

# **Structural and Functional Studies of Growth Arrest and DNA-Damage Proteins**

## **Inaugural – Dissertation**

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

**Sarbani Bhattacharya**

from Chandannagore, West Bengal, India

April, 2009

1<sup>st</sup> Reviewer: Prof. Dr. Udo Heinemann

2<sup>nd</sup> Reviewer: Prof. Dr. Ferdinand Hucho

Date of Defence: 25.06.09

# **SELBSTSTÄNDIGKEITSERKLÄRUNG**

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe.

Berlin, den 06 April 2009

Sarbani Bhattacharya

**Success is not final, failure is not fatal: it is the courage to continue that counts**

**-Winston Churchill**

## ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Dr. Udo Heinemann who gave me the opportunity to work in the fascinating field of crystallography. I appreciate his support, ideas and scientific discussion which have immensely contributed to this work. His knowledge, patience, vision and guidance have provided me with life-time benefits.

My special thanks to Dr. Oliver Daumke for lending me his help whenever needed and for critical revision of this manuscript.

I am grateful to Dr. Klaas Max for his help, suggestions and ideas both in my scientific life and outside laboratory. Thank you for everything.

I would like to take this opportunity to thank Dr. Juergen-Joachim Müller who taught me the basics of structure solving, Dr. Yvette Roske for her help with crystallization, Dr. Eva Müller and Dr. Albrecht Otto for all the mass spectrometry analysis, Prof. Dr. Joachim Behlke for Analytical Ultracentrifugation service, Dr. Dmitri Svergun and Dr. Haydn Mertens for SAXS data collection and analysis.

Dr. Anja Schütz, I would like to thank for her help, ideas and constructive criticism.

I feel lucky to be a part of Heinemann group and would like to thank all the present and past members who kept the working environment not only congenial but also stimulating and helped me in innumerable way. I acknowledge the support, both scientifically and as friends from members – Kerstin Böhm, Bettina König, Harald Striegl, Chengcheng Wang, Jennifer Hanna, Claudia Maria Haas and Ulrich Gohlke.

My warm gratitudes to Anette for her technical help and to our secretary Birgit who helped me in countless occasions. Special thanks are due to Andreas Knespel whose friendship I would cherish for ever.

I can not thank enough Marlene Feichtinger who made my stay in a foreign land, far away from home pleasant and memorable. Thank you for sticking with me during my need.

I would like to thank Dr. Shreemanta K. Parida, Dr. Mayuri Pandu and Subhamay Bhattacharya for patiently bearing my occasional bouts of frustration and their encouragement and support. Thanks to Reshmi Mukherjee, Siuli Banerjee and Debjani Goswami for their belief in me.

Words are not enough to thank Dr. L. Ravi Kumar without whose substantial help, support, patience and encouragement I would have been nowhere.

I would also like to thank MDC for providing funding and resources throughout this research.

Special thanks to my brother Anirban Bhattacharya for his technical help and for just being there. Last but not the least; I would like to thank my parents for making everything worthwhile.

Thank you all!!

## ZUSAMMENFASSUNG

Die Expression von Genen, die für sogenannte growth-arrest and DNA-damage-inducible 45 (GADD45) Proteinen codieren, kann durch verschiedenen genotoxischen und nicht-genotoxischen Stress induziert werden. Humanes GADD45 spielt eine wichtige Rolle beim Zellwachstum und Proliferationsregulierung mit spezifischen Funktionen in der Zellzykluskontrolle, MAPK Signaltransduktion, Apoptose und Immunantwort.

Bei Säugetieren werden drei Arten von GADD45 unterschieden:  $\alpha$ ,  $\beta$  und  $\gamma$ . Diese drei homologen Proteine sind hauptsächlich im Zellkern zu finden und zeigen eine allgemeine antiproliferative Aktivität. Unterschiede finden sich in ihrem Expressionsmuster und dem Aktivierungsmechanismus. So kann GADD45 $\alpha$  bei DNS-Schäden den Verlauf des Zellzykluses beim Übergang der G2- zur M-Phase arretieren, wohingegen GADD45 $\gamma$  ausschließlich durch IL-2 induziert wird. Außerdem wird die negative Wachstumskontrolle durch physische Interaktion mit dem proliferating-cell nuclear antigen (PCNA) vermittelt. Aufgrund dessen wird GADD45 als ein mögliches therapeutisches Ziel für anaplastischen Schilddrüsenkrebs angesehen. Alle drei GADD45 Proteine weisen bedeutende LXXLL-Motive (X steht für eine beliebige Aminosäure) auf, die wahrscheinlich für die Bindung an Hormonrezeptoren benötigt werden.

Es war bis vor kurzem keinerlei strukturelle Information dieser Proteine verfügbar. In der vorliegenden Arbeit habe ich die Struktur und Funktion dieser für die Zellzykluskontrolle und Autoimmunsuppression wichtigen Gruppe von Proteinen, insbesondere von GADD45 $\gamma$ , charakterisiert. Ich habe die Kristallstruktur des humanen GADD45 $\gamma$  mit einer  $\alpha/\beta$ -plait Topologie bestimmt. Übereinstimmend mit den Ergebnissen aus der analytischen Ultrazentrifugation und analytischen Gelfiltration wurde auch im Kristallgitter ein Homodimer gefunden. Trotz geringer Sequenzähnlichkeit ähnelt die Struktur derer von Proteinen der großen ribosomalen Untereinheit, insbesondere L7Ae, und rechtfertigt die Klassifikation von GADD45 zu der L7Ae/L30e/S12e ribosomalen Proteinfamilie.

Aufgrund der Kristallpackung ist die Bildung von Dimeren über unterschiedliche Grenzflächen möglich. Mittels Mutagenese, Kleinwinkelröntgenstreuung und Fluoreszenzspektroskopie, habe ich die verschiedenen Möglichkeiten Dimerisierung von GADD45 $\gamma$  eingehender untersucht.

Erstaunlicherweise befinden sich alle drei LXXLL-Motive von GADD45 $\gamma$  im Inneren des Proteins und sind für eine mögliche Ligandenbindung, beispielsweise von Kernrezeptoren, nicht

zugänglich. Eine direkte Bindung von GADD45 $\gamma$  mit der Liganden-bindenden Domäne des Kernrezeptors PPAR $\gamma$  konnte nichtsdestotrotz mittels analytischer Ultrazentrifugation und Interaktionsstudien (pull-down assays) gezeigt werden. Wie bereits aus der Struktur ersichtlich zeigt auch die Mutagenese der Leucinreste des LXXLL-Motivs, dass diese Motive nicht für die Rezeptorbindung benötigt werden. Aufgrund dessen ist es sehr wahrscheinlich, dass andere Seitenketten für die Interaktion von GADD45 $\gamma$  mit Kernrezeptoren wie PPAR $\gamma$  verantwortlich sind. Das stöchiometrische Verhältnis von 2:1 (PPAR $\gamma$ : GADD45 $\gamma$ ) wurde mit Hilfe analytischer Ultrazentrifugation ermittelt. Diese Komplexbildung konnte allerdings in anderen biochemischen Experimenten nicht bestätigt werden. Bei dem Versuch der Co-Kristallisation wurde ausschließlich PPAR $\gamma$ -Kristalle erhalten. Vermutlich handelt es sich um eine schwache und/oder transiente Interaktion oder diese wird durch weitere, bislang unbekannte, Faktoren beeinflusst. Aufgrund der erfolgreichen Strukturbestimmung von GADD45 $\gamma$  versuchte ich darauffolgend GADD45 $\alpha$  zu kristallisieren.

Um geeignete Ausgangsbedingungen zu finden, wurde mit Hilfe eines Thermofluor Assays eine Vielzahl von Pufferlösungen und pH-Werten auf eine stabilisierende Wirkung von GADD45 $\alpha$  getestet. Trotz günstiger biophysikalischer Eigenschaften und einer großen Anzahl von stabilisierenden Pufferlösungen, pH-Werten und Additiven konnte keine geeignete Kristallisationsbedingung gefunden werden. Durch limitierende Proteolyse und in silico-Analyse wurde versucht, ein kristallisierbares Konstrukt zu identifizieren. Auch die, aufgrund eines Sequenzvergleichs mit GADD45 $\gamma$  naheliegende Entfernung des N-Terminus, verbesserte die Kristallisierbarkeit nicht. Vermutlich wurde die Kristallisation durch hochmolekulare Aggregate und/oder die Hydrophobizität verhindert. Auch bei dem Versuch einen Komplex aus GADD45 $\alpha$  und PCNA zu kristallisieren, wurden keine Kristalle mit geeigneten Diffraktionseigenschaften gefunden. Obwohl die Komplexbildung in vitro gezeigt werden konnte, wurden ausschließlich PCNA-Kristalle erhalten. Zukünftige strukturelle Studien dieses Proteins können jedoch auf der vorliegenden Arbeit aufbauen.

## ABSTRACT

The expression of genes encoding growth-arrest and DNA-damage-inducible 45 (GADD45) proteins can be induced by various genotoxic and non-genotoxic stresses. The human GADD45 proteins play an important role in cell growth and proliferation regulation with specific functions in cell-cycle control, MAPK signaling, apoptosis, and immune response.

There are three mammalian GADD45 proteins ( $\alpha$ ,  $\beta$ , and  $\gamma$ ). All are primarily localized in the nucleus, display general antiproliferative activity and are highly homologous to each other. They differ in the expression pattern and in their activation mechanism. To highlight differences, it is GADD45 $\alpha$  which can halt the cell-cycle progression at G<sub>2</sub>/M transition upon DNA damage while GADD45 $\gamma$  is the only family member induced by IL-2. Furthermore, negative growth control is mediated by physical interactions with the proliferating-cell nuclear antigen (PCNA). For this reason, it has been suggested as a possible therapeutic target for anaplastic thyroid cancer. Importantly, all three GADD45 proteins contain LXXLL motifs (X is any amino acid), which has been suggested to be required for hormone-receptor binding.

Until recently, there was no structural information on any of the GADD45 proteins. Being important members of cell-cycle regulatory cascades and tumor and autoimmune suppressors, I characterized the structure and function of this group of proteins, especially GADD45 $\gamma$ . I have determined the crystal structure of human GADD45 $\gamma$  which reveals an  $\alpha/\beta$ -plait topology. Moreover it forms a homodimer in the crystal which is in agreement with dimers in solution as confirmed by analytical ultracentrifugation and analytical gel filtration. Despite low sequence similarity the structure shows striking resemblance to large-subunit ribosomal proteins, especially L7Ae, and therefore justifies its classification under the L7Ae/L30e/S12e ribosomal protein superfamily.

Inspection of the crystal packing revealed more than one possible interaction of monomers that could generate the expected dimer. Using mutagenesis, small angle X-ray scattering and fluorescence spectroscopy, I tried to identify the correct dimer interface.

Surprisingly, the three LXXLL motifs present in the GADD45 $\gamma$  sequence and implicated in nuclear receptor binding are buried inside the dimer and inaccessible to ligand proteins. Nevertheless, a physical interaction of GADD45 $\gamma$  with the ligand-binding domain of nuclear receptor protein PPAR $\gamma$  was demonstrated by analytical ultracentrifugation as well as pull-down assays. Mutagenesis of leucines in the LXXLL motif supported the prediction based on the crystal



structure that this motif plays no role in nuclear receptor binding. Therefore, it is most likely that the interaction of GADD45 $\gamma$  with nuclear receptors like PPAR $\gamma$  is mediated by residues other than LXXLL. Using analytical ultracentrifugation, a binding stoichiometry of 2:1 (PPAR $\gamma$ : GADD45 $\gamma$ ) was also established. However, this complex formation was not observed in other biochemical experiments, and co-crystallization experiments yielded crystals containing only PPAR $\gamma$ , thereby suggesting that this interaction might be weak and/or transient, occurring at cellular level only or maybe stabilized by further factors currently unknown.

The successful crystal structure analysis of GADD45 $\gamma$  led me to attempt the crystallization of GADD45 $\alpha$ . Using a thermofluor assay, a wide array of buffers and pH conditions were screened for their stabilizing effect on GADD45 $\alpha$ . Although it displayed favorable biophysical properties this protein proved recalcitrant to crystallization under various buffer, pH, additive or redox conditions. Therefore, limited proteolysis and analysis with *in silico* tools were employed to identify crystallizable protein forms. The removal of the N-terminus as suggested by a structural alignment with GADD45 $\gamma$  did not render this protein amenable to crystallization. The presence of high-molecular-weight aggregates as well as the protein's hydrophobicity could explain the non-crystallizability of GADD45 $\alpha$ . In further experiments I attempted to co-crystallize GADD45 $\alpha$  with its binding partner PCNA without obtaining diffraction-quality crystals. Complex formation could be shown *in vitro*, but co-crystallization experiments yielded crystals that contained only PCNA. Future attempts to gain structural insight into this protein could benefit from this study.

# TABLE OF CONTENTS

## 1. Introduction

1.1	Genotoxic and non-genotoxic stresses resulting in DNA damage-----	1
1.2	Cellular response to stress and GADD45-----	2
1.3	Isoforms of GADD45-----	3
1.4	Mode of action: GADD45 in cellular stress responses-----	6
1.4.1	Cell-cycle arrest-----	6
1.4.2	Crosstalk with p21-----	8
1.4.3	DNA repair-----	8
1.4.4	Apoptosis-----	11
1.4.5	Coactivators of nuclear hormone receptors -----	11
1.5	Medical significance: GADD45 in tumor progression and cancer-----	13
1.6	High-resolution structure determination by X-ray crystallography-----	15
1.7	Crystallization-----	15
1.8	Principles of X-ray diffraction-----	18
1.8.1	Why X-ray?-----	18
1.8.2	Collection of X-ray diffraction data-----	18
1.8.3	Data processing-----	21
1.8.4	Structure factor equation-----	22
1.8.5	Data quality indicators-----	23
1.8.6	Calculation of the electron density-----	24
1.8.7	Solution of the crystallographic phase problem-----	25
1.8.8	Phase improvement-----	27
1.8.9	Model building-----	27
1.8.10	Refinement and model validation-----	28
1.9	Goals of this study-----	31

## 2. Materials----- 32

2.1	Chemicals-----	32
2.2	Enzymes-----	33
2.3	Kits-----	33
2.4	Columns-----	33
2.5	Strains, clones and plasmids-----	34
2.6	Media and buffers-----	35
2.7	Synthetic Oligonucleotides-----	38

## 3 Methods----- 40

3.1	Molecular biology methods-----	40
3.2	Biochemical methods-----	43
3.2.1	Protein production-----	43
3.2.2	Protein purification-----	44
3.2.3	Production of recombinant human GADD45 $\gamma$ -----	48
3.2.4	Protein analysis-----	50
3.3	Spectroscopic methods-----	53
3.3.1	Analytical ultracentrifugation (AUC)-----	53
3.3.2	CD spectroscopy-----	54
3.3.3	Dynamic light scattering (DLS)-----	56

3.3.4	Fluorescence spectroscopy-----	57
3.3.5	Small-angle X-ray scattering (SAXS)-----	58
3.3.6	Thermal shift assay-----	59
3.4	Crystallography methods-----	60
3.4.1	Crystallization and data collection-----	60
3.4.2	Crystal structure determination-----	61
4.	<b>The architecture of human GADD45 <math>\gamma</math></b> -----	62
4.1	<b>Results</b> -----	62
4.1.1	GADD45 $\gamma$ is a dimer in solution-----	62
4.1.2	GADD45 $\gamma$ is folded and predominantly helical-----	63
4.1.3	Determining the crystal structure of GADD45 $\gamma$ -----	64
4.1.4	The structure of GADD45 $\gamma$ : Monomer-----	66
4.1.5	Common fold-----	68
4.1.6	Dimerization interface(s)-----	70
4.1.7	Identification of the correct interface by mutagenesis-----	74
4.1.8	Analysis of local environment of GADD45 $\gamma$ by tryptophan fluorescence----	78
4.1.9	Identification of the correct dimer interface using SAXS-----	79
4.1.10	Binding to nuclear receptor is sterically blocked-----	82
4.1.11	Nuclear receptor binding confirmed by pull-down analysis and AUC-----	83
4.1.12	Role of LXXLL motifs in binding: Mutational analysis-----	85
4.1.13	Mutant binding is comparable to wild type-----	88
4.1.14	Crystallization of a GADD45 $\gamma$ /PPAR $\gamma$ complex -----	89
4.2	<b>Discussion</b> -----	91
4.2.1	Homologous structures-----	91
4.2.2	Dimerization is a common characteristic of GADD45 proteins-----	93
4.2.3	Homodimerization-----	93
4.2.4	Comparison with mouse interface(s)-----	95
4.2.5	Sequence conservation identifies interface 3 as the likely interface-----	96
4.2.6	Significance of dimerization involving smaller contact area-----	98
4.2.6	Implication of LXXLL motif inGADD45 $\gamma$ -----	99
5.	<b>Characterization of human GADD45<math>\alpha</math></b> -----	100
5.1	<b>Results</b> -----	100
5.1.1	Protein expression and purification-----	100
5.1.2	Attempts to crystallize GADD45 $\alpha$ -----	102
5.1.3	In search of stable core fragment-----	103
5.1.4	Utility of in silico tools-----	103
5.1.6	Generation and crystallization of N-terminally truncated GADD45 $\alpha$ -----	105
5.1.7	Co-crystallization with PCNA-----	106
5.2	<b>Discussion</b> -----	108
5.2.1	Self-association driven by hydrophobic interactions-----	108
5.2.2	Potential for GADD45 $\alpha$ :PCNA complex-----	109
6.	<b>References</b> -----	110
7.	<b>Abbreviations</b> -----	118
8.	<b>Appendix</b> -----	120

## Introduction

The maintenance of normal cellular homeostasis depends on how accurately cells decipher signals for cell-cycle progression, growth, terminal differentiation, and apoptosis. Regulation of cell cycle and growth is integral for cell survival, and the intricate mechanisms that control proliferation and cell cycle are numerous. Any aberration in this regulation machinery might lead to tumorigenesis. Defects in DNA-repair pathways are known to be an early step in oncogenesis, accelerating its progression [1-4]. Paradoxically, DNA-damage inducing agents that can cause tumors, such as X-ray irradiation and a variety of alkylating agents are frequently used in cancer therapy [5, 6]. In recent years it has become increasingly evident that both the molecular basis for the initial susceptibility of malignant cells to anti-cancer drugs and the development of treatment resistance originate from genetic lesions that alter the function of genes playing roles in normal cell homeostasis, notably in determining cell-cycle progression and apoptotic set points [5]. Thus, understanding the molecular genetic pathways that mediate negative growth control is of high priority from both a basic science and cancer therapeutic point of view.

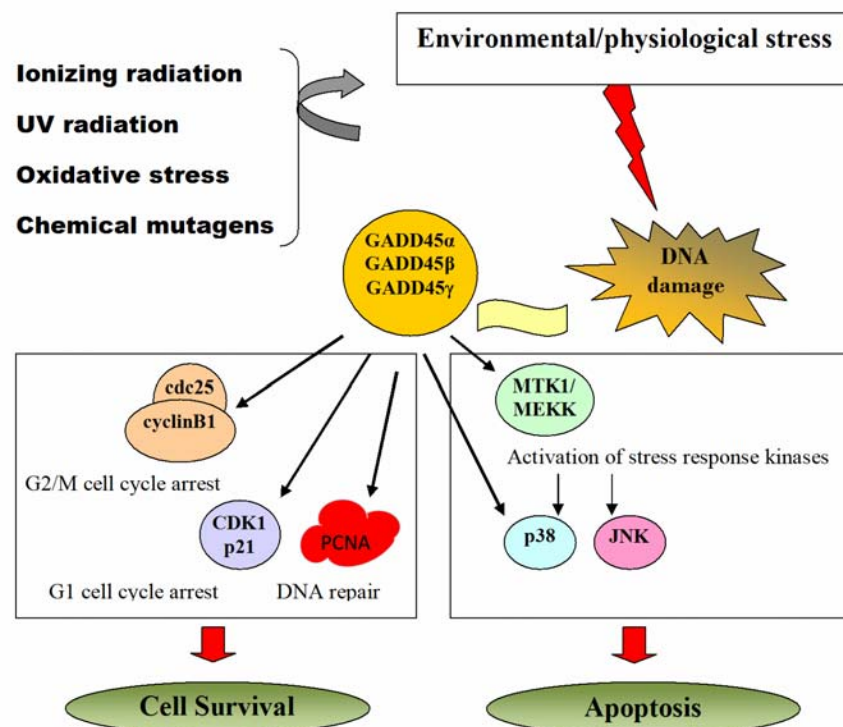
### 1.1 Genotoxic and non-genotoxic stresses resulting in DNA damage

Genotoxic stress, resulting in DNA damage, is an inescapable part of life. Throughout their lives cells are exposed to both endogenous and exogenous sources of DNA-damaging agents. Free radicals and peroxides generated during normal physiological processes and during inflammation constitute a source of endogenous DNA-damaging agents, while exposure to numerous chemical and physical agents from the environment is the exogenous source. Genes induced as a result of exposure to adverse conditions are dependent on the nature of the stress [7]. Heat stress, oxidative stress and DNA damage induce different sets of genes that are particular for each type of stress, but some of these genes can be induced by more than one agent. Both environmental agents as well as physiological processes involving activated oxygen species and other reactive agents can damage DNA [8] [9]. Two common environmental agents that cause cellular injury by damaging DNA are ionizing radiation, such as X-rays, and nonionizing UV radiation [8]. X-rays damage DNA primarily by generating hydroxyl radicals, also produced by oxidative stress, which alter the structure of bases, and the resulting DNA damage is repaired by various base-excision repair mechanisms []. Lesions caused by UV radiation are primarily cyclobutane pyrimidine dimers that are repaired by nucleotide-excision repair. Alkylating agents, such as methyl methanesulfonat

(MMS), represent a major class of chemical agents that damage DNA [10]. Damage caused by alkylating agents is repaired by either base-excision repair mechanisms as well, or by *in-situ* repair [8].

## 1.2 Cellular response to stress and GADD45

Cellular response following DNA damage includes the induction of genes that may have a role in protection, altered mutagenesis and growth inhibition [11]. In both bacteria and eukaryotes, induction of genes involved in cell-growth delays or cell-cycle checkpoints is a common response to DNA damage. Mammalian cells have evolved an elaborate defense mechanism to maintain genomic integrity under genotoxic stress. This includes activation of cell-cycle-arrest checkpoints at the G<sub>1</sub>/S and G<sub>2</sub>/M transitions [12-14] and/or activation of programmed cell death [5-6].



**Figure 1.1: Schematic diagram of the induction of *GADD45* and its subsequent effects.** A variety of environmental stress factors can lead to the upregulation of *GADD45* genes which in turn can activate cell-cycle arrest, DNA-excision repair and apoptosis by physically interacting with other cellular proteins. Modified from – Hoffman B, J Mol Signal (2008) 12, 3-15.

Whether a particular cell undergoes either cell-cycle arrest, to allow damaged DNA to be repaired, or rapid apoptosis depends on the extent of genomic damage and the cell type

For example, hematopoietic cells readily undergo apoptosis following exposure to doses of  $\gamma$ -irradiation which induce cell-cycle arrest in fibroblasts [15]. The p53 tumor suppressor pathway [16-18] and the p38/JNK kinase pathways [19-22] have been identified to play major roles as mediators of cellular response to genotoxic stress. However, the molecular/genetic circuitries of stress response pathways and their mode of action have not yet been fully determined. For example, how such pathways interact to signal either cell-cycle arrest or programmed cell death is still not clearly understood.

Of late, the growth-arrest and DNA-damage 45 (*GADD45*) family of genes and its encoded proteins have emerged as key players in cellular stress responses. *GADD45* genes have been implicated in stress signaling in response to physiological or environmental stressors, which results in cell-cycle arrest, DNA repair, cell survival or senescence, or apoptosis, and their role as stress sensors is mediated by a complex interplay of physical interactions with other cellular proteins that are involved in cell-cycle regulation and cellular response to stress (Fig. 1.1).

### 1.3 Isoforms of *GADD45*

GADD stands for Growth Arrest and DNA Damage. In mammals, the *GADD45* family of genes includes three isoforms, *GADD45 $\alpha$* , *GADD45 $\beta$* , and *GADD45 $\gamma$* , each of them encoding a small (~18 kDa), evolutionarily conserved protein. The *GADD45* proteins are highly homologous to each other (55-57% overall identity at the amino-acid level), highly acidic, and primarily, but not exclusively, localized within the cell nucleus [23, 24].

*GADD45 $\alpha$* , also termed *GADD45*, was originally identified on the basis of rapid induction in Chinese hamster ovarian cells by UV radiation [25]. *GADD45 $\beta$* , originally termed MyD118, was identified as an immediate-early responsive gene induced by IL-6 in the murine myeloid cell line M1 [26, 27]. *GADD45 $\gamma$* , also known as CR6, which was isolated using an MyD118 (*GADD45 $\beta$* ) cDNA fragment to screen for other potential members of the *GADD45* gene family was originally cloned as an immediate-early response gene in T cells stimulated by interleukin-2 [28].

#### **Isoform-specific characteristics**

The *GADD45* proteins appear to serve similar, but not identical, functions along different stress response pathways. Different levels of expression for each of the three *GADD45* genes is observed in multiple murine tissues including heart, brain, spleen, lung, liver, skeletal muscle, kidney and testes [25], yet knockout mice (-/-) for each of these genes are viable [29, 30]. It has

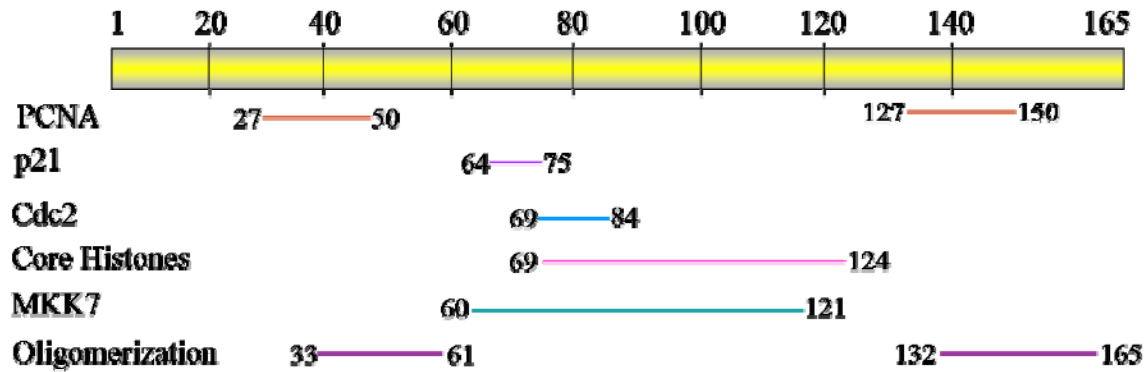
been found that individual members of the *GADD45* family are differentially induced by a variety of genotoxic and environmental stress agents [31-35] indicating that each gene is induced by a distinct subset of environmental stresses. Each *GADD45* gene has a distinctive pattern of expression [31, 36] consistent with each *GADD45* family member playing a different role in response to each source of stress. Consistent with the distinct expression patterns, regulation of expression for each *GADD45* gene is also distinct. For instance, *GADD45 $\beta$*  is a primary response gene to both IL-6 and TGF- $\beta$ , and *GADD45 $\gamma$*  is induced as a primary response to IL-2 and IL-6 whereas *GADD45 $\alpha$*  is the only member of the family whose up-regulation has a p53 component, although its induction can also be p53 independent [23, 28, 31, 37, 38]. Amongst the isoforms, there is specialization in terms of function. *GADD45 $\alpha$*  is the only member of the GADD gene family that is frequently inducible by ionizing radiation [39], and it has been shown that appreciable induction of *GADD $\alpha$*  by ionizing radiation occurred only in cells having a wt p53 phenotype [40]. *GADD45 $\alpha$* , but not *GADD45 $\beta$*  or *GADD45 $\gamma$* , is a direct downstream target of p53 [41]. *GADD45 $\alpha$*  inhibits Cdc2 kinase and can induce G<sub>2</sub>/M arrest, while neither *GADD45 $\beta$*  nor *GADD45 $\gamma$*  universally function in this manner. Therefore, to what extent the function of each of the GADD45 proteins is unique or overlaps with the functions of the other proteins, remains to be clarified.

### **Protein-protein interaction**

Despite the isoform-specific distinctions described above, all three GADD45 family members follow a common module of action. Overexpression of *GADD45* has been shown to induce cell-growth arrest [24] and apparently apoptosis [42]. Short-term transfection assays have revealed that the three GADD45 proteins can synergize in suppression of colony formation by several different human tumor cell lines [43, 44]. They also function to modulate tumor formation in response to oncogenic stress [45].

The physiological functions of GADD45 proteins are mediated by a complex interplay of physical interactions with other cellular proteins that are implicated in cell-cycle regulation and the cellular stress response including PCNA, cdk1, p21, MEKK4, and p38 (Fig. 1.2).

Proliferative-cell nuclear antigen (PCNA) is an indispensable component of the DNA replication/repair machinery [46-48], and it has been shown that all three GADD45 proteins interact with PCNA [43, 49-53]. This interaction is also suggested to impede UV-induced cellular apoptosis [51, 52].



**Figure 1.2:** Regions of GADD45 (shown here is GADD45 $\alpha$ ) implicated in interaction with other cellular proteins and self-association.

All GADD45 isoforms interact with the cyclin-dependent kinase inhibitor p21 (Cip1, Wip1, Cdkn1A) [43, 54], a particularly important cell-cycle regulator protein, and compete with p21 for binding to PCNA. Conversely, p21 blocks the ability of GADD45 to bind to PCNA. The role for this interaction in cell-cycle control remains to be elucidated.

Under physiological conditions, p21 and PCNA are found in quaternary complexes that also include cyclins and cyclin-dependent kinases (cdks). All three GADD45 proteins specifically interact with the cdc2 (cdk1)/cyclinB1 mitosis-promoting complex and inhibits its kinase activity, which is crucial for G<sub>2</sub>-M progression [55-58].

It has also been demonstrated that GADD45 proteins interact with MEKK4 which is an upstream activator of the p38/JNK pathway associated with apoptosis [32, 41, 59]. The role of GADD45s in repair may also be realized through their ability to bind histones and modify the accessibility of DNA on damaged chromatin [60]. GADD45 proteins can promote genome stability as they have a role in DNA demethylation [61].

The role(s) of GADD45 proteins in these fundamental cellular processes are supported by several observations made both *in vitro* and in a mouse model. Studies of the GADD45 $\alpha^{-/-}$  mouse show genomic instability [29], reduced nucleotide-excision repair as well as increased levels of mutations and chemically induced tumorigenesis [62]. These interactions and expression patterns therefore strongly suggest that GADD45 proteins have an active role in the maintenance of genomic stability. In the following the role of GADD45 proteins in response to their interacting partners are discussed in more detail.



## 1.4 Mode of action: GADD45 in cellular stress responses

### 1.4.1 Cell cycle arrest

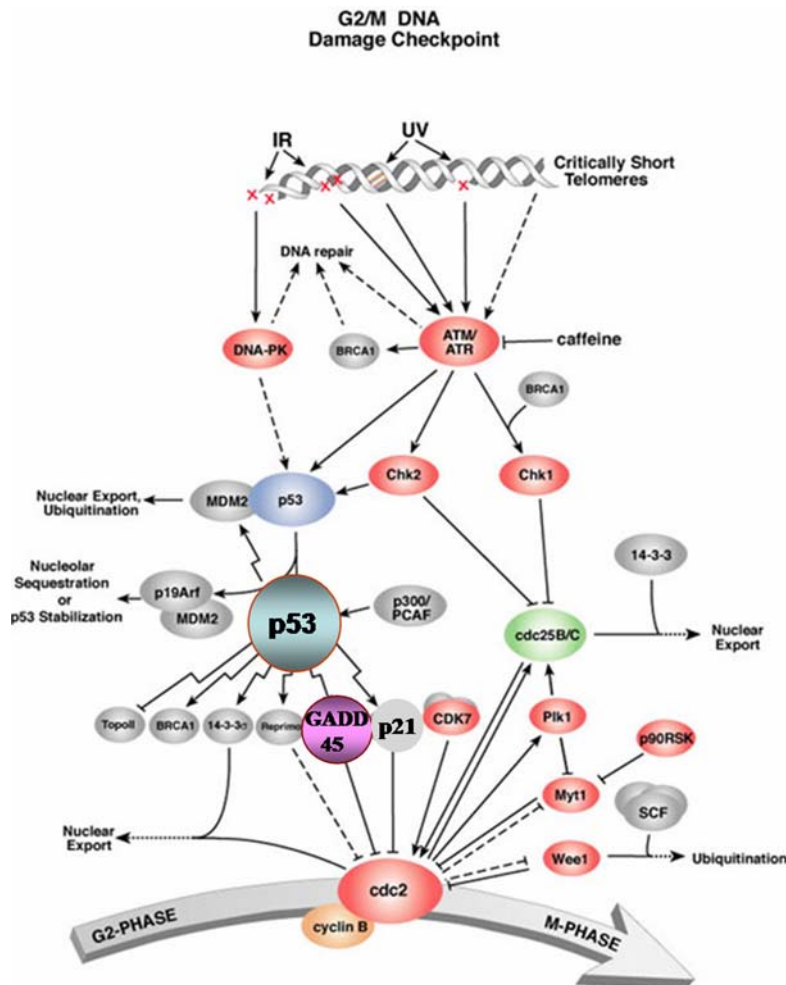
The cellular response to genotoxic stress in mammalian cells is complex and may involve multiple cellular processes, including cell-cycle checkpoint and apoptosis. These processes are thought to be the important biological events in maintaining genomic fidelity. Their inactivation would result in genomic instability, and in turn lead to cell transformation and tumorigenesis. Cell-cycle checkpoints are essential for the maintenance of genomic stability in response to DNA damage. Checkpoints provide a transient growth arrest in cell-cycle progression, presumably to allow time for the cells to repair damaged DNA before replication or mitosis occurs.

