INR genes[40]. Although the dimerisation partners of MAX usually exhibit a short half-life, in case of the MYC family in the area of 20-30 minutes, it was

shown that factors such as Ras are able to stabilise MYC, hence increasing the overall protein level within the cell[39].

Contrary to Myc, Max is able to form homodimers. However, these dimers do not recruit a chromatin remodelling complex and cause the loss of the cooperative PolII activation. Therefore, the Max concentration acts as an indirect inhibition factor of the E-box regulated genes. This is achieved through the decrease of free Max concentration but also through blocking the E-box recognition sites[40]. This behaviour could be proven through the introduction of misssense mutations in MAX, which lead to an increase in E-box gene expression[42].

Apart from Myc, several members of the MAD family, to be specific Mad1, Mxi1 and later Mad3 and Mad4, are able to interact with MAX[43]. The dimer formed by this interaction in general represses E-box associated expression, which suggests an antagonistic relationship between the two dimers Myc-Max and Mad-Max[39]. The MAX-Mad dimer hereby promotes cell differentiation. Towards differentiation, Myc levels decrease and increasing amounts of Mad are found to be bound to Max[44]. Apart from Myc and Mad, other transcription factors are also able to dimerise with MAX, including Mnt, Mga and Mlx[40].

2.5 NMR

NMR-spectroscopy is by now a well established and powerful tool in chemistry as well as biology. The first observations of the phenomenon of nuclear magnetic resonance in solution took place as early as 1945[45]. Just as every other form of spectroscopy, this tool observes the interaction of electromagnetic radiation and matter. To be specific NMR-spectroscopy enables to elucidate the relation between the magnetic field of nuclei, aligned in an external magnetic field, and electromagnetic waves in the frequency range of radio-transmission. NMR-spectroscopy is, among others, a common approach for the investigation of biomolecules. In comparison to e.g. X-ray crystallography or cryo-EM, NMR-spectroscopy is hereby not limited to observation of a static molecule, but is able to determine dynamic behaviour and explore the conformational space of a specimen.

2.5.1 Protein NMR – HSQC

A major drawback of NMR-spectroscopy is its limitation in its detection of atoms exhibiting an intrinsic magnetic moment and a consequential nonzero spin. Despite hydrogen possessing a nonzero spin, multidimensional measurements often necessitate the artificial introduction of isotopes into the studied molecule, as natural abundance of e.g. ^{13}C or ^{15}N is relatively low. In praxis this isotope-labeling is done through expression in an isotopically enriched medium or through the use of isotopically labelled amino acid precursors. Although this may complicate sample preparation, NMR-spectroscopy still secured its position as major tool within structural biology. This is partially due to the possibility of measurements in solution and therefore in an environment and state similar to the natural conditions of the sample, but also because of the diverse applicability and plethora of experiments. The measurement of biomolecules, such as proteins, often struggles with the resolution of separable peaks. To overcome this obstacle, structural biology employs multidimensional experiments. The basis for most multidimensional experiments,

including the experiments employed in this study, is heteronuclear single quantum coherence (HSQC).

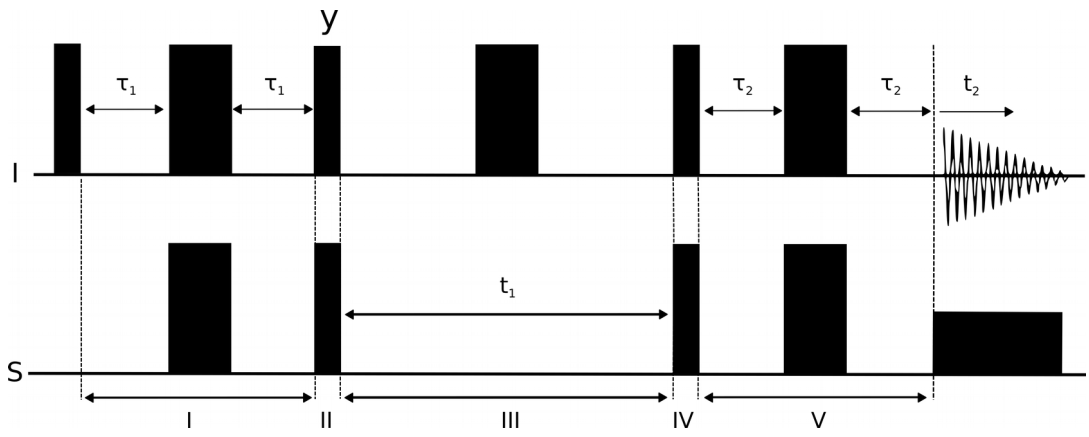

Figure 2.5.1.1: This illustration is a simplified representation of a pulse sequence describing the HSQC experiment. The vertical oriented black blocks hereby represent RF(radio-frequency)-pulse. The slim blocks stand for 90° (π/2) pulses, whereas the wide blocks signify 180° (π) pulses. The horizontal black bar at the end of the pulse sequence depicts decoupling of the heteronucleic channel. The data acquisition on spin I is symbolised through a stylised free induction decay (FID). The intervals τ_1 and τ_2 are ideally chosen to have a duration of $1/4J_{IS}$.The pulse sequence is temporally structured through the numbers I to V.

HSQC-measurements enable the acquisition of one-bond correlation spectra through the transfer of magnetisation from hydrogen nuclei to a different coupled NMR-active nuclei (e.g. ^{13}C, ^{15}N). This is accomplished through the use of other basic experiments within the pulse sequence. The most elementary of these features is the spin echo. Through the succession of a delay of time τ followed by an 180° pulse about X and another delay τ, it is possible to refocus the offset of preceding spins independent to their precession frequency[46]. This feature is depicted in *figure 2.5.1.1* within areas I and V. To achieve the transfer of polarisation to a *J* coupled nucleus, HSQC employs the method “insensitive nuclei enhanced by polarization transfer”,

commonly known as INEPT, which employs the above mentioned spin echo. This enhancement is due to the higher gyromagnetic ratio of hydrogen, which is transferred to the heteronucleus and further used for detection. The essential INEPT sequence is comprised of an initial 90° pulse only applied on the proton, followed by a spin-echo on both nuclei, which refocuses the offset while the coupling evolves, and a final 90° pulse on both nuclei. All of these pulses are about the X axis with exception of the final 90° pulse in the proton channel, which is about the Y axis. These final 90° pulses transfer the anti-phase state from the proton to the heteronucleus and therefore effectively pass on the equilibrium magnetisation of the hydrogen spin to carbon or nitrogen. To gain in-phase magnetisation, a second spin-echo is commonly applied after the magnetisation-transfer-pulse[45]. As shown in *figure 2.5.1.1* the HSQC sequence is initialised by an INEPT-like phase ranging from I-II. During period III only the heteronuclear magnetisation can evolve, as the centred 180° pulse refocuses all evolution due to coupling. Therefore, the magnetisation is frequency-labelled during this duration. The segments IV..V, which are frequently named inverse INEPT, transfer the anti-phase back to the I spin through two 90° pulses about X. Although it would be possible to detect the FID immediately after IV, this would cause the detection of an anti-phase term. This can be avoided through another spin-echo element in combination with S-spin decoupling[4647]. This representation does not fully depict the experiments carried out during this study, it nevertheless shows the basic mechanics of HSQC. A factor which has been neglected in this analysis is coherence selection. In praxis, a HSQC experiment is usually augmented through phase cycling or pulse field gradients[48].

2.5.2 Relaxation

As this study focuses on the evaluation of forces influencing the relaxation of nuclear spins, it is crucial to briefly discuss the underlying concept. As excitation of a spin by a pulse of appropriate frequency poses perturbation of the equilibrium state, the spin population, as also valid for any other physical system, will aim to again reach the initial state of equilibrium. This behaviour of spins to eventually realign with the

external magnetic field is termed relaxation. NMR-spectroscopy differentiates between the transverse (x,y) and longitudinal (z) partition to magnetisation, as longitudinal magnetisation is aligned to the static magnetic field B_0. This distinction is crucial, due to only magnetisation in the xy-plane being observable. To describe the time-dependent change in magnetisation we use the same methodology, distinguishing between longitudinal (spin-lattice) and transverse (spin-spin) relaxation[45].

To quantify longitudinal relaxation we employ the measure of T_1, which is the time constant affecting the transition from M_z back towards the equilibrium magnetisation M_{eq} in a fashion of: $M_z(t) = M_{z,equ}\ (1\text{-}e^{-t/T_1})$. Through this time constant we can further define the longitudinal relaxation rate R_1, which is defined through $1/T_1$. This value is of direct importance for the planning of experiments, as the relaxation delay has to be set to at least $5*T_1$ to ensure the sufficient recovery of the equilibrium magnetisation. In a similar fashion the values T_2 and R_2 have been established, which are defined through: $M_{xy}(t) = M_{xy}(0)\ e^{-t/T_2}$. The time constant T_2 affects resolution of peaks, as it describes the rate of decay of the FID and therefore the time frame of detectability and consequentially the peak width. Due to T_1 describing the re-establishment of the equilibrium state, whereas T_2 only gives measure to the decline of the transverse magnetisation components, in the majority of cases T_1 is longer than T_2.

Relaxation necessitates ways to transfer quanta of excitation energy, the most prominent being fluctuating fields caused through molecular motions, termed dipole-dipole relaxation. Hereby, a fluctuating field at larmor frequency (ω_0) of the spin is needed to allow relaxation. This mechanism affects both longitudinal as well as transverse relaxation, through translational and rotational motion. To quantify the range and distribution of possible reorientation frequencies we employ the spectral density function *J(v),*which is dependent on the size and shape of the molecule. Hereby, size indirectly correlates to the range of possible reorientation frequencies. To simplify calculations, we use the correlation time τ_c to describe the function, whereas this measure is defined as the average tumbling time of a molecule. Due to the change in frequency distribution of the fluctuating field, an increase in τ_c, causes a decrease in T_1, but only up to the minimum at $\tau_c = \omega_0^{-1}$. There upon T_1 increases, as the probability of motion at ω_0 decreases due to the spread of frequencies. In case of

transverse relaxation we observe a continuous decrease with an increase in correlation time. This difference is due to the fanning out of magnetisation in the xy-plane, meaning a phase-inhomogeneity of the spins caused by slow tumbling. This effect is also the reason for the size limitation in NMR-spectroscopy[45 49]. Dipole-dipole interaction is however not the sole influence on relaxation. It is further affected by effects such as CSA (chemical shift anisotropy).

To observe and measure longitudinal relaxation, one can employ the inversion recovery experiment. This simple pulse sequence (180° - τ - 90° - FID) allows to calculate T_1 through a series of measurements differing in the delay τ and fitting an exponential function to the acquired peak heights. Transverse relaxation can be measured through an experiment, which uses the already mentioned spin echo (90° - τ - 180° - τ - FID). In theory, already one measurement suffices to define T_2, as the amplitude of the fourier transformed echo is proportional to $e^{-2\tau/T_2}$. This however can be avoided through repeated measurements differing in the delay.

2.5.3 Paramagnetism

Paramagnetism is a physical property of atoms possessing unpaired electrons. It hereby represents a specific form of magnetism, as opposed to diamagnetism. This differentiation can be illustrated through the interaction with an externally applied magnetic field. Whereas diamagnetic species are repelled, paramagnetic atoms or molecules are weakly attracted by the magnetic field. Paramagnetic species are of high interest regarding NMR-spectroscopy, as the distinct effect they have on nuclear spins can be exploited to gain various types of data. These effects can be roughly categorised into paramagnetic relaxation enhancement (PRE), pseudo-contact shifts (PCSs) and residual dipolar couplings (RDCs), whereas the latter two can only be observed if a paramagnetic centre with anisotropic *g*-tensor is present. All of these three observables can be exploited to gain long-range structural information.

2.5.3.1 Paramagnetic probes

As every molecule possessing an unpaired electron consequentially contains a paramagnetic centre, the occurrence of paramagnetic effects is quite common. Specifically metal complexing proteins may experience natural paramagnetic effects. Hereby, metals such as Mn^{2+}, Fe^{2+}, Co^{2+},Ni^{2+} or Cu^{2+} act as paramagnetic centre. Although studies have been performed employing these natural paramagnets for structural analysis[50] to translate paramagnetic NMR-spectroscopy into a method of general applicability, it was essential to design paramagnetic compounds, which can specifically interact with biomolecules. We distinguish between two major groups of paramagnetic probes, the one being nitroxide stable radicals, the other being metal chelators, binding metals with high affinity. To introduce paramagnetic centres into proteins, these probes are usually conjugated to cysteine residues by either a disulfide linkage, as is the case regarding the used MTSL, or via a C – S bond, relevant for iodo- or bromoacetamide-derivatives[49]. Apart from binding specificity and stability, another factor which has to be considered is the necessity for a diamagnetic reference sample. In case of metal chelators this can be accomplished via substitution with a diamagnetic metal ion or as relevant for this study by quenching of MTSL through the addition of ascorbic acid. Not only do these paramagnetic entities differ in their strength, but as mentioned above, can we also differentiate between anisotropic and isotropic centres. Anisotropy of a paramagnetic centre is given through its χ-tensor or the closely related *g*-tensor. Hereby, variance in the magnetic moment of the paramagnetic centre due to a change in orientation of the molecule within the external magnetic field implies anisotropy and can be quantified through the measure $\Delta\chi$. The paramagnetic probe MTSL, which was used in this study, belongs to the group of isotropic paramagnets[51].

2.5.3.2 Paramagnetic Relaxation Enhancement

As mentioned above, there are three distinct effects caused by the introduction of paramagnetic centres. As this study exclusively employed the isotropic probe MTSL, the sole effect observable is paramagnetic relaxation enhancement and we shall

therefore focus on this effect. PRE accelerates both longitudinal as well as transverse relaxation. Therefore, a change in the respective relaxation rates R_1 and R_2 of nuclei within range of effect is apparent. This effect can be directly observed in a decrease in peak amplitude and increase in peak width. To quantify PRE, the difference between the diamagnetic and paramagnetic relaxation rates is measured, which in case of an amide proton is symbolised through $^1H^N–\Gamma_1$ or $^1H^N–\Gamma_2$ respectively. In case of nitroxide-labels, this effect on relaxation is mainly due to dipole-dipole interaction between the unpaired electron X and the proton H^N and can be described through the Solomon-Bloembergen equations. Although PRE is dependent on r^{-6}, whereby r stands for the distance between the electron and affected spin, it is possible to observe the effect up to a distance of 20-30 Å. These long range interactions are possible due to the strong magnetic field of the unpaired electron, which consequentially also causes residues within close proximity to relax too fast to observe[45].

$$(1)\quad \Gamma_2=\frac{1}{15}\left(\frac{\mu_0}{4\pi}\right)^2 \gamma_I^2 g^2 \mu_B^2 S(S+1)\{4J_{SB}(0)+3J_{SB}(\omega_I)\} \qquad (2)\quad J_{SB}(\omega)=r^{-6}\frac{\tau_c}{1+(\omega\tau_c)^2}$$

Formula 2.5.3.2.1/2: (1) Formula defining the parameters PRE is dependent on. Hereby, g is the g-factor, $\omega_I/2\pi$ the proton Larmor frequency, γ_I the proton gyromagnetic ratio, μ the magnetic moments. (2) Definition of the generalized spectral density function for the reduced correlation function $J_{SB}(\omega)$. r stands for the distance between the paramagnetic electron and the proton, τ_c the correlation time defined as $(\tau_r^{-1}{}_{\text{(rotational correlation time)}} + \tau_s^{-1}{}_{\text{(effective electron relaxation time)}})^{-1}$.