The cellular response to genotoxic stress involves cell-cycle arrest at G<sub>1</sub>-S and G<sub>2</sub>-M checkpoints [40, 63]. The p53 tumor-suppressor gene has been shown to play a critical role in cellular responses following DNA damage by controlling the cell cycle. Upon stress, p53 becomes transcriptionally active and up-regulates the transcription of downstream effector genes, which contain p53-recognition sites in their regulatory regions. Among the genes activated are *GADD45* [40, 64] and *p21* [65], each of which independently exhibits growth-suppressive activity. Induction of *GADD45* by ionizing radiation is dependent on the normal cellular p53 function via the p53-binding site located at the third intron of the *GADD45* gene [41, 64]. However, p53 is not required for the induction by non-ionizing radiation DNA-damaging agents such as UV radiation (UV) and methylmethane sulfonate (MMS), but it can contribute to these cellular responses.

Cell-cycle check points are regulated by several cyclins and cyclin dependent kinases (cdks). The transition from G<sub>2</sub> to M is regulated by the G<sub>2</sub>-specific kinase consisting of Cdc2 and cyclin B1 [12, 66]. After DNA damage is induced by either UV radiation or alkylating agents, GADD45, in a p53-dependent or independent manner, activates the G<sub>2</sub>/M checkpoint and halts the cell-cycle progression by directly inhibiting the kinase activity of the Cdc2-cyclin B1 complex by physically interacting with Cdc2, but not with cyclin B1 [29, 55] (Fig 1.3). This inhibition is specific, because GADD45 has no inhibitory effect on the kinase activity of the G<sub>1</sub>-specific Cdk2/cyclin E complex [55]. Experiments in multiple cell types have demonstrated that GADD45 proteins function in cell-cycle regulation. In human endothelial and fibroblast cells,

inhibiting endogenous *GADD45* expression by antisense *GADD45* impaired the G<sub>2</sub>/M checkpoint following exposure to either UV radiation or MMS [58, 67].

In addition, microinjecting a *GADD45α* expression vector into primary human fibroblasts



**Figure 1.3: Pathway showing *GADD45* activation and G<sub>2</sub>/M arrest.** [www.cellsignal.com/reference/pathway](http://www.cellsignal.com/reference/pathway)

In contrast, *GADD45γ* inhibits *cdc2/cyclinB1*-kinase activity without disrupting the complex [58]. Moreover, *GADD45α* shows a much greater inhibitory effect on *Cdc2/cyclin B1* kinase activity when compared to *GADD45β* and *GADD45γ* [68]. For *GADD45α*, it has been shown that the central region (residues 50-76) is required for binding to *cdc2* [56]. It appears that this region contains a unique acidic motif DEDDDR that may play a key role in the inhibition of *Cdc2-cyclin B1* kinase activity and in the induction of a G<sub>2</sub>/M arrest, because mutating the acidic residues to alanines abolished these activities. It will be interesting to examine whether one can exploit this acidic motif therapeutically to inactivate the G<sub>2</sub>/M checkpoint mediated by

arrested the cells at the G<sub>2</sub>/M boundary of the cell cycle [55]. This arrest was attenuated by the overexpression of cyclin B1 and *Cdc25C*. Therefore, these findings imply that *GADD45* growth suppression is mainly mediated via cell-cycle G<sub>2</sub>-M arrest which is modulated through inactivation of the *Cdc2-cyclin B1* kinase. However, not all three isoforms interact with and block *cdc2/cyclinB1* with equal vigor. Association of either *GADD45α* or *GADD45β* proteins with *cdc2/cyclinB1* results in dissociation of the complex which, in turn, inhibits *cdc2*-kinase activity [58].

GADD45. Such a strategy would potentially enhance cancer chemotherapy because inactivation of a G<sub>2</sub>/M checkpoint can sensitize cancer cells for DNA damage-induced apoptosis [69].

### 1.4.2 Crosstalk with p21

p21 is an universal kinase inhibitor of many cyclin-dependent kinases which is induced following cell exposure to genotoxic stresses, including ionizing radiation, MMS, and UV [10, 21]. The interaction of GADD45 with p21 appears to be interesting, as there are extensive similarities between *p21* and *GADD45* response to DNA damage. Both are induced by a wide spectrum of DNA-damaging agents and exhibit p53-dependent and -independent regulation. Both cause growth suppression and cell-cycle arrest when overexpressed [4, 5, 11, 13, 21]. However, unlike p21 which is responsible for the control of the G<sub>1</sub>-S checkpoint [19, 20] by binding to cyclin E/Cdk2 kinase and inhibiting the kinase activity [24-26], the GADD45 inhibition of kinase activity has been shown to be specific for the Cdc2/cyclin B1 complex which “drives” the cell from G<sub>2</sub> to mitosis and has been implicated in the control of cell cycle G<sub>2</sub>-M arrest [16].

The central region spanning amino acids 65–84 of the GADD45 $\alpha$  has been shown to be required for p21 binding. Interestingly, this region also encodes the Cdc2-binding activity. Therefore the central region of the GADD45 protein might serve as an important core for the interaction of GADD45 with other cell-cycle regulators, and those interactions can delicately coordinate the regulation of cell-cycle checkpoints.

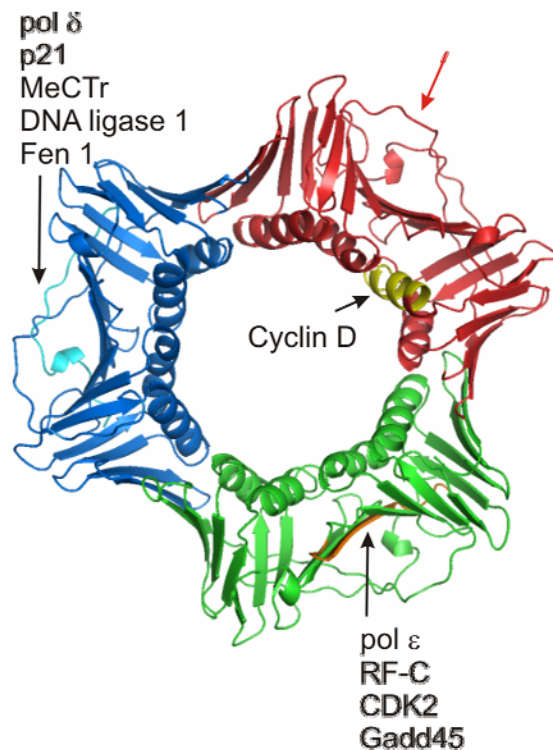
At present, the biological significance of GADD45 interaction with p21 remains unclear and requires further investigation. However, the interaction between p21 and GADD45 suggests a cross-talk between these two important cell-cycle regulators. Induction of both *p21* and *GADD45* following DNA damage is probably required for the initiation of cell-cycle G<sub>1</sub>-S and G<sub>2</sub>-M checkpoints. The establishment of these checkpoints would ensure the completion of DNA repair before cells enter S phase for DNA replication and mitosis for chromosomal segregation.

### 1.4.3 DNA repair

Interaction of the GADD45 proteins with PCNA promotes nucleotide excision repair (NER) of DNA [51, 52, 70] and thus delays cell-cycle progression when damage is detected. This reaction has been shown to rescue the cells from apoptosis. When mutants of GADD45 that lack the PCNA-interaction domain are ectopically expressed, they induce apoptosis more efficiently [51]. This led to the conclusion that the interaction of GADD45 proteins with PCNA triggers negative

growth control. PCNA was originally characterized as a DNA sliding clamp for replicative DNA polymerases which performs the essential function of providing replicative polymerases with the high processivity required to duplicate an entire genome. Because of its ability to interact with multiple partners from several metabolic pathways, PCNA plays a key role in controlling several reactions through coordination and organization of different partners [71]. It is a component of multiple quaternary complexes, which include the cycling CDKs and the CDK inhibitor p21<sup>WAF1/CIP1</sup> [72-74] that play a pivotal role in cell-cycle regulation [72, 75, 76], DNA replication [77], and repair of damaged DNA [78].

The structure of PCNA shows that it forms a ring-shaped trimeric complex which can encircle double-stranded DNA and move freely along it, providing a sliding platform to mediate protein-DNA interactions (Fig. 1.4).



**Figure 1.4: The crystal structure of human DNA clamp, a trimer of the protein PCNA.** In the assembled replication complex, the clamp encircles the DNA, which passes through the central pore. The monomers are colored red, green and blue, respectively. The interdomain connecting loop of one monomer is shown by a red arrow. The main regions of the PCNA ring involved in protein-protein interactions are highlighted along with the relevant partners. These are: (1) the interdomain connecting loop (cyan); (2) the inner side  $\alpha$ -helices at the N-terminus (yellow); and (3) the C-terminal tail (orange). For clarity, only one region per each monomer is highlighted. – modified from Hubscher U, J Cell Sci 2003 116: 3061-3068

The binding sites on PCNA for many of its partners have been mapped. The interdomain connecting loop (residues 121-132), was implicated as the major interaction site for many proteins including p21 [79]. PCNA interacts with both p21 and GADD45, and it has been shown that p21 and GADD45 compete for PCNA binding [80]. This picture is further complicated by the ability of GADD45 to interact with p21. The biological significance of these mutually exclusive interactions involving PCNA, p21 and GADD45 remains to be elucidated.

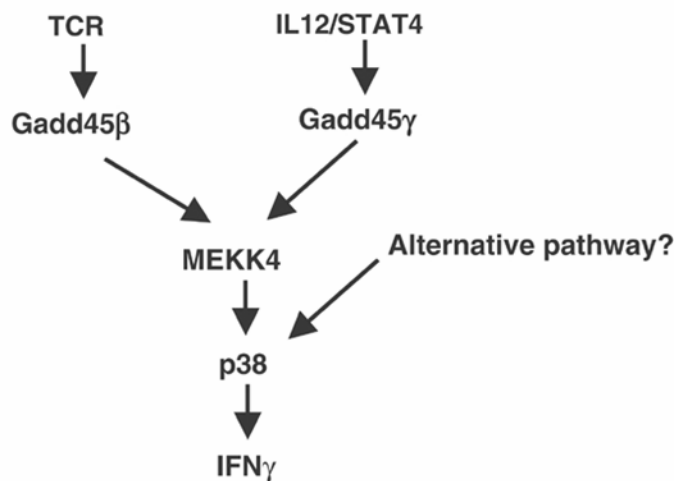
The interaction with PCNA is one of the best documented cellular interactions of GADD45 proteins. Several studies have mapped the interacting regions of both proteins using deletion mutants and peptide binding and have suggested that the stoichiometry of the GADD45:PCNA interaction is 2:1. Although some consistencies emerge from these different investigations, contradictions exist regarding the regions of GADD45 that are proposed to be required for PCNA binding [51-53]. On one hand, the N-terminus of GADD45 has been suggested to be involved in PCNA binding because PCNA-interacting proteins such as p21 have the motif QXXhXXaa (where h indicates a hydrophobic amino acid and a indicates an aromatic amino acid), and GADD45 $\gamma$  contains the sequence QGCLTAGVY (36-44) which is similar to the PCNA-binding motif [81]. This motif is well conserved in GADD45 $\alpha$  and  $\beta$ , too. It was reported that the N-terminal half of GADD45 $\alpha$  which contains this motif is sufficient for PCNA binding and that GADD45 binding resides on the C-terminus of PCNA [53] (Fig. 1.3). These results therefore suggest that binding of the GADD45-family members to PCNA is mediated by the N-terminal conserved motif around 30-45.

On the other hand, using yeast-two-hybrid (YTH) analysis, *in vitro* transcription/translation and transient expression in various cells, PCNA-interacting domains have been localized to the C-terminus of GADD45 isoforms (residues 95-165, 114-156 and 76-159 for GADD45 $\alpha$ ,  $\beta$  and  $\gamma$ , respectively), and GADD45 binding was mapped on the N-terminal (residues 1-46) and middle region of PCNA involving the interdomain connecting loop [51-52]. Structural elucidation is therefore necessary to solve this ambiguity.

The interaction of GADD45 and p21 which are part of the p53-response pathway with PCNA may represent a link between the DNA damage response and the regulation of DNA replication and repair. In response to DNA damage, cells will normally arrest cell-cycle progression to allow DNA repair to take place. If the mechanisms co-ordinating these events fail, the resulting loss of genetic integrity can lead to the accumulation of mutations that will disrupt the normal growth and cell-cycle controls.

### 1.4.4 Apoptosis

GADD45 proteins have a pro-apoptotic function. It was observed that blocking *GADD45 $\beta$*  by antisense expression in M1 myeloblastic leukemia cells impaired TGF $\beta$ -induced cell death, thereby



**Figure 1.5:** Diagram for the regulation of GADD45/MEKK4-mediated p38 apoptotic pathway in CD4<sup>+</sup> T cells – Author, X., EMBO J (2004) 23, 1576-1586

by implicating GADD45 $\beta$  as a positive modulator of TGF $\beta$ -induced apoptosis [37, 38]. TGF $\beta$ -induced apoptosis is mediated by GADD45 $\beta$  via p38 activation in primary hepatocytes from wild-type mice and was blocked in hepatocytes from GADD45 $\beta$ <sup>-/-</sup> mice [38]. Furthermore, ectopic expression of all three GADD45 genes was shown to induce apoptosis in HeLa cells [32]. P38 and JNK are stress-responsive MAP-kinase pathways which are involved in cell-cycle regulation and apoptosis [82].

They are activated by various anti-mitogenic stimuli such as environmental or genotoxic stress, TGF $\beta$ , and the inflammatory cytokines TNF $\alpha$  and interleukin-1 (IL-1) [83, 84]. P38 and JNK are activated by signaling cascades involving a MAPK kinase (MAPKK) which is responsible for phosphorylation of appropriate MAPK, and a MAPK kinase kinase (MAPKKK/MAP3K) that phosphorylates and activates MAPKK [85].

A human MAP3K, MTK1 (mouse MEKK4), is an upstream activator of both p38 and JNK pathways. It has been found that all three GADD45 proteins bind to an N-terminal domain of MTK1, and the resulting stimulation of its kinase activity has been implicated in stress-induced p38 and JNK activation [32] and the production of a signature cytokine, IFN $\gamma$  [86] (Fig. 1.5). That might be the reason, why T-helper type-1 (Th1)-effector cells deficient in GADD45 $\gamma$  show reduced IFN $\gamma$  production and p38/JNK activation.

### 1.4.6 Coactivators of nuclear hormone receptors

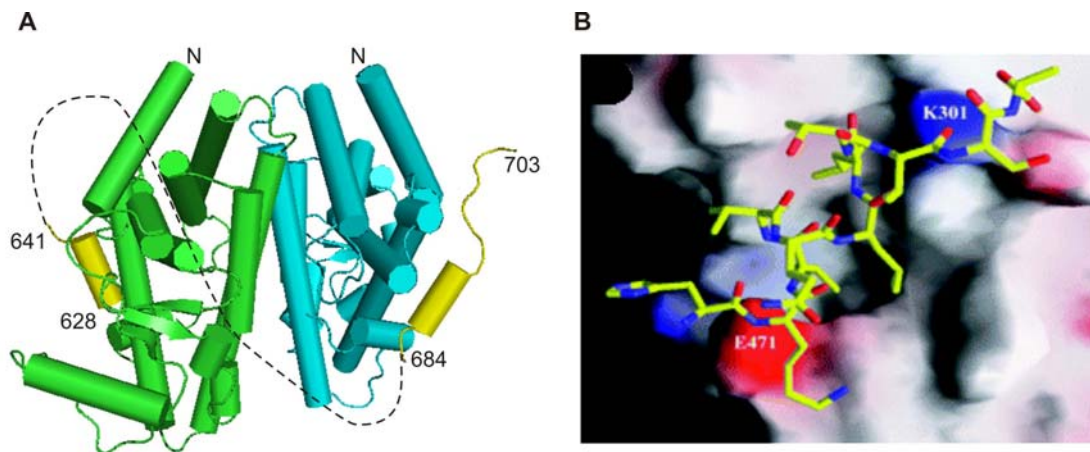
Nuclear hormone receptors (NR) are ligand-activated transcription factors that regulate diverse biological processes by stimulating or repressing target gene expression [87]. More than half of

the gene family's members are 'orphan' receptors i.e have not been associated with physiologically relevant ligands like steroids, thyroxine or retinoids. However, several essential fatty acids, oxidized lipids and prostaglandin derivatives can bind and activate the receptors.

The peroxisome proliferator-activated receptors (PPAR) are well known examples of NRs which function as key regulators of glucose, lipid and cholesterol metabolism. In human three subtypes,  $\alpha$ ,  $\gamma$  and  $\delta$ , have been identified, each with specific roles and tissue distribution [88] and associated with selective ligands [89, 90]. PPAR $\gamma$ , which is the best studied member of the PPARs, is expressed at high levels in adipose tissues and macrophages and can strongly induce adipogenesis [91-93].

NRs contain a central DNA-binding domain and a C-terminal ligand-binding domain (LBD) which mediates dimerization and transactivation functions. Upon binding ligands, NRs act as dimeric transcription factors to activate or repress target gene expression by binding to specific DNA sequences termed hormone-response elements [94]. All nuclear receptors are known to require the ligand-dependent recruitment of co-activator proteins to effectively stimulate gene transcription [95, 96]. The steroid receptor co-activating factor-1/SRC-1 is one of the best studied of the co-activators. The nuclear-receptor-interaction domain of all known and studied co-activators is highly conserved and known to contain three repeated motifs of the consensus sequence LXXLL (where X is any amino acid), each of which is sufficient for ligand-dependent interaction with several nuclear receptors [97, 98].

Indeed the crystal structure of the ternary complex containing PPAR $\gamma$ -LBD along with its ligand rosiglitazone and co-activator SRC-1 shows that glutamate and lysine residues that are highly conserved in LBDs of NRs form a 'charge clamp' with the backbone atoms of the LXXLL helices of SRC-1 especially the terminal leucines [99] (Fig. 1.6). This result shows that LXXLL motifs are vital in mediating cooperative assembly of co-activator complexes on homodimeric and heterodimeric nuclear receptors.



**Figure 1.6:** **A)** A schematic drawing of the PPAR $\gamma$ -LBD dimer and SRC-1, including the ligand rosiglitazone (in a space-filling representation). The two PPAR $\gamma$  monomers are blue and green, and the two SRC-1 interacting helices are yellow. Amino acids 642–669 of SRC-1 which are not observed in the crystal structure due to disorder are shown as a dashed line. **B)** The binding of SRC-1 (amino acids 628–642) to the LXXLL-binding site of PPAR $\gamma$ -LBD. SRC-1 is shown as yellow sticks on an electrostatic surface of PPAR $\gamma$  showing the coactivator-binding site. E471 and K301 side chains of PPAR $\gamma$ -LBD contribute to the red (negative) and blue (positive) charges on the surface of the coactivator-binding site and form the charge clamp with the LXXLL binding motif. – Milburn, VM, *Nature* **395**, September 1998, 137-143

It has been experimentally shown that all three GADD45 family proteins interact with several nuclear-hormone receptors including PPARs *in vitro* [100]. Moreover, each *GADD45* gene was found to contain transcription modulation activity for NRs which is a common feature of many known coactivators [110]. Sequence analysis reveals that all GADD45 proteins contain at least one complete and two incomplete LXXLL motifs. Because this short sequence motif, present in the co-activators, is necessary to mediate the binding of these proteins to nuclear-hormone receptors, it is very likely that GADD45 proteins bind to and act as co-activators of nuclear receptors using this motif.

## 1.5 Medical significance: GADD45 in tumor progression and cancer

Evidence was obtained that GADD45 proteins play a role in modulation of tumor development. GADD45 $\alpha^{-/-}$  and GADD45 $\beta^{-/-}$  mice displayed increased mutation frequency and susceptibility to ionizing radiation and chemical carcinogenesis. Loss of GADD45 $\alpha$  accelerates Ras-driven mammary tumor formation which is linked to the decrease in JNK activation and a decrease in senescence, correlated with a decrease in p38-kinase activation [45]. Although GADD45 genes are not frequently mutated in a variety of neoplasms, mutations are often found in invasive



pancreatic ductal carcinomas [41]. Moreover, there is a correlation of survival in patients with this cancer when mutation of *GADD45α* is combined with p53 mutant status. In addition, *GADD45α* expression is reduced in a number of cancer types, with promoter methylation in the majority of breast cancers resulting in reduced expression when compared to normal breast epithelium [42]. Additional to the regulation by p53, *GADD45* is known to be regulated by other tumor suppressor genes and oncogenes that have roles in genome stability. For instance, *Brcal* which is mutated in breast and ovarian cancer causes up-regulation of *GADD45α* via p53. Increased expression of the proto-oncogene *c-myc*, which is a known proliferation signal in tumor cells, leads to decreased *GADD45α* expression and stimulates proliferation of damaged cells [101]. Furthermore, NFκB-mediated repression of both *GADD45α* and *GADD45γ* is essential for escape from programmed cell death [42, 112], and it has been found that up-regulated expression of NFκB, also down-regulates *GADD45α* in many tumor types, such as prostate cancer [102]. *GADD45γ* is also a powerful growth suppressor controlling pituitary cell proliferation, and it represents the first identified gene whose expression is lost in the majority of pituitary tumors [103]. Its expression is also reduced in anaplastic thyroid cancer where its re-expression results in apoptosis [104]. *GADD45α* up-regulation serves as a marker for glioblastoma and some forms of pancreatic cancers [105]. Interestingly, increased basal *GADD45α* levels have been found to correlate with an increased survival rate for both of these tumor types [115]. These results suggest that a diagnosis of cancer may be assessed by examining particular cell-cycle markers such as *GADD45α* and in the future may be useful for determining prognosis.

## 1.6 High resolution structure determination by X-ray crystallography

The knowledge of accurate molecular structures is a prerequisite for rational drug design and for structure-based functional studies to aid the development of effective therapeutic agents and drugs. Crystallography can reliably provide the answer to many structure-related questions, from global folds to atomic details of bonding and has emerged as the most powerful and accurate method for determining protein structures. It exploits the fact that X-rays are diffracted by crystals. X-rays have an appropriate wavelength (in the Ångström range,  $1 \text{ Å} = 10^{-8} \text{ cm}$ ) to be scattered by the electron cloud of an atom of comparable size. Based on the diffraction pattern obtained from X-ray scattering off the periodic assembly of molecules or atoms in the crystal, the electron density can be reconstructed. Additional phase information must be extracted either from the diffraction data or from supplementing diffraction experiments to complete the reconstruction (the phase problem in crystallography). A model is then progressively built into the experimental electron density, refined against the data resulting in a quite accurate structural model.

Each crystallographic protein structure analysis requires several important steps. Amongst these i) the high amounts of highly pure and soluble protein required, ii) the crystallization itself that might take several months, iii) the diffracting quality of the crystals, and iv) solving the phase problem are crucial and limiting steps. However, over the past decade, automation in X-ray crystallography concerning protein expression and purification, high-throughput micro-crystallization and documentation, crystal shipping to high-end synchrotron radiation beamlines and automated data collection from the home source have tremendously accelerated the tedious and time-consuming process of X-ray crystallography. To date, about 85% of all protein structures deposited in the Protein Data Bank (PDB) are solved by this method.

In the present study the crystal structure of GADD45 $\gamma$  (PDB id - 2wal) was solved and will be reported. In this chapter, a general introduction into structure analysis by X-ray crystallography is given. For more information refer to common textbook [106].

## 1.7 Crystallization

### Why crystals?

X-ray scattering from a single molecule is incredibly weak and at present impossible to detect above the noise level of the experimental data, which would include scattering from air and water. A crystal arranges huge number of molecules in a defined orientation, so that scattered

waves can add up in phase and raise the signal to a measurable level. In a sense, a crystal acts as an amplifier. Therefore, the crucial part of any protein crystallography studies is protein crystallization. Despite very wide knowledge about protein crystallization it is still impossible to predict any conditions for protein crystallization.

## Growing the crystal

The principle of crystallization is to slowly reduce protein solubility by increasing the precipitant and protein concentrations. As a result, individual protein molecules align themselves in a

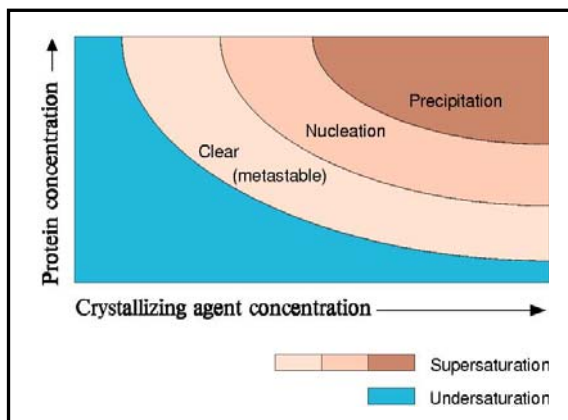


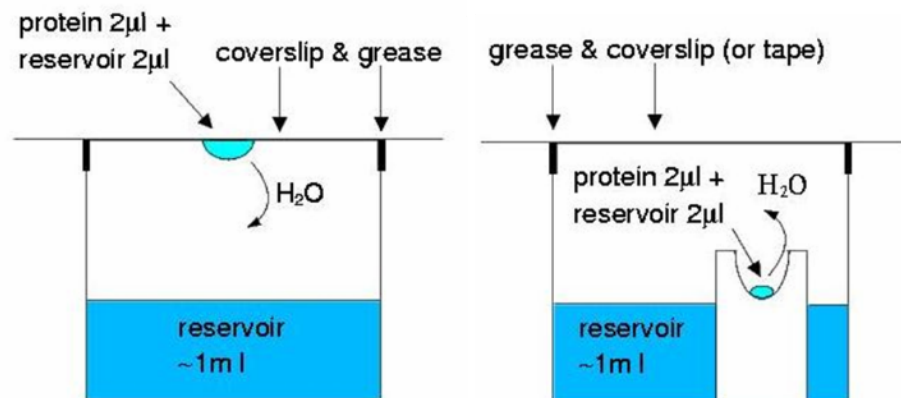
Figure 1.7: Phase diagram of protein crystallization.

Airlie J McCoy, University of Cambridge.

repeating series of "unit cells" by adopting a consistent orientation. Under the right condition a super-saturated state can be reached, where few nuclei form and grow to become crystals. If the conditions are not optimal, the protein will precipitate or form microcrystals at too high concentrations, or the solution may stay clear if protein and precipitant concentration are too low. This can be illustrated with a phase diagram (Fig. 1.7).

Depending on protein and precipitant concentration the system is in an undersaturated or a supersaturated state. Crystals will not grow but dissolve in the undersaturated region. For supersaturation it can be differentiated between the labile state where nuclei can spontaneously form and grow and the metastable state where nuclei can grow but not form. Nucleation and crystal growth in the supersaturated phase are always in competition with amorphous precipitation, which will occur when protein and/or precipitant concentration are too high.

The most commonly used technique for growing crystals is the vapor-diffusion method which can again be divided into sitting-drop and hanging-drop crystallization. Both involve a droplet containing purified protein, buffer, and precipitant being allowed to equilibrate with a physically separated larger reservoir containing similar buffers and precipitants in higher concentrations in a closed system. The position of the drop is different in the two methods (Fig. 1.8).



**Figure 1.8:** Protein crystallization setups showing hanging-drop (left) and sitting drop (right) vapour diffusion methods.  
www.structmed.cimr.cam.ac.uk/

The hanging drop is mixed on a cover slide that is used upside down to cover the reservoir well. Silicon grease is used to seal the cover. The sitting drop is mixed in an elevated pit above the reservoir, and the well is sealed. Due to the lack of physical contact between the drop and the reservoir, the most volatile component (usually water) will evaporate from the drop and condense at the reservoir until equilibrium is reached. Hence, the volume of the droplet will be reduced and the concentration of the macromolecules and precipitating agents will slowly increase in the drop. Several parameters influence the formation of well-ordered crystals. These include homogeneity of the materials to be crystallized, purity (absence of other molecules which can impair crystal-lattice formation) as well as solvent conditions, precipitant concentration, pH, additives, ligands, cofactors, salts, precipitants and temperature. The most common precipitants are PEGs (polyethylene glycols), salts (especially ammonium sulfate, Na/K phosphate and sodium chloride) and organic reagents (mainly 2-methyl-2,4-pentanediol MPD). Sparse-matrix screens are the best way to screen the multi-dimensional crystallization space. These screens are formulated on the basis of precipitant solutions that have been successfully used in crystallization trials before and therefore more or less cover the whole range of crystallization conditions. Once initial crystallization conditions are found, a systematic optimization with grid screens is required, where each parameter is varied so that the best possible crystal can be grown.

### Freezing protein crystals

To collect high-resolution data, high-intensity X-ray synchrotron radiation is used which will cause radiation damage of the crystals. To avoid this, crystals are vitrified and cooled in a liquid nitrogen stream. This has following advantages: 1) A decrease in temperature also decreases the atomic motion and results in a better signal-to-noise ratio of the diffraction data; 2) At lower temperatures the speed of damaging reactions (e.g. decarboxylations of glutamates and

aspartates) inside the crystal which arise from ionizations of chemical groups and radical formation by X-rays is decreased. This results in a better consistency of diffraction data from different images.

However, vitrification in liquid nitrogen can lead to the formation of ice crystals which give rise to strong diffraction spots or rings overlapping with the protein diffraction pattern and reducing the amount of usable data. Since the vitrification process may even destroy the protein crystal, optimized cryo-conditions must be established. If the mother liquor the crystal has grown in is not suitable for vitrification, cryo-protectants (e.g. glycerol, low-molecular-weight PEGs, MPD) have to be added to the solution. The crystal has to be transferred into the cryo-solution and can then be frozen.

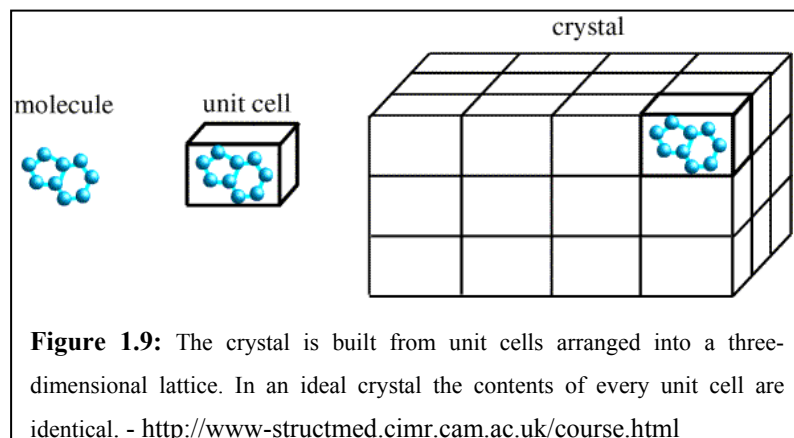
## 1.8 Principles of X-ray diffraction

### 1.8.1 Why X-rays

The use of electromagnetic radiation to visualise objects requires the radiation to have a wavelength comparable to the smallest features that one wishes to resolve. X-rays are electromagnetic radiation, with a wavelength shorter than visible light. X-rays emitted from copper targets bombarded with high energy electrons have several characteristic wavelengths: the one which is used in protein crystallography is called  $\text{CuK}\alpha$  and has a wavelength of  $1.5418 \text{ \AA}$ . This is very similar to the distance between covalently linked carbon atoms in organic molecules and therefore, it is well suited to study molecular structure.

### 1.8.2 Collection of X-ray diffraction data

A crystal is regular arrangement of molecules in 'unit cells' that are arranged by translation in an



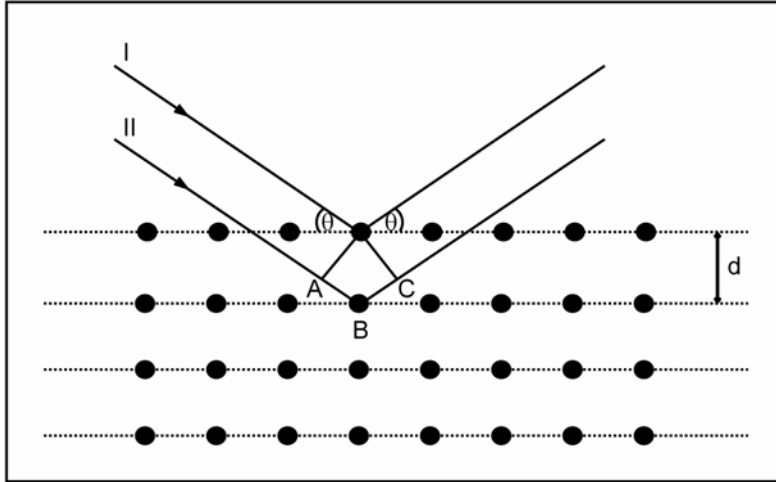
ordered array in the same orientation (Fig. 1.9). A unit cell defines the crystal coordinate system by three edges ( $a$ ,  $b$ ,  $c$ ) and three angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ), that can all be different or follow some restrictions depending on the crystal system. If the unit cell

possesses internal symmetry, the cell can be further dissected into asymmetric units. The asymmetric unit is the smallest entity that is sufficient to build up the complete unit cell by applying the symmetry operations. A 2-fold axis reduces the asymmetric unit to half of the unit cell, a 3-fold axis to a third etc., and symmetry elements in different directions multiply. In a protein crystal, only rotational symmetry (rotation or screw axes) can occur, because, due to the chirality of the amino acids, no mirror planes and inversion centers are possible. That reduces the number of possible space groups from 230 to 65, which are listed in the International Tables for Crystallography [T. Hahn, 2005].