Apart of the above described effect also the Curie-spin relaxation (CSR) may emerge due to dipole-dipole. CSR is caused through the interaction between the time-averaged magnetisation of the unpaired electron and a nuclear spin. In case of nitroxide labels this effect can be neglected, as $\tau_s \approx \tau_r$, which causes a cancellation of CSR[52]. In contrast to PCSs, PRE does not induce a shift of the affected residue

peaks. Therefore, a new assignment, as is often necessarily needed.

$$\Gamma_{2,CSR}=\frac{1}{5}\left(\frac{\mu_0}{4\pi}\right)^2\frac{\omega_H^2 g^4 \mu_B^4 S^2(S+1)^2}{(3kT)^2 r^6}\left(4\tau_r+\frac{3\tau_r}{1+(\omega_H\tau_r)^2}-4\tau_c-\frac{3\tau_c}{1+(\omega_H\tau_c)^2}\right)$$

Formula 2.5.3.2.3: Change of transverse relaxation rate due to curie spin relaxation. The meaning of the used symbols corresponds to the use in Formula 2.5.3.2.1/2. The Boltzmann constant is signified by k, the temperature in Kelvin by T.

2.5.3.3 Paramagnetic Relaxation Interference

As was first employed studying correlated structural fluctuations in intrinsically disordered proteins, it is possible to gain additional structural information beyond a static model through the augmentation of the conventional paramagnetic NMR-techniques. By introducing two paramagnetic centres, the two dipoles X(1)-$^1H^N$ and X(2)-$^1H^N$ emerge, whereas X signifies an unpaired electron. Cross-correlated relaxation emerges due to interference of different relaxation mechanisms. This is valid and was already observed regarding the interaction of dipoles and or chemical shift anisotropy. A novel source of cross-correlation hereby represents the interference of X(1)-$^1H^N$ and X(2)-$^1H^N$. To quantify this effect we first have to measure $^1H^N$-Γ_2[X(1)], the change of the transverse relaxation rate caused by the unpaired electron X(1), and $^1H^N$-Γ_2[X(2)], followed by a measurement of a molecule exhibiting both paramagnetic centres $^1H^N$-Γ_2[X(1)+X(2)]. The PRI can be calculated through: Δ^1H^N– Γ_2 = $^1H^N$-Γ_2[X(1)+X(2)] - {$^1H^N$-Γ_2[X(1)] + $^1H^N$-Γ_2[X(2)]}. The magnitude of Δ^1H^N–Γ_2 as a function of residue can hereby elucidate structurally correlated regions within the protein[53].

$$\frac{\Delta^1 H^N - \Gamma_2}{\sqrt{{}^1H^N - \Gamma_2[X(1)] * {}^1H^N - \Gamma_2[X(2)]}} \approx \frac{3\cos^2\varphi - 1}{2}$$

Formula 2.5.3.3.1: This formula shows the relationship between the normalised interference term and φ, defined as the angle enclosed by the two vectors X(1)-$^1H^N$ and X(2)-$^1H^N$.

As the interference term $\Delta^1H^N - \Gamma_2$ is dependent on the value *φ*, the angle between the vectors originating from the amide proton to the respective unpaired electrons X(1) and X(2), we can quantify this relationship according to *formula 2.5.3.3.1.* This relation implies a range of definition for the normalised interference term -0.5..+1. As a consequence we note the possibility for positive as well as negative interference, but also stringent requirements for the acquired R_2 rates. Applying this method to a globular and static biological system may enable to deduce angles of a three-dimensional structure from relaxation time measurements.

3 Aim of the work

Our goal has been to apply and evaluate the recently introduced PRI experiment[50] on globular proteins. The main focus of this work is to demonstrate the presence of negative as well as positive interference between paramagnetic effects and further elucidate and validate the application of PRIs as a tool for structural analysis of proteins. As explained in the paragraph assigned to the topic of PRE and PRIs, the interference term is dependent on the angle enclosed by two vectors originating from spin to respective labels. This might be an original approach for structural characterisation. A major obstacle, regarding the implementation of this method, is the requirement of exact data. Small changes in relaxation rates, or misinterpretation caused through mobility of the label itself, can already lead to drastic changes in the derived angular results. The experimental procedure consists of T_2 relaxation rate measurements of MTSL-labelled proteins. At least the measurements of two single mutants and their associated double mutant are needed to gather sufficient data to derive angular information. To test the applicability of CCR (Cross Correlation Relaxation), different biological systems were utilised. GB1 was chosen because of its well known three-dimensional structure and simple expression and purification. Owing to the folded nature and defined architecture, it was possible to calculate all theoretical angles within the protein. These results could be consulted for a direct comparison with the angles determined from the experimentally obtained data. For this purpose GB1 mutants were devised carrying cysteine residues on specific positions within the proteins. These residues were chosen to obtain a maximum overlap of the two effective PREs as well as a variety in derivable angles to better demonstrate the correlation of experimental data to the structurally calculated. Therefore Ala 47 and Lys 9 were mutated to Cysteines, effectively constructing two single mutants and the corresponding double mutant. These mutations allow the attachment of paramagnetic labels through disulfide bonds, which makes the measurement of the change in transverse Relaxation between the active and quenched labelled protein possible. As another example of a folded protein with characterised structure, NGAL, also referred to as LCN2, was chosen. Though the physiological and medicinal importance of this protein shall not be neglected, in this

study the sole focus lies on the retrieval of structural information and the comparison to calculated data. NGAL hereby was chosen due to its significantly larger size, which reduces the influence of the label flexibility but also increases the correlation time. The latter also increases the observable paramagnetic effect. As in the case of GB1, cysteine mutations were designed to measure transverse relaxation rates. Therefore Lys 62 and Ser 87 were exchanged to cysteines. For the purpose of these measurements, additional mutants were created (Ala 40, Glu 147). The expression however was abandoned, due to issues regarding the purification of the NGAL protein. To test the method of CCR on a fundamentally different system, a small, mainly unstructured peptide originating from the sequence of Brca1, was chosen. In case of this protein, not the intramolecular PREs were measured, but a binding partner was chosen to carry the paramagnetic labels. For this purpose, isotopically unlabelled MAX mutants were employed. Prior studies have already shown significant binding of this Brca1 peptide to MAX dimers. To measure the different relaxation profiles, homodimers of Arginin – 5 and Glycin – 35 as well as a heterodimer of both mutants bound to the Brca1 peptide.

4 Materials

4.1 Cellstrains

E.coli DH5alpha
E.coli BL21
E.coli BL21(DE3)pLysS
E.coli Rosetta pLysS

4.2 Buffers and media

4.2.1 NGAL

10x Tris pH 7.4

Tris 200 mM

NaCl 500 mM

dissolve in 1 L ddH2O, adjust pH to 7.4, filtrate and degas

before use dilute 1:10 with ddH_2O

10x PBS

NaCl 1.37 M

KCl 27 mM

Na_2HPO_4 100 mM

KH_2PO_4 18 mM

dissolve in 1 L of ddH2O, adjust pH to 7.4, filtrate and degas

before use dilute 1:10 with ddH2O

HS PBS

NaCl 1.5 M
Imidazole 20 mM
dissolve in desired amount of 1x PBS, adjust pH to 7.4, filtrate and degas

High Imidazole

Imidazole 500 mM
dissolve in desired amount of 1x PBS, adjust pH to 7.4, filtrate and degas

TEV-Buffer

DTT 1 mM
EDTA 0.5 mM
add to 100 mL of Tris-buffer and adjust pH to 7.4

Guanidiniumchloride-buffer

1xTris + 6M GndHCl pH=7.4

4.2.2 Brca1

Lysis buffer

Tris 25mM
NaCl 300mM

pH 7.4

β-ME (Mercaptoethanol) 1mM (add only before use)

low-imidazole buffer

---

lysis buffer + 20mM imidazole (pH 8)

high-imidazole buffer

---

lysis buffer + 400mM imidazole (pH 8)

MES-buffer

---

20 mM MES

50 mM NaCl

pH 5.5

4.2.3 MAX

Standard-Buffer

20mM $NaPO_4$ pH=7.0

100mM NaCl (g/l), 1mM EDTA

High Salt

Standard with 1M NaCl.

4.2.4 GB1

PBS buffer and derivatives were produced according to NGAL protocol.

GB1 measurement buffer

NaCl 22 mM
KCl 2.7 mM
Na2HPO4 10 mM
KH2PO4 1.8 mM
pH = 7.4

4.2 Media

LB

LB 20 g/L
dissolve in 1 L ddH_2O and autoclave

M9

Na2HPO4 6.0 g/L
KH2PO4 3.0 g/L
NaCl 0.5 g/L
$^{14}N/^{15}N$-NH4Cl 1.0 g/L
dissolve in 970 mL ddH2O and autoclave

add after autoclaving and before expression:

20 % Glucose (12C or 13C) 20 mL

Trace Elements 10 mL

Antibiotic 1 mL (each)

1M $MgSO_4$ 2 mL

$CaCl_2$ 300 µL

Reaction buffer (MTSL)

100mM NaPi

pH=8

1mM EDTA

20% EtOH

0.5 mM $NiCl_2$

1 mM EDTA

6 M Guanidinium HCl

4.3 Primers GB1

GB1 mut Fwd
5’–CTTCACGGTAACCGAATAATCTGGCAGTGGTTCTG–3’

GB1 mut Rev
5’–CAGAACCACTGCCAGATTATTCGGTTACCGTGAAG–3’

GB1 amp NcoI Fwd
5’–ACTGATCCATGGCGCAGTACAAGCTTATC–3’

GB1 amp XhoI Rev
5’–AGCAGTCTCGAGTTATTCGGTTACCGTGAAG–3’

GB1 K9C Fwd
5’–GTACAAGCTGAACGGTTGCACCCTGAAAGGTGAAAC–3’

5’–GCTTATCCTGAACGGTTGCACCCTGAAAGGTGAAAC–3’

GB1 K9C Rev
5’–GTTTCACCTTTCAGGGTGCAACCGTTCAGCTTGTAC–3’

5’–GTTTCACCTTTCAGGGTGCAACCGTTCAGGATAAGC–3’

GB1 A47C Fwd
5’ – GGACCTACGACGACTGTACCAAAACCTTCACGG – 3’

GB1 A47C Rev
5’–CCGTGAAGGTTTTGGTACAGTCGTCGTAGGTCC–3’

4.4 Equipment

Äkta pure
HiLoad 16/60 Superdex 75 prepgrade
HiTrap Chelating HP
Varian Inova 500
Varian Direct-Drive 600
Mastercycler Gradient
NanoDrop
Amicon Ultra-15 Centrifugla Filter Units
PD-10 Desalting Columns

4.5 Software

NMRPipe Version 9.1 Rev 2017.241.15.09
Sparky Version 3.113
NMRPipe Conversion Utility Version
LibreOffice 5.2.5
VNMRJ
Rstudio 0.98.1074
ApE – A plasmid Editor v2.0.49
Unicorn 7.0

5 Methods

5.1 Transformation

100µL of competent bacterial cells are thawed on ice and plasmid is added (~100 ng). The cells are then incubated for 15 minutes on ice and afterwards held at 42°C for 90 seconds to permeabilize the cells through heat-shock. After another incubation time of 2 minutes on ice, 500 µL sterile LB medium is added. To stimulate growth in the culture, the sample is placed in a shaker at 37°C for 1 hour. 100µL of sample are now plated on an agar plate containing the antibiotics specific for the plasmid and cell strain. The plates are kept overnight at 37°C until colonies have formed. A negative control of non-transformed bacterial cells plated on an agar plate sharing the same antibiotics can increase the certainty of proper execution. This method was implemented in all biological systems listed, only differing in bacterial strain and respective antibiotics.

5.2 NGAL

5'_MKHHHHHHPMSDYDIPTTENLYFQGAMGQDSTSDLIPAPPLSKVPLQQNFQDN
QFQGKWYVVGLAGNAILREDKDPQKMYATIYELKEDKSYNVTSVLFRKKKCDYWIR
TFVPGSQPGEFTLGNIKSYPGLTSYLVRVVSTNYNQHAMVFFKKVSQNREYFKITLY
GRTKELTSELKENFIRFSKSLGLPENHVFPVPIDQCIDG_3'

MW: 23,85 kDa
Theoretical pI: 8.70

Table 5.1: The sequence shown above represents the sequence of the expressed NGAL construct. Hereby, the chosen sites of mutation (A40, K62, S87, E147) are indicated by red letters. Also shown, is the introduced His_6- tag, located at the 5' end of the transcript. The TEV recognition pattern is represented in green.

5.2.1 Expression

E.coli/BL21-DE3-pLysS cells are transformed with the plasmid pETM-11 containing the sequence coding for the desired NGAL mutants. Prior to expression a preculture is grown. 10 mL LB per Liter of culture are inoculated with a colony of transformed cells. To avoid contamination, we add the antibiotics chloramphenicol and kanamycin (1 mM each), which were as well used in the preparation of the agar plates needed for transformation. The preculture is either grown overnight at 37°C or over the weekend at room temperature, placed in the shaker. For the expression of NGAL, first, 10 mL of preculture are added to 1 L LB, again containing chloramphenicol and kanamycin. To enhance the yield of NGAL, we use 4 L of this culture and grow shaking at 37°C until an OD_{600} of 0.8 is reached. To rid the cells of LB, the culture is spun down for 15 minutes at 4500 rpm. The pellet is resuspended in M9-medium and further grown for 45 minutes at 37°C. The culture is now induced by adding IPTG in a concentration of 0.8 mM and expressed at 30°C overnight.

5.2.2 Purification

Following expression the cells were harvested by spinning them down (6000 rpm/4°C/15 minutes). The pellet is resuspended in PBS, containing 1 mM β-mercaptoethanol, to reach a final volume of 30-40 mL. Lysis of the cells is performed through sonication. Therefore the samples are put on ice and sonicated three times at 50% amplitude for three minutes. Between each sonication step we wait approximately 5 minutes to avoid overheating of the sample. The cell lysate is centrifuged again, this time at 18.000 rpm/4°C/25 minutes to rid the sample of insoluble cell components. To remove remaining cell debris from the sample, it is filtered through a 0.45 µM filter. Subsequently, the sample is applied to an already equilibrated Hitrap Chelating HP (5 mL) which has been loaded with 0.5 mM $NiCl_2$. After loading the sample, the column is washed with 5 CVs (column volumes) of PBS

– 5 CVs High Salt PBS – 5 CVs PBS to ensure specificity of binding. Finally, the protein is eluted with a gradient of PBS to High Imidazole PBS (45 min, 100%). After determining the protein containing fractions through SDS Page, they are pooled and dialyzed in 500 mL of Tris pH 7.4. The dialyzed protein is washed with 30 mL of TEV-buffer using a centrifugal filter tube (4000 rpm 4°C) and concentrated to approximately 500 µL. To cleave the His-tag off the protein, the protease TEV is added in a proportion of 1 mg TEV : 50 mg protein. The cleavage is executed overnight, rotating the sample tube at 4°C. Next, NGAL is separated using a GE-HiLoad Superdex 75 16/60. This step is essential to rid the sample of the protease as well as the cleaved tag. After evaluation of the fractions via SDS-Page, the fractions of interest are pooled and concentrated to 500 µL. As NGAL is bound to the coexpressed bacterial ligand enterobactin, we have to unfold the protein. Herefor, 0.56 g of guanidinium HCl are added to the sample and it is placed at 70°C for 20 minutes. By testing different techniques to separate NGAL from its ligand enterobactin, the conclusion was made that the most efficient way is to run the sample through a centrifugal filter tube using Gnd-HCl-Tris to further wash the sample. Initially we used PD-10 Desalting Columns to separate protein from ligand. This however removed enterobactin insufficiently. As a folded protein is needed for the planned experiments, we refold the protein in dialysis. Therefore the concentrated sample is diluted with and dialysed in Tris overnight. The dialysis buffer is exchanged twice, each time dialysing the sample overnight. The refolded protein is centrifuged at 18.000 and 4°C for 25 min. The desired protein, located in the supernatant, is concentrated and prepared for measurements. As the precipitate also contains NGAL, it can be denatured and refolded to enhance yield.

5.2.3 Enterobactin

As mentioned in *section 5.2.2* the expression of NGAL in E.coli leads to coexpressed enterobactin. This ligand poses a problem as it is bound to ferric iron, a paramagnetic species. Although we have developed a strategy to rid the sample of enterobactin, another issue arises. The binding site of enterobactin is able to interact with the

paramagnetic label MTSL. Hence it is necessary to block the binding site to avoid protein aggregation. Therefore, we express the ligand enterobactin itself, but substitute the paramagnetic iron for diamagnetic gallium. To express enterobactin we use the cell strain E.coli W3110 Dfur::cat, which over-expresses the ligand. A preculture is prepared as usual and used to inoculate M9 medium. The culture is then grown at 37°C for 24 hours, followed by a centrifugation step (6000 rpm, 15 min, 4°C). The desired ligand is found in the supernatant. First the pH is adjusted with hydrochloric acid to inhibit the complexing interaction of enterobactin with iron. This reaction can be monitored through a significant change in colour, as the supernatant changes from a reddish hue to a clear solution. The ligand is extracted with ethyl acetate and the extract washed with ddH_20, citric acid buffer and again ddH_2O using a separatory funnel. The extract is rid of its liquid via rotavapor and again dissolved in a few mL of ethyl acetate. The concentrated sample is mixed with 40 mL of n-hexane and left on ice. This mixture is centrifuged
(4000 rpm, 4°C, 15 min) and the precipitate dried for 30 minutes at room temperature. The sample is subsequently dissolved in methanol and stored at -20°C. Ligand concentration can be determined through the addition of Fe^{3+} and KOH and absorption measurement at A_{495}. As only enterobactin bound to a metal ion interacts with NGAL, we load enterobactin with a 2-fold excess of gallium before adding it to the protein samples.

5.3 GB1

5.3.1 Cloning

To construct a vector expressing the desired mutants of GB1, we used pET-GB1 as template. Through site directed mutagenesis a stop codon is introduced at the end of the GB1 transcript. The strategy chosen is a restriction digest cloning. Therefore the restriction sites NcoI and XhoI are introduced into the sequence of pET-GB1. The resulting amplicon and the vector pET-M11 are both digested with the respective restriction enzymes. Following the digest, the DNA is purified using the QIAquick PCR Purification Kit and insert and vector are ligated. After every PCR step the product is digested with DPN1 to rid the sample of the paternal strain, which does not contain the wished mutation, as well as purified using the QIAquick PCR Purification Kit. To ensure proper amplification agarosis gels are used to compare concentration and size of the product. Further, these plasmids are used to transform E.coli DH5alpha. Colonies of these plates are used to grow o/n cultures which are spun down and lysed to extract DNA. The harvest of bacterial plasmid was performed using the QIAprep Spin Miniprep Kit. After successful introduction of the GB1 gene into the vector pET-M11 we performed site directed mutation on the plasmid. Through primers encoding for point mutations, we mutate the site Lys 9 and Ala 47 into cysteines.

5'_MKHHHHHHPMSDYDIPTTENLYFQGAMAQYKLILNG**K**TLKGETTTEAVDAATAEK
VFKQYANDNGVDGEWTYDD**A**TKTFTVTE_3'

MW: 9.42 kDa
Theoretical pI: 5.32

Table 5.2: Shown above is the protein sequence of GB1. Located at the 5' end is a His_6- tag symbolised by blue letters. The TEV recognition site is

shown in green and the introduced mutations are indicated by red letters (Lys 9, Ala 47).

5.3.2 Expression

For expression of GB1, E.coli BL21 were transformed with pET-M11-GB1-His. The preculture was spun down and the pellet directly transferred to minimal media M9. The culture is grown at 37°C until an OD_{600} of 0.8 is reached. The culture is induced with 0.8 mM IPTG and expresses o/n at 28°C.

5.3.3 Purification

The harvest and an affinity chromatography were carried out in the same method shown in the purification of NGAL. To cleave the histidine-tag off the protein, GB1 is dialysed overnight while TEV (1:50 w/w) is added to the dialysis tube. To further purify the protein and separate the tag as well as the protease, a GE-HiLoad Superdex 75 16/60 size exclusion is used.

5.4 BRCA1

5’_MKHHHHHHPMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLE
EKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYN
GKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTW
PLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEA
AFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASP
NKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGE
IMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTPGSLEVLFQGPCTLNSSIQ
KVNEWFSRSDELLGSDDSHDGESESNAKVADVLDVLNEVDEYSGSSEKID_3’

MW: 6364.6
Theoretical pI: 3.96

Table 5.3: Shown above is the protein sequence of the chosen Brca1 peptide. Located at the 5’ end is a His_6-tag indiciated by blue letters, followed by a MBP tag (magenta). The recognition pattern of the used 3C protease is shown in green.

5.4.1 Expression

Rosetta pLysS competent cells are transformed with H6_MBP_3C_Brca1_pep01_pETM-44. After preparing a preculture, four litres of LB with added antibiotics, are inoculated and grown at 37°C until an OD_{600} of 0.5-0.7 is reached. The culture is centrifuged at 4000 rpm and 4°C for 15' and the pellet is resuspended in M9 and transferred to one litre of M9. After 45 minutes at 37°C, the culture is induced at a concentration of 0.4 mM IPTG and the temperature of the shaker is set to 28°C for around 2 hours, followed by 37°C o/n.

5.4.2 Purification

After centrifugation (4000 rpm, 15 min, 4°C) the pellet is resuspended in lysis Buffer. Sonication, centrifugation and filtration were conducted as already described in the section covering NGAL. The sample is then loaded on a His-Trap equilibriated with Tris, washed with a high salt Tris, and eluted with a gradient from low to high imidazole-buffer. This is followed by size-exclusion chromatography, using a GE-HiLoad Superdex 75 16/60 equilibrated and run in Tris-buffer. The pooled fractions are concentrated to 0.5-1 mL and 3C protease is added in a proportion of 1:50 (w/w). The sample is incubated o/n at 4°C. First, the cleaved protein is run through a centrifugal filter tube of 10 kDa to separate the cleaved H6-MBP tail. The Brca1 peptide is located in the flow-through of the centrifugal filter tube. Using a 3 kDa filter tube, we wash and transfer the sample into MES-buffer.

5.5 MAX

5’_MADKRAHHNALERKRRDHIKDSFHSLRDSVPSLQGEKASRAQILDKATEYIQ
YMRRKNHTHQQDIDDLKRQNALLEQQVRALEKARSSAQLQA_3’

MW: 10927.2
Theoretical pI: 9.98

Table 5.4: Shown above is the protein sequence of the expressed MAX. Hereby indicate red letters the sites of mutation (Arg 5,Gly 35)

5.5.1 Expression

BL21-DE3-pLysS E.coli are transformed with pET-3d Max plasmid. The preculture and LB culture were executed in the same fashion as Brca1. The cells, resuspended in M9, are further grown for
25 minutes at 30°C and induced at a concentration of 0.4 mM IPTG. MAX is expressed overnight at 30°C.

5.5.2 Purification

The protein harvest was carried out the same way as Brca1, only differing regarding the used lysis buffer, for this purpose Standard-Buffer was used. After filtration, Max is further purified through two ammonium sulphate precipitation steps. We first bring the solution to a concentration of 55% v/v and further to 77% v/v. At each of these steps the solution is kept on ice and stirred for 30 minutes. The solution is further centrifuged for 30 minutes, 4°C, 4000 rpm. Whereas the first precipitation step leads to the desired protein being in the supernatant, the second addition of ammonium sulphate leads to precipitation of MAX. To rid the sample of the used salts, we dialyse the, in Standard-Buffer resuspended, pellet overnight. To increase purity, the protein

is run on a Resource S and eluted through a High Salt gradient. As final purification step the protein is run on a GE-HiLoad Superdex 75 16/60 in Standard-Buffer.

5.6 MTSL-Tagging

First, a protein solution of up to 2.5 mL is prepared. A 2.5 fold excess of DTT is added and the sample is incubated for 15 minutes at room temperature. To remove the DTT, which would compete with the tagging process, we use a PD-10 desalting column. Before the sample is loaded on the column, it is equilibrated with 30 mL of reaction buffer. The initial flow through is discarded as it is within the dead volume. To elute the protein, 3.4 mL of reaction buffer are applied to the column and the flow through collected. To establish that all the cysteines are freely accessible, the free thiol concentration is measured using DTNB as a reporter. The accessible thiol concentration can be measured through an increase in OD_{412} and is directly observable through yellow colouration. As the extinction coefficient is known, we can calculate the concentration of free cysteines. A 5-times excess of MTSL is added to the sample, which is kept shaking at 37°C for three hours. To determine the extent of protein being labelled, the free thiol concentration is measured again after incubation.

6 Results

6.1 GB1

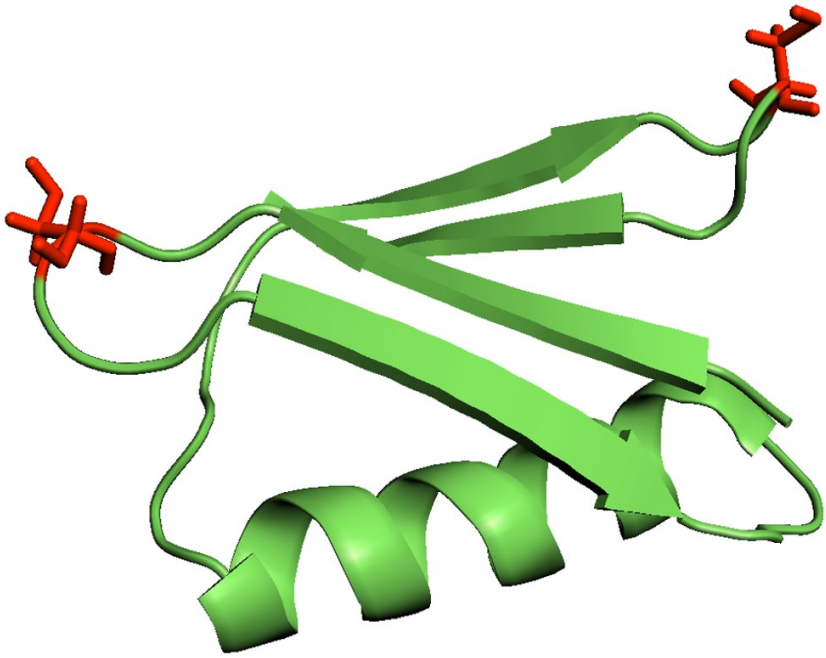

Figure 6.1.1: Three dimensional model depicting the structure of GB1 and the introduced mutations. This representation is based on a high resolution NMR structure[54].

In this study, the small protein GB1 is ^{15}N-labelled and expressed to observe and evaluate the margins of paramagnetic relaxation interference. The selected positions were chosen not to interfere with structural motifs and were therefore positioned within loop regions. Apart from structural integrity, another factor which has to be considered is the area of paramagnetic effect. To allow interference of PREs, the system obviously necessitates spins affected by both paramagnetic sites, originating from the labels. Due to the distance dependence of the relaxatory effect, there is not only an upper limit of observability, but also residues too close to the label are not evaluable. The disappearance of peaks is due to a relaxation rate too high to be still

observable. Another factor which has to be considered is the distance between the two labels. Not only would the range of possible angles be limited if the labels are too close to each other, but also would this further limit observability of proximate nuclear spins. To proof the proposed hypothesis, there has to be a distribution of positive and negative interference, thus a variety of angles. Through a given three-dimensional structure one can calculate the diversity of angles, specific to the chosen label positions. As already stated in the section dedicated to PRI, interference of paramagnetic enhancement is dependent on the angle of X(1)-$^{1}H^{N}$-X(2), as described in *formula 2.5.3.3.1*. Therefore, zero crossing occurs at $(2\pi n \pm \cos^{-1}(-1/3))/2$, in other words positive interference is observed in the range of 0°-54.74° and 125.26°-180°, whereas negative interference only arises in the range of 54.74°-125.26°.

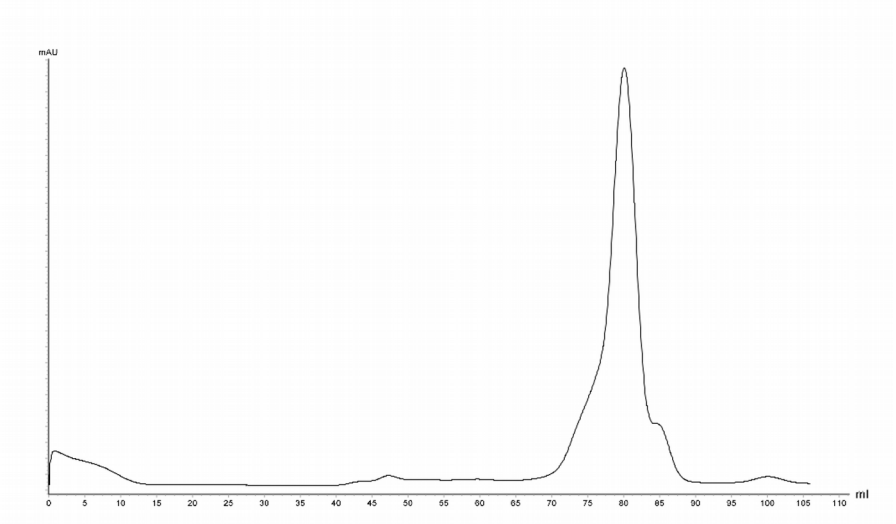

Figure 6.1.2: Chromatogram of a size exclusion chromatography performed on GB1. The black line hereby represents the relative absorbance at wavelength 280 nm (A_{280}). The most prominent peak represents the wanted productand is situated around 80 mL retention volume. The small peak at 47 mL retention volume is caused through the protease TEV, the peak around 100 mL retention volume is the cleaved HIS-tag.

To achieve this goal, three different species of both single mutants and the equivalent double mutant are expressed and purified. To enable two-dimensional NMR experiments, all samples were expressed ^{15}N-labelled. To ensure conformity of all prepared samples, we chose to investigate the matter employing the classic ^{1}H-^{15}N-HSQC experiment. The measurements were performed at 25°C/pH = 7.4/GB1 measurement buffer/0.6 mM. The result should overlap with a given NMR spectrum of the wt GB1, with the exception of the mutated and their neighbouring residues. *Figure 6.1.3* shows the comparison of the double mutant and wt GB1 and as expected, we observe distinct shifts around the site of mutation. However, the overall spectrum still highly resembles the wild type, therefore we can assume that structural integrity could be maintained. To ensure that the MTSL label does not lead to structural change or precipitation, this comparison was performed on all samples and their labelled versions.