Within each crystal, sets of equivalent, parallel lattice planes can be defined based on their orientation in the unit cell. The Miller indices  $h, k$  and  $l$  define how often a set of parallel lattice planes intersects with the unit cell edges  $a, b, c$ . Lattice planes are a helpful concept to understand scattering of X-rays by a crystal.

When an electromagnetic wave interacts with a crystal, in which the spatial arrangement of the atoms is highly ordered, it is diffracted into secondary waves whose frequencies are either lower as that of the primary wave (Compton effect) or identical to it (elastic scattering or Thomson effect). Protein crystallography is based on elastic scattering of X-rays by crystals, which has been first discovered by Max von Laue in 1912, who predicted that diffraction would occur due to same dimensions of incident beam wavelength and atom distances in the crystal.

Elastic scattering is the outcome of the interaction between an oscillating electromagnetic field vector and electrons of atoms. By that, the electrons are excited and behave as oscillating electrical dipoles, and re-radiate secondary waves of the same wavelength. Due to phase differences between secondary waves, two phenomena are observed: constructive and destructive interference. Constructive interference occurs if waves have the same orientation and are in-phase. This results in an increased intensity. During destructive interference the waves extinguish each other as they are out-of-phase. Fig. 1.10 describes a diffraction event (2-dimensional section of a crystal lattice) with parallel lattice planes within a crystal producing diffracted X-rays.



**Figure 1.10: Scheme of a 2-dimensional crystal lattice.**

Two incident X-rays (I and II, left side) are diffracted at the angle  $\theta$ . Lattice points are displayed as dark circles, the distance between two lattice planes is given as  $d$ .

([www.structmed.cimr.cam.ac.uk](http://www.structmed.cimr.cam.ac.uk))

For the two parallel incident waves (I and II in Fig. 1.10), the path difference of  $AB + BC$  has to be equal to  $2d \sin \theta$  ( $d$ : distance between planes,  $\theta$ : angle of reflection). Constructive interference of two waves can only occur if the path difference  $AB + BC$  is an integer multiple of their wavelength. This diffraction condition was first described in 1913 and is called Bragg's law:

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

**Equation 1.1**

$\theta$  angle of reflection

$\lambda$  wavelength of incident X-ray beam

$d$  distance between lattice planes of one set,

$n$  integer called the order of the reflection

Each atom in the crystal scatters X-rays in all directions, and only those that positively interfere with one another, according to Bragg's law, give rise to diffracted beams that can be recorded as distinct diffraction spots above background. Each diffraction spot is the result of interference of all X-rays with the same diffraction angle emerging from all atoms.

In Bragg's law, as the angle increases,  $d$  must become smaller for the path length to remain equal to one wavelength. Rearrangements of the previous equation to:

$$\sin(\theta)/\lambda = n/2d$$

**Equation 1.2**

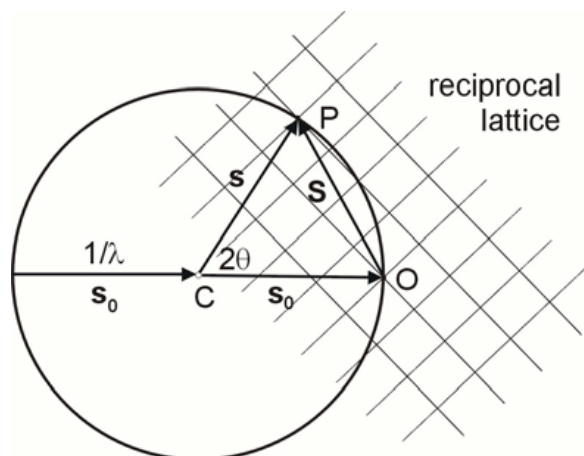
$$d = n\lambda/(2\sin(\theta))$$

**Equation 1.3**

can help to understand the concept of reciprocal space: the bigger the angle of diffraction the smaller the spacing to which the diffraction pattern is sensitive. The observed intensity of the diffraction spots can be thought of as corresponding to the size of the reciprocal lattice point.

A most useful means to understand the occurrence of diffraction spots is the Ewald construction [107].

The Ewald construction offers a geometrical way of visualizing the conditions under which re-



**Fig. 1.11: Ewald construction**

(For description, see text). - Drenth J, 1999, principles of X-ray crystallography

radiated X-rays will constructively interfere in reciprocal space (Fig.1.11). A sphere with the radius  $1/\lambda$  around the crystal passing through the origin of the reciprocal space is called Ewald's sphere. The incident X-ray ( $s_0$ ) wave passes through the center (C) of the sphere and is diffracted at the crystal lattice by  $2\theta$ . The origin of the reciprocal lattice (O) lies on the surface of the

Ewald sphere such that the incident X-ray passes through O. The difference vector  $S$  between incoming ( $s_0$ ) and scattered ( $s$ ) beam is called

diffraction vector, which lies orthogonal to the plane of the reciprocal lattice. If the endpoint of  $S$  intersects the Ewald sphere, P will be the correspondent point of the reciprocal lattice.

In other terms, for each point of the reciprocal lattice, which is lying on the Ewald sphere, Bragg's law is satisfied. As the crystal is rotated in the X-ray beam, various reciprocal lattice points come into contact with this sphere and more reflections  $h\ k\ l$  are measured. Because the  $h\ k\ l$  plane equals the  $-h\ -k\ -l$  plane, the intensity of  $h\ k\ l$  is the same as that of  $-h\ -k\ -l$  (Friedel's law). Thus, the diffraction image is centro-symmetric even though the crystal is not and the data that have to be measured are cut by half. From the reciprocal cell dimensions observed in the diffraction pattern the dimensions of the real unit cell can be calculated. In addition, the symmetry of the unit cell is preserved in the diffraction pattern, and systematic absences of Bragg reflections occur in the diffraction pattern that helps to assign the space group.

### 1.8.3 Data processing

To obtain a complete set of diffraction data, i.e a measurement of all reflections  $h\ k\ l$ , a data collection strategy is employed based on the orientation and space group of the crystal. The diffraction patterns are indexed, that is  $h$ ,  $k$  and  $l$  are assigned to the individual reflections. The same reflection might be recorded several times on different images as the corresponding



reciprocal lattice point might lie on Ewald's sphere at different orientations of the crystal. This can be used to scale the individual images, as the intensities for  $h\ k\ l$  (and  $-h\ -k\ -l$ ) should be the same in each frame. Also, pairs of reflections related by crystallographic symmetry (Bijvoet pairs) can be averaged to improve the accuracy of data. The scaling will compensate for some errors arising from crystal damage during measurement, X-ray absorption and scattering or anisotropy of the crystal.

### 1.8.4 Structure factor equation

X-rays are diffracted by the electrons of the atoms. The intensity of each diffracted beam and its phase shift with respect to the incident beam, however, depend on all atom positions and element types in the crystal's unit cell with respect to the planes of reflection. The overall intensity of the diffracted beams is defined by addition of individual phase shifts caused by atoms in the unit cell. Each diffracted beam can be fully described by its structure factor, a complex number that contains information about the amplitude of the beam as well as its phase shift

$$\vec{F}_{hkl} = |\vec{F}_{hkl}| \cdot e^{i\varphi_{hkl}} \quad \text{Equation 1.4}$$

$\vec{F}_{hkl}$  - structure factor of a diffracted beam, originating from lattice plane  $hkl$ , which defines its reflection condition.

$|\vec{F}_{hkl}|$  - structure factor amplitude,  $\varphi$  - phase shift

The structure factor can be expressed in terms of Miller indices and fractional atomic coordinates of the atoms in the crystallographic unit cell:

$$\vec{F}_{hkl} = \int_{j=1}^n f_j e^{2\pi i(hx_j + ky_j + lz_j)} \quad \text{Equation 1.5}$$

$n$  - number of atoms in the unit cell of a crystal,  $f_j$  atomic scattering factor of atom  $j$ ,  $x_j, y_j, z_j$  - fractional atomic coordinates of atom  $j$  within the unit cell,  $hkl$  - lattice plane.

### 1.8.5 Data quality indicators

The quality of X-ray data is assessed by four different parameters. One of them is the symmetry  $R$  value ( $R_{\text{sym}}$ ). The second quantity is the ratio of recorded intensities and its standard deviation  $I/\sigma(I)$  and the third one being the data redundancy i.e. how often on average a given reflection and/or one of its symmetry related reflections have been observed. The fourth quantity is the completeness of the data set both overall and in the highest resolution shell. The quantity  $R$  results from merging symmetry-related intensities.

It is defined as:

$$R_{\text{sym}} = \frac{\sum_{hkl} \sum_i |\bar{I}(hkl) - I(hkl)_i|}{\sum_{hkl} \sum_i I(hkl)_i} \quad \text{Equation 1.6}$$

Here  $I(hkl)_i$  is the intensity of a given measurement and  $\bar{I}(hkl)$  is the average of that reflection from all measurements. The sums are over all individual measurements of the same reflection  $i$  and over all reflections ( $hkl$ ).

Diederichs and Karplus (1997) proved that  $R_{\text{sym}}$  is seriously flawed, because it is inherently dependent on the redundancy of the data. They proposed an adjusted  $R$  called  $R_{\text{meas}}$  because it should accurately reflect the reliability of a diffraction data set, independent on multiplicity. The robust  $R_{\text{meas}}$  is given by:

$$R_{\text{meas}} = \frac{\sum_{hkl} \sqrt{\frac{\eta_h}{(\eta_h - 1)}} \sum_i |\bar{I}(hkl) - I(hkl)_i|}{\sum_{hkl} \sum_i I(hkl)_i} \quad \text{Equation 1.7}$$

where  $\eta_h$  is the multiplicity.

Based on the volume of the asymmetric unit and molecular weight of the protein, the number of molecules in the asymmetric unit can be estimated (Matthews, 1968). The Matthews coefficient  $V_M$  is given by the following equation:

$$VM = \frac{V}{MW \cdot Z} \quad \text{Equation 1.8}$$

$MW$  - molecular weight of the monomer in Dalton

$V$  - volume of the asymmetric unit in  $\text{\AA}^3$

$Z$  - number of molecules in the asymmetric unit.

A Matthews coefficient of a protein crystal of  $2.5 \text{ \AA}^3/\text{Da}$  corresponds to a solvent content of 50%.

The solvent content  $x_s$  of a crystal can be estimated by

$$X_s = 1 - \frac{1}{V_M \cdot N_A \cdot \rho_P} \approx 1 - \frac{1.23 \text{ \AA}^3/\text{Da}}{V_M} \quad \text{Equation 1.9}$$

$V_M$  = Matthews coefficient

$N_A$  = Avogadro's number

$\rho_P$  = protein density  $\sim 1.35 \text{ g/cm}^3$

### 1.8.6 Calculation of the electron density

Diffraction occurs one photon at a time. The probability that the photon will be scattered in any particular direction is given by the square of the amplitude of the sum of the scattered waves. Therefore, when the intensity of a diffraction spot (which is proportional to the number of photons in that spot) is measured, one takes the square root as part of determining the amplitude for the electron density calculations. The structure factors are related to the electron density of the macromolecule in the crystal by a Fourier transform. Electron density maps show the distribution of electrons within a crystal. Peaks in the electron density maps indicate atom positions. It can be described as a three-dimensional Fourier series of three-dimensional waves.

The structure factor  $\vec{F}_{hkl}$  which contains the whole information of a diffraction experiment is the Fourier transform of the electron density ( $\rho$ ). Therefore, the electron density at any given position  $xyz$  within the unit cell can be calculated from the structure factors by performing a Fourier transformation according to following equation:

$$\rho_{x,y,z} = \frac{1}{V} \sum_h \sum_k \sum_l \vec{F}_{hkl} \cdot e^{-2\pi i(hx+ky+lz)} \quad \text{Equation 1.10}$$

$V$  – volume of the unit cell

$hkl$  – indices of independent reflections.

The amplitude  $|\vec{F}_{hkl}|$  of every structure factor is proportional to the square root of the measured intensity  $I_{(hkl)}^{1/2}$ . Because  $h k l$  are discrete lattice points, the function is a sum rather than an integral. The structure factors are composed of amplitude and a phase contribution (Equation 1.4). Whereas the structure factor amplitudes  $|\vec{F}_{hkl}|$  can be derived from the intensities in the diffraction pattern (Equation 1.11), phase information is lost because the phase angle  $\alpha(hkl)$  of any reflection  $hkl$  in a diffraction experiment is unassigned and has to be obtained by methods discussed below.

### 1.8.7 Solution of the crystallographic phase problem

One of the bottlenecks in protein crystallography is the determination of accurate phases from the measured intensities in order to reconstruct the electron density of the unit cell. The structure factor amplitudes  $|\vec{F}_{hkl}|$  are proportional to the square roots of the intensities ( $I$ ) and can be extracted from the recorded intensities as:

$$|\vec{F}_{hkl}| \cong \sqrt{I_{hkl}} \quad \text{Equation 1.11}$$

$I_{hkl}$ . Intensity of a beam diffracted by lattice plane  $hkl$ .

$|\vec{F}_{hkl}|$  - its structure-factor amplitude.

The problem of recovering the missing phases, when only intensities are available, is known as the crystallographic phase problem. It is the fundamental problem in crystal structure determination.

Four major methods exist to solve the phase problem:

- a) **Direct methods** in which mathematical calculations based on probability theory can be used to derive phase information. Although these models work perfectly well for small molecule crystals, it is not easy to extend them successfully to protein crystals.
- b) **Multiple Anomalous Diffraction (MAD)** in which phase information can be obtained from the information contained in the scattering by an atom whose natural absorption frequency is close to the wavelength of the incident radiation.
- c) **Multiple Isomorphous Replacement (MIR)** in which heavy atoms are introduced into a light atom structure and are used as marker atoms to provide phase information.

d) **Molecular Replacement (MR)** is used when a very similar structure model is available. Initial phasing is based on phase approximation taken from the data of the comparable structure.

In this work, MAD method have been used which is briefly described in the following.

### Principles of MAD phasing

The method takes advantage of the tunability of synchrotron radiation X-ray sources (0.5-2.0 Å) and the presence of anomalous scatterers in the crystal that have absorption edges in that wavelength range. When the energy of incident X-rays approaches the absorption edge energies of an atom, resonance occurs which results in anomalous scattering. The absorption edges of C, O, N, S and H are far away from the accessible energy range and therefore are not suitable for MAD phasing. MAD phasing can be carried out using proteins in which methionine residues are replaced by selenomethionine [108].

### Choice of wavelengths

The choice of wavelengths is critical when performing a MAD experiment, as wavelengths are selected so as to maximize the  $f''$  component of the anomalous scattering (equation 10). This in turn minimizes the  $f'$  component, which is the source of dispersive differences between the data collected at different wavelengths. The total scattering factor can be described by two correction terms:

$$f = f_0 + f' + f'' \quad \text{Equation 1.12}$$

Where  $f$  is total scattering,  $f_0$  is the normal or Thomson scattering and  $f'$  and  $f''$  are the real and imaginary components of the anomalous scattering.

The form and the position of the absorption edge depends on the chemical environment of the heavy atom in the crystal, therefore one cannot take the theoretically calculated values of  $f'$  and  $f''$  of isolated atoms in vacuum. An X-ray absorption spectrum can be measured from the protein crystal containing the anomalous scattering element.

Typically diffraction data are collected at three wavelengths:

1. Absorption edge or peak ( $\lambda_1$ ) maximizes the  $f''$  component of the anomalous scattering and produces the largest differences in Friedel pairs.
2. Inflection point of the absorption edge ( $\lambda_2$ ) minimizes the  $f'$  component.

3. Remote wavelength ( $\lambda_3$ ) is usually collected above the absorption edge (smaller wavelength/higher energy).

### **Methods for locating anomalously scattering atoms**

Most MAD structures have many anomalously scattering atoms, and this makes the Patterson function too complicated for manual interpretation. Two methodologies have been developed for solving the substructure. The first approach is based on automated interpretation of the Patterson function in combination with difference Fourier techniques and is implemented in the program SOLVE [109]. The other approach is to use crystallographic Direct Methods [110, 111].

### **1.8.8 Phase improvement**

#### **Density modification**

Protein crystals usually contain 30-70% solvent, organized in channels of unordered water molecules. In solvent flattening the electron density is constrained towards a flat solvent region, and this real-space density modification is iterated with a phase combination step in reciprocal space. A similar iterative procedure is used in histogram matching where prior information in the form of expected density histograms is applied as constraints on the electron density map. Knowledge of non-crystallographic symmetry (NCS) can be used to modify the electron density by averaging over independent molecules. NCS averaging has proven to be very powerful in extending the available phase information when multiple copies of the same molecule are present in the same asymmetric unit. Another development in the field of density modification is the implementation of maximum-likelihood theory in the RESOLVE program [112].

### **1.8.9 Model building**

An electron density map obtained from initial phasing and density modification has to be correctly interpreted. In this process prior knowledge of the amino-acid sequence is of great importance. Therefore, visualization programs for manual model building like O [113] make extensive use of databases of commonly observed main-chain and side-chain conformations. Major advancement has been achieved in more automated ways of map interpretation. Pattern recognition methods have been implemented in semi-automated model building programs like RESOLVE (Terwilliger, 2002). Iteration of model building with refinement has been implemented in RESOLVE.

To improve the initial model, weighted electron density maps are calculated that facilitate the interpretation of the data and the comparison with the current model. The  $2F_o - F_c$  map is usually contoured at  $1\sigma$  and gives an electron density that enhances differences between measured and calculated structure factor amplitudes and thus reduces model bias.

$$\rho_{x,y,z} = \frac{1}{V} \sum_{hkl} \left( 2 \left| \vec{F}_{hkl,obs} \right| - \left| \vec{F}_{hkl,calc} \right| \right) \cdot e^{-2\pi i(hx+ky+lz)+i\varphi_{calc}} \quad \text{Equation 1.13}$$

The  $F_o - F_c$  map is termed difference electron density map and is displayed at lower contour levels (usually  $3\sigma$ ).

$$\rho_{x,y,z} = \frac{1}{V} \sum_{hkl} \left( \left| \vec{F}_{hkl,obs} \right| - \left| \vec{F}_{hkl,calc} \right| \right) \cdot e^{-2\pi i(hx+ky+lz)+i\varphi_{calc}} \quad \text{Equation 1.14}$$

It shows maxima where the model is still incomplete (peaks at positive difference density) or minima where atoms have been incorrectly placed (peaks at negative difference density).

### 1.8.10 Refinement and model validation

Refinement is an iterative process to improve the initial phases, to improve model fit to the electron density and to correct the geometry of the model. The initial model will contain errors, which can be reduced by minimizing the energy of a geometrical and crystallographic energy term. The geometrical term of a model contains bond lengths, bond angles, torsion angles, non-bonded interactions, hydrogen bonds, van-der-Waals interactions, planar restraints and chiral center restraints of the input model. The crystallographic term deals with the amplitudes of the measured reflections  $F_{obs}$ . From the refined model new structure factors  $F_{calc}$  are calculated. The difference between measured and calculated structure factors is expressed as  $R$  value, residual disagreement, which is 0.0 for exact agreement and around 0.59 for total disagreement. The  $R$  value is the quantity traditionally used for defining the quality of model fit and accuracy.

$$R = \frac{\sum_{(hkl)} \left| \left| \vec{F}_{obs}(hkl) \right| - k \left| \vec{F}_{calc}(hkl) \right| \right|}{\sum_{(hkl)} \left| \vec{F}_{obs}(hkl) \right|} \quad \text{Equation 1.15}$$

Where  $h, k, l$  define the reciprocal lattice points of the crystal,  $\vec{F}_{obs}(hkl)$  and  $\vec{F}_{calc}(hkl)$  are the observed and calculated structure factor amplitudes, respectively, and  $k$ , is a scale factor. Using the calculated phases and the measured reflection intensities, an improved electron density map can be calculated.

In spite of all the stereochemical restraints like restricting bond angles, bond lengths, torsion angles and so on to stereochemically acceptable values; it is possible to overfit the diffraction data, i.e. very low  $R$  values can be obtained from an incorrectly refined model.

Brünger in 1992 defined a statistical quantity  $R_{free}$  that measures the agreement between observed and computed structure factor amplitudes for a test set (T) of reflections (usually ~5-10%) that is omitted in the modeling and refinement process. The remaining reflections included in the refinement are known as the working set.

$$R_{free} = \frac{\sum_{(hkl) \in T} \left\| \vec{F}_{obs}(hkl) - k \vec{F}_{calc}(hkl) \right\|}{\sum_{(hkl) \in T} \left| \vec{F}_{obs}(hkl) \right|} \quad \text{Equation 1.16}$$

The  $R_{free}$  value, unlike the  $R$  factor, cannot be decreased by refining a false model, because the reflections on which it is based are excluded from this process.  $R_{free}$  is only expected to decrease during the course of a successful refinement. Consequently, a high value of  $R_{free}$  and a low  $R$  value may indicate an inaccurate model. The use of  $R_{free}$  is thus a valuable guide to the progress of refinement.

Each refinement step is accepted, if an overall decrease in energy is achieved. In the first round of automated refinement, a rigid-body refinement is performed where the starting model is fitted to the experimental data by tuning the position and orientation of the whole structural model within the unit cell without allowing individual atoms to move. Rigid-body refinement is followed by refining atomic displacement parameters which account for individual atomic motion as well as static disorder or dynamic movement of atomic groups in the crystal.

Depending on the resolution of the diffraction data, atomic displacement parameters are either refined isotropically, using spherical models, or anisotropically, using ellipsoid models for describing the displacement of individual atoms. The isotropic description of atomic displacements requires one additional parameter, the  $B$  factor (mean atomic displacement or Debye-



Waller correction) for structure factor calculation:

$$\vec{F}(h,k,l) = \int_{j=1}^n f_j e^{-\frac{B_j \sin(\theta)^2}{\lambda^2}} e^{2\pi i(hx_j + ky_j + lz_j)}$$

**Equation 1.17**

$B_j$ , the “ $B$ -factor” is the surface area of a sphere describing the mean displacement of an atom  $j$  from its coordinates.

The  $B$ - (or temperature-) factor describes the statistical motion of an atom within a sphere with the radius  $r$ , and the radius of displacement can be calculated from the  $B$ -factor:

$$B = 8\pi^2 r^2$$

**Equation 1.18**

The radius  $r$  defines a sphere of mean displacement.

The  $B$ -factor is included in the calculation of structure factors as a weighting term that reduces the contribution of atoms with high  $B$ -factors. Thus, a high  $B$ -factor of an atom might also indicate an error in the model.

Anisotropic refinement of atomic displacement can be carried out with TLS (translation, libration, screw-motion) parameters assigned to individual groups of the model. This requires fewer parameters than a full anisotropic atomic  $B$ -factor refinement and is suitable for data with 1.5 Å to 2.5 Å maximal resolution. Residues of the molecule are grouped and treated as a rigid body with 6 translation, 6 libration and 8 screw-motion degrees of freedom in displacement [114, 115]. Each cycle of automated refinement is followed by the calculation of difference density maps and a manual inspection and optimization of the refined model in real space.

A further criterion to judge a protein model is the monitoring of stereochemical parameters that are also optimized during refinement. The root-mean-square deviations (rmsd) from standard bond lengths, angles and planarity should be within a reasonable range. Furthermore, the torsion angles between the peptide planes at the  $C^\alpha$  atom of an amino acid are restricted to defined combinations, summarized in the Ramachandran plot.

## 1.9 Goals of this study

*GADD45* genes have been implicated in stress signaling, which results in cell-cycle arrest, DNA repair and cell survival, or apoptosis. Which factors dictate whether GADD45 proteins function in either cell survival or apoptosis is unclear. Although the potential roles of GADD45-family proteins have been illustrated based on the function of their associated proteins, none of their biological significance is clear yet. The mechanism of how the stress-sensor functions of GADD45 proteins co-ordinate the response of cells to environmental and physiological stressors needs better understanding. Because of GADD45 proteins' ability to induce apoptosis, new therapies targeting up-regulation of *GADD45* expression in cells may lead to specific cancer cell killing and improved current drug regimens for which structural insight is mandatory. In the beginning of this Ph.D work, no structural information on any of the GADD45 proteins was available.

The goal of my study was to characterize the structure and function of human GADD45 proteins. Therefore the purification, crystallization and structural characterization of individual GADD45 isoforms namely GADD45 $\gamma$  and GADD45 $\alpha$  were pursued. Structure-based characterization of protein-protein interactions and the preparation of relevant complexes for structural analysis was also attempted.

## 2 Materials

### 2.1 Chemicals

---

Agar, granulated	BD Bioscience
Agarose, electrophoresis grade	Invitrogen
Ammonium persulfate (APS)	Serva
Ampicillin sodium sulfate	Roth
Bugbuster protein extraction reagent	Novagen
Chloramphenicol	Roth
Complete mini EDTA-free protease inhibitor tablets	Roche
Complete protease inhibitor tablets	Roche
Desthiobiotin (DTB)	SIGMA
Isopropyl-thiogalactoside (IPTG)	AppliChem
Kanamycin sulfate	Roth
N,N,N',N'-tetramethylethylenediamine (TEMED)	Serva
Ni-Nitrilotriacetic-acid agarose	Qiagen
Nitrocellulose membrane	Schleicher & Schuell
Overnight express instant TB medium	Novagen
Peptone	BD Bioscience
Phenylmethylsulfonyl fluoride (PMSF)	SIGMA
Polyvinylidene-fluoride membrane	Roth
Rotiphorese acrylamide/bisacrylamide (37.5:1, 30 %) soln	Roth
Seleno-methionine	Acros Organics
Strep-Tactin MacroPrep resin	IBA
Talon Superflow metal-affinity resin	BD Bioscience
Tryptone	AppliChem
T9 overnight expression medium	Novagen
Yeast extract	Roth
anti-His horseradish peroxidase (HRP) conjugate	Qiagen
anti-rabbit GADD45 $\gamma$ antibody (9E10)	Santa Cruz

---

All other fine-chemicals used in this study were purchased in analytical grade from Merck, Roth or Sigma-Aldrich.

## 2.2 Enzymes

---

Benzonase, DNase-I, lysozyme	Roche
DNA polymerases Pfu, Turbo Pfu	Stratagene
DNA polymerase Pfu Plus	Roboklon
Restriction enzymes	NEB
T4 DNA ligase	Promega, Roboklon
TEV protease	Konrad Büssow
Thrombin	Serva

---

## 2.3 Kits

---

QIAprep Spin Miniprep Kit	Qiagen
QIAGEN Plasmid Maxiprep Kit	Qiagen
QIAquick PCR Purification/Gel Extraction Kit	Qiagen
QIAGEN Ni-NTA Spin Kit	Qiagen
ECL plus Western Blotting Detection System	Amersham

---

## 2.4 Columns

---

HisTrap HP columns, 1 ml/5ml	GE Healthcare
HiTrap S cation-exchange column	GE Healthcare
HiTrap Q anion-exchange column	GE Healthcare
PD10 Fast desalting column	Amersham
Superdex 75 HiLoad 16/60	GE Healthcare
Superdex 75 HiLoad 26/60	GE Healthcare
Superdex 75 Analytical 10/300	GE Healthcare
Superdex 200 HiLoad 16/60	GE Healthcare
Superose 12	GE Healthcare
Strep-Tactin Superflow column	IBA GmbH
PoRos Talon Superflow column	Applied Bioscience
QIAGEN Ni-NTA Spin column	Qiagen

---

## 2.5 Strains, clones and plasmids

NAME	GENOTYPE	SOURCE (REFERENCE)
<i>E. coli</i> strains		
DH5 $\alpha$	F $^-$ $\Phi$ 80lacZ $\Delta$ M15 $\Delta$ (lacZYA-argF) U169 recA1 endA1 hsdR17 (rK $^-$ , mK $^+$ ) phoA supE44 $\lambda^-$ thi-1 gyrA96 relA1	Invitrogen
JM109 (DE3)	endA1, recA1, gyrA96, thi, hsdR17 (rK $^-$ , mK $^+$ ), relA1, supE44, $\lambda^-$ , $\Delta$ (lac-proAB), [F', traD36, proAB, lacIqZ $\Delta$ M15], IDE3	Promega
BL21(DE3)	F $^-$ ompT hsdSB(rB $^-$ , mB $^-$ ) gal dcm y(DE3)	Novagen
B834 (DE3)	F $^-$ ompT hsdSB(rB $^-$ , mB $^-$ ) gal dcm met(DE3)	Novagen
NEB 10-beta		NEB

**Table 2.1:** Name, genotype and source (reference) for the bacterial strain used

For cloning purpose DH5 $\alpha$ , and/or JM109 (DE3) were used. For high-yield protein expression, the plasmids were transformed in *E. coli* BL21 (DE3) and/or BL21-CodonPlus (DE3)- RIL. For selenomethionine incorporation, a methionine auxotroph expression strain B834 (DE3) was employed. NEB 10-beta electrocompetent cells were used for the transformation of mutants.

### Template clones

GADD45 $\gamma$	Konrad Büsow
GADD45 $\alpha$	Patrick Umbach
PCNA	Christina Cardoso
PPAR $\gamma$	PSPF

### Plasmids

pET-28a	<i>E. coli</i> expression vector with kanamycin resistance, T7/lac promoter, N-terminal His-tag, thrombin cleavage site	Novagen
pCDF Duet	<i>E. coli</i> expression vector with streptomycin resistance, T7/lac promoter, N-terminal His-tag for MCS-I, no tag for MCS-II, thrombin cleavage site	Novagen
pQTEV	<i>E. coli</i> expression vector with ampicillin resistance, T7/lac promoter, N-terminal His-tag, TEV protease cleavage site	Novagen

## 2.6 Media and buffers

### Culture media

Culture media were autoclaved, and antibiotic stock solutions were sterile-filtrated after preparation. 50 µg/ml of ampicillin (Amp), 100 µg/ml of chloramphenicol (Cm), 30 µg/ml of kanamycin (Kan) and/or 50 µg/ml streptomycin (Sm) were added to each medium prior to use. For preparation of agar plates, 15 g of agar were dissolved in 1 l millipore H<sub>2</sub>O and autoclaved. For protein expression using the pET system, isopropyl-thiogalactoside (IPTG) was added to cultures to a final concentration range varying from 1.0-0.5 mM. Media were generally stored at 4°C and, if not indicated specifically, antibiotic stock solutions at -20 °C.