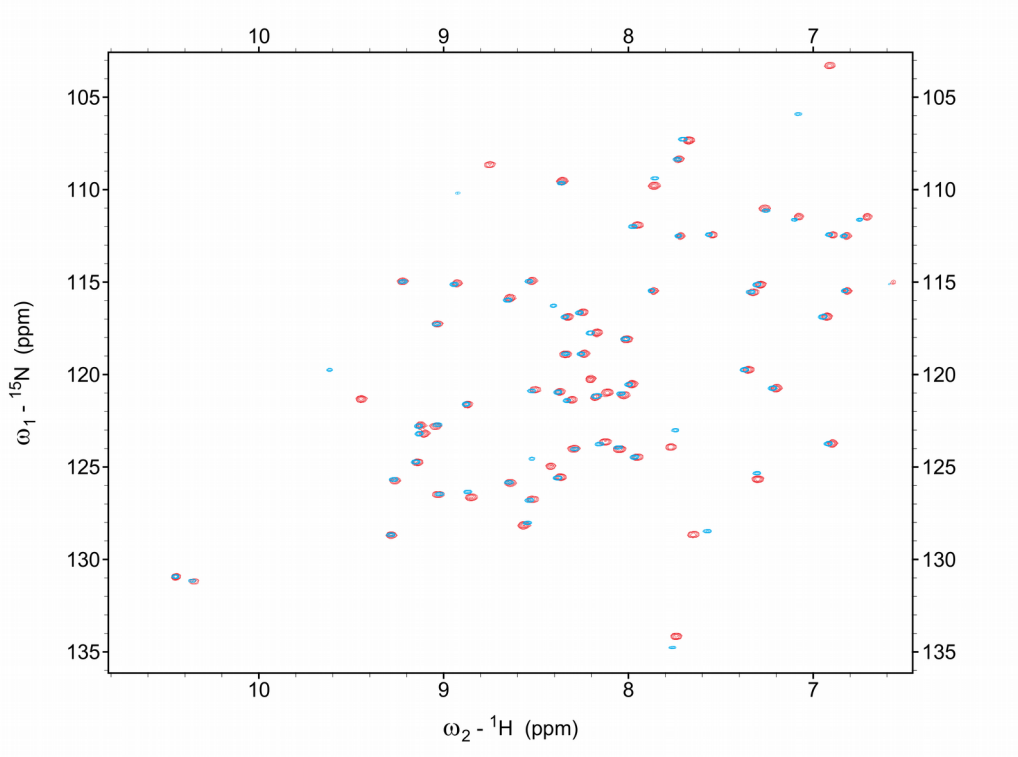

Figure 6.1.3: ^{1}H-^{15}N HSQC of GB1 wt in red, in comparison with ^{1}H-^{15}N HSQC of GB1 K9C:A47C

The critical information needed for this study is the proton T_2 relaxation time, which was acquired through multiple two dimensional spectra, varying in their delays. The peak height of these different points in time should follow an exponential decay. The same analysis was performed on the quenched samples. The parameters used for the measurements were 25°C/pH = 7.4/GB1measurement buffer/0.6 mM. To resolve the relaxation curve we chose the delays: 0.01, 0.025, 0.04, 0.05, 0.075, 0.1 (sec). Though initially we used the built-in function of the NMR analysis tool Sparky to fit the data to an exponential decay, the approach was discarded because the signal-to-noise ratio is not directly integrated in the calculation. To avoid this issue we manually extract the base RMSD value from each specific spectrum. The logarithm of this data is now taken to produce a linear fit of each residue. This is represented in *figure 6.1.4,* through the residue Asn 36. To include the signal-to-noise ratio to the fit we randomly sample within the according range of signal ± RMSD, which can be observed through the bar height. Hereby, all combination of three data points were taken in account and chosen according to the quality of the fit. The calculated slope now can be transformed into the relaxation rate R_2.

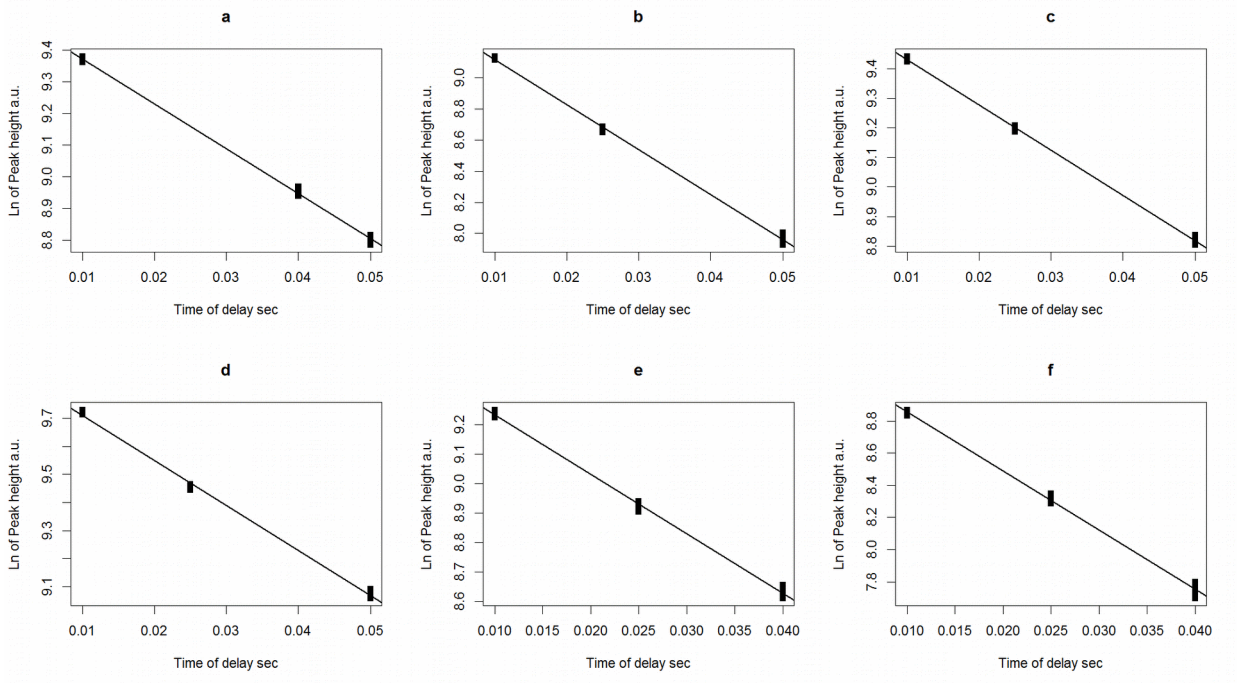

Figure 6.1.4: These graphs show the fit of logarithmic peak heights of N36 plotted against their corresponding time. The height of the bars represent

the peak height ± the spectral RMSD. (a) quenched K9C, (b) K9C, (c) quenched A47C, (d) A47C, (e) quenched double mutant, (f) double mutant

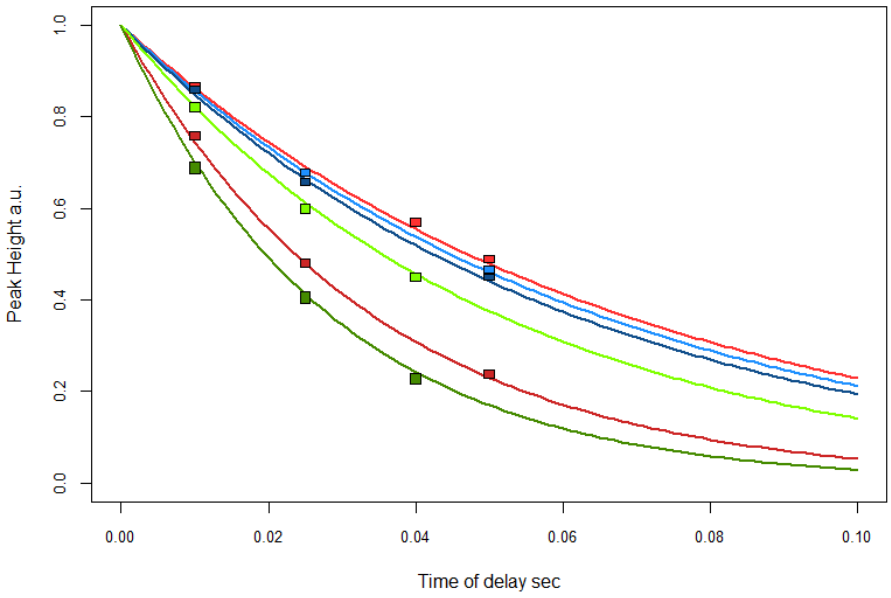

Figure 6.1.5: Peak height of N36 plotted against time. Barheight equals peak height ± the spectral RMSD. Quenched K9C (light red), K9C (dark red), quenched A47C (light blue), A47C (dark blue), quenched double mutant (light green), double mutant (dark green)

To enable the comparison between the angles derived from a pre-existing three-dimensional structure and the experimental data, it was first necessary to calculate the difference of the paramagnetic and quenched measurement. This can be seen in the *figure 6.1.5,* displaying the enhanced relaxation and their quenched counterpart of residue Asn 36. The relaxation profile of all residues is shown in *figure 6.1.6*. Here, we depict the rates $^{1}H^{N}$-Γ_2 of both single and the double mutant and their respective error bars. Not accounting for the interference effect, it is expected that the double labelled profile is comprised of the features inherent to both single labelled species. A significant increase of the relaxation rate is evident in residues surrounding the label position, but also other loci in structural proximity. The most prominent of these features is situated in the region of Val 38 and observable in the mutant K9C as well

as the double mutant. The theorised interference is however not directly observable through this representation.

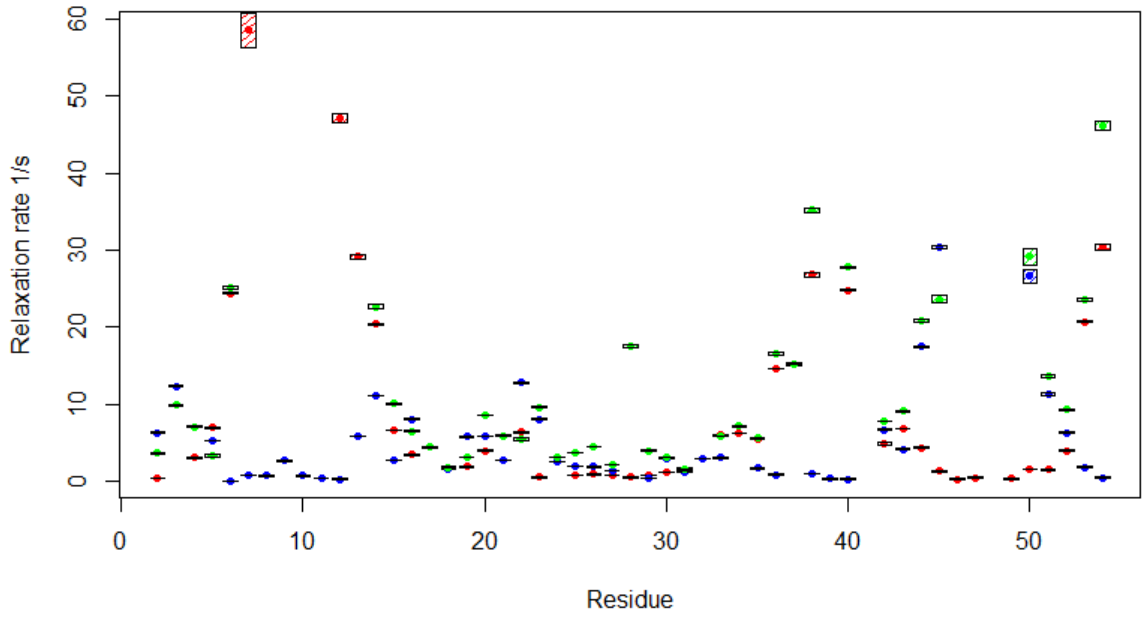

Figure 6.1.6: Graph depicting corrected relaxation rates against residue number. K9C (red), A47C (blue), K9C A47C (green).

To evaluate the impact of interference between the distinct PREs, we calculate the term $\Delta^1H^N\text{-}\Gamma_2$ divided by the square root of $[^1H^N\text{-}\Gamma_2(X1) * {}^1H^N\text{-}\Gamma_2(X2)]$, as discussed in *section 2.5.3.3*. These results were then compared to $(3*\cos(\phi)-1)/2$, ϕ being the angle derived from the three-dimensional structure (*formula 2.5.3.3)*. To directly compare the distribution of both the experimental and theoretical interference values, we portray an overlay of the respective contribution of frequencies in intervals of 0.5. During evaluation it is already apparent that only a fraction of calculated interference rates lie within definition range. This issue may arise due to incomplete consistency within measurements or minor disparity between samples. Another factor adding to this problem is the miniscule extent of relaxation change, impairing the significance of consequent changes in relaxation rate. The histogram seen in *figure 6.1.7* below

emphasizes this issue, which in general discredits the quality and validity of the acquired results. Therefore the shapes of overall distributions do not resemble each other, but only visualizes the difference in range of interference rates.

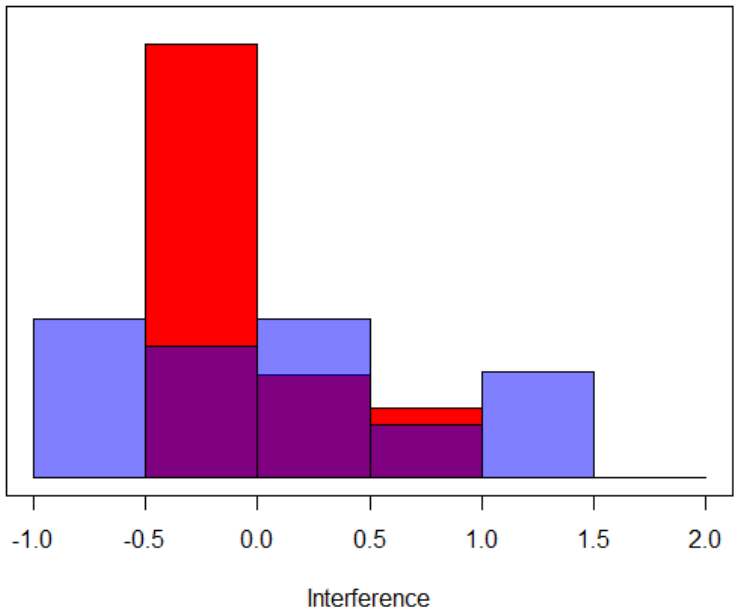

Figure 6.1.7: Histogram of relative theoretical (red) and experimental (blue) interference rates

To effectively juxtapose theoretical and experimental data, we plot both interference rates along the according sequence *(figure 6.1.8)*. It is evident that even after stringent selection of appropriate experimental values, coherence is still insufficient to encourage validity of the proposed correlation. Therefore, we searched for additional parameters which could have influenced this poor outcome. Due to the rather small size of the protein GB1, the specific position of paramagnetic labels greatly influences the overall intramolecular angles. To account for the flexibility of the MTSL label and the side chain of the cysteine it is attached to, we chose to sample combinations of distance restricted randomised coordinates, utilising them as possible label positions. We then calculate the intramolecular angles for all these combinations and optimised for maximal coherence between experimental and theoretical data. This improved interference pattern is also represented within the

figure 6.1.8. Although there is improvement concerning single residues, the overall results are still not convincing.

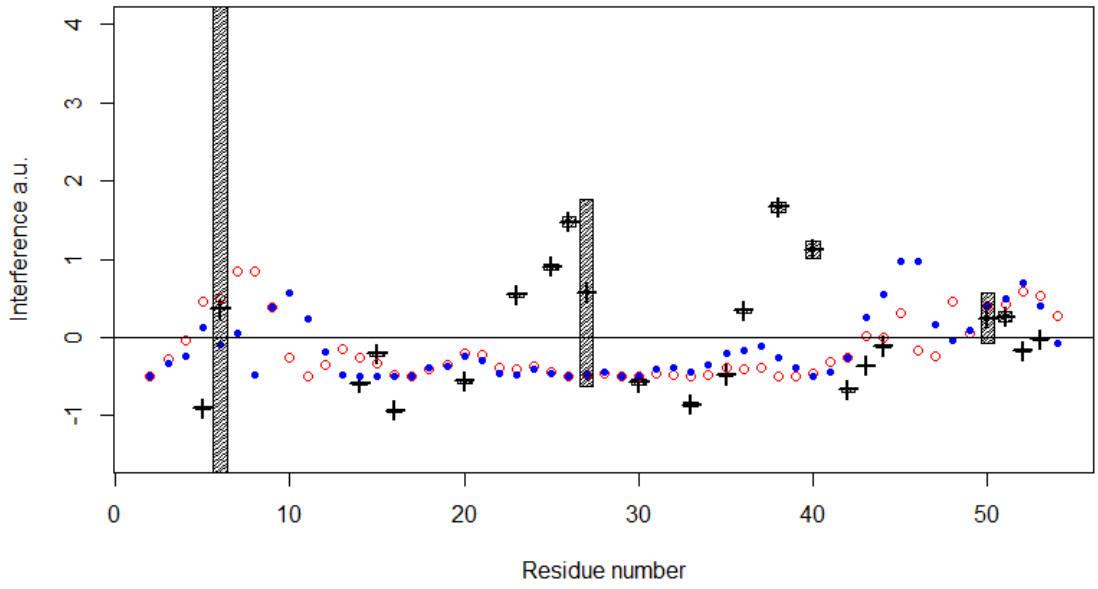

Figure 6.1.8: Theoretical interference rates plotted against residue number. Original calculated interference in red, rates enhanced through random sampling(blue). Experimental values are represented by crosses, the bars enclosing them show the standard error.

Apart from the research approaching the topic of PRI, we also used GB1 to investigate the incorporation of a novel precursor of lysine, which would allow selective labelling. Despite high yield and several trials, the precursor could not be observed in carbon isotope specific NMR experiments and therefore was not successfully taken up in the expressed protein.

6.2 NGAL

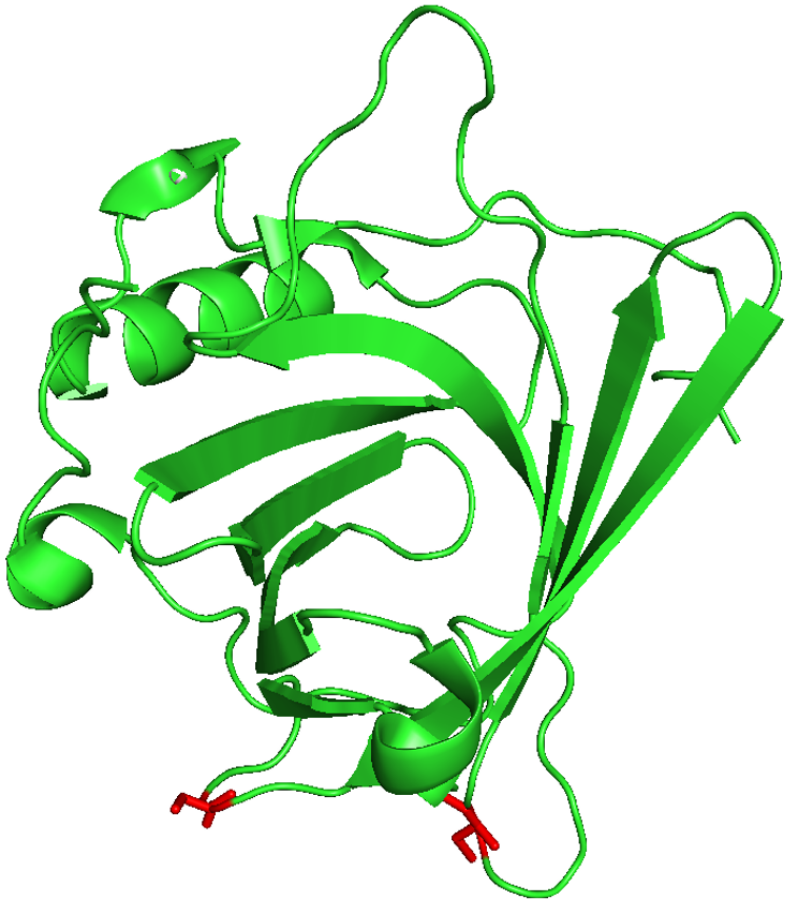

Figure 6.2.1: Three dimensional model of NGAL(LCN2) based on a enterobactin bound crystal structure[55].

Due to the marginal changes in relaxation rate experienced on GB1, we chose a larger experimental system, which consequentially exhibits a longer correlation time. To allow for the measurement of PRI, we introduced cysteine mutations in the protein. We placed mutations at position Lys 62 and Ser 87 and expressed the single mutants and double mutants respectively. As mentioned above alternative positions were chosen as well, but eventually not expressed. All of these positions were chosen to fulfil the requirements stated in *section 6.1*, concerning the experimental planning of GB1. Although expression and purification of NGAL are methods already established, several problems presented themselves, the most significant being the

coexpressed bacterial ligand enterobactin. This poses a problem, as enterobactin binds ferric iron which acts as a paramagnetic entity. A crucial factor, aiding to the prevention of this issue, is to start the expression culture in LB medium and not as initially proposed directly in M9 medium. As enterobactin is released by bacteria as stress response to salvage ferric iron, we argued that the minimal medium M9 enhances the expression of this ligand. This could be directly observed through intense colouration caused by the ligand bound iron which significantly decreased upon change to LB. Another factor which heavily influenced quality of samples is the final separation of residual enterobactin bound to NGAL. As mentioned in the methods concerning NGAL purification, we denaturate the protein followed by a purification step and subsequently refolding. Regarding this separation step we investigated different strategies and finally settled at the method mentioned in *section 5.2.2*.

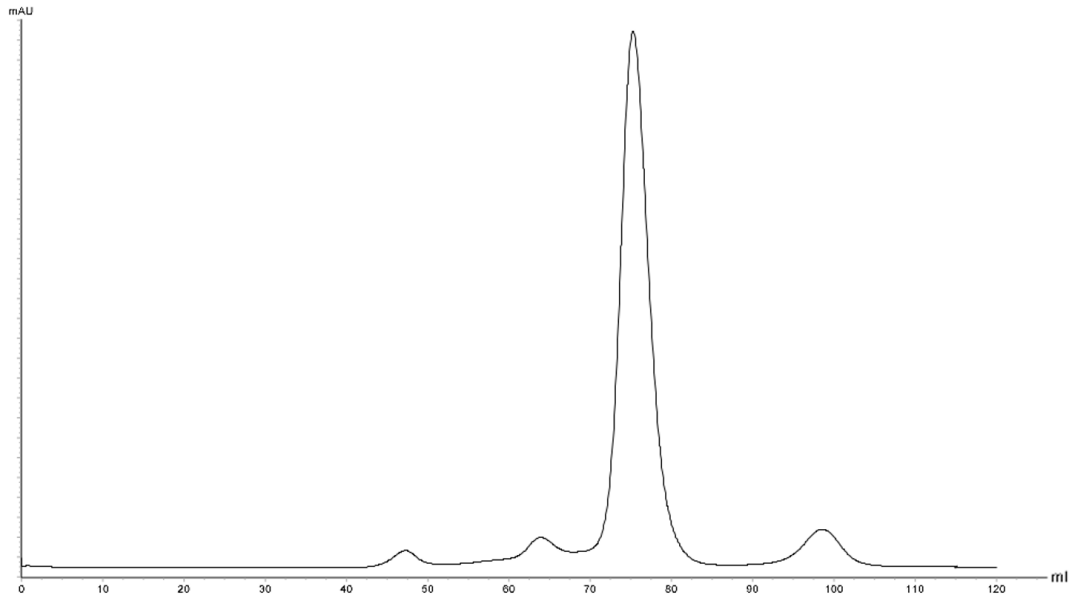

Figure 6.2.2: Size exclusion chromatography of NGAL.. The black line hereby represents A_{280}. The main peak (76 mL retention volume) represents cleaved NGAL whereas the peak around 65 mL retention volume is due to residual uncleaved protein. The protease is located at around 47mL and the Tag at around 100mL retention volume.

To rule out structural changes induced through the introduced mutations, but also to investigate the quality of the sample, we measured and compared the HSQC shifts of the mutants and wild type. The spectra were measured at 25°C/pH = 7.4/Tris/0.4 mM and are shown in *figure 6.2.3*. To account for the effect of wt enterobactin bound to the protein, another spectrum is included, representing a comparison between the double mutant and a wild type, ligand-bound spectrum. Already through these measurements it was apparent that residual wt ligand is still bound to NGAL and affecting peak height and width. Although we tried to continuously improve our approach to rid the protein of its ligand, residual populations of enterobactin-complexed iron still distort the measured relaxation times and overall appearance of the spectrum. The spectra shown in *figure 6.2.3* exhibit a relative low signal-to-noise ratio, apparent through the unfavourable peak shape and separation. Due to this, we already questioned the suitability of the measured samples to extract reasonable relaxation rates from. Despite the unfavourable results, we still propose structural integrity based on the overall conformance of chemical shifts, excluding the positions of mutation. *Figure 6.2.4*, shows a comparison between the purified double mutant and an iron loaded enterobactin bound sample of NGAL. The spectra were measured at 25°/pH = 7.4/Tris/0.4 mM (double mutant)/0.5 mM (enterobactin bound). This shows the profound effect the paramagnetic species has on the observability of peaks. Peaks close to the site of ligand interaction experience a drastic decrease in peak height due to the increase in relaxation rate. The representation emphasizes the importance of a sample free of unwanted paramagnetic elements. This however could not be achieved to the desirable degree within the scope of this project.

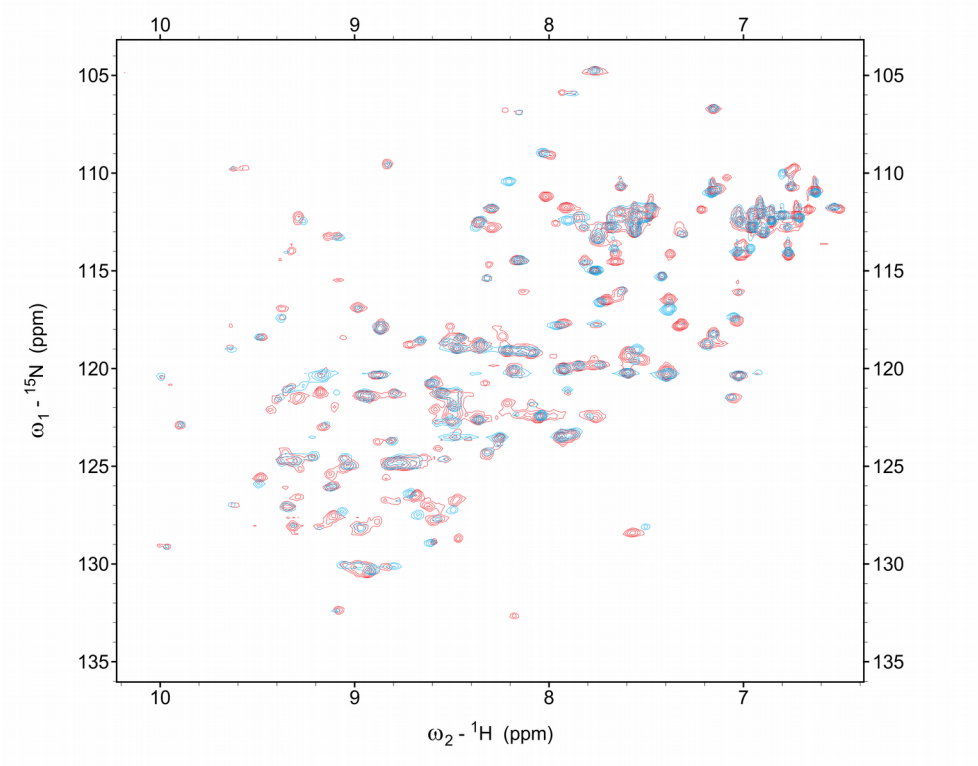

Figure 6.2.3: ^{1}H-^{15}N HSQC spectra of the wild type (red) and double mutant (blue) of NGAL.

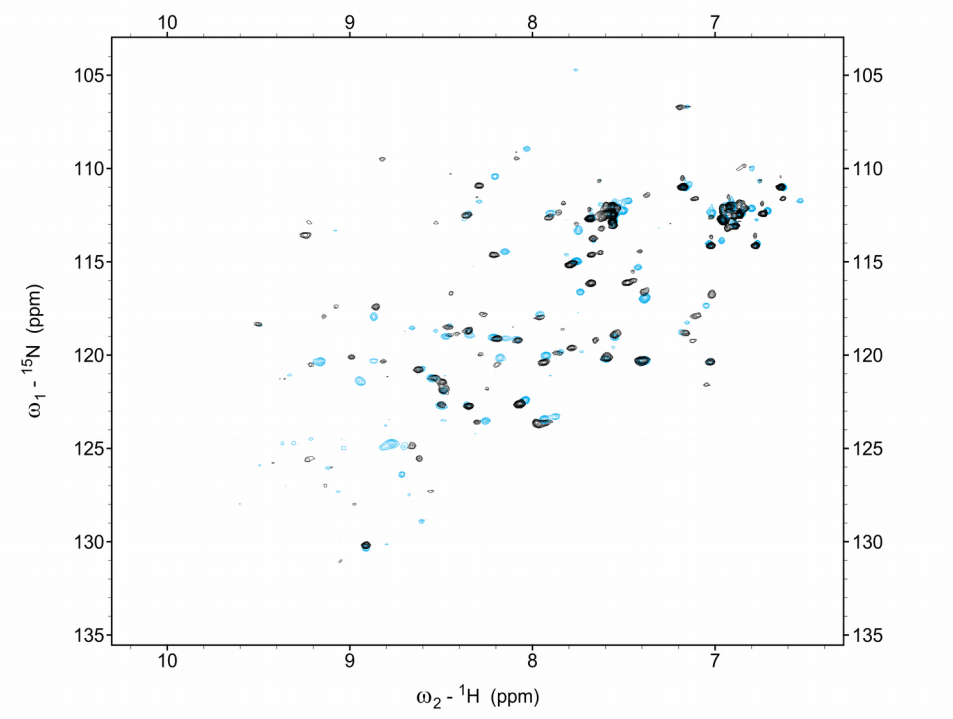

Figure 6.2.4: ^{1}H-^{15}N HSQC spectra of the NGAL double mutant (blue) and wtNGAL bound to ferric enterobactin (black).

Before tagging the sample, it is essential to occupy the binding site with the diamagnetic species of enterobactin to avoid intermolecular interaction between MTSL and the binding site. As described within *section 5.2.3* we expressed and purified enterobactin in a separate experiment and chose the diamagnetic element gallium as substitute for iron. After successful tagging of the samples, we measured the transverse relaxation rate, employing the same experiment as used in the case of GB1, only differing in the number and magnitude of chosen intervals. The spectra were recorded at 25°C/pH = 7.4/Tris/0.3 mM. To resolve the transverse relaxation curve of NGAL we chose the time points: 0.001, 0.004, 0.009, 0.02, 0.03 (sec). It can be directly seen that the signal-to-noise ratio is rather poor in comparison to GB1. This results in a higher uncertainty, regarding the calculated rates and further hinders a reasonable continuation of evaluation. As six different rates are needed to compute the interference rate, already one flawed value can drastically affect the outcome. This possibly leads to rates being outside the range of definition and therefore a loss of utilisable data. *Figure 6.2.5* shows the fit of the logarithmic peak heights of residue Leu-159, against time, which in turn reveals the relaxation rate of the specific residue.

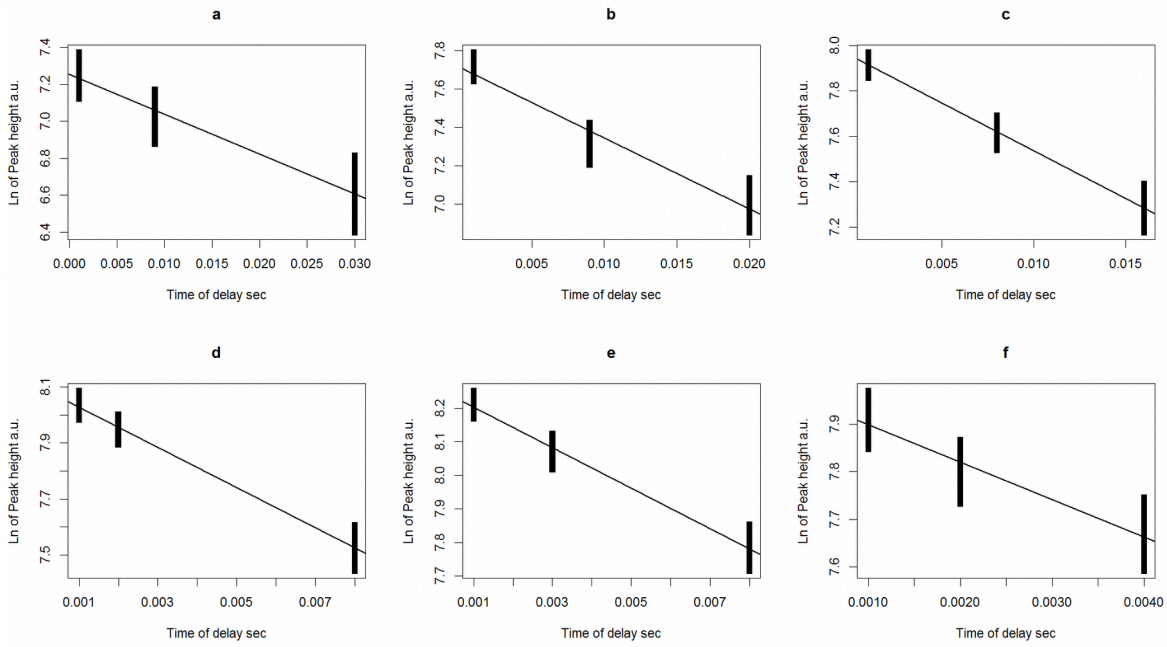

Figure 6.2.5: The graphs shown represent the fit of logarithmic peak heights of residue L159 plotted against their corresponding time of delay.