NAME	CHEMICAL	AMOUNT
LB medium	Tryptone	10 g
	Yeast extract	5 g
	NaCl	10 g
	ddH <sub>2</sub> O	up to 1 l
SB medium	Yeast extract	24 g
	Tryptone	12 g
	Glycerol	3.75 g
	KH <sub>2</sub> PO <sub>4</sub>	2.3 g
	K <sub>2</sub> HPO <sub>4</sub>	12.5 g
	ddH <sub>2</sub> O	up to 1 l
LB agar	agar	15 g
	LB	up to 1 l

### Molecular biology buffers

NAME	CHEMICAL	AMOUNT
TAE buffer	Tris HCl	40 mM
	Sodium acetate	5 mM
	EDTA	1 mM
	Acetic acid	To pH 8
Ethidium bromide solution	TAE buffer with ethidium bromide	10 mg ml <sup>-1</sup>
DNA molecular weight marker	1 kb ladder (NEB)	0.5 µg µl <sup>-1</sup>
DNA sample buffer	5x EDTA	100 mM
	Xylencyanol	0.05% (v/v)
	Glycerol	30% (v/v)
	Bromphenol blue in TAE-buffer	0.05% (w/v)

**Protein purification buffers**

NAME	CHEMICAL	AMOUNT
His-lysis buffer pH 8	NaH <sub>2</sub> PO <sub>4</sub>	50 mM
	NaCl	300 mM
	Imidazole	10 mM
	β-ME	2 mM
His-wash buffer pH 8	NaH <sub>2</sub> PO <sub>4</sub>	50 mM
	NaCl	300 mM
	Imidazole	20 mM
His-elution buffer pH 8	NaH <sub>2</sub> PO <sub>4</sub>	50 mM
	NaCl	300 mM
	Imidazole	300-500 mM
TALON-lysis buffer pH 7.5	Tris	20 mM
	NaCl	300 mM
	Imidazole	10 mM
	β-ME	2 mM
Talon wash buffer pH 7.5	Tris	20 mM
	NaCl	300 mM
	Imidazole	20 mM
Talon elution buffer pH 7.5	Tris	20 mM
	NaCl	300 mM
	Imidazole	250 mM
Cation-exchange buffer (low salt) pH 7.5	Tris	50 mM
	DTT	1 mM
Anion-exchange buffer (high salt) pH 7.5	Tris	50 mM
	NaCl	1 M
	DTT	1 mM
Gel-filtration buffer pH 7.5 – 8.0	Tris	20 mM
	NaCl	200 mM
	DTT	2 mM
	EDTA	0.5 mM

**SDS-PAGE Buffers**

Electrophoresis buffer	Tris	25 mM
	Glycine	190 mM
	SDS	0.1% (w/v)
Stacking gel buffer (4x)	Tris	0.5 mM
pH 6.8	SDS	0.4% (w/v)
Stacking gel	30% acrylamide /	0.4 ml
	0.8 % (w/v)	
	Bisacrylamide stock	
	Stacking gel buffer	0.75 ml
	dH <sub>2</sub> O	1.85 ml
	APS 10% (w/v)	30 µl
	TEMED	5 µl
Separation gel buffer (4x)	Tris	1.5 M
pH 8.8	SDS	0.4% (w/v)
Separation gel 15% (w/v)	30 % acrylamide /0,8 %	5 ml
	(w/v) Bisacrylamide	
	separation gel buffer	2.5 ml
	dH <sub>2</sub> O	2.5 ml
	APS 10% (w/v)	75 µl
	TEMED	10 µl
SDS-sample buffer	2x SDS	3% (w/v)
	Glycerol	20% (w/v)
	β-mercaptoethanol	3% (v/v)
	Bromphenol blue	0.05% (w/v)
	EDTA	10 mM
Coomassie stain	Coomassie R-250	0.25% (w/v)
	Methanol	45% (v/v)
	dH <sub>2</sub> O	45% (v/v)
	Acetic acid, glacial	10% (v/v)
Destain solution	Methanol	40% (v/v)
	dH <sub>2</sub> O	50% (v/v)
	Acetic acid, glacial	10% (v/v)

**Western blot buffers**

Semi-dry transfer buffer	Tris HCl	25 mM
pH 8.3	Glycine	150 mM
	Methanol	10%
PBS-T	PBS	1 l
	Tween-20	1 ml



## 2.7 Synthetic Oligonucleotides

Synthetic oligonucleotides were ordered from MWG Biotech AG.

### Vector-specific primers

Primer sequences (5' 3') vectors

Name	Sequence	Vector
T7 promotor	taa tac gac tca cta tag gg	pET
T7 terminator	gct agt tat tgc tca gcg g	pET
pQE for	gta tca cga ggc cct ttc gtc t	pQTEV
pQE rev	cat tac tgg atc tat caa cag gag	pQTEV
DuetUP2	ttg tac acg gcc gca taa tc	pETDuet
DuetDOWN1	gat tat gcg gcc gtg tac aa	pETDuet

### Oligos for cloning and site-directed mutagenesis

Only forward primers are listed here. Reverse primer have reverse complementary sequence.

Primer sequences (5' 3')

Name	Sequence	Application
GADD45gamma-1-fw	cgc ttt gcg gcc cgc cat gac tct gga aga agt ccg c	Cloning of GADD45 $\gamma$ into pET 28a
GA-28a-NdeI-fw	cgc ttt cat atg act ttg gag gaa ttc tcg	Cloning of GADD45 $\alpha$ into pET 28a
Pcnaduet-sacI-tev-fw	ggg aat tga gct cgg aaa acc tgt att ttc aga tgt tcg agg cgc gcc tg	Cloning of PCNA (at MCS1) in to CDFDuet-1 vector.
G45gduet-NdeI-fw	ccg ttt cat atg act ctg gaa gaa gtc cgc g	Cloning of GADD45 $\gamma$ (at MCS2) in to CDFDuet-1 vector
26-30 all_Ala_fw	atg cag ggt gcc ggg aaa gcg gcc gct gcc gcg gcc ctg tcg gcg cag cgt cag ggc	Substitution of <sup>26</sup> LHELL <sup>30</sup> with alanines
26-30 leu_Ala_fw	atg cag ggt gcc ggg aaa gcg gcc cac tgc gcg gcc ctg tcg gcg cag cgt cag ggc	Substitution of leucines of <sup>26</sup> LHELL <sup>30</sup> with alanines
101-105 all_ala_fw	cgc gtg ggc gat gtg cag cgg gcc gcg gct gcg gcc ggc gcc ggc gag gag gcg ggt	Substitution of <sup>101</sup> LAAIV <sup>105</sup> with alanines
101-105 leu_ala_fw	cgc gtg ggc gat gtg cag cgg gcc gcg gct atc gtg ggc gcc ggc	Substitution of leucines of <sup>101</sup> LAAIV <sup>105</sup> with alanines

117-121 all_ala_fw	gag gcg ggt gcg ccg ggc gac gcc gct gcc gcg gcc att tcg aac ccc aac gag gac	Substitution of <sup>117</sup> LHCIL <sup>121</sup> with alanines
117-121 leu_ala_fw	gag gcg ggt gcg ccg ggc gac <b>gcc</b> cac tgc atc <b>gcc</b> att tcg aac ccc aac gag gac	Substitution of leucines of <sup>117</sup> LHCIL <sup>121</sup> with alanines
E69_Ala_fw	ttc tgt gtg ctg gct gcg ggt gcg gag gac gcg ggc gac atc gcg ctg cag	Substitution of E69 with alanine
T79E_fw	atc gcg ctg cag atc cat ttt <b>gag</b> ctg atc cag gct ttc tgc tgc	Substitution of T79 with glutamate
L80E_fw	gcg ctg cag atc cat ttt acg <b>gag</b> atc cag gct ttc tgc tgc gag	Substitution of L80 with glutamate
Q82E_fw	cag atc cat ttt acg ctg atc <b>gag</b> gct ttc tgc tgc gag aac gac	Substitution of Q82 with glutamate
D150A_fw	gag gag agc cgc agc gtt aac <b>gcc</b> tgg gtg ccc agc atc acc ctc	Substitution of D150 with alanine
W150F_fw	gag agc cgc agc gtt aac gac <b>ttt</b> gtg ccc agc atc acc ctc ccc	Substitution of W150 with phenylalanine
GA_QCM_10_fw	ggc ctg gtg ccg cgc ggc agc cag aag acc gaa agg atg gat	Deletion of N-terminal 10 residues of GADD45 $\alpha$

### 3 Methods

General molecular-biology methods used in this work refer to protocols from the text book Molecular Cloning [116]. People who were involved in cloning and production of human GADD45 $\gamma$  native protein are Dr. Claudia Quedenau and Dr. Frank Götz. Dr. Yvette Roske helped in crystallisation, Dr. Andrew P. Turnbull analysed the native dataset of human GADD45 $\gamma$  and Dr. Jürgen J. Müller assisted in solving structure.

#### 3.1 Molecular biology methods

##### Construction of plasmids

Double-stranded DNA fragments were amplified using the polymerase-chain reaction [117] from a template with the help of a pair of 20-30 bp long oligonucleotides (primers) that are complementary to the 5' and 3' end, respectively, of the amplified sequence.

A template fragment flanked by the primer sequences is produced in large quantities by repetitive cycles of DNA melting, annealing of primers and strand synthesis with a thermostable DNA. Gene-specific PCR primers were designed to add recognition sites for restriction nucleases to the 5' and 3' end of the amplified DNA sequence (Table 3.1). To allow directional cloning, recognition sites for different restriction nucleases were added with either primer. Vector-specific primers used for colony PCRs and sequencing are listed in Table 2.2. Plasmids were generated by amplifying cDNA sequences with Pfu polymerase that possesses proofreading activity. Colony PCRs were used to identify positive clones after ligation. Cells from a bacterial colony were directly transferred into tubes containing a Taq polymerase reaction mix with pipette tips (Table 3.2). The following guideline was used for setting up the reaction mixture:

Component	Final concentration
10x Ultra Pfu buffer	1x
Primers	0.5 $\mu$ M each
Template DNA	10-30 ng
dNTPs	200 $\mu$ M each
Ultra Pfu polymerase	2.5 U
H <sub>2</sub> O	to 50 $\mu$ l

**Table 3.1:** Standard PCR

Component	Final concentration
10x Taq buffer	1x
Primers	0.25 $\mu$ M each
Template DNA	colony
dNTPs	200 $\mu$ M each
Taq polymerase	0.5 U
H <sub>2</sub> O	to 15 $\mu$ l

**Table 3.2:** Colony PCR

Segment	Step	Temperature [°C]	Duration [min:sec]
1	melting	94	2:00
2	melting	95	0:30
3	annealing	56	1:00
4	extension	72	2:00
Cycles - 25			
5	final extension	72	10:00
	On hold	4	Forever

**Table 3.3:** Standard PCR program.

The PCR amplicon was gel extracted, incubated with restriction enzymes in the recommended buffers and cloned into plasmid by ligation to the restriction endonuclease sites of the polylinker. Table 3.4 summarises the standard reaction mixture used for restriction digestion and ligation.

Compound	Amount
Purified PCR reaction or plasmid preparation	50 µl or 2-3 µg
10x reaction buffer	6 µl
Restriction enzyme I	1 µl (10-20 U)
Restriction enzyme II	1 µl (10-20 U)
100x BSA	0.6 µl
H <sub>2</sub> O	To 60 µl

Compound	Amount
Insert (~500 bp)	4 ng
Vector (~4000 bp)	10 ng
T4 ligase	1 U
10x reaction buffer	1.5 µl
H <sub>2</sub> O	to 15 µl

**Table 3.4: Left,** standard restriction digestion reaction. The reaction was incubated 8 h or overnight at 37 °C. **Right,** standard ligation reaction

Reactions were incubated at room temperature for 1-2 h or overnight at 8 °C using T4 ligase and then transformed into competent *E. coli* DH5α cells. Competent cells were prepared according to Hanahan [118]. Clones on ligation plates were checked with colony PCR to identify plasmids containing the desired insert. Positive colonies were used to inoculate LB cultures for mini (2 ml) or maxi (150 ml) plasmid preparations.

## Site-directed mutagenesis

Mutations and internal deletions were introduced using site-directed mutagenesis. To maximize mutagenesis efficiency and minimize the incidence of primer dimer formation, a two-stage protocol described by Wang and Malcolm [119] was used. In this reaction both strands of a template plasmid are fully amplified from primers which contain the desired mutations. By carrying out the amplification of the two strands separately using each primer in a separate reaction, formation of primer dimers is prevented during the first rounds of amplification, which otherwise can profoundly decrease the efficiency of mutagenesis. After the initial steps of amplification both reactions are mixed and additional rounds of amplification are performed. Upon amplification, plasmids containing unmodified DNA strands are cleaved by adding Dpn-I restriction enzyme (New England Biolabs), which only cleaves methylated DNA. Mutated nicked plasmids generated in the amplification cycles do not carry methyl groups. After cleavage, aliquots of the digests were transformed into *E. coli* Top10 cells, and functional plasmids were selected by appropriate antibiotics.

Two primers were generated for each mutagenesis reaction, which cover 18 to 21 nucleotides upstream to the area of modification, the desired nucleotide changes, and 21 nucleotides downstream to the area of modification, for both strands of the template DNA. Two single-primer PCR reactions preceding the standard QCM procedure was performed (see below).

Component	Final conc.
10x Ultra Pfu buffer	5 $\mu$ l (1x)
one Primer (forward or reverse)	0.25 $\mu$ M
Template DNA	10-50 ng
dNTPs	200 $\mu$ M each
Ultra Pfu polymerase	1 $\mu$ l (2.5 U)
H <sub>2</sub> O	to 50 $\mu$ l

**Table 3.5:** Two-stage QCM PCR

Immediately before the mutagenesis reaction was started, 1  $\mu$ l (5 u) of Pfu polymerase (Stratagene) was added to both tubes, the tubes were inserted into a thermocycler and the following protocol was started:

Segment	Time	Temperature	Comment
Initialization	30 s	95 °C	initial denaturation of DNA
	30 s	95 °C	denaturation of DNA
Step 1: 10 cycles	30 s	58 °C	primer annealing
	300 s	68 °C	primer elongation
Combination			After the first round of amplification, the content of both tubes was mixed.
	30 s	95 °C	denaturation of DNA
Step 2: 16 cycles	30 s	58 °C	primer annealing
	300 s	68 °C	primer elongation
Termination		4 °C	storage

**Table 3.6:** Protocol for site-directed mutagenesis of bacterial plasmids.

Following temperature cycling, 10 units of DpnI restriction enzyme was directly added to the PCR reaction mixture and incubated for 1 h at 37 °C. During incubation the methylated template DNA from *E. coli* is destroyed whereas the synthetic DNA from PCR remains intact. Upon digestion, 1 - 3 µl of the reaction mix were used to transform chemo-competent cells of *E. coli* strain Top10. The clones for correct introduction of the desired mutation were screened by sequencing (MWG Biotech) using vector-specific primers.

## 3.2 Biochemical methods

### 3.2.1 Protein production

#### Protein synthesis

The expression constructs were transformed into *E. coli* BL21 (DE3) or BL21-(DE3) – Rosetta cells. An overnight culture was inoculated in SB medium or LB medium containing appropriate antibiotics (100 µg ml<sup>-1</sup> ampicillin; or 34 µg ml<sup>-1</sup> chloramphenicol; or 30 µg ml<sup>-1</sup> kanamycin). The pre-culture was diluted 50-fold in the same medium. Cell cultures were grown at 37 °C with a rotary shaker until an OD<sub>600</sub> of ~1.5 (SB medium) or ~0.6 (LB medium) was reached. Induction was initiated by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were harvested by centrifugation at 5000 x g for 10 min and pellets were stored at -80 °C. For expression at 20 °C, cells were grown to an OD<sub>600</sub> of ~0.6 at 37 °C and cooled down to 20 °C

prior to induction with 1 mM IPTG. Growth was continued at 20 °C over night. Cells harvested 4 h later were spun down flash frozen and stored at -80 °C.

### **Production of seleno-methionine labeled protein**

For the seleno-methionine (SeMet) labeling, the protein was expressed in *E. coli* B834 (DE3) and cells were grown essentially as described by Budisa et al. [120]. A pre-culture in LB medium was grown for one day, and cells were harvested and washed with new minimal medium (NMM). The culture was diluted 1:100 with NMM containing 20% methionine and 80% SeMet (25 µg ml<sup>-1</sup> total concentration) and grown over night. Cells were again washed with NMM and diluted 1:100 in NMM containing 100% SeMet. The culture was induced with 1 mM IPTG with an OD<sub>600</sub> of 0.4-0.8 (approximately 24 h later).

### **3.2.2 Protein purification**

All steps of protein purification were carried out at 4 °C unless stated otherwise. Fast protein liquid chromatography (FPLC) was carried out on either a VISION chromatography or Äkta Explorer low pressure system. All columns were equilibrated in corresponding buffers before use and filled with 20% ethanol for long-time storage

### **Cell lysis**

All steps were carried out at 4 °C, unless stated otherwise. Cell pellets were thawed on ice and suspended in appropriate ice-cold lysis buffer (3 ml per g of cell pellet) supplemented with additives (see 3.2.5) containing one Complete EDTA-free Protease Inhibitor Cocktail tablet, 2 mM β-mercaptoethanol and DNaseI or Benzonase. The suspension was passed 3 times through a French Press at a pressure of 1,200 psi. The lysate was centrifuged for 60 min at 20,000 rpm at 4 °C and the supernatant passed through a 0.45 µm filter.

### **Immobilized metal affinity chromatography**

Affinity chromatography using Ni Sepharose or Talon was the first step in the chromatographic purification. It was performed to remove most cellular proteins from recombinantly expressed 6x His tagged proteins. The initial step of each purification protocol was carried out either as batch purification or using an FPLC system.

**Batch binding with Ni-nitrilotriacetate (NTA) agarose**

Fusion proteins carrying an oligo-histidine tag (6x-His tag or 7x-His tag) can be purified with Ni-nitrilotriacetate (Ni-NTA) matrix. NTA is coupled to Sepharose CL-6B and chelates  $\text{Ni}^{2+}$  ions. The remaining coordination sites of Ni(II) can form a complex with histidine residues and thus bind His-tags. Elution of His-tagged protein is achieved by the addition of the histidine analog imidazole which occupies the binding sites for histidine when present in excess.

After cell lysis and subsequent centrifugation, the supernatant containing His-tagged fusion protein was incubated for 1-2 h at  $4^{\circ}\text{C}$  with Ni-NTA matrix according to manufacturer's instruction (QIAGEN). The suspension was then transferred into a polypropylene column. The affinity matrix was washed extensively with His-wash buffer and eluted with His-elution buffer (300 mM imidazole) with gravity flow through the self-poured column.

**Purification with Ni-NTA agarose HisTrap HP columns**

The supernatant of the cell lysis was applied to HisTrap HP columns and then washed with 5 column volumes of washing buffer A. Further removal of non-specifically bound proteins was increased by washing with 5-10% of His-elution buffer (in total 65-90 mM imidazole concentration). Finally, bound protein was eluted by washing with 100% of His-elution buffer (500 mM imidazole).

**Purification with Talon resin**

Alternatively to Ni-NTA, matrix fusion proteins carrying a His-tag can be purified with Talon resin. Talon resin works analogous to Ni-NTA but has  $\text{Co}^{2+}$  ions chelated resulting in a better specificity for histidine tags.

Cells expressing His-tagged fusion protein were lysed in Talon-lysis buffer supplemented with 1 mM PMSF, one protease tablet Complete Mini EDTA-free, 2 mM  $\beta$ -mercaptoethanol and DNase I. The cleared lysate obtained by centrifugation and filtration (0.45  $\mu\text{m}$  filter) was incubated with Talon resin for 1 h. The affinity matrix was applied to a polypropylene column, washed extensively with Talon-wash buffer and eluted with Talon-elution buffer.

**Buffer exchange**

Buffer was exchanged using a PD10 column wherever necessary. Proteins differ greatly in size from salts and small molecules. In this process, separation of the two groups is achieved as high-



molecular-weight substances (>5 kDa) are excluded from the column resin and thus elute first, whereas low-molecular-weight substances enter the pores of the resin freely and thus elute later.

### **Thrombin cleavage**

N-terminal 6x-His fusion tags were removed by recombinant thrombin using the RECOMT Thrombin CleanCleave Kit according to user instructions. The thrombin is covalently bound to agarose beads for easy removal. The buffer is supplemented with 10 mM CaCl<sub>2</sub>. The cleavage reaction was performed overnight at 4 °C with slow agitation on a shaker to ensure homogeneous distribution of the thrombin beads in the protein solution. Once the digestion is completed, as monitored by SDS-PAGE, the beads were sedimented by low-speed centrifugation (600 x g, 5 min, 4°C). The protein solution was filtered through 0.2 µm filter to rid the protein of any residual thrombin beads. Additionally, 1 mM EDTA and Pefabloc were added to the protein solution.

### **TEV digestion**

To cleave the affinity tag from the recombinant protein, specific recognition sites for proteases are introduced in the expression constructs. The His-tag of pQTEV can be removed with tobacco etch virus (TEV) protease at its cleavage site.

Elution fractions from affinity purification containing recombinant protein were incubated with the required protease at a ratio of 40:1 (fusion protein: protease, w/w) at 4 °C until cleavage was complete (usually over night). To separate cleaved protein from the tag and uncleaved protein, TEV digest reactions were dialysed against Talon-wash buffer without imidazole and either 1) loaded onto Ni-NTA batch matrix which captures TEV and uncleaved protein. Unbound fractions were then collected and further purified; or, 2) In case of native GADD45γ which has a pI of 4.3, digested samples were loaded on a HiTrap S cation-exchange column equilibrated with 20 mM Tris pH 7.5, 1 mM DTT using VISION chromatography system and eluted with 0 to 1 M linear NaCl gradient in same buffer at approximately 0.5 M NaCl. Because of the charge difference, elute fraction contained TEV protease whereas protein came out in the flowthrough which was then purified further.

### **Ion-exchange chromatography**

Proteins have numerous functional groups that can have both positive and negative charges. Using ion-exchange chromatography, proteins can be separated according to their size and net

charge, which is dependent on the composition of the mobile phase. By adjusting the pH or the ionic concentration of the mobile phase, various protein molecules can be separated. For example, if a protein has a net positive charge at pH 7, then it will bind to a column of negatively charged beads, whereas a negatively charged protein would not. By changing the pH so that the net charge on the protein is negative, it too will be eluted.

Elution by changing the ionic strength of the mobile phase has a more subtle effect - it works as ions from the mobile phase will interact with the immobilized ions in preference over those on the stationary phase. This "shields" the stationary phase from the protein, (and *vice versa*) and allows the protein to elute.

Affinity-purified protein samples were loaded on MonoQ and/or MonoS HR 5/5 columns equilibrated in low-salt buffer (20 mM Tris, pH 7.5 or 8.0). This buffer was also used for washing until the OD<sub>280nm</sub> at the column outlet returned to base level. Protein was eluted in a 600 ml gradient from 0% to 60% high salt buffer (500mM NaCl) and collected in fractions, which were analyzed by SDS page. Fractions containing protein of the correct size were pooled.

### **Size-exclusion chromatography**

The last step of protein purification involved polishing by gel-filtration chromatography, also known as size-exclusion chromatography as it relies on the separation of particles based on their size and shape. The column material is composed of inert polymer beads with a defined pore diameter. When applied to such a column, molecules in the sample are separated according to their size and shape. Fractionation of the molecules is achieved as they move through a bed of porous beads, diffusing into the beads to greater or lesser degrees. Smaller molecules being able to enter the pores travel a longer way through the column and are eluted later than bigger particles. This step is required to separate aggregates and high molecular-weight impurities from the protein which is difficult to achieve by any other technique.

Every column has a characteristic void volume that is the minimum volume a big particle needs to pass through without entering any pore. The column volume gives the total amount of buffer that is required to equilibrate the system. Columns are calibrated by determining the elution volume of standard proteins. A linear relation exists between the logarithm of the molecular weight of the compounds and their elution volume. Thus, the molecular weight of a sample protein can be calculated based on the elution volume observed in gel-filtration chromatography.

In this study, gel-filtration experiments were performed with an Äkta Explorer, and the size-exclusion column was equilibrated with at least 1 column volume of buffer. Superdex-75 and Superdex-200 columns were used, having separation ranges of 3-70 kDa and 10-600 kDa, respectively. The sample volume was 1-2 ml for 16/60 columns and up to 5 ml for columns with dimensions of 26/60 (diameter in mm/length in cm), and purification was carried out at flow rates of 1 ml/min for the 16/60 column and 2 ml/min for the 26/60 column at 4 °C with a maximal pressure set to 0.3 MPa. For analytical purposes, a Superdex-75 10/300 column was used which has a column volume of 25 ml and was run at a flow rate of 0.8 ml/min with a maximal pressure of 1.8 MPa. The elution profile was monitored measuring the absorption at 280 nm, and peak fractions were analyzed with SDS-PAGE.

### **3.2.3 Production of recombinant human GADD45 $\gamma$**

#### **Cloning & expression**

The human cDNA encoding for GADD45 $\gamma$  (GenBank CAH71791) was cloned between the BamHI and NotI sites of the vector pQTEV (GenBank AY243506) [121]. The resulting plasmid was introduced into *E. coli* SCS1 Rosetta cells carrying the helper plasmid pSE111 [122].

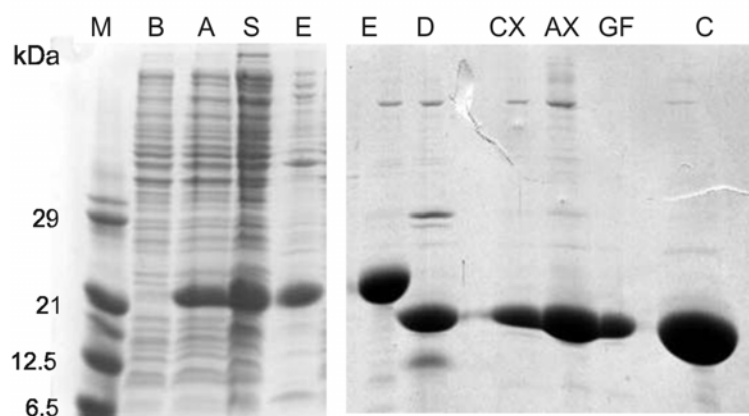
The clone that was used for protein production contains two point mutations and a deletion introduced accidentally in the cloning procedure. The amino acid Ala2 was changed to Thr and Thr58 to Ala. The recombinant protein contains two additional N-terminal residues (Gly-Ser) encoded by the BamHI site.

An overnight culture was inoculated in LB medium containing 100  $\mu\text{g ml}^{-1}$  ampicillin. The cultures were grown at 37 °C to an OD<sub>600</sub> of ~0.6 and induced with 1 mM IPTG. Growth was continued for 4 h, and cells were harvested by centrifugation at 5000 g for 10 min.

For production of selenomethionine-labeled protein, two point mutations L29M, L101M and one double mutation L73&79M were introduced (QuikChange Site-Directed Mutagenesis Kit, Stratagene). GADD45 $\gamma$  L101M crystals were obtained under similar conditions as native crystals and were of the same size and shape. Therefore GADD45 $\gamma$  L101M was selected for selenomethionine labeling. The resulting plasmid was used to transform *E. coli* B834 (DE3) cells (Novagen) carrying the helper plasmid pSE111, and cells were grown essentially as described by [120]. After induction with 1 mM IPTG, growth was continued for 4 h, and cells were harvested by centrifugation at 5000 g for 10 min.

## Purification

All steps were performed at 4 °C. Cells from a 2-liter culture expressing 6×His-GADD45γ were resuspended in 40 ml lysis buffer containing 20 mM Tris buffer, pH 7.5, 300 mM NaCl, 10 mM imidazole, 2 mM β-mercaptoethanol and supplemented with 1 mM PMSF, one protease-inhibitor tablet (Complete EDTA-free, Roche) and 250 U Benzonase (Merck) and disrupted by sonication on ice. The lysate was cleared by centrifugation at 75000 g for 20 min, sterile filtered through 0.45μm filter (Millipore) and applied onto a PoRos Talon Superflow column (Clontech) using a Vision chromatography system. After washing with 20 mM Tris buffer, pH 7.5, 300 mM NaCl, 10 mM imidazole, the protein was eluted with 250mM imidazole. The eluted protein was supplemented with DTT and EDTA to give final concentrations of 0.5 mM and 0.1 mM, respectively. The His-tag was removed after incubation with TEV protease overnight at 4 °C at a protease: protein-ratio of 1:40 (w/w). After 10-fold dilution with 20 mM Tris-HCl, pH 7.5, the protein was applied to a cation-exchange column (POROS 20 S, Applied Biosystems). The flow-through containing GADD45γ was collected, applied to an anion-exchange column (POROS 20 HQ, Applied Biosystems) and GADD45γ eluted in a NaCl gradient. The protein was further purified on a Superose 12 16/50 column (Amersham Biosciences), equilibrated in 15 mM Tris-HCl, pH 7.5, 50 mM NaCl, 2mM DTT and 0.1 mM EDTA followed by concentration with Biomax concentrators (Millipore, 5-kDa cutoff) to a final concentration of 30 mg/ml. After confirming monodispersity with dynamic light scattering (Laser-Spectroscatter 201, RiNA GmbH, Berlin) the protein was used for crystallization. Se-Met labeled GADD45γ was purified accordingly (Fig. 3.1).



**Figure 3.1: SDS-PAGE of samples from the purification of GADD45γ.**

M: marker lanes; molecular weight of marker proteins is given in kDa B: lysate of cells before induction; A: after induction; S: supernatant; E: elution from Talon affinity purification; D: TEV protease digest; CX: pooled fractions from cation-exchange flow-through; AX: pooled

elute fractions from anion-exchange column; GF: peak fractions of Superose 12 gel filtration; C: concentrated sample used for crystallization;

## Production of GADD45 $\gamma$ mutants

Cells expressing mutants of human GADD45 $\gamma$  were expressed in 2 litre culture medium. Cells were lysed in Talon-lysis buffer. The cleared lysate obtained by centrifugation and filtration (0.45  $\mu$ m filter) was incubated with 3 ml Talon resin for 1 h at 4 °C. The affinity matrix was applied to a polypropylene column, washed extensively with Talon-wash buffer and eluted with Talon-elution buffer. After supplementing with 0.5 mM DTT and 0.1 mM EDTA, samples were incubated with TEV protease overnight at 8 °C at a protease: protein ratio of 1:40 (w/w). Buffer was exchanged to 20 mM Tris (pH 7.5), 300 mM NaCl by PD10 fast desalting column and further incubated with Talon/Ni-NTA affinity matrix to remove TEV protease (recombinant TEV carries an N-terminal His tag). The flow-through containing cleaved protein was pooled, concentrated and fractionated on Superdex 75 10/300 or 16/60 columns (GE Healthcare) in either 25 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2 mM DTT and 0.1 mM EDTA or 20 mM Hepes (pH 7.5), 150 mM NaCl, 2 mM DTT, 1 mM EDTA.

### 3.2.4 Protein analysis

#### Determination of protein concentration

Concentration of protein preparations was determined measuring the UV spectrum of the samples. The absorbance at 280 nm was used to calculate the protein concentration with Beer's law based on the molar extinction coefficient ( $\epsilon$ ) calculated with ProtParam [71].

Beer's law:  $T = I/I_0 = 10^{-\epsilon \cdot c \cdot d}$  or  $A_\lambda = \epsilon_\lambda \cdot c \cdot d$

T: transmission of light

$I_0$  and  $I$ : intensity of the incident light and that after the material, respectively

$\epsilon_\lambda$ : the molar extinction coefficient [ $M^{-1} \text{ cm}^{-1}$ ] at a given wavelength

c: concentration [M]

d: path length [cm]

A: absorbance

#### Sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE)

The purity of protein samples was analyzed by discontinuous SDS-PAGE. The gels were produced by co-polymerization of acrylamide with the crosslinker N,N'-methylenebisacrylamide

in the presence of the radical starter ammonium peroxodisulfate (APS) and the catalyst tetramethylethylenediamine (TEMED). The pore size of the gel is determined by the crosslinking which in turn is regulated by variations in the concentration and the ratio of both monomers. SDS, an anionic detergent denatures the secondary and non-disulfide-linked tertiary structures and binds in a ratio of approximately 1.4:1 (w/w) to protein. Thereby the charges of the proteins are masked and their migration velocity in an electric field is approximately proportional to the logarithm of their molecular mass.

Protein samples are first focussed in the top stacking gel with wide pores at pH 6.8, and then separated according to their apparent molecular sizes within a range of 1 to 100 kDa in the bottom separation gel. The comparison to a marker lane with proteins of known size allows an estimation of the size of the sample. In the present study, 15% SDS gels were used. Prior to SDS-PAGE, samples were mixed with SDS-sample buffer and boiled for 5 min. Electrophoresis was carried out with electrophoresis buffer at 120 V (stacking gel) until the samples stacked onto the separation gel and was switched to 150 V. Gels were stained with staining solution for 30 min, and destained until protein bands could clearly be distinguished from the background.

### Staining of PAGE gels

For regular staining, Coomassie Brilliant Blue R-250 based staining solution was used. For silver staining, a rapid protocol by Nesterenko et al. [123] was adopted. Procedure is described below.

Steps	Solution	Time
Fixation	60 ml acetone stock; 1.5 ml TCA stock; 25 $\mu$ l 37% HCOH	5 min
Rinse	H <sub>2</sub> O	3 $\times$ 5 s
Wash	H <sub>2</sub> O	5 min
Rinse	H <sub>2</sub> O	3 $\times$ 5 s
Pretreat	60 ml acetone stock	5 min
Pretreat	100 $\mu$ l Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub> stock	1 min
Rinse	H <sub>2</sub> O	3 $\times$ 5 s
Impregnate	0.8 ml AgNO <sub>3</sub> stock; 0.6 ml 37% HCOH; 60 ml H <sub>2</sub> O	8 min
Rinse	H <sub>2</sub> O	3 $\times$ 5 s
Development	1.2 g Na <sub>2</sub> CO <sub>3</sub> ; 25 $\mu$ l HCOH; 25 $\mu$ l Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub> stock; 60 ml H <sub>2</sub> O	10 - 20 s
Stop	1% glacial acetic acid in H <sub>2</sub> O	30 s
Rinse	H <sub>2</sub> O	10 s

**Table 3.8:** A short protocol for silver staining of polyacrylamide gels

**Solutions**

Acetone stock	50% acetone in H <sub>2</sub> O.
TCA stock	50% TCA in H <sub>2</sub> O.
AgNO <sub>3</sub> stock	20% AgNO <sub>3</sub> in H <sub>2</sub> O (store in dark).
Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub> stock	10% Na <sub>2</sub> S <sub>2</sub> O <sub>3</sub> in H <sub>2</sub> O.

**Limited proteolysis**

For limited proteolysis purified samples were digested with thermolysin, trypsin and chymotrypsin on ice and at 25 °C with protein at a concentration of 1 mg/ml in 20 mM Tris pH 7.4, 150 mM NaCl for digestion with thermolysin and similar buffer at pH 8.0 in case of digestion with trypsin and chymotrypsin. The enzyme-to-substrate ratio was 1:100 (by weight) in each case. At different time intervals, aliquots (15 µl) were sampled and boiled immediately in 10 µl 2x SDS sample buffer for 5 min to stop the proteolysis and stored at -20 °C. The protein samples were analyzed on SDS-PAGE.

**Mass spectrometry**

Proteins were identified by mass spectrometry in the core facility for protein analysis at MDC, Berlin. Bands of interest were excised from polyacrylamide gels, digested with trypsin and the resulting peptides analyzed by MALDI-MS peptide-mass ‘fingerprinting’ under reducing conditions. The identity of the peptides was confirmed by MALDI-TOF tandem-mass spectrometry as described [124].

**Western transfer**

Proteins that were separated by SDS-PAGE were transferred onto nitrocellulose or polyvinylidene fluoride (PVDF) membranes (Western blotting) for immunodetection. PVDF membranes were activated by soaking in methanol prior to usage. SDS gel and activated membrane were soaked in semi-dry transfer buffer and placed between presoaked filter paper. Electrotransfer of the proteins was performed for 1 h at a constant current of 1 mA/cm<sup>2</sup> in a semi-dry blotting chamber.

**Immunodetection**

Proteins immobilized on membranes were detected with specific antibodies against the epitope tags. The membrane was blocked with PBS-T with 3% BSA to saturate unspecific binding sites. Primary antibodies diluted in PBS-T with 1.5% BSA (α-GADD45γ 1:800) were incubated with the blocked membranes for 1 h at room temperature or over night at 4 °C. The primary antibody

solutions were collected and re-used. After washing three times for 15 min with PBS-T, the membrane was incubated for 1 h at room temperature with secondary antibody (diluted 1:10000 in PBS-T with 1.5% BSA) that is coupled with horseradish peroxidase (HRP).

For detection of His-tags, an  $\alpha$ -His-HRP conjugate (Qiagen) was used according to the manufacturer's instructions. Membranes were again washed three times for 15 min with PBS-T. Visualization was performed with Enhanced Chemiluminescence (ECL).

### **Association studies**

For His-pulldowns, the His-tagged ligand-binding domain of PPAR $\gamma$  was pre-incubated with untagged GADD45 $\gamma$  protein (WT and mutants) in 1:5 molar ratios at 4 °C for 2 h. Afterwards, the mixture was incubated with pre-washed Talon resin for 3 h at 4 °C. Beads were washed 3 times with Talon-wash buffer and eluted with Talon-elution buffer. Beads after elution were also collected to run as a control. Samples were then subjected to SDS-PAGE and Western blot analysis.

## **3.3 Spectroscopic methods**

Special care was taken that all measurements were performed immediately after gel filtration. In case that the sample had to be shipped to collaboration partners, the protein was kept at 4 °C and sterile-filtered (0.05  $\mu$ m) or centrifuged before the measurements to remove any aggregates.

Analytical ultracentrifugation was performed by Dr. Joachim Behlke at MDC, Berlin.

Small-angle X-ray scattering (SAXS) experiments were performed and analyzed by Dr. Dmitri Svergun and Dr. Haydn Mertens at EMBL-DESY, Hamburg.

### **3.3.1 Analytical ultracentrifugation (AUC)**

AUC is used to study macromolecules in solution under the influence of a strong gravitational force [125]. Most macromolecules have a density different from the solvent surrounding them and will sediment in a sufficiently strong field. Analyzing the dynamic behavior by determining sedimentation velocity until sedimentation equilibrium is reached provides information about size, shape, density and conformational changes in proteins. The main use of sedimentation velocity experiments is to examine whether a protein sample is homogeneous in solution or a mixture of forms (e.g. monomer/dimer or higher oligomers). The study of sedimentation equili-



rium is particularly useful for the determination of molecular masses or subunit stoichiometry in solution.

Sedimentation equilibrium experiments were performed in a XL-A type analytical ultracentrifuge (Beckman, Palo Alto). About 70 µl of protein sample (loading concentration 0.2-0.5 mg/ml in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl) was centrifuged in externally loaded six-channel cells with a 12 mm optical path, for 2 h at 26,000 rpm (overspeed), and then for 26 h at an equilibrium speed of 22,000 rpm at 10°C. Radial absorbance distributions of each sample at sedimentation equilibrium were recorded at 275, 280 and 285 nm respectively and fitted globally to equation

$$A_r = A_{rm} e^{MF} \quad \text{Equation 3.1}$$

with

$$F = \frac{(1 - \rho v) \omega^2 (r^2 - r_m^2)}{2RT} \quad \text{Equation 3.2}$$

with the program POLYMOLE (Behlke et al., 1997). In these equations,  $M$  is molecular mass,  $\rho$  is the solvent density,  $v$  is the partial specific volume,  $\omega$  is the angular velocity,  $R$  is the gas constant and  $T$  the absolute temperature.  $A_r$  is the radial absorbance and  $A_{rm}$  represents the corresponding value at the meniscus position. When proteins adopt monomer–dimer equilibrium, the molecular mass  $M$  can be treated as a weight average parameter ( $M_w$ ). In the case of GADD45γ, this value is a composite of the molecular mass values  $M_m$  (here 17) or  $M_d$  and the partial concentrations of monomers,  $c_m$ , and dimers,  $c_d$ , according to equation

$$M_w = \frac{c_m M_m + c_d M_d}{c_m + c_d} \quad \text{Equation 3.3}$$

from which the equilibrium constant  $K_d = c_m^2 / c_d$  can be determined.

### 3.3.2 CD spectroscopy

Circular dichroism (CD) is a spectroscopic technique that allows the estimation of the secondary structure content of a protein in solution and requires protein amounts in the microgram range [126]. It relies on the interaction of circularly polarized light with optically active (chiral) compounds, as it is the case for proteins because of the peptide bonds. CD measures the wave-

length-dependent differential absorbance of left and right circularly polarized light, because the absorbance of left circularly polarized light differs from that of right circularly polarized light in all optically active substances:

$$\Delta A_{\lambda} = A_{\lambda,l} - A_{\lambda,r} \quad \text{Equation 3.4}$$

$A$  - absorbance, l and r – indicate left and right circular polarization at a given wavelength  $\lambda$  [nm].

For studying structural changes using CD spectroscopy, the molar dichroism ( $\Delta \varepsilon_{\lambda}$ ) is used, which is concentration-independent :

$$\Delta A_{\lambda} = (\varepsilon_{\lambda,l} - \varepsilon_{\lambda,r})cl \quad \varepsilon - \text{molar absorption coefficients,} \quad \text{Equation 3.5}$$

$$\Delta \varepsilon_{\lambda} = \varepsilon_{\lambda,l} - \varepsilon_{\lambda,r} = \frac{\Delta A_{\lambda}}{cl} \quad c - \text{concentration [} \frac{\text{mol}}{\text{l}} \text{]} \quad \text{Equation 3.6}$$

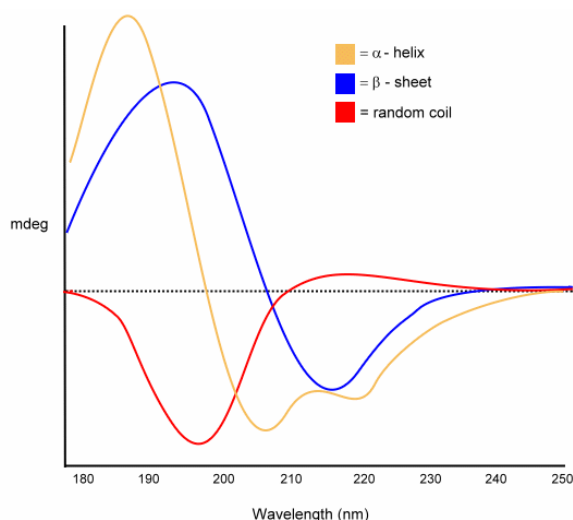
$l$  - optical path length [cm]

For historical reasons, the molar dichroism ( $\Delta \varepsilon_{\lambda} = \varepsilon_{\lambda,l} - \varepsilon_{\lambda,r}$ ) is reported in degrees of ellipticity.

Dichroism and ellipticity values can be interconverted:

$$\theta_{\lambda} = 3298 \cdot \Delta \varepsilon_{\lambda} \quad \Delta \varepsilon_{\lambda} - \text{molar dichroism, } \theta_{\lambda} \text{ molar ellipticity.} \quad \text{Equation 3.7}$$

There are two types of optically active chromophores in proteins, side groups of aromatic amino acid residues and peptide bonds. CD spectra in the far-UV region (peptide part of the spectrum, 180–250 nm) reflect the symmetry of the peptide bond environment and indicate the content of secondary structure in the protein molecule. In the far-UV region, the different forms of regular secondary structure found in proteins i.e  $\alpha$ -helix,  $\beta$ -sheet, and random coil each give rise to a characteristic CD spectra. This is illustrated by the figure 3.2, which shows spectra for poly-lysine in these three different conformations. Far-UV CD also enables discrimination between random coils and premolten globules, based on the ratio of the ellipticity values at 200 and 222 nm.



**Figure 3.2:** Secondary structure profile of helix, sheet and random coil as determined by CD spectroscopy in the "far-uv" spectral region (190-250 nm). At these wavelengths the chromophore is the peptide bond, and the signal arises when it is located in a regular, folded environment.

The approximate fraction of each secondary structure type that is present in any protein can thus be determined by analyzing its far-uv CD spectrum as a sum of fractional multiples of such reference spectra for each structural type.

CD spectra covering a range of 200 to 250 nm (far UV) were recorded using a spectropolarimeter (J-720, JASCO) in  $1 \times 1$  cm quartz cuvettes of 1 cm optical path length, filled with a total volume of 3.4 ml. The recording speed was 20 nm / min, the filter bandwidth was 1 nm at a resolution of 1 nm, the damping was set to 1 s. The protein concentration was adjusted to 100  $\mu$ M and buffer was exchanged to PBS. Spectra were recorded at 15  $^{\circ}$ C and processed by subtracting buffer spectra recorded under the same conditions.

### 3.3.3 Dynamic light scattering (DLS)

Dynamic light scattering (DLS) is a method used to determine the size distribution profile of small particles in solution. When a colloidal dispersion is irradiated with light having a wavelength much larger than the size of the particles to be measured, light is scattered uniformly in all directions (Rayleigh scattering). Depending on their size, the particles undergo Brownian motion. As a consequence, the distance between particles varies over time, leading to changes in interference of scattered light, thus fluctuations in the scattering. These fluctuations can be measured when coherent and monochromatic (laser) light is used, and allows the determination of the diffusion coefficient  $D$  of the particle. DLS measures the intensity of light scattered by molecules in solution, i.e the translational diffusion coefficient. This is related to the hydrodynamic radius ( $R$ ) of the molecule by the Stokes-Einstein equation. If aggregation occurs between molecules, it is recognizable by an increase in  $R$ .

$$R = \frac{6\pi\eta D}{kT} \quad \eta - \text{viscosity of the medium} \quad \text{Equation 3.8}$$

The DLS profile of a protein is highly predictive of its crystallizability. Proteins with singular distribution have a higher probability of crystallization while proteins with non-specific aggregates are less apt to crystallize and the crystals are of poor quality. In the present study, DLS was used for sensing the presence of aggregated protein. Prior to crystallization all protein samples were checked with the Laser-Spectroscatter 201, RiNA GmbH, Berlin.

### 3.3.4 Fluorescence spectroscopy

Fluorescence from the amino acid tryptophan has long been known to be sensitive to the polarity of its local environment. Tryptophan is intrinsically fluorescent due to its aromatic character and can be directly used as fluorophore to detect conformational changes. Among the properties used are changes in the fluorescence intensity, wavelength maximum ( $\lambda_{\text{max}}$ ), band shape, anisotropy, fluorescence lifetimes, and energy transfer. They are applied to study folding/unfolding, substrate binding, external quencher accessibility, etc. Tryptophan can be selectively excited at 295 nm and, depending on the polarity of the local environment, has an emission peak ranging from ca. 300 to 350 nm which roughly correlates with the degree of solvent exposure of the chromophore. Tryptophan fluorescence is more useful than tyrosines and phenylalanines, as these display lower extinction coefficients and quantum yields. In addition, proteins tend to contain fewer tryptophan residues than tyrosines and phenylalanines, and therefore, more specific local information can be attained. However, tryptophan fluorescence is strongly influenced by the proximity of other residues (i.e., nearby protonated groups such as Asp or Glu can cause quenching of Trp fluorescence). Also, energy transfer between tryptophan and the other fluorescent amino acids is possible, which would affect the analysis. The use of intrinsic fluorescence for the study of protein conformation is in practice therefore limited to cases with few (or perhaps only one) tryptophan residues. The intensity of fluorescence is not very informative in itself. The wavelength of the emitted light is a better indication of the environment of the fluorophore. Free single tryptophan residues that are exposed to water, have maximal fluorescence at a wavelength of about 340-350 nm, whereas totally buried residues fluoresce at about 330 nm. Protein fluorescence is often used to probe protein stability and conformational dynamics [127]. As it is highly sensitive it demands relatively small amounts of material. Finally, the time scale of

fluorescence emission is in the nanosecond range, which allows fast data acquisition and makes it especially suitable for kinetic experiments.

Tryptophan fluorescence for GADD45 $\gamma$  was measured using a spectrofluorimeter equipped with a thermostate cuvette (FP 6500, JASCO). All experiments were carried out at 10°C. The excitation and emission wavelengths were set to 280 nm and 340-400 nm, respectively, and their filter bandwidths were set to 5 nm and 10 nm. The photomultiplier voltage was set to manual mode and a range of 750 to 1000 V was used, depending on the required sensitivity of the spectrofluorimeter. The protein concentration was adjusted within a range of 270  $\mu$ M – 270 nM, based on the  $K_D$  value. For lower concentrations, higher voltage had to be used in order to improve the fluorescence signal. Experiments were carried out in a 1  $\times$  1 cm quartz cuvette, filled with a total volume of 1,750  $\mu$ l of measuring buffer. For each concentration, the fluorescence was recorded in duplicate, each batch 3 times over a period of 10 s. Average values of fluorescence were calculated, which were corrected for the dilution. Fluorescence data were corrected by subtracting buffer which was determined in separate experiments in the absence of the protein. Corrected fluorescence data were normalized to initial protein fluorescence.

### 3.3.5 Small-angle X-ray scattering (SAXS)

Small-angle X-ray scattering (SAXS) is a tool for structure analysis of native biological macromolecules in solution, from individual proteins to large complexes [128],[129]. In contrast to high-resolution structure determination by X-ray crystallography, which is limited in obtaining good quality crystals, and NMR spectroscopy, whose limitation is the analysis of the proteins bigger than  $\approx$  30 kDa (limit for high resolution), SAXS is a technique for obtaining global size and shape information of native biological macromolecules in solution with sizes ranging from a few kDa to several mDa [130]. Especially concerning large macromolecular assemblies or polymers, whose high-resolution subunit structures are known, SAXS offers complementary information on structure assembly in the aggregation state. In addition, it is a non-destructive method, in contrast to electron microscopy techniques, and require a homogeneous dilute solution of macromolecules in a near-physiological buffer without special additives. In a SAXS experiment, the sample is exposed to X-rays and the scattered intensity,  $I(s)$ , is recorded as a function of the momentum transfer  $s$  ( $s = 4\pi\sin\theta/\lambda$ , where  $2\theta$  is the angle between the incident and scattered radiation) [131]. The obtained scattering curves can then be fitted to ideal scattering curves of structures with known geometries. The scattering of X-rays at small angles (close to the

primary beam) provides low-resolution structural information on the overall shape and internal structure in the absence of crystals.

All experimental data were recorded at the EMBL beamline X33 on the DORIS storage ring at DESY in Hamburg. The exposure time for all data collections was 120 s and the sample used for this experiment was in a vacuum cuvette with two 25  $\mu\text{m}$  thick mica windows. The sample thickness of this cuvette is 1 mm and the sample volume approximately 50  $\mu\text{l}$ . The cuvette was cleaned between each measurement with ethanol, HCl, rinsed with distilled water and dried using nitrogen gas. The corresponding buffer, which consisted of 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 2 mM DTT, was measured before and after each protein sample. The experimental data were radially integrated to give a 1D X-ray scattering profile of intensity against scattering angle and normalised for changes in intensity caused by the decay of the circulating current in the storage ring. The average of the scattering from the buffer measured before and after each sample was subtracted from the scattering of the protein solution leaving only scattering from the protein under investigation.

Since the samples could be damaged by synchrotron radiation during the collection time, 2 mM DTT were added to all samples before measurement in order to minimise the effects. The degree of dose dependent aggregation caused by radiation damage was checked for each protein complex using multiple exposures of the same sample and shown to be minimal for the exposure time used in this experiment. For each sample a concentration series was measured in order to verify that the solutions were monodisperse and to check whether concentration-based interactions were observable in the data. The sample concentrations were determined directly at the beamline using the nanodrop spectrophotometer. Calibration of the extrapolated scattering at zero angle with molecular weight was checked using a 4 mg/ml solution of bovine serum albumin (BSA).

### **3.3.6 Thermal shift assay**

The thermal shift assay is a protein stability assay that uses an environmentally sensitive fluorescent dye to monitor protein thermal unfolding and directly measures stability changes due to different buffer conditions or binding of compounds. Since protein stability varies with solution conditions, either due to a binding event or due to a bulk solvent, various buffers were tested to find out the best conditions for protein stability.

The thermal stability of proteins was measured with the 7500 Fast Real-Time PCR System (Applied Biosystems) in a 96-well plate using buffers with varying concentrations of NaCl (50-500 mM), glycerol (2-10 %), different pH values (5-9) and 2 additives (Additive Screen, Hampton Research). Each well contained 9  $\mu$ l buffer, 1  $\mu$ l protein (1-2  $\mu$ g/ $\mu$ l) and 1  $\mu$ l of 50x SYPRO Orange dye (Invitrogen). The program consisted of three steps, whereas step 1 was a pre-incubation for 1 min at 20 °C and step 2 and 3 were cycles comprising the temperature increase of 1 °C within 20 s. The temperature gradient proceeded from 20 to 90 °C and included 70 cycles.

### **3.4 Crystallography methods**

#### **3.4.1 Crystallization and data collection**

Crystals of both native and Se-Met derivatized GADD45 $\gamma$  were obtained by the sitting-drop method of vapor diffusion using a 96-well Greiner plate (Crystal Quick™ low profile) at 20 °C and drops containing 400 nl of protein (30 mg ml<sup>-1</sup>) plus 400 nl of reservoir solution equilibrated against 75  $\mu$ l of reservoir solution [132]. All pipetting steps were carried out using the Hydra Plus One crystallization robot (Matrix Technologies, Hudson, NH, USA). The best native crystals were obtained in 2.0 M Na-malonate pH 5.5, 0.125 M Na-benzoate pH 5.5 and grew within 7 to 14 days. SeMet-derivatized crystals were grown from a protein solution at 32.9 mg/ml containing 400 nl protein and 400 nl of 2.2 M Na-malonate pH 5.5, 0.05 M Na-benzoate pH 5.5 at 20 °C. Crystals were transferred to a cryo-protectant solution consisting of 2.2 M Na-malonate pH 5.5, 0.1 M Na-benzoate and rapidly flash-cooled in a liquid-nitrogen stream at 100 K. Diffraction data to 2.3 Å resolution (native) were collected on a MAR345 imaging plate detector using the synchrotron source at a wavelength of 0.90832 (beamline PSF-ID14.2 at BESSY, Free University, Berlin [133]). Multiwavelength anomalous diffraction data at the Se edge were collected at 100 K to 2.97 Å on the MARCCD detector at beamline PSF-ID14.1, BESSY, Berlin. The measured data were integrated, scaled and merged using the program HKL2000 [Otwinowski & Minor, 1997].

### 3.4.2 Crystal structure determination

The phase problem was solved by MAD phasing using datasets of Selenomethionine substituted crystals. All 3 selenium atom sites expected in the asymmetric unit were assigned using SHELXD [134]. Initial phases were calculated using SOLVE [109]. The phases were subsequently improved by density modification implemented in RESOLVE [14]. Model building was improved using the program O [113] and later COOT [135]. After preliminary refinement of the partial structure, the model was placed into the isomorphous unit cell of the high-resolution native protein data and subsequently refined using REFMAC5 [136]. After density modification and solvent flattening structural refinement was carried out by simulated annealing with Crystallography and NMR system or CNS [137] and then used as a starting model for autotracing of additional amino acids using the program ARP/wARP [138].  $2F_o - F_c$  and  $F_o - F_c$  difference electron density maps were calculated based on atomic coordinates and X-ray diffraction data using the FFT program [139]. Maps were extended using the Extend program [139] and converted into the BRIX map format using the mapman program [140]. The O software [141] was used to display density contour maps and atomic models in real space. TLS refinement followed, using the TLSANAL program from the CCP4 suite [139]. PROCHECK [142] and WHATCHECK [143] programs were used to check the model for further structural inconsistencies (short interatomic distances, unsatisfied hydrogen bonding, improper geometries of functional groups, overall packing quality, backbone torsion-angle distribution and quality). All molecular graphics in this work have been generated using the Pymol software [144]. Electrostatic surface potentials were calculated using the adaptive Poisson-Solver (APBS) [145]. Molecular properties such as interatomic distances, hydrogen-bonding patterns, van-der-Waals contacts, symmetry-related interactions, *etc.*, were evaluated using the EBI-PISA server [158].



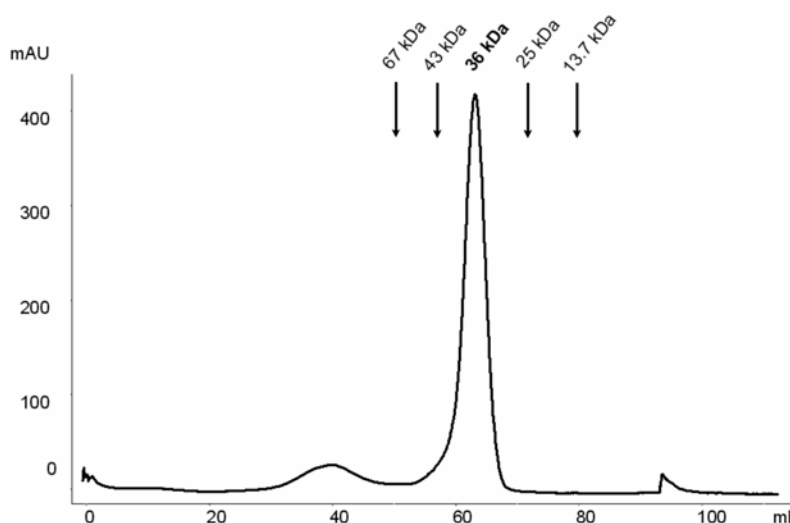
## 4. The architecture of human GADD45 $\gamma$

### 4.1 Results

#### 4.1.1 GADD45 $\gamma$ is a dimer in solution

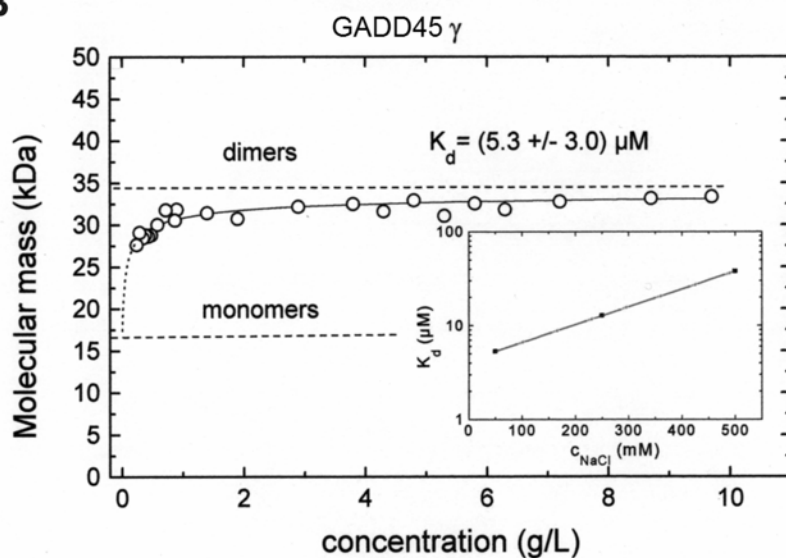
The molecular mass of the GADD45 $\gamma$  monomer is 17 kDa. The apparent molecular mass as determined by analytical size exclusion chromatography gave a value between 34 - 36 kDa which is suggestive of a dimer (Fig. 4.1A). The dimeric arrangement is also consistent with the finding from analytical ultracentrifugation (Fig 4.1B).

**A**



**Figure 4.1: GADD45 $\gamma$  is dimeric in solution.** A: Elution profile of GADD45 $\gamma$  from analytical gel-filtration column (Superdex -75) calibrated with the following marker proteins – albumin ( $M_r$  67,000), ovalbumin ( $M_r$  43,000), chymotrypsin ( $M_r$  25,000) and ribonuclease ( $M_r$  13,700).

**B**



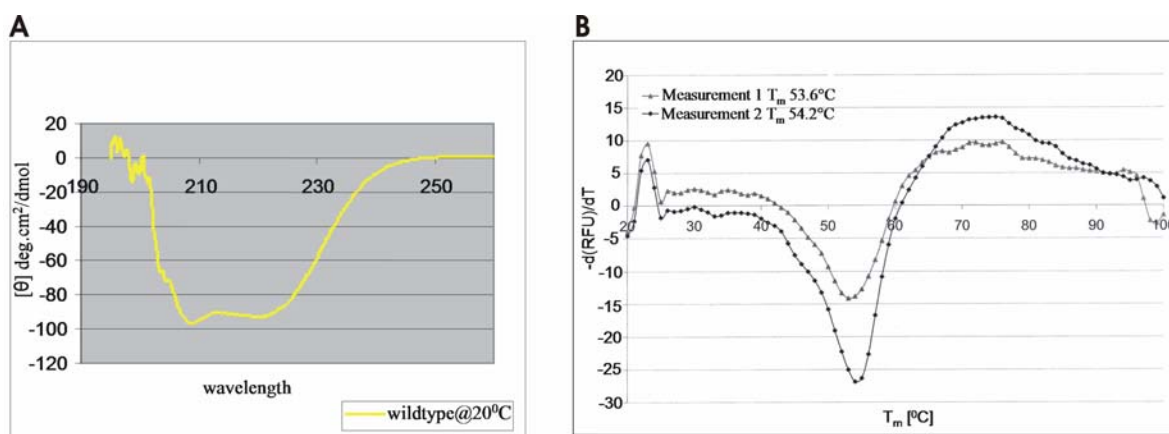
**B:** Analytical sedimentation equilibrium ultracentrifugation of molecular mass versus protein concentration. The dashed horizontal lines indicate the theoretical molecular masses of the dimer. **Inset:** Dissociation constants of GADD45 $\gamma$  at different NaCl concentrations.

In the concentration range examined, the apparent molecular mass was found to be 28–33 kDa. These values lie between the theoretical molecular masses for the monomer (17 kDa) and the dimer (34 kDa) and, therefore, support monomer-dimer equilibrium. An average dimer dissociation constant of  $K_d = 5.3 \pm 3.0 \mu\text{M}$  derived from the concentration-dependent  $M_r$  measurements was obtained at 50 mM NaCl, indicating predominant dimeric arrangement in this concentration range. The binding constant was found to be dependent on ionic strength. The dissociation constant was determined as a function of the salt concentration, and a pronounced correlation was found (Figure 4.2B, inset). The dissociation constant is increased eight fold – from 5  $\mu\text{M}$  to 40  $\mu\text{M}$  – when the salt concentration is increased from 50 to 500 mM. This is consistent with the observation of stabilizing charge-charge interactions between the subunits, which would be screened and thereby weakened by increasing the ionic strength.

However, the deduced dissociation constant of  $5.3 \pm 3.0 \mu\text{M}$  is quite high and implies that GADD45 $\gamma$  dimerization is weak and perhaps transient in the cell.

#### 4.1.2 GADD45 $\gamma$ is folded and predominantly helical

To assess the secondary structure of GADD45 $\gamma$ , CD spectra in the far-UV region were recorded which revealed that GADD45 $\gamma$  is structured (Fig. 4.2A). The spectrum exhibited two negative peaks at 209 and 219 nm indicative of a predominantly folded structure with high  $\alpha$ -helical content ( $\sim 45\%$ ). The structure was also highly stable as shown by determination of the melting temperature which resulted in a  $T_m$  of  $\sim 54^\circ\text{C}$  (Fig. 4.2B).



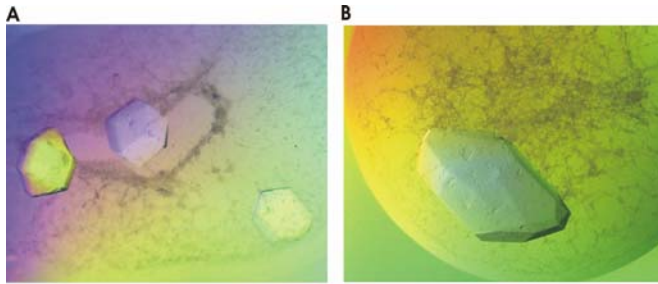
**Figure 4.2:** CD spectra and melting point determinations of human GADD45 $\gamma$ .

**A)** The molar ellipticity as measured by CD analysis shows the protein to be folded and predominantly helical.

**B)** de convolution of the first derivative of the melting curve as determined by thermofluor analysis. The temperature was increased at a rate of  $1^\circ\text{C}$  per minute. The protein is highly stable with a  $T_m$  of about  $54^\circ\text{C}$ .  $T_m [^\circ\text{C}]$  – melting temperature,  $-d(\text{RFU})/dT$  – change in relative fluorescence unit with changing temperature. A sample's  $T_m$  value is calculated as the temperature at which the melt curve has the lowest (i.e the most negative)  $-d(\text{RFU})/dT$  value.

### 4.1.3 Determining the crystal structure of GADD45 $\gamma$

Both native and Se-Met derivatized crystals of recombinant GADD45 $\gamma$  were obtained by sitting-drop at 20 °C with 1:1 protein:reservoir solution ratio. Native crystals were grown against 2.0 M Na-malonate pH 5.5 and 0.125 M Na-benzoate pH 5.5. While SeMet labelled GADD45 $\gamma$  was crystallized in 2.2 M Na-malonate pH 5.5, 0.05 M Na-benzoate pH 5 (Fig. 4.3).



**Figure 4.3:** A. Crystals of native GADD45 $\gamma$  grown in 2 M malonate, 0.1 M benzonate, pH 5.5  
B. Crystals of SeMet-GADD45 $\gamma$  grown in 2.2 M malonate, 0.05 M benzonate, pH 5.5

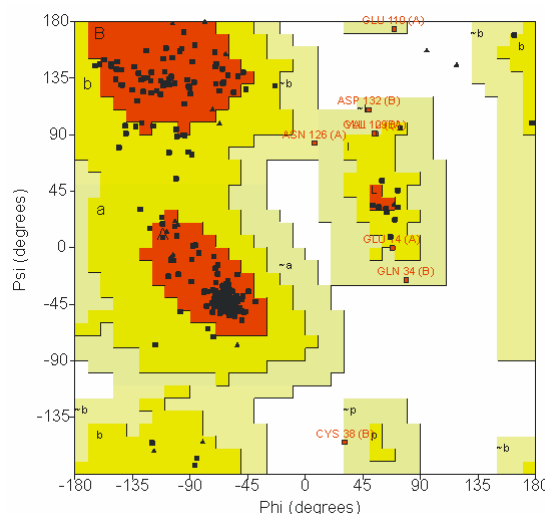
Diffraction data from a native crystal to 2.3 Å and SeMet labeled crystal to ~3 Å were collected at BESSY (beamlines BL14.1 and 14.2), Berlin. Both native and SeMet labeled crystals belonged to space group P3<sub>2</sub>21 with cell parameters  $a = b = 108.72$  Å and  $c = 73.37$  Å. At the time of this work, no known homology model was available. Therefore multiple-wavelength anomalous diffraction (MAD) phasing was used to solve the phase problem; datasets were collected at peak, inflection point, and remote wavelengths for MAD phasing. The solvent content was calculated to 66% ( $V_M = \text{Å}^3/\text{Da}$ , Equation 1.9) assuming 2 protein molecules in the asymmetric unit. The data sets were processed and scaled using the program HKL2000. Data collection statistics are listed in Table 4.1.