The height of the bars hereby represent the peak height ± the spectral RMSD. (a) quenched K62C, (b) K62C, (c) quenched S87C, (d) S87C, (e) quenched double mutant, (f) double mutant

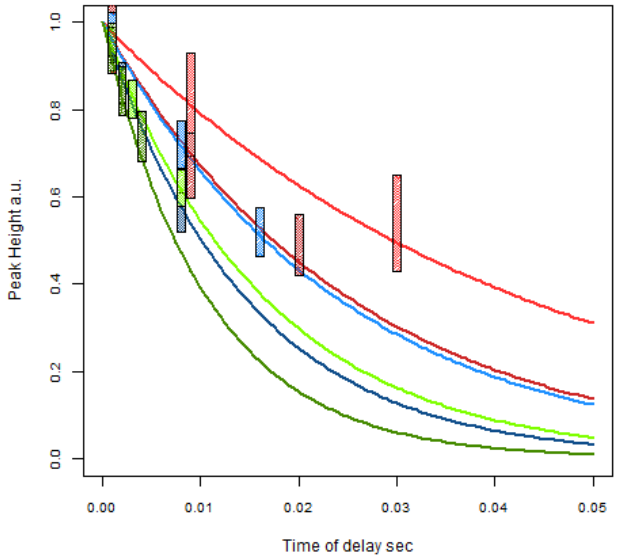

Figure 6.2.6: Plot of peak heights of residue L159 against according time, whereas the bar height represents peak height ± the spectral RMSD. Quenched K62C (light red), K62C (dark red), quenched S87C (light blue), S87C (dark blue), quenched double mutant (light green), double mutant (dark green)

Figure 6.2.6 shows a comparison of the enhanced relaxation of Leu 159 and their quenched counterpart, as well as the margin of error originating from the signal-to-noise ratio of the acquired spectra. During evaluation it already became evident that, not only is the exhibited margin of error too substantial to allow significant calculated rates, but the fraction of utilisable data is severely limited in comparison to GB1. However, all reasonable data is guided through the same procedure as in case of GB1. To calculate a measure of interference we follow the calculation also utilised in *section 6.1* and use the comparison between theoretical and experimental values according to *formula 2.5.3.3.* In *figure 6.2.7* we plot the distribution of experimentally determined interference rates in comparison with the back-calculated interference

rates, derived from the 3D structure. There is however only little information we can gain through this depiction due to insufficient sample-size.

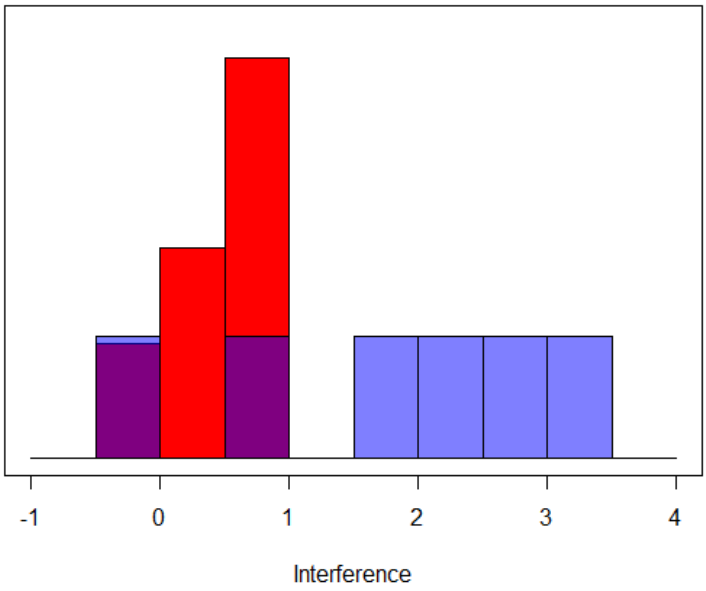

Figure 6.2.7: Histogram of relative theoretical (red) and experimental (blue) interference rates.

Figure 6.2.8 shows a comparison between all interference rates. Again we used the method created to generate an adapted interference pattern based on randomly generated label positions described in *section 6.1*. Although the range of definition is only between -0.5 : 1.5 we included experimental results in the range of -1 : 4. This shall only visualise the flawed significance of the acquired results, not overestimating the data points within range. The only noteworthy data point is corresponding to the residue Cys 76. The results of this experiment are therefore as a whole negligible and far off any results which could support the proposed hypothesis.

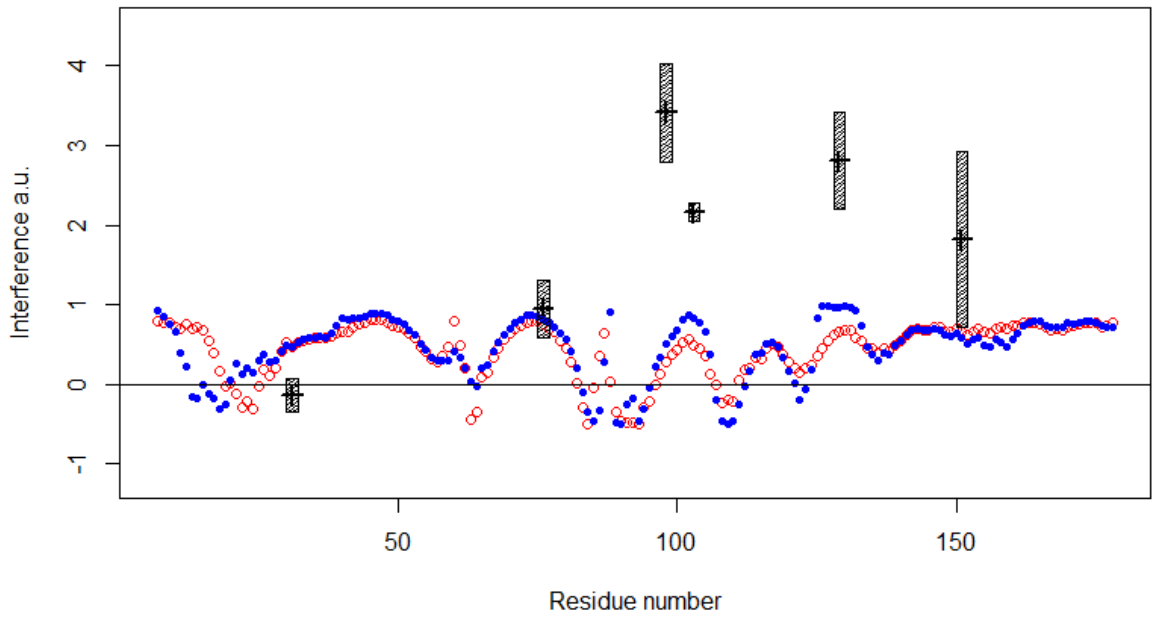

Figure 6.2.8: Plot represents comparison of interference rates. Original theoretical (red) and randomly enhanced (blue) values set against experimental values ± their respective standard error.

6.3 Max-Brca1

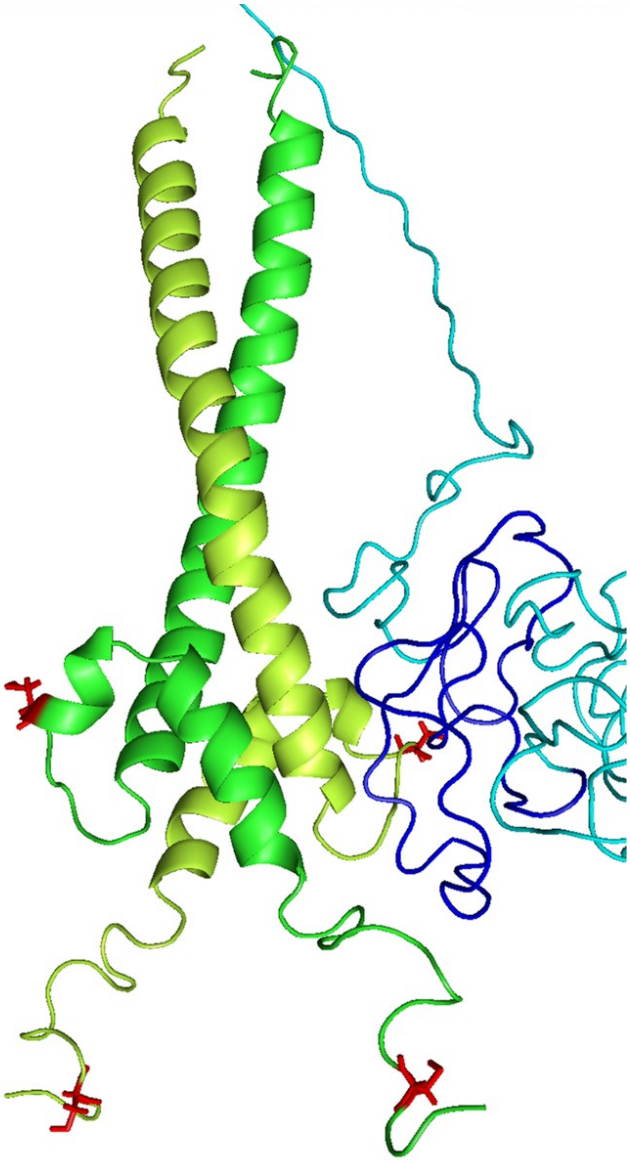

Figure 6.3.1: NMR structure of a MAX dimer (light and dark green) bound to Brca1. The peptide, used in this study is portrayed in dark blue. Mutations on Max are signified through red colouration. Structure is taken from (Kurzbach et. Al, unpublished).

As depicted in the three-dimensional model above, we used cysteine mutants of Max in this experiment. These cysteines were used to label the protein with the paramagnetic entity MTSL. We chose positions which would not interfere with the

integrity of the structure and uphold the other requirements mentioned in *section 6.1.* Additionally, the inserted mutation and the paramagnetic label should not interfere with dimerisation of MAX. Contrary to the experimental systems above, Max and Brca1 were used to examine intermolecular paramagnetic phenomena. To achieve this, the Brca1 peptide was expressed ^{15}N-labelled, whereas its binding partner, the Max dimer, bears the paramagnetic species. This of course implies that we cannot rely on NMR techniques to examine the quality of the produced MAX mutants and therefore solely examined the samples via chromatographic methods.

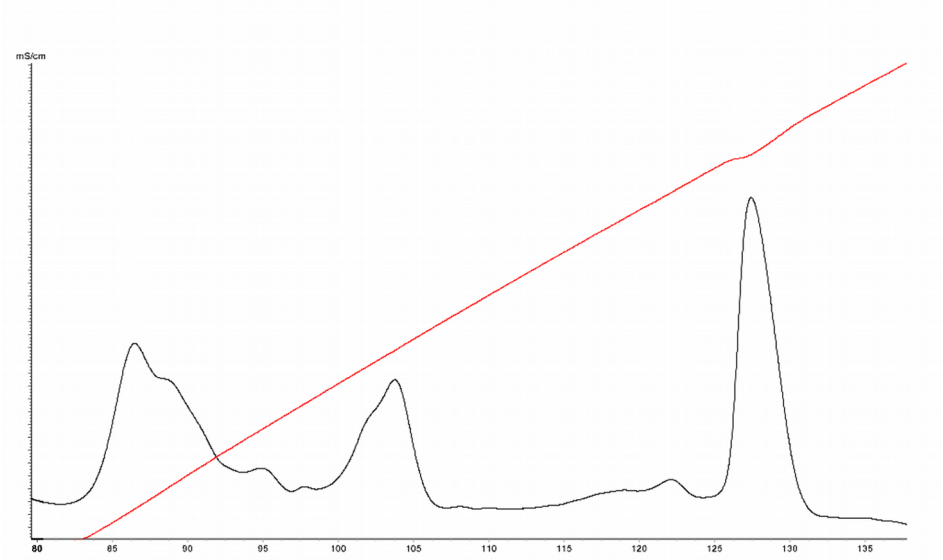

Figure 6.3.2: Chromatogram of Max run on a Resource S column. The peak of interest is located around 128 mL elution volume. The black line represents A_{280} and the inclining red line represents the change in conductivity upon elution.

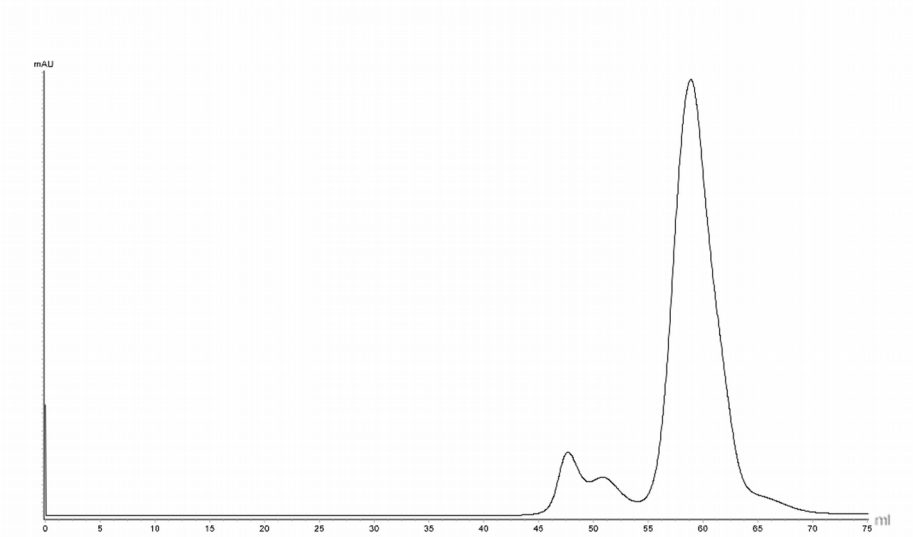

Figure 6.3.3: Size exclusion chromatography of Brca1 (A_{280}).

Whereas interference in GB1 and NGAL was measured through comparison of single to double mutants, we used MAX homodimers, exhibiting the same mutation site compared to a mixed heterodimer of both MAX mutants. As performed in all other experimental systems, we utilised the HSQC experiment to observe the quality of the sample. Additionally, we monitored the change upon binding of the Max dimer and therefore repeated the measurement. The peptide exhibits great changes in chemical shifts and overall spectral character upon binding of the MAX dimer. Additionally, the acquired spectrum exhibits a decrease in discernibility due to spectral overlap. These factors lead us to the inevitable necessity of determining a new assignment on the peptide. We therefore measured HNCA and HNCOCA of a $^{13}C/^{15}N$ labelled Brca1 bound to a MAX dimer. These measurements were performed at 20°/MES-buffer/pH = 5.5/0.7 mM MAX/0.7 mM Brca1.

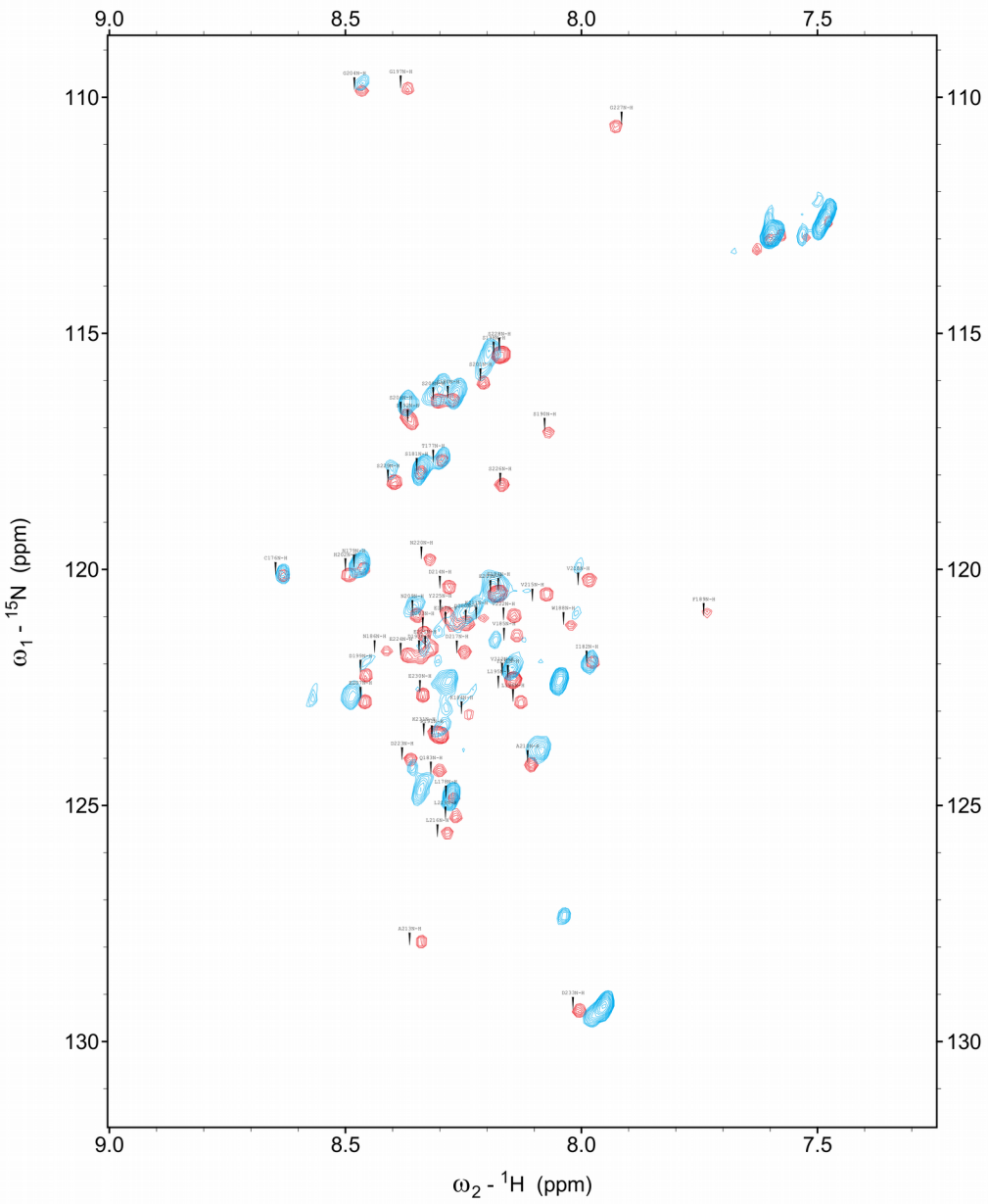

Figure 6.3.4: ^{1}H-^{15}N HSQC spectra of Brca1 unbound (red) and bound to MAX (blue)

As already stated, this first evaluation is followed by relaxation measurements of the sample. The parameters of these measurements are 20°C/MES-buffer/pH = 5.5/0.5 mM MAX/0.5 mM Brca1. Due to the already observed low spectral quality, which

might be due to the increased correlation coefficient, occuring while bound to the Max dimer, also the signal-to-noise ratio is inadequate to deduct accurate rates. However, it was possible to extract some reasonable rates from acquired data and further use them in our calculations. *Figure 6.3.5* shows the fit of logarithmic peak heights against time of the residue Asn 179. The graphs (a) to (f) hereby display the six fits needed to calculate the interference rate of one residue. *Figure 6.3.6* displays a comparison of the calculated fits and the initial peak heights of the same residue. In both figures the bar width represents the RMSD, specific for each spectrum.

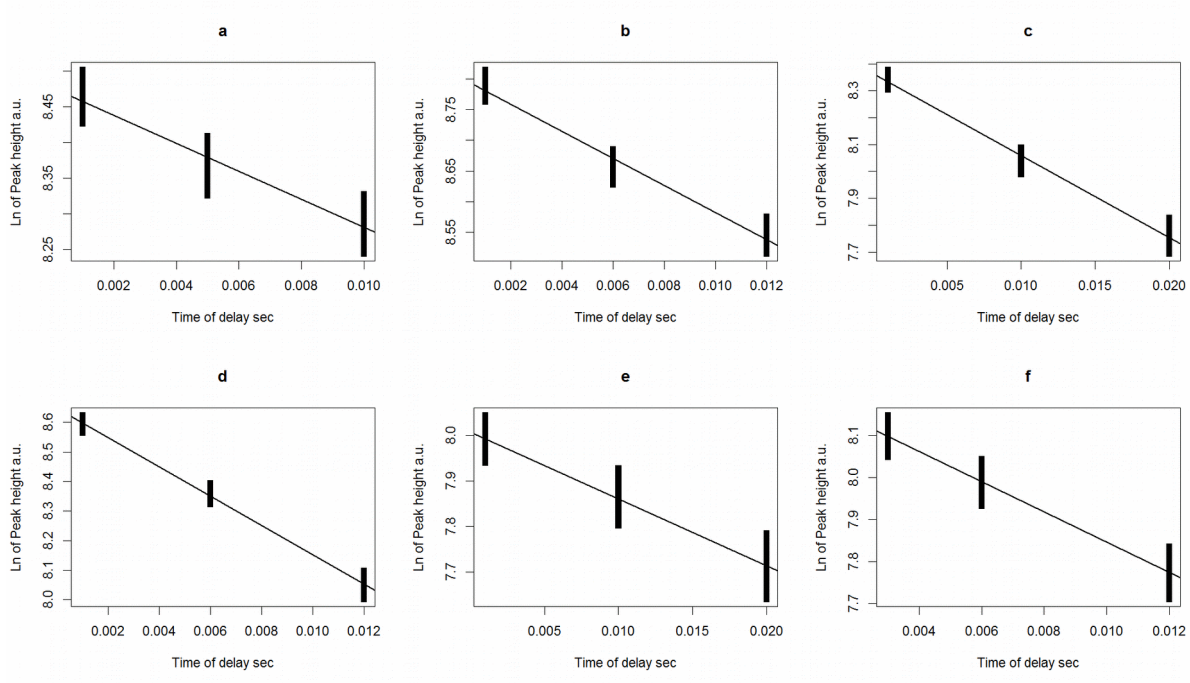

Figure 6.3.5: The shown graphs show the fit of logarithmic peak heights of the residue N179 plotted against their corresponding time. The height of the bars represent the peak height ± the spectral RMSD. (a) quenched R5C, (b) R5C, (c) quenched G35C, (d) G35C, (e) quenched double mutant, (f) double mutant

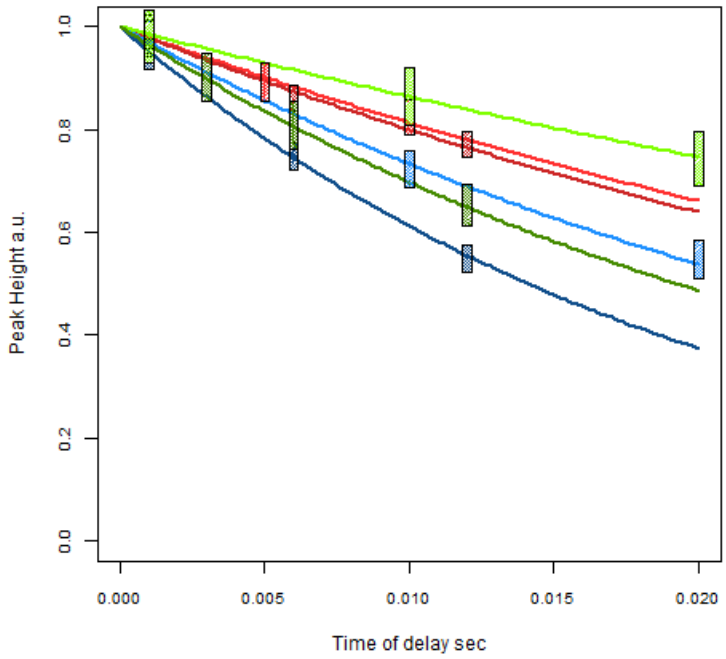

Figure 6.3.6: Plot of peak heights of residue N179 against according time, whereas the bar height represents peak height ± the spectral RMSD. Quenched R5C (light red), R5C (dark red), quenched G35C (light blue), G35C (dark blue), quenched double mutant (light green), double mutant (dark green)

Although the proton relaxation measurements themselves are identical the ones performed for GB1 and NGAL, evaluation differs to account for the partition of the dimer. After calculation of the change in relaxation rates due to the introduction of the paramagnetic entity, we formulate the difference between the heterodimer and the arithmetic mean of homodimers. This measure however can't be directly compared to the angle dependent term we utilised in the systems above. First we ought to calculate the theoretical contributions to the paramagnetic effect given through the distances between labels and spin. Due to the fact that in every measurement Brca1 is affected by two paramagnetic centres, we need to calculate both distances respectively, but also include a term describing the interfering effect, which is dependent on the calculated angle as can be seen in *formula 6.3.7.1/2/3*. After characterising the effects present at each different dimer, we formulate the difference

between the average mean of homodimers and the heterodimer $(E^{X(1)}+E^{X(2)})/2 - E^{X(1):X(2)}$ and compare it to the measured values, processed the same way.

$$(1) \quad E^{X(1)} = \frac{1}{r_{X(1)}^{-6}} + \frac{1}{r_{X(1)'}^{-6}} + \frac{1}{r_{X(1):X(1)'}} * \frac{3\cos^2\varphi_{X(1):X(1)'} - 1}{2} \qquad (2) \quad \frac{1}{r_{X(1):X(1)'}} = \frac{1}{r_{X(1)}^{-3}} * \frac{1}{r_{X(1)'}^{-3}}$$

$$(3) \quad E^{X(1):X(2)} = \frac{\frac{1}{r_{X(1)}^{-6}} + \frac{1}{r_{X(2)'}^{-6}}}{2} + \frac{\frac{1}{r_{X(2)}^{-6}} + \frac{1}{r_{X(1)'}^{-6}}}{2} + \frac{\frac{1}{r_{X(1):X(2)'}} * \frac{3\cos^2\varphi_{X(1):X(2)'} - 1}{2} + \frac{1}{r_{X(2):X(1)'}} * \frac{3\cos^2\varphi_{X(1):X(2)'} - 1}{2}}{2}$$

Formula 6.3.7.1/2/3: Alternative formula to calculate the participating rates regarding the measurement of Brca1.(1) The value $E^{X(1)}$ describes the experienced change of relaxation rate due to proximity of the unpaired electrons X(1) and X(1)'. The prime in X(1)' hereby indicates the location of the paramagnetic centre on the paired MAX molecule. The respective distances of $^1H^N$-X(1) and $^1H^N$-X(1)' were derived from a given three-dimensional structure and are indicated with $r_{X(1)}$ and $r_{X(1)'}$. The term $r_{X(1):X(1)'}$ is defined through the supplemented formula (2). Formula (3) describes the calculation of the theoretical change of relaxation rate in case of a mixed dimer accounting for both the possible dimers X(1):X(2)' and X(2):X(1)'. To achieve this we proceed in the same fashion as in formula (1), but substituting each term by the arithmetic mean of both possible dimers.

Due to this difference in calculation the correlation coefficient does not cancel itself out and has to be accounted for in a different way. This however poses a problem as none of the produced samples is only affected by one paramagnetic site, which could be used to deduct the necessary value through distance calculation. As a result, we chose to account for the correlation coefficient through matching of the calculated

rates and acquired rates. This may distort the significance of the results, as the calculated factor is not homogeneous across the measured residues. As the possibility of negative interference is the main goal of this thesis, we display the distribution of interference values compared to the calculated equivalent.

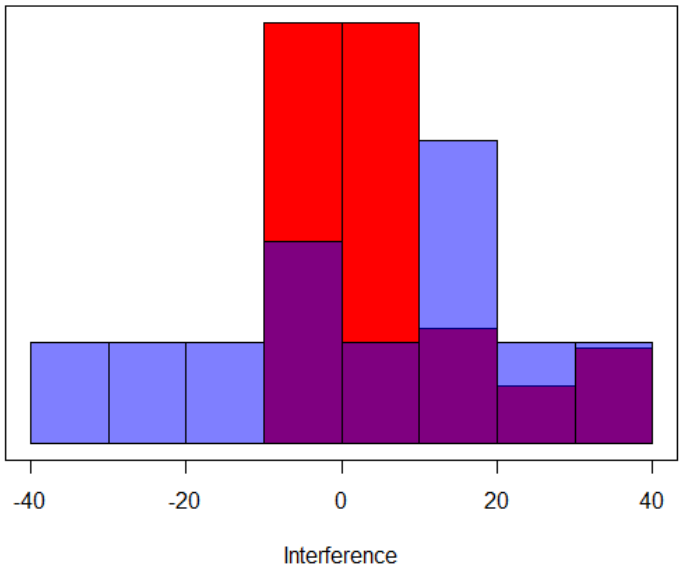

Figure 6.3.8: Histogram of relative theoretical (red) and experimental (blue) interference rates.

The comparison of theoretical and experimental data shown in figure 6.3.9 shows minor correlation of specific residues, but as the interference rates themselves took part in the calculation of the correlation coefficient, the significance is highly questionable. However, as this coefficient only applies as a multiplier we can argue that a correlation of algebraic signs is present.

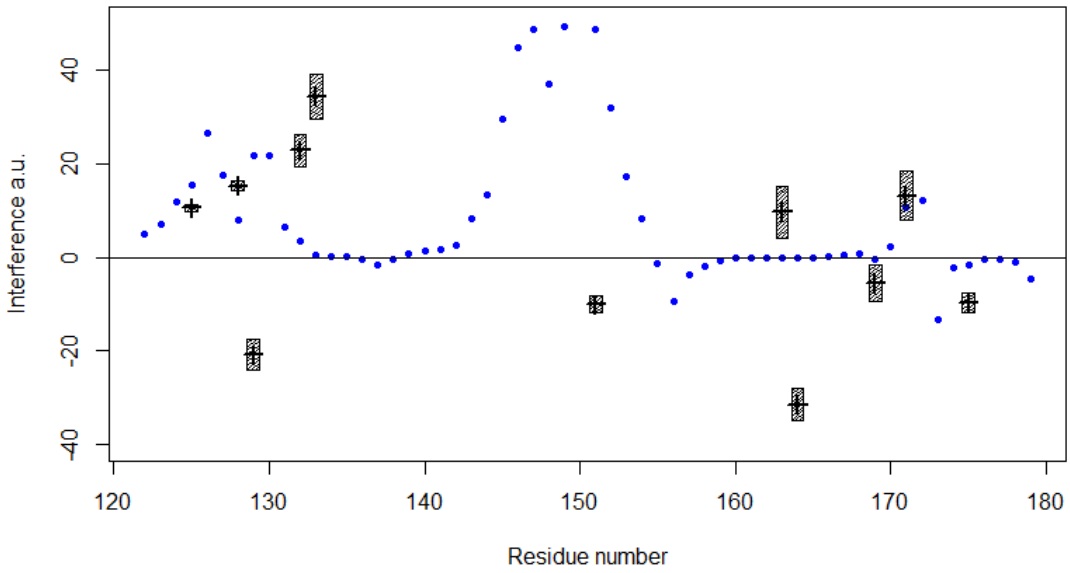

Figure 6.3.9: Plot of theoretical interference rates (blue) against residue number, overlaid with experimental interference ± the respective standard error.