	Native	peak	inflection	remote
Wavelength [Å]	0.90832	0.97968	0.97980	0.9184
Space group	P3 <sub>2</sub> 21	P3 <sub>2</sub> 21	P3 <sub>2</sub> 21	P3 <sub>2</sub> 21
Resolution [Å]	2.3	2.97	2.78	2.97
Observed reflections	153516	148991	139987	117133
Independent reflections	21647	10481	12784	10505
$\langle I/\sigma(I) \rangle$	18.7	18.41 (8.47)	27.9 (2.8)	38.9 (7.04)
Averaged redundancy	7.09	14.2	11.0	11.2
Completeness ov./l. s.	93.0 (80.3)	99.5 (99.9)	99.2 (96.3)	99.4 (99.9)
Cell a [Å]	108.50	108.77	109.08	108.89
b [Å]	108.50	108.77	109.08	108.89
c [Å]	73.02	73.84	74.33	74.06
Rsym [%]:ov./l. s.	6.6 (66.4)	6.3 (33.5)	5.6 (54.1)	5.3 (36.1)

**Table 4.1:** Data collection statistics for native and SeMet-GADD45 $\gamma$  crystals. Values in parenthesis refer to the outer shell of reflections

After initial phase calculation and improvement by density modification, an initial model was built as described in method 3.4.2. The model was interactively improved using the program O [113] and COOT [136]. Structural refinement was carried out by REFMAC5 [136]. When the refinement of the protein structure could not be advanced any further, automated and manual searches of water molecules were initiated. Water positions were checked for consistency and maintained in the refinement model only, if the following criteria were fulfilled: 1) spherical electron density maxima are present at each potential water position in both difference electron density maps -  $2F_o - F_c$  map contoured at  $1\sigma$  and the  $F_o - F_c$  map contoured at  $3\sigma$ ; 2) each water molecule is connected to other solvent molecules or to polar protein groups by hydrogen bonds within a distance of 2.4 - 3.6 Å, and 3) their  $B$ -factors were not exceedingly high. Some molecules initially identified as waters could be assigned as ions, based on distance criteria, geometrical coordination of surrounding groups, and occurrence in the crystallization setups. In the structure, 14 atoms found in the protein-solvent interface were assigned as malonate. Several rounds of iterative model building and refinement using TLS refinement followed, using data between 19.8-2.4 Å resolution. 10 N-terminal residues of the recombinant construct were not observed in the electron density. The final model contains 149 residues in one monomer and 137 in the other with 54 water molecules and an  $R$ - and free  $R$ -factor of 21.5 and 25.05 %, respectively. The model has good geometry without outliers, as shown by the program PROCHECK [142]. In the Ramachandran diagram, all residues of the model are located in the most favorable or additionally allowed regions (89.5 and 10.5%, respectively) (Fig. 4.4). This proves that the weighting of the geometry term during refinement was high enough. The refinement statistics are summarized in following table.

Refinement	
Resolution [Å]	19.8-2.4
No. of reflections	18783
R <sub>work</sub> /R <sub>free</sub> (%)	21.5/25.05
No. atoms	2210
Protein	2142
Malonic acid	14
water	54
r.m.s. deviations	
Bond lengths [Å]	0.11
Bond angles (°)	1.31
Torsion angles (°)	6.97
Ramachandran plot (%)	
	82.5 (core)
	14.2 (allowed)
	3.3 (gener.)



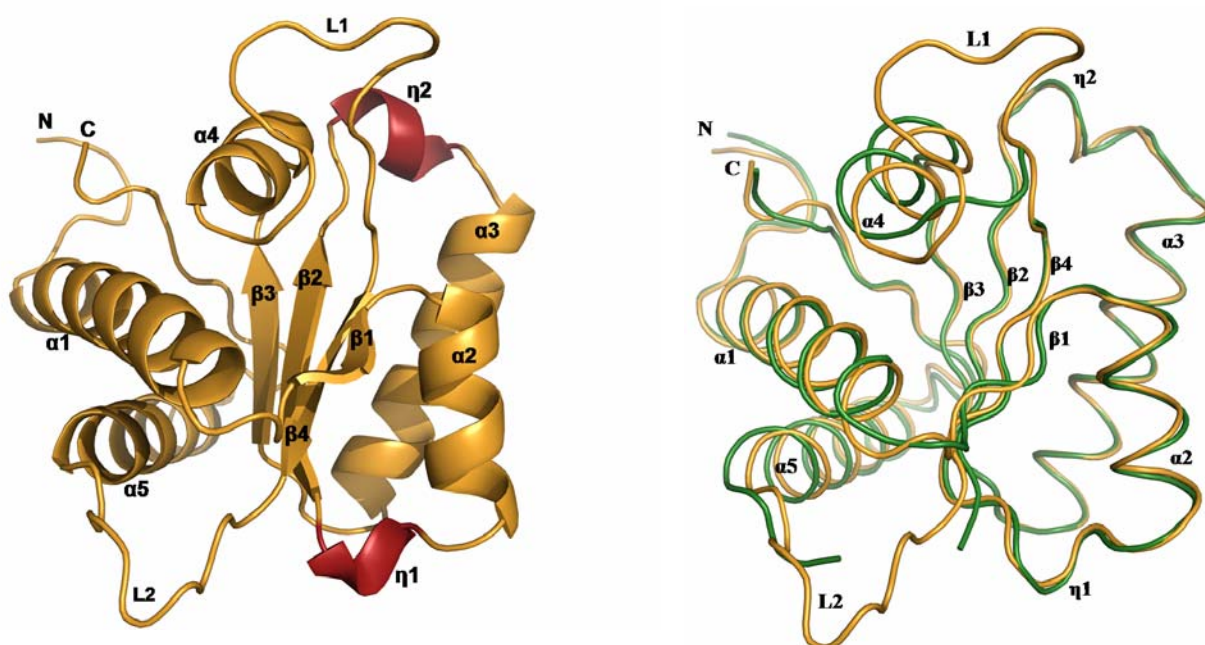
**Figure 4.4: Left) – Table showing refinement statistics of GADD45 $\gamma$  model.**

**Right) - Ramachandran plot showing the torsion angles of all peptide bonds of the GADD45 $\gamma$  model.** Triangles represent glycine and proline residues, squares all other amino acids. 82.5% of all amino acids have torsion angles in the most favoured region (red background), 14.2% in the additional allowed region (yellow background) and 3.3% in the generously allowed region (red squares). None of the residues has a phi-psi combination in the disallowed region. The plot was generated using the program Procheck (Laskowski et al., 1993).

#### 4.1.4 The structure of human GADD45 $\gamma$ : Monomer

The secondary structural elements of the monomer are arranged in an  $\alpha$ - $\beta$ - $\alpha$  sandwich fold [146],[147] with alternating  $\alpha$ -helices and  $\beta$ -strands, which account for 43 and 9.5% of all residues, respectively, separated by coiled regions (Fig 4.5). In agreement with CD analysis, the monomer is predominantly helical with 5  $\alpha$ -helices. The central core of the  $\alpha$ - $\beta$ - $\alpha$  sandwich is composed of three parallel  $\beta$ -strands ( $\beta$ 1- $\beta$ 3) and one antiparallel  $\beta$ -strand ( $\beta$ 4) topologically spaced between the  $\beta$ 1 and  $\beta$ 2 strands. In addition, there are two short  $3_{10}$  helices which are located within the  $\alpha$ 2- $\beta$ 2 and  $\beta$ 2- $\alpha$ 3 connecting loops (Fig. 4.5).

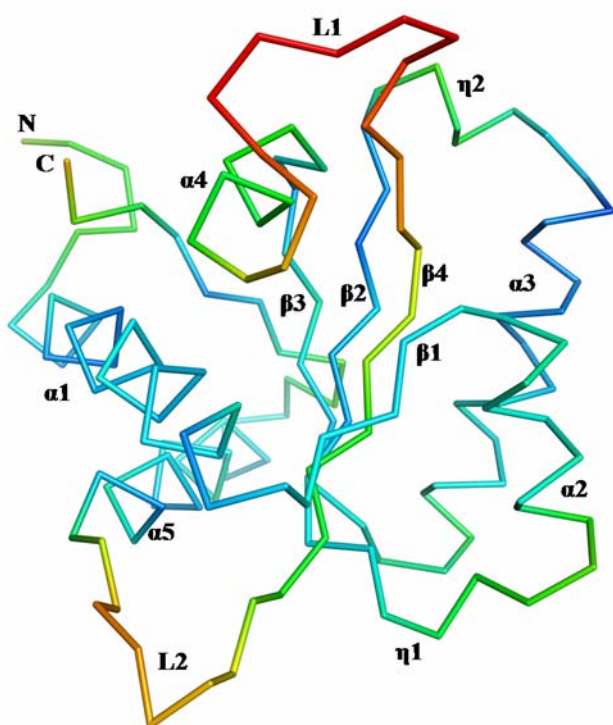
Helices  $\alpha$ 4 and  $\alpha$ 5 are parallel to one another and roughly orthogonal to the parallel pair  $\alpha$ 2 and  $\alpha$ 3. Helices  $\alpha$ 2 and  $\alpha$ 3 pack on one side of the  $\beta$ -sheet with the helix axes running roughly parallel to the  $\beta$ -strands, whereas helices  $\alpha$ 1,  $\alpha$ 4, and  $\alpha$ 5 pack on the opposite side of the  $\beta$ -sheet with helix axes running across the  $\beta$ -strands.



**Figure 4.5:** **A)** Cartoon representation of the GADD45 $\gamma$  monomer. Two  $3_{10}$  helices are colored red. **B)** Superposition of the two GADD45 $\gamma$  molecules present in the asymmetric unit of the crystal. Molecule A is colored orange and B molecule is colored green. Note the difference in N- and C-termini and in the loops.

Of the 159 residues, the 10 N-terminal residues had no electron density in both molecules. Two prominent loop regions, L1 (residues 105-118) and L2 (residues 123-132) which connect  $\alpha4$  with  $\beta4$  and  $\beta4$  with  $\alpha5$ , respectively, are not visible in the electron density map of molecule B and poorly presented in molecule A for which segmental flexibility must be assumed. There are certain differences in the conformation of both monomers, mainly due to different crystal contacts of the chain. The root-mean-square deviation (r.m.s.d.) after a superposition of the two molecules in the asymmetric unit with the program LSQKAB (Kabsch, 1976) is 0.87 Å for 125 matching  $C^\alpha$  pairs. The main conformational change occurs at the N- and C- termini and in the loops (Fig. 4.5B) which are highly dynamic as indicated by the crystallographic  $B$  values.

The temperature factor or B-factor can be thought of as a measure of how much an atom oscillates or vibrates around the position specified in the model. High B-values indicate parts of the molecule which are vibrationally active. In this way, from B-factors computed during refinement, it can be learnt which atoms in the molecule have the most freedom of movement



**Fig. 4.6:** Model of GADD45 $\gamma$  monomer colour coded according to B-factor: red (“hot”) for high B-values indicating vibrationally active; blue (“cold”) for low values of B indicating parts that are particularly rigid.

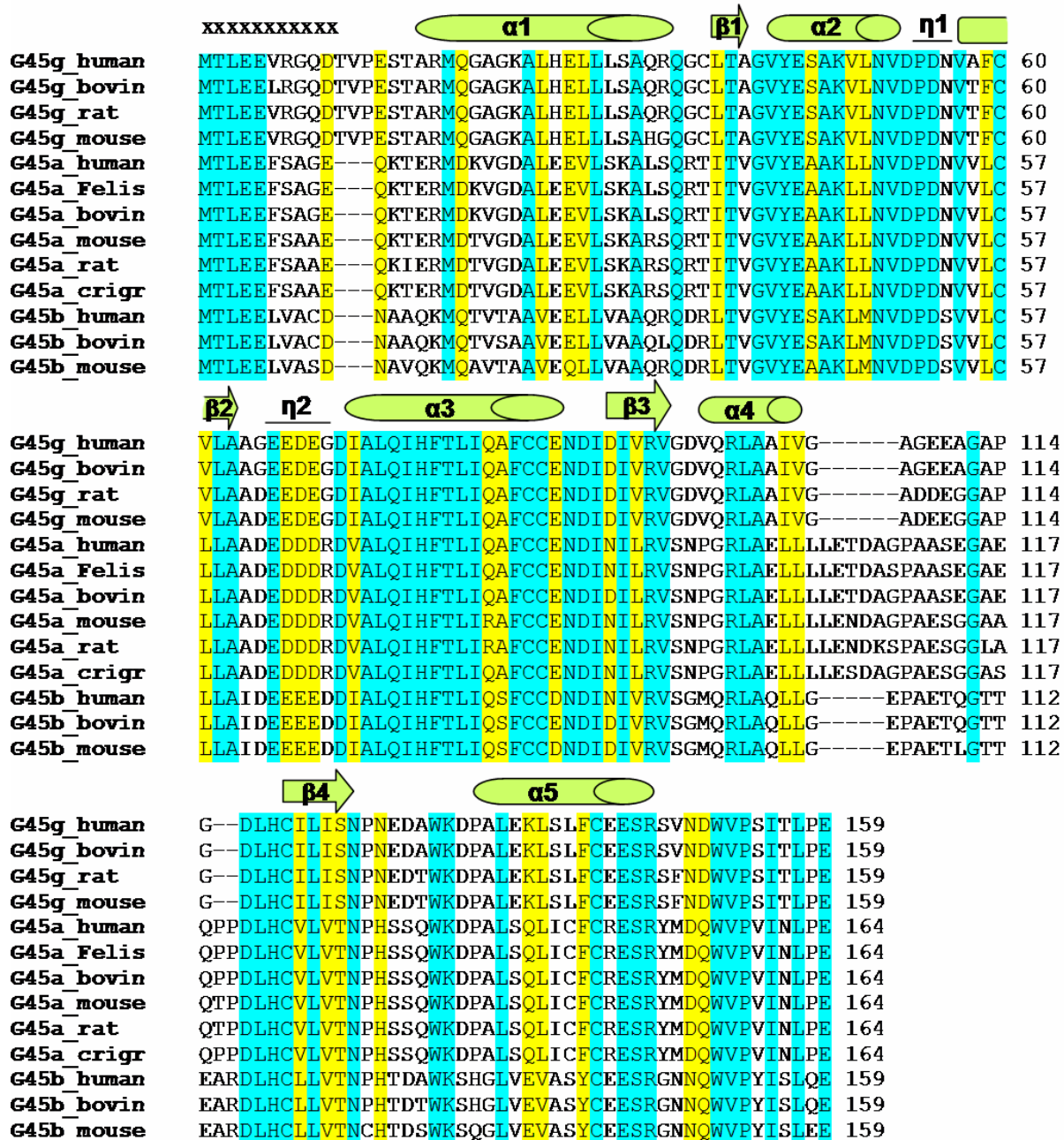
As illustrated in figure 4.6, temini and the loops in the GADD45 $\gamma$  model have high B-values indicating flexibility of these regions. Structural flexibility in these regions might acquire an order by GADD45 $\gamma$  binding to ligand proteins. The following discussion of the GADD45 $\gamma$  structure is based on molecule A which is more complete and appears more rigid than molecule B.

#### 4.1.5 Common fold

The pattern of sequence conservation shows that residues that are highly conserved among GADD45 isoforms are concentrated mostly in helices  $\alpha 2$ ,  $\alpha 3$ , the  $3_{10}$  helices and the residues in the  $\beta$ -sheet with which these helices interact (Fig 4.7).

From the sequence alignment it is clear that the isoforms differ only in the length of the sequence and principal variations are mostly concentrated in coiled regions. Given the generally high degree of sequence conservation, it is very likely that the  $\alpha/\beta$ -plait fold may represent the common fold for all GADD45 proteins and that structural variations are confined to the loop regions.





**Figure 4.7: Multiple sequence alignment of GADD45 proteins.** The secondary structural elements of GADD45 $\gamma$  shown above the aligned sequences are in agreement with figure 4.5. Identical and strongly similar residues are highlighted in cyan and yellow, respectively.  $\eta 1$  and  $\eta 2$  represent two  $3_{10}$  helices. xxx- marks missing residues in both monomers of GADD45 $\gamma$ . Multiple sequence alignment as obtained from CONSURF was prepared using CLUSTALW (Higgins et al., 1992).



#### 4.1.6 Dimerization interface(s)

The asymmetric unit of the crystal contains two monomers of GADD45 $\gamma$ . Inspection of the crystal packing showed several lattice contacts between monomers that could generate dimer. (Fig. 4.8. Example AB/BC/AH/GH etc). Some of these dimers simply result from crystal packing and do not contain two fold rotational axis which would be expected for a symmetric dimer. These types of dimers, therefore, can not exist in solution.

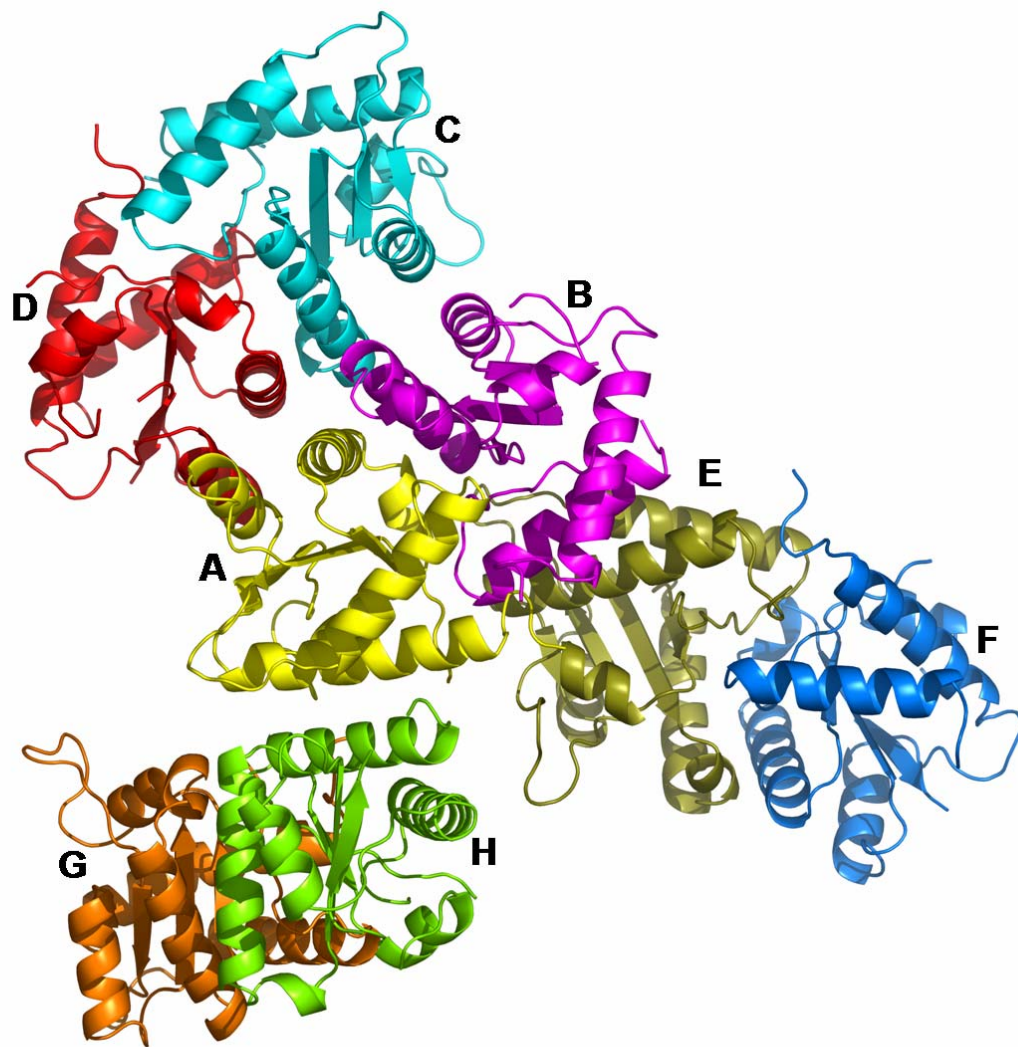


Figure 4.8: Crystal packing of GADD45 $\gamma$  showing more than one possibility of dimers

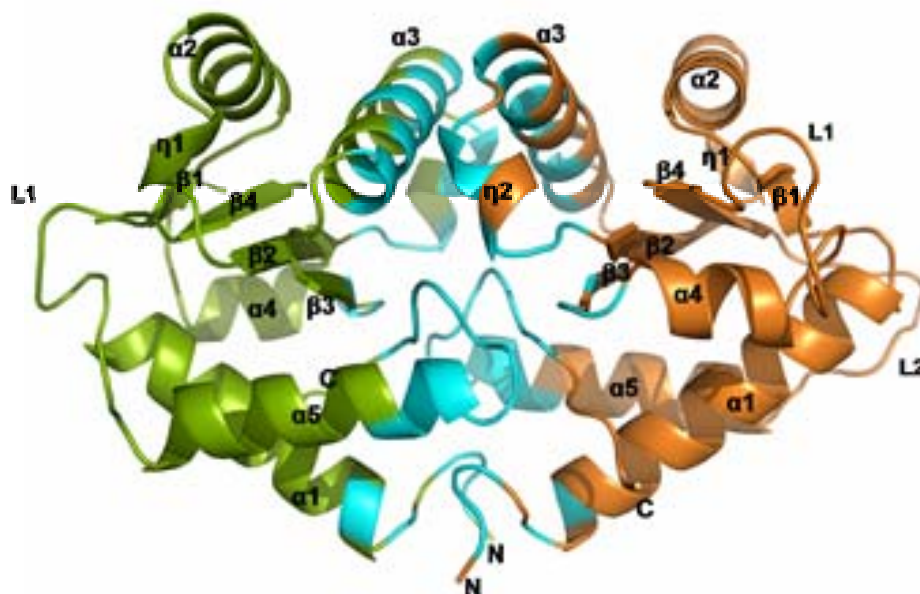
This observation was supported by a detailed analysis obtained from the PISA server from EMBL-EBI [148] which revealed that there are at least seven possible interfaces which can generate a dimer between GADD45 $\gamma$  monomers in the crystal lattice (Table 4.2). Out of which interfaces 4 - 7 (PISA reference numbering) cannot represent proper (closed) dimers of C<sub>2</sub> point symmetry.

NN	Structure 1			Structure 2			Interface area, Å <sup>2</sup>	$\Delta^iG$ kcal/mol	$\Delta^iG$ P-value	N <sub>HB</sub>	N <sub>SB</sub>	N <sub>DS</sub>	CSS
	Range	<sup>i</sup> N <sub>at</sub>	<sup>i</sup> N <sub>res</sub>	Range	<sup>i</sup> N <sub>at</sub>	<sup>i</sup> N <sub>res</sub>							
1	B	124	36	A	134	36	1304.0	-11.9	0.313	10	2	0	1.000
2	B	55	12	A	63	15	499.2	-3.2	0.553	4	6	0	0.000
3	B	44	10	A	39	10	413.8	-6.8	0.127	2	2	0	0.296
4	A	10	3	A	10	3	108.5	-3.3	0.058	0	0	0	0.060
5	A	12	4	A	14	4	101.5	0.2	0.701	1	3	0	0.000
6	B	11	3	B	11	3	90.2	-0.6	0.536	0	0	0	0.000
7	B	6	2	B	6	2	48.4	-1.5	0.155	0	0	0	0.029

**Table 4.2: Protein-protein interfaces found in the GADD45 $\gamma$  crystal; obtained by PISA web server analysis** ([http://www.ebi.ac.uk/msd-srv/prot\\_int/pistart.html](http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html)) NN - row number. Range - the selection range for the respective interfacing structure. <sup>i</sup>N<sub>at</sub> and <sup>i</sup>N<sub>res</sub> - the number of interfacing atoms and interfacing residues in the corresponding structure, respectively. Interface area in Å<sup>2</sup> -calculated as difference in total accessible surface areas of isolated and interfacing structures divided by two.  $\Delta^iG$  - the solvation free energy gain upon formation of the interface, in kcal/M.  $\Delta^iG$  P-value - the P-value of the observed solvation free energy gain; P-value is a measure of interface specificity, showing how surprising, in energy terms, the interface is. N<sub>HB</sub> indicates the number of potential hydrogen bonds across the interface. N<sub>SB</sub> indicates the number of potential salt bridges across the interface. N<sub>DS</sub> indicates the number of potential disulfide bonds across the interface. CSS stands for the Complexation Significance Score, which indicates how significant for assembly formation the interface is.

### Interface 1

The solvent-accessible surface (SAS) area buried in interface 1 (PISA reference numbering; henceforth denoted interaction 1) is 1316 Å<sup>2</sup>, corresponding to 16% of the total monomer surface. The dimer is primarily formed by antiparallel contacts involving  $\alpha 3$  and  $\eta 2$  (second 3<sub>10</sub> helix), the N- terminal segment of  $\alpha 1$  and the residues following  $\alpha 5$  (Fig. 4.9). The two monomers of the dimer are related by a two-fold rotational axis. The interface is stabilised by 10 hydrogen bonds, listed in Table 4.3.



**Figure 4.9: Cartoon representation of model 1.** Chain A is coloured orange and chain B is coloured green. Residues involved in this particular interface are shown in cyan.

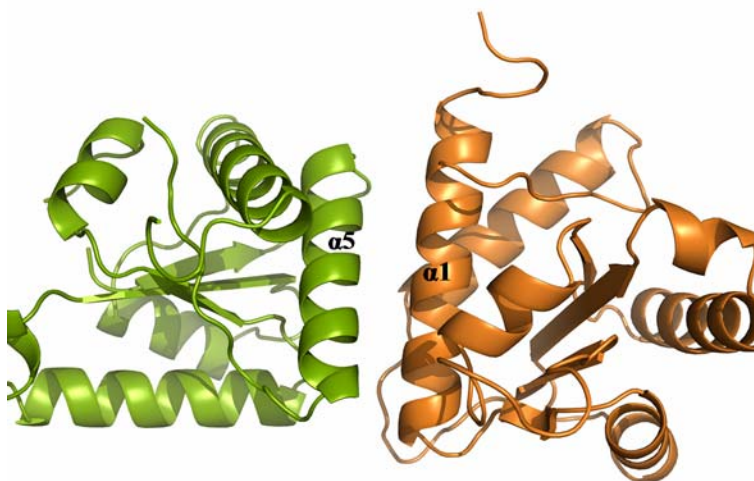
The subunit interaction is further stabilized by two monodentate salt bridges involving residues Arg94 (chain B) and Asp150 (chain A) which do not follow the symmetric nature of the dimer i.e this salt bridge interaction is seen only once in the dimer interface 1. Some of the intermolecular interactions are also found to be asymmetric. However, local asymmetry in a symmetric dimer is not uncommon especially when protomers are related by non-crystallographic (NCS) symmetry.

Subunit A	Subunit B	Type of interaction
Ala64 O	Trp151 NE1	H bond
Glu69 OE1	Gln82 NE2	H bond
Thr79 OG1	Thr79 OG1	H bond
Ser144 OG	Glu13 OE2	H bond
Asn149 ND2	Asp97 OD2	H bond
Asp150 OD1	Arg94 NH1	H bond
Trp151 NE1	Gly96 O	H bond
Trp151NE1	Ala64 O	H bond
Ser154 OG	Asp150 OD2	H bond
Thr156 OG1	Val148 O	H bond
Asp150 OD1	Arg94 NH1	Salt bridge
Asp150 OD1	Arg94 NH2	Salt bridge

**Table 4.3:** Residues involved in hydrogen bonds and salt bridges in interface 1.

## Interface 2

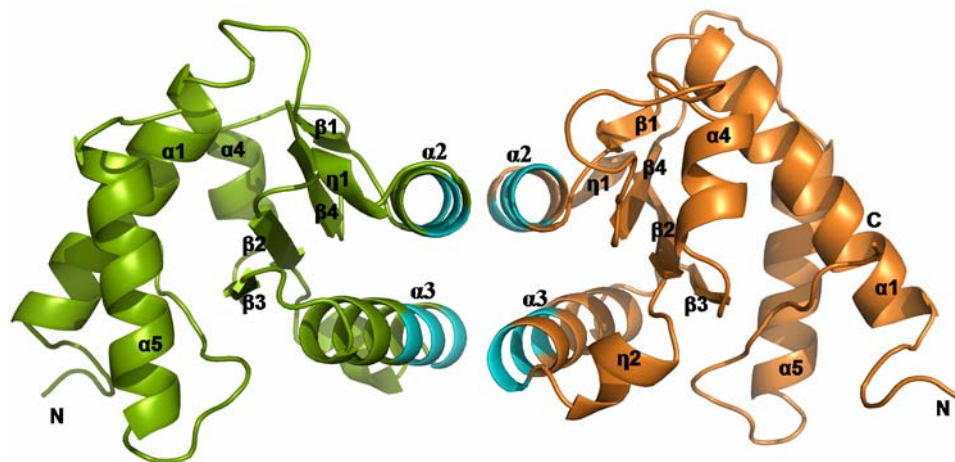
Interface 2 (PISA reference numbering) cannot occur in a proper (closed) dimer, as the dimer resulting from this contact is asymmetric. The interface shows that  $\alpha 1$  of chain A interacts with  $\alpha 5$  of chain B in a way which leaves free interacting surfaces on both chains which would lead to aggregation (Figure 4.10). Therefore, this crystal contact does not support dimerization in solution.



**Figure 4.10:** Cartoon representation of interface 2.

## Interface 3:

Interface 3 / interaction 3 (PISA reference numbering) has a buried SAS per monomer of  $\sim 400$  Å. This interface is a four-helix bundle consisting of  $\alpha 2$  and  $\alpha 3$  with predominantly hydrophobic interactions. The two monomers are related by a twofold symmetry axis that is perpendicular to the helix axes (Fig. 4.11).



**Figure 4.11: Cartoon representation of interface 3.** Chain A is colored orange and chain B is colored green. Residues involved in this particular interface are shown in cyan. The twofold rotational axis relating the monomers is oriented vertical.

Two symmetric hydrogen bonds and salt bridges contribute to the interfacial contact area (Table 4.4).

Subunit A	Subunit B	Type of interaction
Tyr44OH	Asn51 OD1	H bond
Asn51 OD1	Tyr44 OH	H bond
Glu45 OE2	Lys48 NZ	Salt bridge
Lys48 NZ	Glu45 OE2	Salt bridge

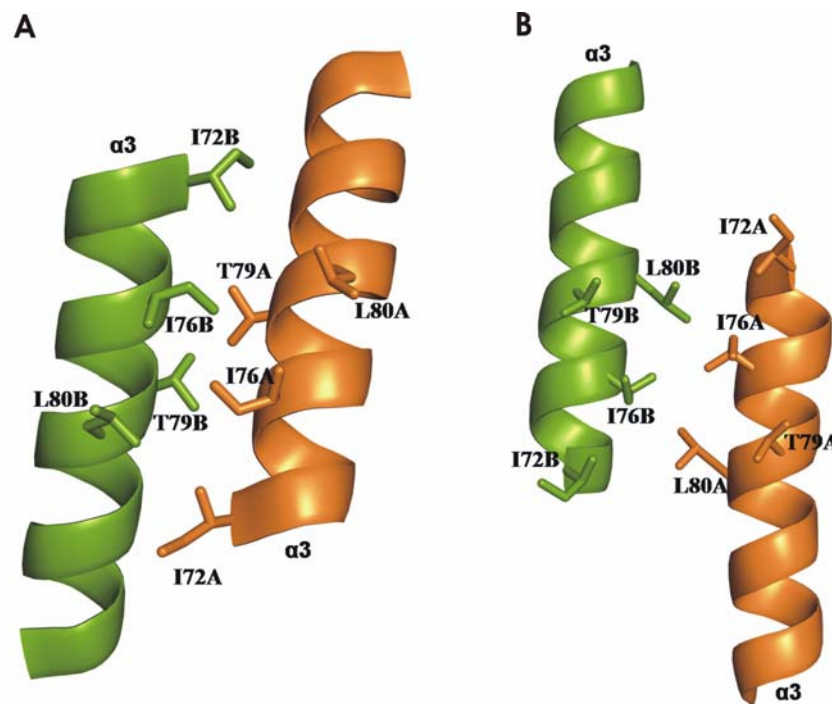
**Table 4.4:**Residues involved in hydrogen bonding and salt bridges in the interface 3

The largest interfacial contact area and highest number of non-covalent interactions are found in interface 1 indicating that it is the most probable mode of dimerization. However, both interaction 1 and 3 contain highly conserved residues and have merits of their own, so both models were considered further to find out the correct dimerization assembly.

#### 4.1.7 Identification of the correct interface by mutagenesis

##### Designing interface disrupting specific point mutants

To elucidate the correct mode of dimerization, both interfaces (1 and 3) were examined in detail to identify residues which, upon mutation, would disrupt one dimer interface, but not the other. Both interfaces involve  $\alpha 3$  helix which is highly conserved in GADD45 proteins. Close inspection of the  $\alpha 3$  helix revealed that the proximity and position of two residues, Thr79 and Leu80 makes them ideal candidates for point mutation that aim at preventing one mode of dimerization without influencing the other because Leu80 is important for interaction 3 only, whereas Thr79 stabilizes interaction 1 (Figure 4.12). Therefore, the individual replacement of Thr79 and/or Leu80 with a residue like glutamate should disrupt only the respective interface either sterically by introducing negative charges in a predominantly hydrophobic environment or by electrostatic repulsion.



**Figure 4.12: Close-up view of the interaction between  $\alpha 3$  helices in (A) interface 1 and (B) interface 3.**

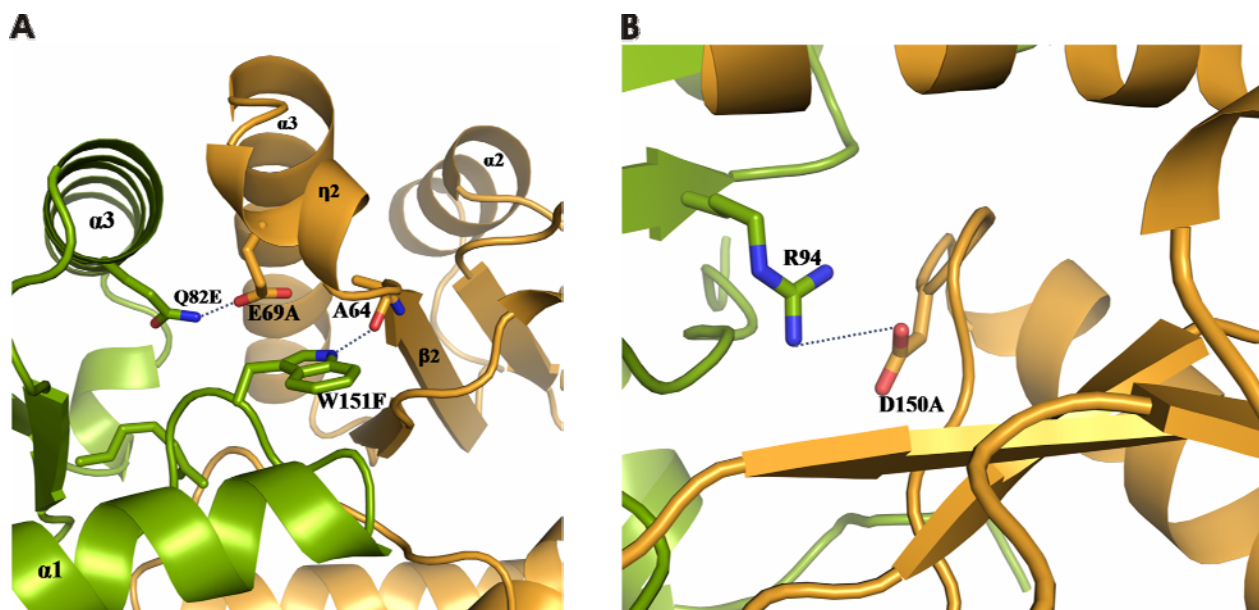
In interface 1, Leu80 residues are pointing outside whereas Thr79 are more into the interface; therefore, modifying Thr79 with a residue like glutamate would compromise this interface only. On the other hand, in the interface 3, Leu80 residues are involved in interfacial contacts whereas Thr79 are not. Thus, mutating Leu80 to an unfavourable residue like glutamate should block interface 3.

In the recently determined structure of mouse GADD45 $\gamma$  [149], the authors also observed more than one dimer interfaces and identified the physiological dimer interface as interface 3 by mutating Thr79 and Leu80, respectively. To cross-check the published result and to see the effect on dimerization of human GADD45 $\gamma$ , T79E and L80E mutants were therefore generated.

Evolutionarily conserved residues which were involved in hydrogen bonding or salt-bridge formation in the interface 1 (eg. Glu69, Gln 82, D150, W151) and therefore might disrupt dimerization were also mutated.

Glu69 which forms a hydrogen bond with Gln82 is part of the evolutionarily conserved DEDDDR motif that was reported to play a key role in the inhibition of Cdc2/cyclin B1 kinase activity and in the induction of a G<sub>2</sub>/M arrest [56]. To disrupt this hydrogen bond and to test its effect on dimerization, Glu69 was mutated to alanine and Gln82 was mutated to glutamate to examine the effect of reversing the charge at this position.



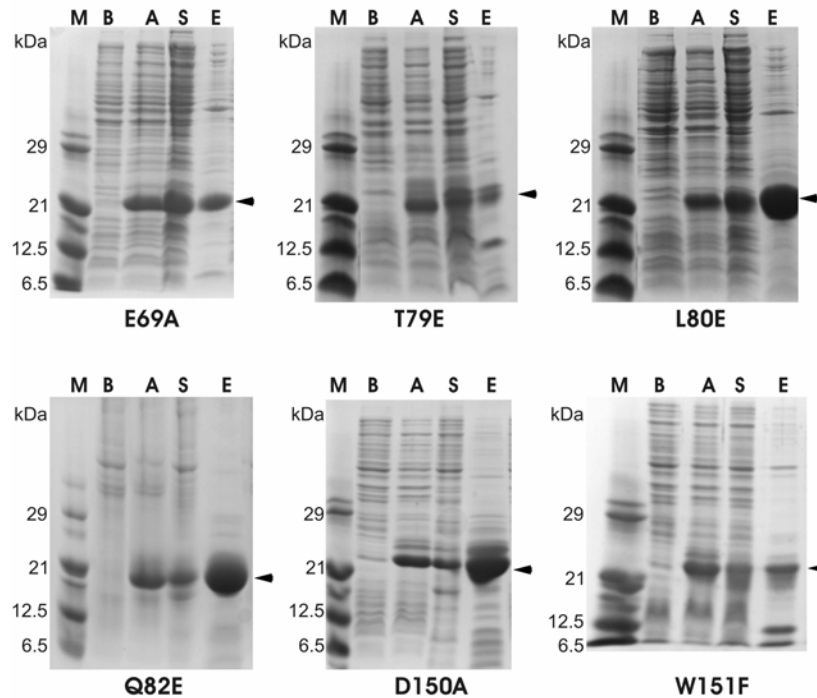


**Figure 4.13: Substitution of residues from interface 1 by site-directed mutagenesis.** A) E69, Q82 and W151 are involved in hydrogen bonding and are substituted with alanine, glutamate and phenylalanine, respectively. B) Salt bridge between R94 and D150. The latter was mutated to alanine.

Similarly, to disrupt the salt bridge and possible dimerization, D150 was mutated to alanine and W151, another residue from interface 1 involved in hydrogen bonding, was changed to phenylalanine. In total 6 single point mutations E69A, T79E, L80E, Q82E, D150A and W151F were introduced (Fig. 4.13).

### Mutant generation and analysis

Point mutants were generated following the method described in 3.1.10. The presence of mutations was confirmed by gene sequencing. The genes encoding mutants or wild-type proteins were expressed and purified following the protocol described in method 3.2.3. All mutants were soluble. (Fig. 4.14). Mutants and wild-type proteins were purified in both 1) 25 mM Tris-HCl (pH 7.5), 50 mM NaCl, 2 mM DTT, 0.1 mM EDTA and 2) 20 mM Hepes (pH 7.5), 150 mM NaCl, 2 mM DTT, 1 mM EDTA, since mutants from mouse homologs were purified in high-salt containing HEPES buffer. The oligomerization state of each mutant was analyzed by analytical gel filtration. The concentration range used in these experiments was 0.5 – 10 mg/ml. Wild-type GADD45 $\gamma$  was dimeric in all the concentrations tested.

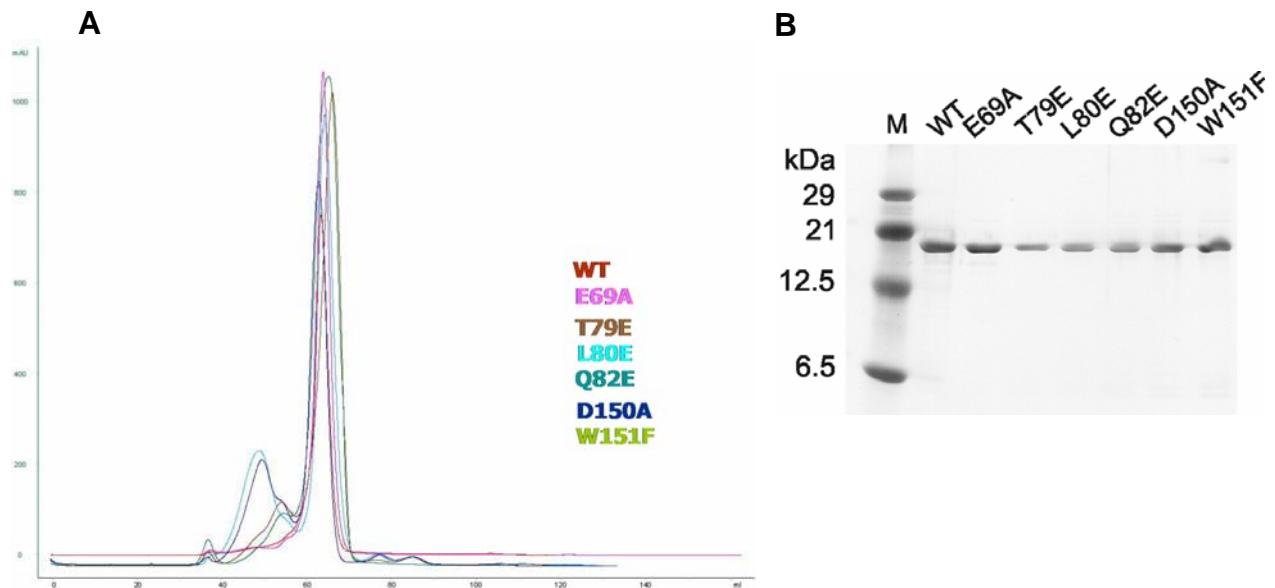


**Figure 4.14: Recombinant synthesis and initial affinity-purification of six GADD45 $\gamma$  mutants.**

M: marker lane, B: Lysate before induction; A: after induction; S: supernatant; E: elution from Talon-affinity matrix. Eluted mutants are marked by arrows.

Mutants were further purified by gel-filtration as described in methods 3.5.1.

None of the mutants could disrupt dimerization and produce a GADD45 $\gamma$  monomer (Fig. 4.15), including the mutant L80E which had been reported to abolish dimerization completely in the case of mouse GADD45 $\gamma$  [159].



**Figure 4.15: A) Mutation of interface residues has no effect on dimerization.** Gel-filtration elution profiles of wild-type GADD45 $\gamma$  and mutants. Mutation exerts no effect on elution behavior and, like the wild type, all mutants elute as dimers. **B)** – Final purity of WT and six mutants used for gel filtration analysis.



Dimerization could not be compromised using different protein concentrations or high-salt buffer. The following table summarizes molecular-weight estimates obtained by gel filtration (Table 4.5).

Construct	Calculated molecular weight kDa	Estimated molecular weight kDa Human			Estimated molecular weight kDa Mouse
		Tris		Hepes	
		10 mg/ml	0.5 mg/ml	1mg/ml	
WT	17.12	35	34	32	33
E69A	17.06	34	34	32	
T79E	17.15	37	35	30	27
L80E	17.13	35	34	30	20
Q82E	17.12	39	32	32	
D150A	17.09	36	36	32	
W151F	17.05	32	30	32	

**Table 4.5:** Molecular weight estimates of WT and mutant human GADD45 $\gamma$  obtained from size-exclusion chromatography. The molecular-weight estimates of mutant L79E under different conditions has been highlighted in red, as this mutant is reported to be monomeric in mouse GADD45 $\gamma$  [159].

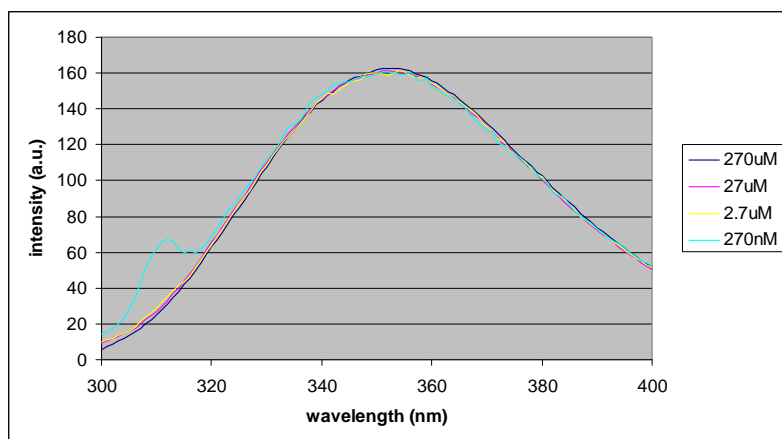
Residues thought to have the potential to disrupt any one mode of interaction could not affect dimerization. It cannot be excluded that the human GADD45 $\gamma$  is more stable than its mouse homolog. Therefore, mutations which affect dimerization in mouse GADD45 $\gamma$  did not show any effect on human GADD45 $\gamma$ . It is also possible that single point mutations were insufficient to disrupt the dimeric assembly. Therefore, it would be interesting to probe combination of mutants.

#### 4.1.8 Analysis of local environment of GADD45 $\gamma$ by tryptophan fluorescence spectroscopy

Tryptophan fluorescence is widely used as a tool to monitor changes in proteins and to make inferences regarding local structure and dynamics [126]. The intensity, quantum yield and wavelength of fluorescence at the maximum of tryptophan emission is solvent-dependent. The fluorescence spectrum shifts to shorter wavelengths and the intensity of the fluorescence increases as the polarity of the solvent surrounding the tryptophan residue decreases (blue-shift emission). On the other hand, if a tryptophan from hydrophobic protein interior is exposed to solvent, emission spectra is expected to shift to longer wavenegth (red-shift emission) [150].

Tryptophan residues which are buried in the hydrophobic core of proteins can have spectra which are shifted by 10 to 20 nm compared to tryptophans on the surface of the protein. GADD45 $\gamma$  has two intrinsic tryptophan residues, Trp130 and Trp151. Trp151 is buried in the interface 1 whereas

Trp130 is solvent-exposed. At a concentration which would facilitate dissociation of the dimer, buried Trp151 would be expected to be exposed and therefore emit a higher fluorescent signal, and the wavelength maximum ( $\lambda_{\max}$ ) would shift to longer wavelength (red-shift). In contrast, both tryptophans are completely solvent-exposed in interface 3. At low protein concentration favoring monomers, the protein therefore would not be expected to show any change in fluorescence. Based on this observation, fluorescence spectra was recorded in order to identify the correct dimerization assembly. Based on the equilibrium dissociation constant ( $K_D$ ) of  $\sim 5 \mu\text{M}$ , tryptophan fluorescence was measured at a concentration range covering  $270 \mu\text{M}$ ,  $27 \mu\text{M}$ ,  $2.7 \mu\text{M}$  and  $270 \text{ nM}$ . At lower concentrations (down to  $27 \text{ nM}$ ) the fluorescence signal contained too much noise for a valid evaluation. As shown in figure 4.16, GADD45 $\gamma$  has a maximal fluorescence emission at  $\sim 350 \text{ nm}$  which did not shift with changing concentrations.



**Figure 4.16:** Fluorescence spectra indicating the absence of change in fluorescence emission of GADD45 $\gamma$  at concentrations from  $270 \text{ nM}$  to  $270 \mu\text{M}$ . Note that concentrations lower than  $270 \text{ nM}$  could not be evaluated since the signal intensity was too low.

These results would be consistent with interface 3, as in this interface both the tryptophans are solvent-exposed. Hence a change in the oligomerization status would not affect fluorescence. In agreement with this argument, no additional spectral emission was observed even at concentration where the molecule was predominantly monomeric.

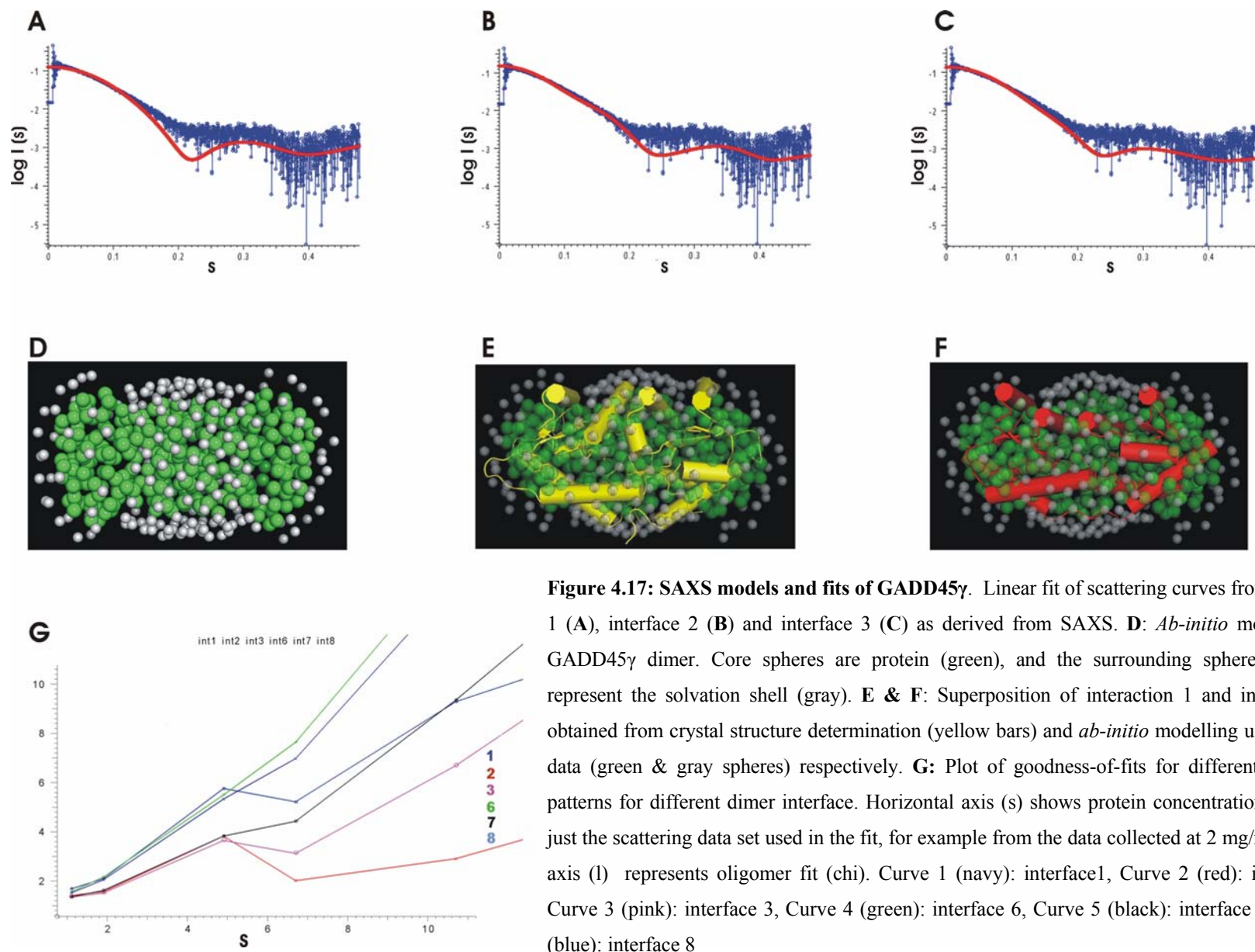
#### 4.1.9 Identification of the correct dimer interface using SAXS

Small angle X-ray scattering is a powerful method to assess protein dimension and shape and was used in the present study to identify the correct physical dimer interface. At the beginning of each experiment the sample was tested for aggregation caused by radiation damage using multiple exposures of the same sample. The results showed no differences in the scattering from multiple measurements indicating that the samples did not suffer from radiation damage. Measurements performed at different protein concentrations indicated strong inter-particle repulsion in the

protein solution at high protein concentrations ( $>7$  mg/ml). The fits provided from the modeling using the GADD45 $\gamma$  crystal structure and the experimental scattering curves were plotted. The calculated scattering from the crystal structure of GADD45 $\gamma$  did not yield an ideal fit to the experimental data, indicating that the quaternary structure in the crystal may be affected by packing constraints.

In agreement with the crystal structure, the SAXS results confirmed the dimeric structure of GADD45 $\gamma$ . The data showed the presence of a dynamic equilibrium between monomeric and dimeric state as suggested by analytical ultracentrifugation (Fig. 4.1). Equilibrium mixture is not suited for modelling as SAXS requires homogenous sample. At concentrations  $\geq 10$  mg/ml significant amounts of dimer were observed, and at lower protein concentration the population of molecules is more heavily weighted towards the monomer. Perhaps this is why GADD45 $\gamma$  was crystallized at  $>10$  mg/ml concentration. However, scattering curves generated at  $\geq 10$  mg/ml could not be used, because the GADD45 $\gamma$  molecules experienced significant repulsive interactions at this concentration which complicated the analysis. The scattering curve at low angles started to be dominated by the inter-particle effects, so the solution at this concentration could not be considered as ideal. Therefore, data fitting was carried out with scattering data obtained at 2 mg/ml concentration (assuming single species, not an equilibrium). Fitting the experimental data using the *ab-initio* and rigid-body modeling by GASBOR (Petoukhov and Svergun 2005; Svergun 1999) gave an improved fit to the data. Since proteins are more flexible in solution compared to the ordered molecules in the crystal structure, the proposed SAXS-model shows the average of all conformations in solution.

The *ab-initio* model shows spheres which do not represent atoms but indicate a region where the electron density corresponds to protein. Neither interaction 1 nor interaction 3 fits the envelope calculated by SAXS very well (Fig. 4.17A & C). Dimer interface 1 fits the overall shape moderately well at low resolution but when analyzing it in more detail it appears to be too compact, and at  $\sim 2$  -1 nm the fit becomes poor. However, the chi value for the fit which is a measure of the discrepancy between the experimental scattering curve and that calculated from a structure model, suggested that interface 2 fitted the data best and interface 3 resulted in the next best fit (in this equilibrium case) (Fig. 4.17G). It was surprising to find that interface 2 turned out to be the best fit, but being asymmetric it can not be valid. From chi value analysis it could be therefore interpreted that between interface 1 and 3, interface 3 is a better fit to the experimental data.

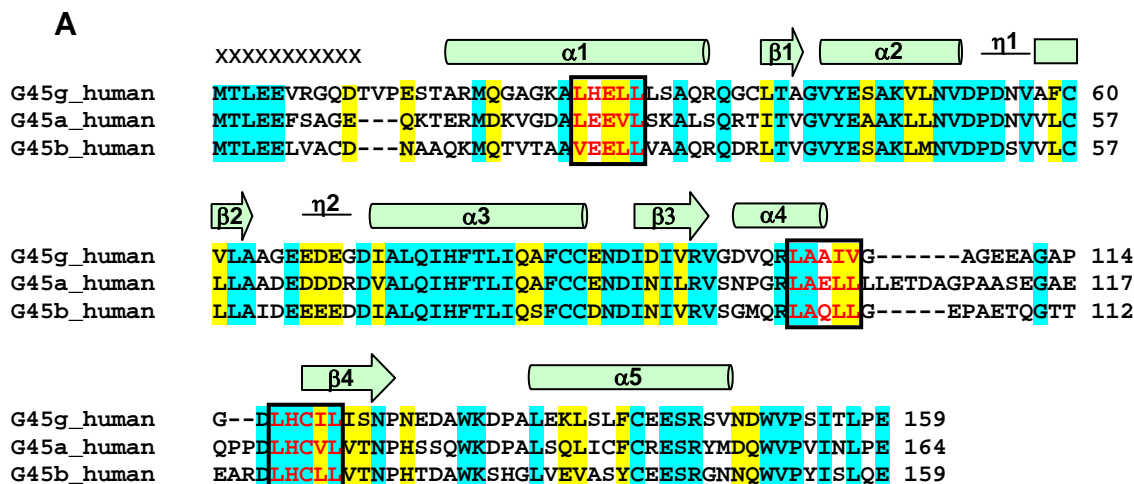


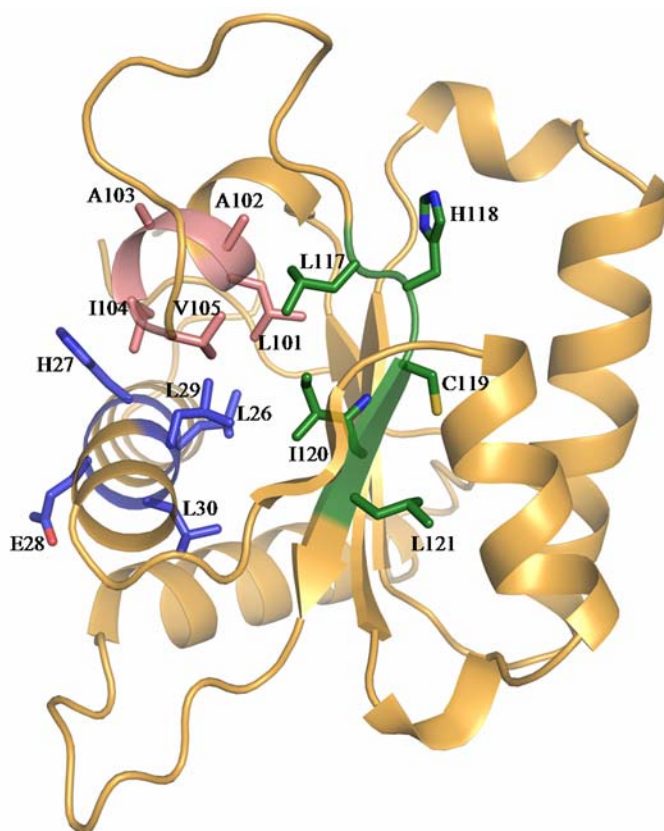
**Figure 4.17: SAXS models and fits of GADD45 $\gamma$ .** Linear fit of scattering curves from interface 1 (A), interface 2 (B) and interface 3 (C) as derived from SAXS. D: *Ab-initio* model of the GADD45 $\gamma$  dimer. Core spheres are protein (green), and the surrounding spheres typically represent the solvation shell (gray). E & F: Superposition of interaction 1 and interaction 3 obtained from crystal structure determination (yellow bars) and *ab-initio* modelling using SAXS data (green & gray spheres) respectively. G: Plot of goodness-of-fits for different scattering patterns for different dimer interface. Horizontal axis ( $s$ ) shows protein concentration, which is just the scattering data set used in the fit, for example from the data collected at 2 mg/ml, vertical axis ( $I$ ) represents oligomer fit ( $\chi$ ). Curve 1 (navy): interface1, Curve 2 (red): interface 2, Curve 3 (pink): interface 3, Curve 4 (green): interface 6, Curve 5 (black): interface 7, Curve 6 (blue): interface 8

#### 4.1.10 Binding to nuclear receptors is sterically blocked

The nuclear-receptor superfamily is a group of ligand-dependent transcription factors which specifically modulate target gene expression. Studies have led to the identification of several coactivators which are required by nuclear receptors for efficient transcriptional activation. All known and studied coactivators are characterized by the presence of LXXLL signature motifs (X – any amino acid) which are considered necessary and sufficient for binding to nuclear-hormone receptors. Interaction between all three GADD45 isoforms and several nuclear hormone receptors were demonstrated using the yeast two-hybrid system as well as pull-down analyses [100]. All three GADD45 isoforms contain in their sequence at least one complete and two incomplete LXXLL motifs whereby it has been presumed that GADD45 proteins can bind to and act as coactivators of nuclear receptors using this motif (Fig. 4.18). Within the motif, especially crucial are the two outer leucines which contact the highly conserved residues especially glutamate and lysine in the binding pocket of nuclear receptors [99] (Fig. 1.6 A).

GADD45 $\gamma$  contains one complete and two incomplete LXXLL motifs in its sequence. The complete LXXLL motif of GADD45 $\gamma$ , 26LHELL30, is located within helix  $\alpha$ 1. Two incomplete motifs, 101LAAIV105 and 117LHCIL121 are located in helix  $\alpha$ 4 and strand  $\beta$ 4, respectively (Fig. 4.18A). Since LXXLL motifs, present in co-activators are believed to be necessary and sufficient to mediate the ligand-dependent interaction with nuclear hormone receptors (eg. peroxisome proliferator-activator receptor [PPAR]), it therefore came as a big surprise to find that the LXXLL motif implicated in nuclear-receptor binding is buried inside the GADD45 $\gamma$  monomer and inaccessible to ligand proteins. The other two, incomplete, motifs were also found to be similarly positioned i.e inaccessible to ligand binding (Fig. 4.18B).



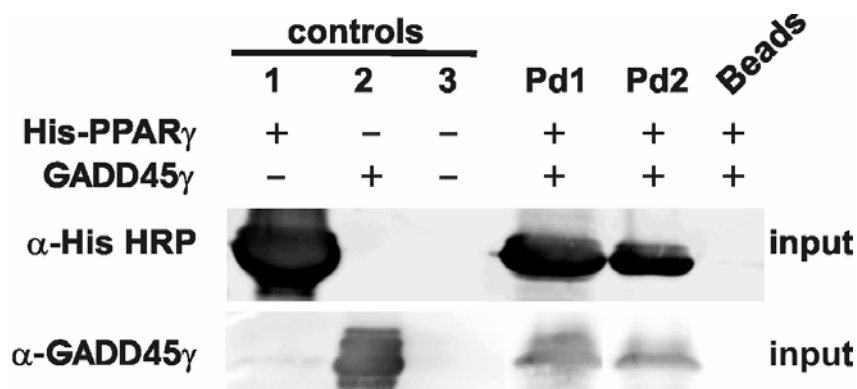


It is obvious that GADD45 $\gamma$ , in the conformation revealed by crystallographic analysis cannot interact with PPAR or other nuclear receptors using the LXXLL motif, especially not via the complete one, <sup>26</sup>LHELL<sup>30</sup>, because of the solvent inaccessibility of the two crucial leucines 26 and 30 which instead point into the interior of the protein (Fig. 4.18 B). Nuclear-receptor binding mediated by this motif would therefore require a dramatic rearrangement of not only the helix but the overall GADD45 $\gamma$  structure so as to avoid steric clash and to fit into the binding pocket. So the obvious question was whether GADD45 $\gamma$  can indeed bind to nuclear hormone receptors like PPAR and if so, whether the LXXLL motifs are essential for binding.

#### 4.1.11 Nuclear receptor binding confirmed by pull-down analysis and AUC

For binding studies with GADD45 $\gamma$ , the ligand-binding domain (LBD) of human PPAR $\gamma$  (the domain where co-activators bind using LXXLL motifs) was used. PPAR $\gamma$ -LBD was produced as N-terminally His-tagged protein in recombinant *E. coli* transformed with an appropriate pET28a-derived vector. The protein was purified as described in [151]. The His-tag was not cleaved for the association study. GADD45 $\gamma$  was purified according to the method described in 3.2.3. Both

proteins were purified on gel filtration columns, separately. Thereafter untagged GADD45 $\gamma$  and His-tagged PPAR $\gamma$ -LBD were used to study pull-down using Talon beads as described in 3.1.19. Samples were then subjected to SDS-PAGE and Western blot analysis in duplicate. Results are shown in figure 4.19.

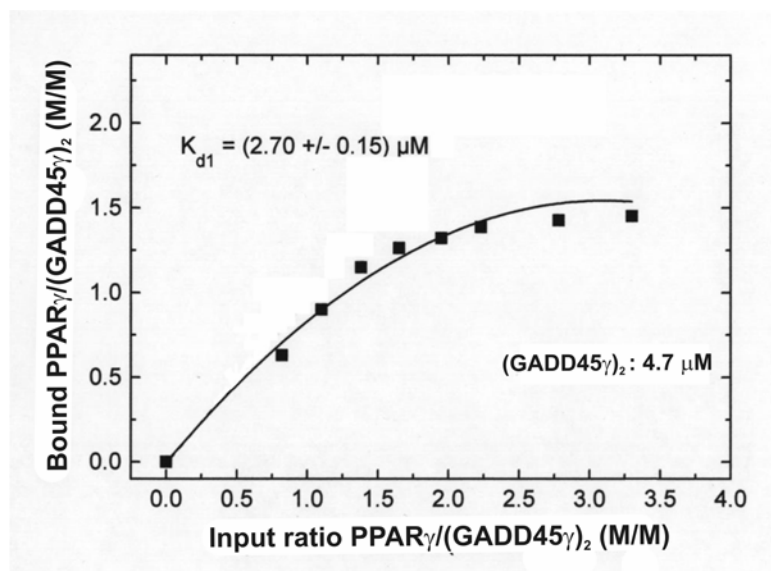


**Figure 4.19: Western blot detection of GADD45 $\gamma$  and PPAR $\gamma$ -LBD in a pull-down experiment.** Equal amounts of proteins were separated in duplicate by SDS-PAGE and subsequently transferred to two polyvinylidene fluoride membranes. Immunodetection for blot A was performed with anti-His HRP conjugate and for blot B anti-GADD45 $\gamma$  rabbit polyclonal IgG was used to detect PPAR $\gamma$  - LBD and GADD45 $\gamma$ , respectively. Ctrl1: His-PPAR $\gamma$ -LBD; Ctrl2: untagged GADD45 $\gamma$ ; Ctrl3: pre-washed Talon beads incubated with GADD45 $\gamma$ ; PD1: pull-down at room temperature; PD2: pull down at 4 °C; Bds: beads after elution. For description, see text.

Pure His-PPAR $\gamma$  LBD was used as control 1 to verify the specificity of anti-His HRP conjugate for His-tag. Purified and cleaved GADD45 $\gamma$  was used as control 2 to check the specificity of anti-GADD45 $\gamma$  IgG. Beads pre-incubated with untagged GADD45 $\gamma$  were used as control 3 to check non-specific binding of GADD45 $\gamma$  with the beads. As an additional control, beads after elution were also tested to rule out the possibility of detecting protein association following precipitation in beads. As is evident from Fig. 4.19, the pull-down experiment indicates that GADD45 $\gamma$  can bind the nuclear receptor PPAR $\gamma$ .

To further verify the GADD45 $\gamma$  association with the PPAR $\gamma$ -LBD, interaction studies by analytical ultracentrifugation were carried out. Complex formation with a dissociation constant of  $K_d = 2.7 \pm 0.15 \mu\text{M}$  was observed (Fig. 4.20), indicating a weak but significant association, and a 1:2 PPAR $\gamma$ : GADD45 $\gamma$  stoichiometry was suggested.