7 Discussion

Foremost it is evident that neither of these three experimental systems produced results which would proof the proposed relationship between spin-angle and interference in a significant way. Although we could show correlation of algebraic sign in specific cases, this interpretation is outweighed by the many data points, either not within range or definition or simply uncorrelated.

GB1 yielded the best ratio of data points within range of definition. Nevertheless, the changes of observed relaxation rates upon quenching of the sample are marginal. This decreases the significance of the acquired results. Despite our efforts to apply this experiment to a larger system, which would in theory return more compelling relaxation rates, both systems tested lead only to insufficient results. This, in case of NGAL, is due to inadequate preparation of the samples, regarding their purity of paramagnetic species and overall quality. If these factors were to be corrected and it was indeed possible in other studies to sufficiently purify NGAL, this system would still pose a plausible strategy to investigate the validity of PRIs.

Regarding Brca1 as our chosen biological system, we could have improved our strategy by restricting the number of labels in a fashion similar to the other examples. To achieve this we would have required both unlabelled and a double cysteine mutant of MAX. With these species it would be possible to create dimers with only one or two labels respectively. This would enable us to perform the same calculations as for the other systems, simplifying the process of evaluation and also accounting for the needed correlation coefficient. Another factor, which could however not as easily be corrected, is the effect that binding of the two proteins has on the measured NMR spectra. As there already have been successful attempts on employing this system for NMR measurements, there certainly is possibility to acquire a better result than the one obtained. The reduction of effective labels on the dimer, may also affect the outcome of the relaxation measurements. Through the consequential increase in relaxation time it could be possible to acquire data points over a longer timeframe.

These results do not falsify the premise stated as they are solely caused through failed experimental planning and execution. Therefore, the experiment itself still possesses great promise as method for structure determination. Hereby we have to note that the accuracy obtained through PRIs may still pose an issue not only for this study. Hence, although the display of correlation between sign of interference and angle may be possible, the initial assumption to use this method as an accurate measurement of angle is questionable. This conclusion is based on the observation that not only do most of the experimentally acquired angles not resemble the theoretical ones, but further is a non-negligible quantity of residues out of range of definition. Hereby we have to note that this evaluation is based on the listed experiments and therefore might underestimate the applicability of the method.

A parameter influencing the results is the flexibility of the label, which depending on size, can hinder the correct interpretation of data. Although the random sampling model proposed in this thesis accounts for these changes in angle, the effectiveness and accuracy could be increased through introduction of restraints. This however necessitates the availability of structural information on the protein in question. If available, it would be possible to compute sterical as well as electrostatic restraints which would restrict the degrees of freedom. This could be enhanced by including parameters based on the properties of the amino acid side chain.

As already discussed regarding GB1, the label position is of importance to optimize the amount of obtainable data. Apart from maintenance of structural integrity, the range of the paramagnetic effect is crucial. To optimize the overlap of affected spins, it is possible to introduce a counting measure. As was initially planned for further experimental systems, one can iterate through all double combination of residues and count all spins within predetermined range of both label positions, excluding positions too close to each other. Through this calculation it is possible to choose the residues which theoretically yield the most significant output.

If this method will lead to plausible results in future studies, it could function as novel approach for structural determination and observation. An argument speaking for this

approach is the already well established use of paramagnetic entities in NMR. Therefore, the experimental adaptations to this method would be minimal. As the solution of protein structures utilising NMR usually depends on distances (PRE, NOE), this angular information could assist structure calculation. These measurements are not limited to angular information, but the distance dependence of PREs can also be deducted from the same measurements. The quality and validity of this method is however still dependent on the acquired accuracy of measurements. Apart from structural determination, it also would be possible to monitor conformational changes of proteins. This could be applied to e.g. measuring change of a ligand upon binding, also leaving the possibility open to label the binding partner and measure indirectly. This would not only yield information on the ligand's structure, but as distance is also contained in the acquired information, the positioning would be elucidated as well.

Conclusion

Although the acquired results do not allow for an accurate calculation of angles, we were we able to show a partial correlation of positive and negative PRIs. This is most apparent in case of GB1 with the exception of the alpha helical region as well as the loop connecting this feature to β-sheet 2 (AA 23..40), as can be seen in *figure 6.1.8*. However, the reason for this behaviour can be found in the distinct motional anisotropy within this region, which has already been characterised[56]. Nevertheless is this system not ideal, as due to its small size the paramagnetic effect and hence the amount of evaluable residues is limited. This partial correlation can also be seen regarding the measurements of Brca1 bound to MAX, as shown in *figure 6.3.9*. and could be refined through a more sensible approach regarding the introduction of paramagnetic species into MAX. Due to this correlation between our data and the proposed theory of angular dependent paramagnetic interference, we can argue for a possible applicability of PRIs, regarding globular proteins. In case of NGAL the poor quality of data simply does not allow for a reasonable evaluation. If quality of samples as well as choice of experimental system are optimised and meet the standards to

produce coherent data, this method may still pose an original strategy for the elucidation of structural information through NMR.

8 Bibliography

1. Kjeldsen, L., Johnsen, A. H., Sengeløv, H. & Borregaard, N. Isolation and primary structure of NGAL, a novel protein associated with human neutrophil gelatinase. *J. Biol. Chem.* **268,** 10425–32 (1993).

2. Chakraborty, S., Kaur, S., Guha, S. & Batra, S. K. The multifaceted roles of neutrophil gelatinase associated lipocalin (NGAL) in inflammation and cancer. *Biochim. Biophys. Acta - Rev. Cancer* **1826,** 129–169 (2012).

3. Flower, D. R. The lipocalin protein family: structure and function. *Biochem. J.* **318 (Pt 1,** 1–14 (1996).

4. Bao, G. *et al.* NIH Public Access. **6,** 602–609 (2011).

5. Fonslow, B. R. *et al.* NIH Public Access. **10,** 54–56 (2013).

6. Hu, L. *et al.* NGAL decreases E-cadherin-mediated cell-cell adhesion and increases cell motility and invasion through Rac1 in colon carcinoma cells. *Lab. Invest.* **89,** 531–548 (2009).

7. Li, C. & Chan, Y. R. Lipocalin 2 regulation and its complex role in inflammation and cancer. *Cytokine* **56,** 435–441 (2011).

8. Jin, M., Jang, E. & Suk, K. Lipocalin-2 Acts as a Neuroinflammatogen in Lipopolysaccharide-injected Mice. *Exp. Neurobiol.* **23,** 155–62 (2014).

9. Ip *et al.* Lipocalin 2 in the central nervous system host response to systemic lipopolysaccharide administration. *J. Neuroinflammation* **8,** 124 (2011).

10. Bi, F. *et al.* Reactive astrocytes secrete lcn2 to promote neuron death. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 4069–74 (2013).

11. Paragas, N. *et al.* NGAL-Siderocalin in kidney disease. *Biochim. Biophys. Acta - Mol. Cell Res.* **1823,** 1451–1458 (2012).

12. Zhao, C. *et al.* Structural characterization of glycoprotein NGAL, an early predictive biomarker for acute kidney injury. *Carbohydr. Res.* **345,** 2252–2261 (2010).

13. Sjöbring, U., Björck, L. & Kastern, W. Streptococcal protein G. Gene structure and protein binding properties. *J. Biol. Chem.* **266,** 399–405 (1991).

14. Zhang, H. *et al.* Immunological characterization and verification of recombinant streptococcal protein G. *Mol. Med. Rep.* **12,** 6311–6315 (2015).

15. Linhult, M., Binz, H. K., Uhlén, M. & Hober, S. Mutational analysis of the interaction between albumin-binding domain from streptococcal protein G and human serum albumin. *Protein Sci.* **11,** 206–213 (2002).

16. Li, Y. D., Lamour, G., Gsponer, J., Zheng, P. & Li, H. The molecular mechanism underlying mechanical anisotropy of the protein GB1. *Biophys. J.* **103,** 2361–2368 (2012).

17. Derrick, J. P. *et al.* N Terminus. 1992 (1993).

18. Sloan, D. J. & Hellinga, H. W. Dissection of the protein G B1 domain binding site for human IgG Fc fragment. *Protein Sci.* **8,** 1643–1648 (1999).

19. Derrick, J. P. & Wigley, D. B. Crystal structure of a streptococcal protein G domain bound to an Fab fragment. *Nature* **359,** 752–4 (1992).

20. Wang, Q., Zhang, H., Kajino, K. & Greene, M. I. BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells. *Oncogene* **17,** 1939–1948 (1998).

21. Clark, S. L., Rodriguez, A. M., Snyder, R. R., Hankins, G. D. V. & Boehning, D. Structure-Function of the Tumor Suppressor Brca1. *Comput. Struct. Biotechnol. J.* **1,** 1–8 (2012).

22. Brzovic, P. S., Rajagopal, P., Hoyt, D. W., King, M.-C. & Klevit, R. E. Structure of a BRCA1– BARD1 heterodimeric RING–RING complex. *Nat. Struct. Biol.* **8,** 833–837 (2001).

23. Shen, Y. & Tong, L. Structural evidence for direct interactions between the BRCT domains of human BRCA1 and a phospho-peptide from human ACC1. *Biochemistry* **47,** 5767–5773 (2008).

24. Paull, T. T., Cortez, D., Bowers, B., Elledge, S. J. & Gellert, M. Direct DNA binding by Brca1. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 6086–6091 (2001).

25. Yamane, K., Katayama, E. & Tsuruo, T. The BRCT regions of tumor suppressor BRCA1 and of XRCC1 show DNA end binding activity with a multimerizing feature. *Biochem. Biophys. Res. Commun.* **279,** 678–684 (2000).

26. Buckley, N. E. & Mullan, P. B. BRCA1 - Conductor of the Breast Stem Cell Orchestra: The Role of BRCA1 in Mammary Gland Development and Identification of Cell of Origin of BRCA1 Mutant Breast Cancer. *Stem Cell Rev. Reports* **8,** 982–993 (2012).

27. Zheng, L., Li, S., Boyer, T. G. & Lee, W. H. Lessons learned from BRCA1 and BRCA2. *Oncogene* **19,** 6159–6175 (2000).

28. Duncan, J. A., Reeves, J. R. & Cooke, T. G. Review BRCA1 and BRCA2 proteins: roles in health and disease. *J Clin Pathol Mol Pathol* **51,** 237–247 (1998).

29. Kote-Jarai, Z. & Eeles, R. A. BRCA1, BRCA2 and their possible function in DNA damage response. *Br. J. Cancer* **81,** 1099–1102 (1999).

30. Li, H., Lee, T. H. & Avraham, H. A novel tricomplex of BRCA1, Nmi, and c-Myc inhibits c-Myc-induced human telomerase reverse transcriptase gene (hTERT) promoter activity in breast cancer. *J. Biol. Chem.* **277,** 20965–20973 (2002).

31. Razandi, M., Pedram, A., Rosen, E. M. & Levin, E. R. BRCA1 Inhibits Membrane Estrogen and Growth Factor Receptor Signaling to Cell Proliferation in Breast Cancer. **24,** 5900–5913 (2004).

32. Yoshida, K. & Miki, Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci.* **95,** 866–871 (2004).

33. Massari, M. E. & Murre, C. MINIREVIEW Helix-Loop-Helix Proteins : Regulators of Transcription in Eucaryotic Organisms. **20,** 429–440 (2000).

34. Nair, S. K. & Burley, S. K. X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors. *Cell* **112,** 193–205 (2003).

35. Wechsler, D. S. & Dang, C. V. Opposite orientations of DNA bending by c-Myc and Max. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 7635–7639 (1992).

36. Landschlz, W. H., Johnson, P. F. & Mcknight, S. L. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science (80-.).* **240,** (1988).

37. Blackwood, E. M. & Eisenman, R. N. Max: A Helix-Loop-Helix Zipper Protein That Forms a Sequence-Specific DNA-Binding Complex with Myc. *Source Sci. New Ser.* **251,** 1211–1217 (1991).

38. Ferré-D'Amaré, a R., Prendergast, G. C., Ziff, E. B. & Burley, S. K. Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **363,** 38–45 (1993).

39. Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. T He M Yc /M Ax /M Ad N Etwork and the T Ranscriptional C Ontrol of C Ell B Ehavior. *Annu. Rev. Cell Dev. Biol.* **16,** 653–699 (2000).

40. Lüscher, B. Function and regulation of the transcription factors of the Myc/Max/Mad network. *Gene* **277,** 1–14 (2001).

41. James, L. & Eisenman, R. N. Myc and Mad bHLHZ domains possess identical DNA-binding specificities but only partially overlapping functions in vivo. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 10429–34 (2002).

42. Comino-M??ndez, I. *et al.* Functional and in silico assessment of MAX variants of unknown significance. *J. Mol. Med.* **93,** 1247–1255 (2015).

43. Hurlin, P. J. *et al.* Mad3 and Mad4: novel Max-interacting transcriptional repressors that suppress c-myc dependent transformation and are expressed during neural and epidermal differentiation. *EMBO J.* **14,** 5646–59 (1995).

44. Ayer, D. E. & Eisenman, R. N. A switch from Myc: Max to Mad: Max heterocomplexes accompanies monocyte/macrophage differentiation. *Genes Dev.* **7,** 2110–2119 (1993).

45. Gunther, H. *NMR Spectroscopy: Basic Principles, Concepts and Applications in Chemistry*. (2013).

46. Keeler, J. Understanding NMR Spectroscopy. *SpringerReference* 211 (2002). doi:10.1007/SpringerReference_67582

47. Rule, G. S. & Hitchens, T. K. *Fundamentals of Protein NMR Spectroscopy*. **5,** (2006).

48. Mandal, P. K. & Majumdar, A. A comprehensive discussion of HSQC and HMQC pulse sequences. *Concepts Magn. Reson. Part A Bridg. Educ. Res.* **20,** 1–23 (2004).

49. Jacobsen, N. E. *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology* (2007). doi:10.1002/9780470173350

50. Jensen, M. R., Petersen, G., Lauritzen, C., Pedersen, J. & Led, J. J. Metal binding sites in proteins: Identification and characterization by paramagnetic NMR relaxation. *Biochemistry* **44,** 11014–11023 (2005).

51. Otting, G. Protein NMR using paramagnetic ions. *Annu. Rev. Biophys.* **39,** 387–405 (2010).

52. Clore, G. M., Tang, C. & Iwahara, J. Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement. *Curr. Opin. Struct. Biol.* **17,** 603–616 (2007).

53. Kurzbach, D., Vanas, A., Flamm, A. G., Tarnoczi, N. & Kontaxis, G. Detection of correlated conformational fluctuations in intrinsically disordered proteins through paramagnetic relaxation interference †. 5753–5758 (2016). doi:10.1039/C5CP04858C

54. Bouvignies, G., Meier, S., Grzesiek, S. & Blackledge, M. Ultrahigh-Resolution Backbone Structure of Perdeuterated Protein GB1 Using Residual Dipolar Couplings from Two Alignment Media. *Angew. Chemie Int. Ed.* **45,** 8166–8169 (2006).

55. Clifton, M. . *et al.* Structural Studies of Human Siderocalin. *To be Publ.* doi:10.2210/PDB3CMP/PDB

56. Shapiro, Y. E. & Meirovitch, E. Slowly Relaxing Local Structure (SRLS) Analysis of 15 N H Relaxation from the Prototypical Small Proteins GB1 and GB3. (2012).