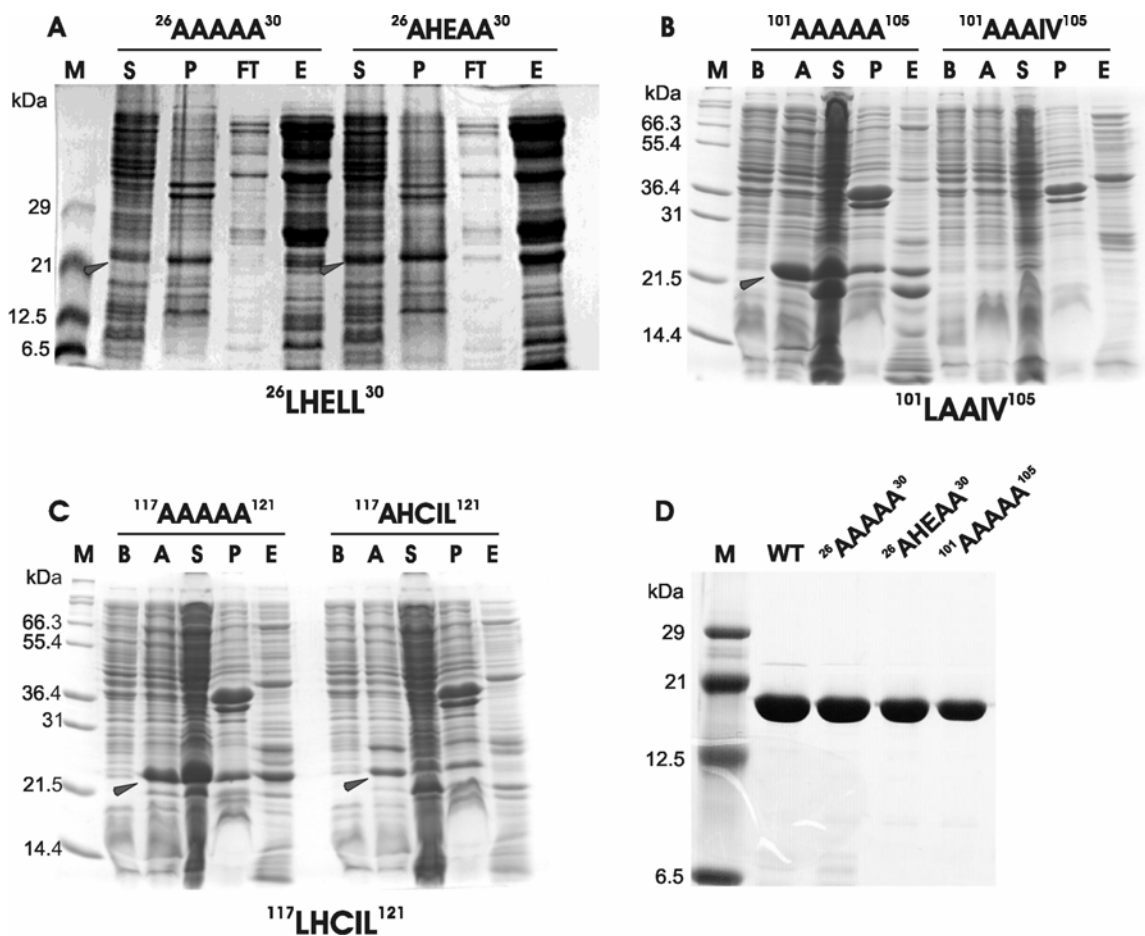
**Figure 4.20:** Analytical sedimentation equilibrium ultracentrifugation plot shows a weak association between PPAR $\gamma$ -LBD and GADD45 $\gamma$ . GADD45 $\gamma$  concentration used for the experiment was 4.7 $\mu$ M – from Prof. Dr. Joachim Behlke

From the association study as well as analytical ultracentrifugation analysis it was therefore concluded that GADD45 $\gamma$  indeed binds to nuclear hormone receptor PPAR $\gamma$ .

#### 4.1.12 Role of LXXLL motifs in binding: Mutational analysis

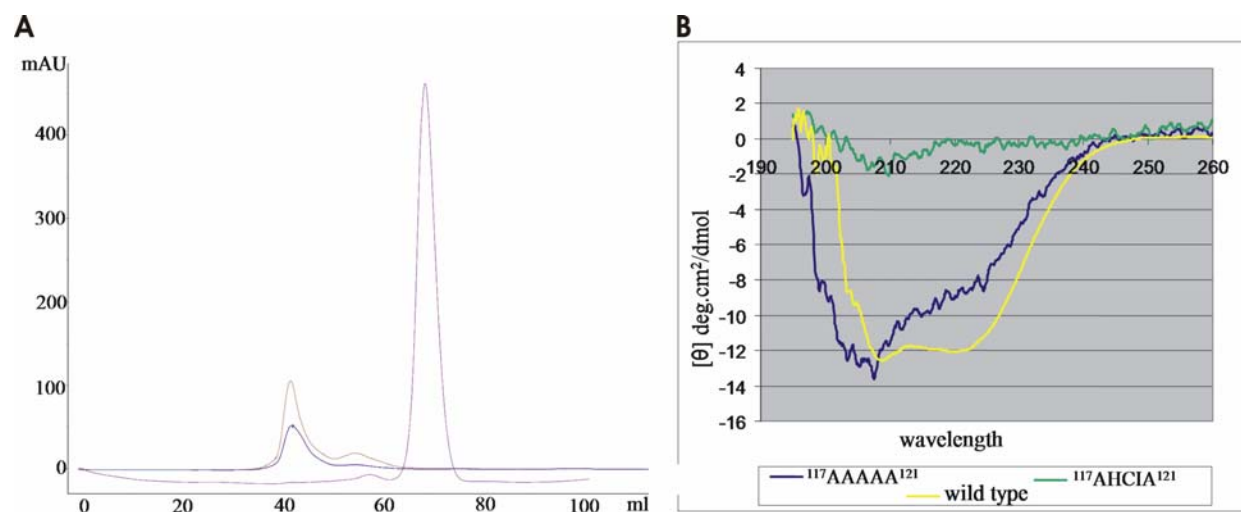
Once binding with nuclear hormone receptor was established, the next task was to find out the role of LXXLL motifs in this binding. For each LXXLL motif (complete or incomplete) two mutations were planned; one where all 5 residues are mutated to alanine and a second where only the leucines are mutated to alanine. In total six mutations were designed. Mutants of  $^{26}$ LHELL $^{30}$  were -  $^{26}$ AHEAA $^{30}$  and  $^{26}$ AAAAA $^{30}$ ; mutants of  $^{101}$ LAAIV $^{105}$  were -  $^{101}$ AAIV $^{105}$ ,  $^{101}$ AAAAA $^{105}$  and mutants of  $^{117}$ LHCIL $^{121}$  were -  $^{117}$ AHCIA $^{121}$  and  $^{117}$ AAAAA $^{121}$  respectively. All six mutants were generated following the QuikChange mutagenesis protocol described in 3.1.10. The presence of mutations was verified by gene sequencing. Mutants were expressed overnight at 20 °C in LB medium with 1 mM IPTG. The  $^{26}$ AHEAA $^{30}$  and  $^{26}$ AAAAA $^{30}$  mutants, however, could not be expressed using this condition. Therefore, autoinducible T9 overnight expression medium was used for overexpression. Despite modulating the expression conditions, medium, temperature or IPTG concentration, the mutant GADD45 $\gamma$  protein bearing the peptide  $^{101}$ AAAAA $^{105}$  could not be obtained. The remaining 5 mutants could be overexpressed and the encoded proteins were soluble (Fig. 4.21).





**Figure 4.21: Expression and solubility profile of A.  $^{26}$ LHELL $^{30}$  B.  $^{101}$ LAAIV $^{105}$  and C.  $^{117}$ LHCIL $^{121}$  mutants.** M: marker lane; the molecular weight of marker proteins is given in kDa. B: lysate before induction; A: after induction; S: supernatant; P: pellet; FT: flow-through; E: elution from Talon affinity matrix. Note:  $^{101}$ AAAIV $^{105}$  could not be expressed. Mutant GADD45γ proteins are marked by arrow. **D.** Final purity of WT and  $^{26}$ LHELL $^{30}$ ,  $^{101}$ LAAIV $^{105}$  and  $^{117}$ LHCIL $^{121}$  mutants. Note:  $^{117}$ AAAAA $^{121}$  and  $^{117}$ AHCIA $^{121}$  eluted in void volumes.

Mutants were purified following the same protocol as used for native GADD45γ. Both set of mutants from  $^{26}$ LHELL $^{30}$  and  $^{101}$ LAAIV $^{105}$  showed comparable elution profile with that of native GADD45γ, thereby confirming the structural integrity of the mutants. However, both  $^{117}$ AHCIA $^{121}$  and  $^{117}$ AAAAA $^{121}$  mutants eluted in the void volume of the gel-filtration column (Superdex 75) thereby suggesting aggregation which was also confirmed by DLS. Furthermore, the CD analysis of these two mutants clearly indicated lack of secondary structure. It could thus be concluded that mutations involving  $^{117}$ LHCIL $^{121}$  rendered the protein unstructured (Fig. 4.22).



**Figure 4.22:** A: Chromatogram showing the elution of  $^{117}\text{AAAAA}^{121}$  (gray dotted line) and  $^{117}\text{AHCIA}^{121}$  (solid blue line) from a superdex 75 (16/60) column overlaid on the corresponding elution profile of native GADD45 $\gamma$  (solid purple line). The mutant proteins eluted in void volume suggesting the presence of aggregate. B: CD spectra of wild-type,  $^{117}\text{AAAAA}^{121}$  and  $^{117}\text{AHCIA}^{121}$  mutants. Molar ellipticity is plotted. Mutations rendered GADD45 $\gamma$  unstructured.

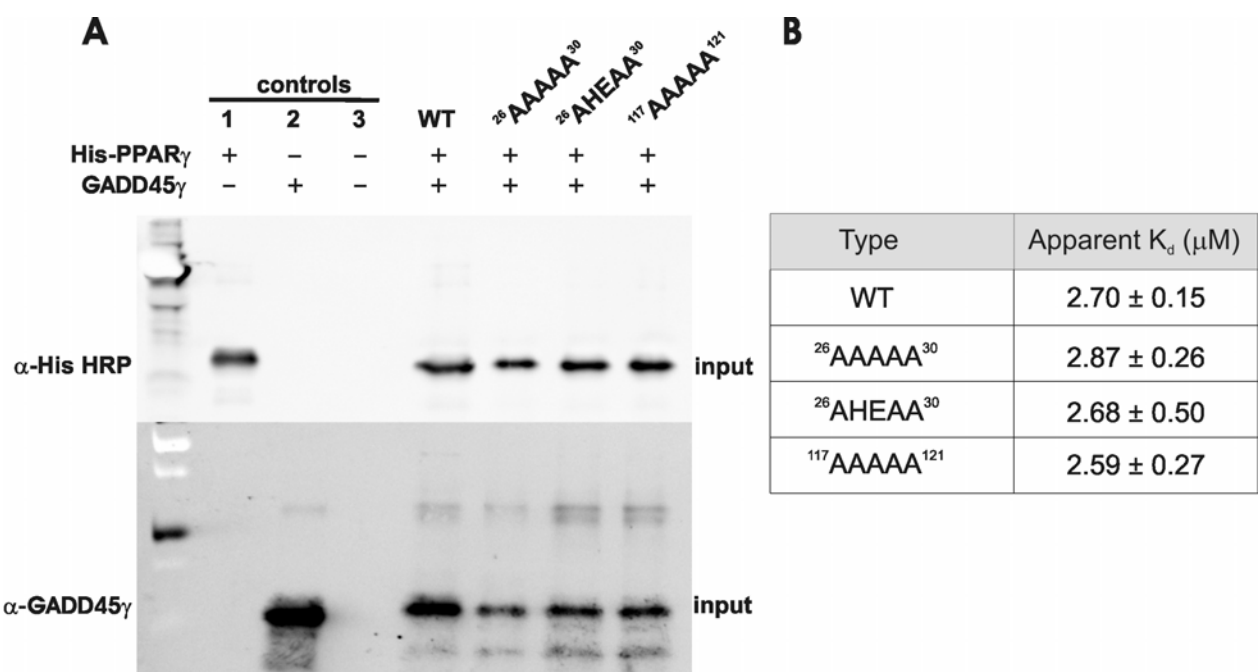
The sequence motif  $^{117}\text{LHCIL}^{121}$  is located on  $\beta 4$ . It is highly conserved and buried deep inside the hydrophobic core of GADD45 $\gamma$ ; it is therefore not surprising that these residues are at least in part essential for protein stability and that their mutation compromised the structural integrity of the protein. Table 4.6 list all mutants and summarize the results.

Mutant	Expression	Solubility	Unfolded/ aggregated
$^{26}\text{AAAAA}^{30}$	yes	yes	no
$^{26}\text{AHEAA}^{30}$	yes	yes	no
$^{101}\text{AAAAA}^{105}$	yes	yes	no
$^{101}\text{AAAIIV}^{105}$	no	—	—
$^{117}\text{AAAAA}^{121}$	yes	yes	yes
$^{117}\text{AAAIIV}^{121}$	yes	yes	yes

**Table 4.6: LXXLL mutants.** Summary showing the results of gene expression, protein purification and structural integrity verification.

4.1.13 Mutant binding is comparable to wild type

Pull-down experiments were performed according to the method described in 3.1.19 using PPAR $\gamma$ -LBD and the three soluble and structured GADD45 $\gamma$  mutants <sup>26</sup>AAAAA<sup>30</sup>, <sup>26</sup>AHEAA<sup>30</sup>, <sup>117</sup>AAAAA<sup>121</sup> as well as the native protein. The binding of the mutants and native GADD45 $\gamma$  to PPAR $\gamma$ -LBD was similar which meant mutation of the motif could not abolish binding to PPAR $\gamma$ -LBD (Fig.4.23A). This experiment therefore proves the ineffectiveness of the LXXLL motifs in nuclear receptor binding by GADD45 $\gamma$ . This result was further verified by analytical ultracentrifugation which showed that all three LXXLL mutants bind to PPAR $\gamma$ -LBD with a binding affinity similar to wild-type GADD45 $\gamma$  (Fig. 4.23B).

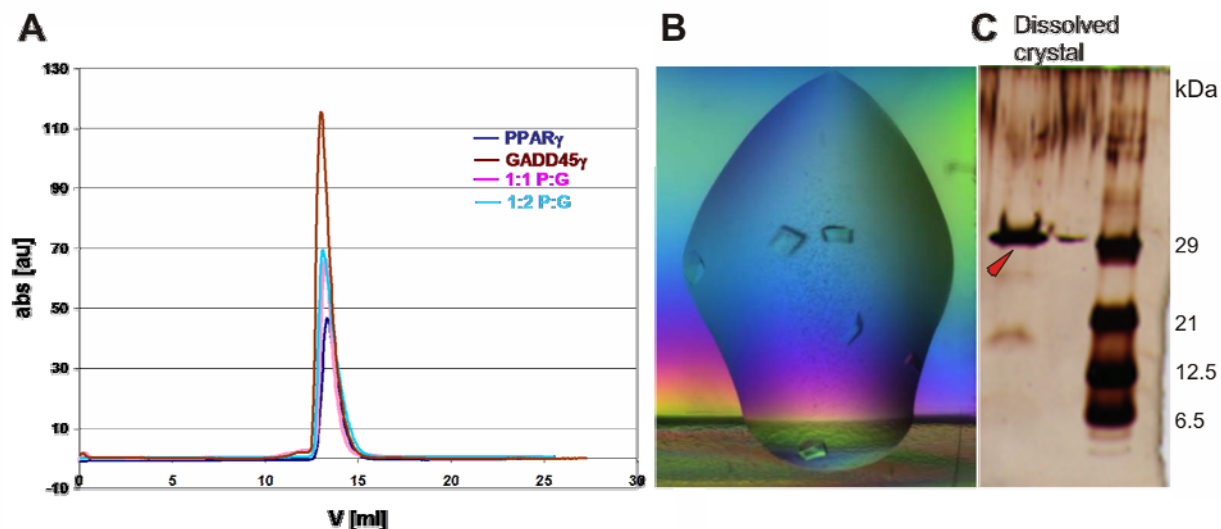


**Figure 4.23: LXXLL mutant binding to PPAR $\gamma$ -LBD is comparable to wild type binding.**  
**A. Western blot detection of GADD45 $\gamma$  WT and mutants pulled down with PPAR $\gamma$ -LBD.** Equal amounts of proteins were separated in duplicate by SDS-PAGE and subsequently transferred to two polyvinylidene fluoride membranes. Immunodetection for blot A was performed with an anti-His HRP conjugate, and for blot B anti-GADD45 $\gamma$  rabbit polyclonal IgG was used to detect PPAR $\gamma$ -LBD and GADD45 $\gamma$ , respectively. **Ctrl1: His-PPAR $\gamma$ -LBD; Ctrl2: untagged GADD45 $\gamma$ ; Ctrl3: pre-washed Talon beads incubated with GADD45 $\gamma$ ; WT: pull down of wild-type GADD45 $\gamma$ ; <sup>26</sup>AAAAA<sup>30</sup>; <sup>26</sup>AHEAA<sup>30</sup>; <sup>117</sup>AAAAA<sup>121</sup>.** **B. Interaction of wild-type and mutant GADD45 $\gamma$  with PPAR $\gamma$  as shown by analytical ultracentrifugation.**  $K_d$ : apparent dissociation constant.

From pull-down experiment as well as analytical ultracentrifugation experiment of wild-type GADD45 $\gamma$  and variants with mutated LXXLL motifs, it was concluded that GADD45 $\gamma$  binds to the nuclear-hormone receptor PPAR $\gamma$  with moderate affinity. The apparent  $K_d$  value of a (GADD45 $\gamma$ )<sub>2</sub>: PPAR $\gamma$  complex is in the range of 2.7  $\mu$ M. This may be taken to imply that this interaction is transient at the cellular level where the effective concentrations of the two proteins are likely to be below the  $K_d$  value. The analysis of GADD45 $\gamma$  proteins with mutated LXXLL motifs clearly shows that the interaction with PPAR $\gamma$  involves residues outside these motifs.

#### **4.1.14 Crystallization of a GADD45 $\gamma$ /PPAR $\gamma$ complex**

The physical interaction of GADD45 $\gamma$  and PPAR $\gamma$ -LBD was demonstrated by analytical ultracentrifugation, and the binding stoichiometry was found to be 2:1 for GADD45 $\gamma$ : PPAR $\gamma$ -LBD. The relatively high apparent  $K_d$  value suggested a rather weak association. In co-crystallization experiments, various stoichiometries were used for the two proteins. To achieve complex formation for co-crystallization, purified GADD45 $\gamma$  and PPAR $\gamma$ -LBD were mixed in both equimolar as well as 2:1 GADD45 $\gamma$ :PPAR $\gamma$ -LBD ratio in 20 mM Tris, pH 7.5, 100 mM NaCl, 2 mM DTT and incubated overnight at 4 °C. Heavy precipitation was observed for the equimolar mixture which was cleared by centrifugation. Both mixtures were then subjected to gel-filtration analysis. PPAR $\gamma$ -LBD runs as a monomer with a molecular mass of about 32 kDa, whereas native GADD45 $\gamma$  runs as a dimer (approximately 34 kDa) and both 1:1 and 1:2 molar mixtures were eluted at a molecular mass of 34 kDa (Fig. 4.24).



**Figure 4.24:** **A:** Purification profiles of PPAR $\gamma$ -LBD (dark blue), GADD45 $\gamma$  (maroon), 1:1 mixture (pink) and 2:1 GADD45 $\gamma$ : PPAR $\gamma$ -LBD mixture (cyan) passed through analytical Superose 12. All peaks elute around 14 ml which corresponds to  $\sim 34$  kDa thereby indicating absence of binding.

**B:** Crystals grown in 0.8 M tri-sodium citrate, 0.1 M sodium cacodylate, 7.5

**C:** Crystals dissolved and analyzed on silver-stained SDS-PAGE show the presence of PPAR $\gamma$ -LBD only (marked by red arrow).

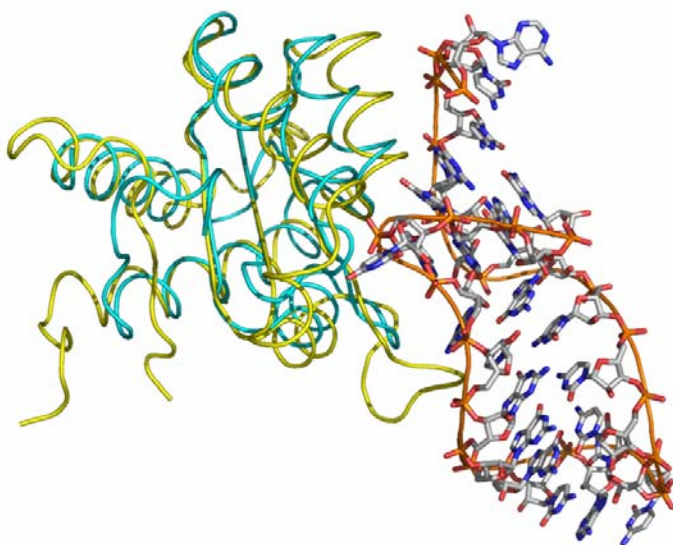
The absence of a gel shift indicated that the two proteins did not form a complex under the experimental conditions. Therefore, the crystals which appeared within 2 days at a 1:2 PPAR $\gamma$ : GADD45 $\gamma$  stoichiometry in screen condition 1 M tri-Na-citrate, 0.1 M Na-cacodylate, pH 6.5, were unlikely to represent the complex. After extensive washing, crystals were dissolved and subjected to SDS-PAGE analysis and stained with silver. It was confirmed that crystals contain PPAR $\gamma$ -LBD only. It can be reasonably assumed that lack of binding and complex formation suggests that this interaction cannot be reproduced in a recombinant system. A missing post-translational modification might be the reason for that; or perhaps this interaction is transient only at the cellular level.

## 4.2 Discussion

### 4.2.1 Homologous structures

A search for proteins with similar 3D structures conducted with DALI (Holm and Sander, 1995) revealed that the nearest structural homologs of GADD45 $\gamma$  (PDB entries 1ffk, 1ck9, 1x70, 1dt9, 1j5e) are mostly large-subunit ribosomal proteins or proteins with RNA-binding and/or processing activity. These proteins belong to the ribosomal L7Ae/L30e/S12e superfamily (Pfam entry PF01248) which includes ribosomal L7Ae from metazoa, L8-A and L8-B from fungi, the 30S-ribosomal protein HS6 from archaeobacteria, the 40S-ribosomal protein S12 from eukaryotes, and L30 from eukaryotes and archaea. A detailed sequence comparison of GADD45 $\gamma$  with its close structural homologs, however, revealed sequence identities less than 15%. Furthermore, there is no demonstrated function that GADD45 $\gamma$  shares with these proteins.

In spite of low sequence similarity, the structure shows striking resemblance to ribosomal protein L7Ae from the 50S-ribosomal subunit of *Haloarcula marismortui* (PDB entry 1ffk). Superposition of GADD45 $\gamma$  and L7Ae (70  $\alpha$ -carbon positions out of 119) with LSQKAB (Kabsch, 1976) yields an r.m.s.d. of 1.27 Å. Differences in the structure are mainly confined to the loop regions, whereas the cores of both proteins show little divergence. Both proteins have alternating  $\alpha$ -helices and  $\beta$ -strands and four-stranded  $\beta$ -sheet cores. The structural similarity is primarily due to the  $\beta$ -sheet cores surrounded by helices. Superposition of GADD45 $\gamma$  on L7Ae complexed with ribosomal RNA shows shape complementarity even when bound to RNA (Fig. 4.25).



**Figure 4.25:** Structural comparison of GADD45 $\gamma$  (gold) with its nearest structural homolog L7ae (cyan) from *Haloarcula marismortui* (PDB 1jj2) with bound RNA (orange, nucleotides shown in sticks with atom colors). Both structures show shape complementarity even when L7Ae is bound to RNA.

It is probable that these proteins share a common ancestor, and during the course of evolution the common fold was adapted to perform function other than RNA binding. Whether GADD45 $\gamma$  is indeed able to bind RNA and which function this interaction might serve is open to speculation and should be subjected to further studies.

### Mouse homolog

Recently, the crystal structure of mouse GADD45 $\gamma$  has been reported [149]. Human GADD45 $\gamma$  shares 95% sequence identity with its mouse homolog. It therefore comes as no surprise that both structures are highly similar with an r.m.s.d. value of 1.2 Å (Fig. 4.26).

**A**

GADD45g_mouse	MTLEEVRGQDTVPPESTARMQGAGKALHELLLSA	GQGCLTAGVYESAKVLNVDPDNVTFC	60
GADD45G_human	MTLEEVRGQDTVPPESTARMQGAGKALHELLLSA	RQGCLTAGVYESAKVLNVDPDNVTFC	60
GADD45g_mouse	VLAADDEDEGDIALQIHFTLIQAFCCENDIDIVRVGDVQRLAAIVGADEEGGAPGDLHCI		120
GADD45G_human	VLAAGEEDEGDIALQIHFTLIQAFCCENDIDIVRVGDVQRLAAIVGAGEEGGAPGDLHCI		120
GADD45g_mouse	LISNPNETWKDPALEKLSLFCEESRSFNDWVPSITLPE		159
GADD45G_human	LISNPNEAWKDPALEKLSLFCEESRSVNDWVPSITLPE		159

**B**



**Figure 4.26:** **A:** Sequence alignment of human and mouse GADD45 $\gamma$  showing a high degree of sequence conservation. Identical residues are on a yellow background, and type-conserved residues on a green background.

**B:** Superimposition of human GADD45 $\gamma$  (gold) and mouse GADD45 $\gamma$  (cyan) shows high degree of similarity.

Both structures are highly similar with the main differences confined to loops, N- and C-termini. Another difference is the presence of a short  $\beta$ -strand comprising only three residues near the C-terminus of murine GADD45 $\gamma$  which is absent in the human ortholog.

#### 4.2.2 Dimerization is a common characteristic of GADD45 proteins

In the crystal, human GADD45 $\gamma$  exists primarily as a dimer assembled through non-covalent bonds. The studies involving gel filtration chromatography as well as analytical ultracentrifugation described above established this fact (Fig: 4.1). SAXS analysis also showed that GADD45 $\gamma$  exists in concentration-dependent monomer-dimer equilibrium. From these results it can therefore be concluded that GADD45 $\gamma$  is a dimer in solution and dimerization is not mere crystallographic artifact.

Dimerization observed in GADD45 $\gamma$  has been reported for other GADD45 proteins. GADD45 $\alpha$  has been reported to oligomerize and form dimeric, trimeric and tetrameric species *in vitro* with dimer being the prevalent form [152]. GADD45 $\beta$  has also been shown to adopt a predominantly dimeric form [51]. GADD45 proteins can also form hetero-oligomers with different family members; however, their role in mediating GADD45 functions has not been fully established [152, 153].

That GADD45 proteins do not necessarily associate as simple dimers and dimerization is necessary for normal functioning of GADD45 proteins can be seen with regard to interacting proteins like PCNA or MKK7. A 2:1 binding stoichiometry of GADD45:PCNA has been suggested [51]. Evidence also supports the existence of a large complex containing an MKK7-GADD45 $\beta$ :GADD45 $\beta$ -MKK7 tetrameric unit [154]. Using analytical ultracentrifugation, in the present study it is also established that binding with nuclear receptor (PPAR $\gamma$ ) involves dimeric GADD45 $\gamma$ . And finally, in case of mouse GADD45 $\gamma$  dimerization has been shown essential for growth inhibition which marks the dimer as physiologically significant [149]. Therefore, dimerization is an important characteristic of GADD45 proteins which is necessary for normal functioning.

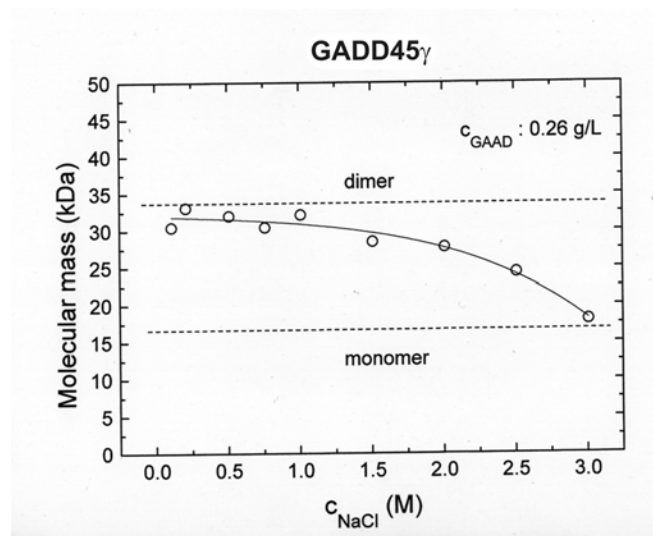
#### 4.2.3 Homodimerization

The crystal structure of human GADD45 $\gamma$  reveals two potential dimerization interfaces from the crystal lattice packing. The dimer 1 interface is extensive whereas the second potential dimerization interface present in the crystal lattice (dimer 3) is less extensive.

Based primarily on deletion mutation, residues in the regions from amino acids 33 to 61 and from 133 to 165 have been reported as the oligomerization interface of GADD45 $\alpha$  [51] which correspond to residues 36-64 ( $\beta$ 1- $\alpha$ 2- $\eta$ 1- $\beta$ 2) and 128-159 ( $\alpha$ 5) of GADD45 $\gamma$ . None of the interfaces formed by crystal contacts in human GADD45 $\gamma$  involve either of these putative



oligomerization regions, but both interfaces involve highly conserved residues and at least one of the putative oligomerization regions (Fig. 4.29). Interface 1 has the largest interfacial contact area which automatically implies that interface 1 is the most likely interaction. Probing chain flexibility and accessibility of exposed residues to solvent by limited proteolysis on human GADD45 $\beta$  shows that the final proteolytic event occurs at Arg91 (Arg94 in human GADD45 $\gamma$ ) which suggests that this residue is either buried within the protein or contained within a rigid structure [53]. In interface 1 of GADD45 $\gamma$ , Arg94 is indeed found buried in the interface and engaged in salt bridge and hydrogen bonds (Fig. 4.14B). A molecular dynamics simulation of human GADD45 $\beta$  also indicates the residues which are engaged in several hydrogen bonds and salt bridges to stabilize the fold of the protein [53]. In the present study of human GADD45 $\gamma$ , some of these residues, notably Arg91, Gln79, Arg146 (Arg94, Gln82 and Arg146 in hGADD45 $\gamma$ ) are indeed involved in hydrogen bonds and/or salt bridges or simply buried in the interface only in case of interface 1. These results therefore, strongly favour interface 1.



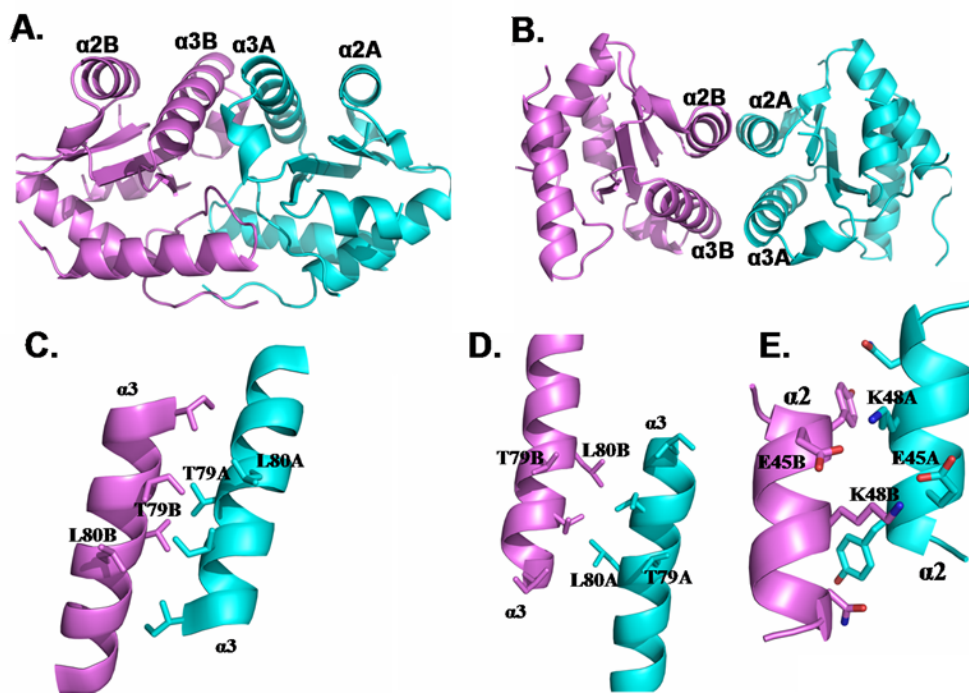
**Figure 4.27:** Salt dependence of GADD45 $\gamma$  dimer stability as shown by analytical ultracentrifugation

However, both interfaces are homodimeric, associated in an antiparallel fashion and form extended hydrogen bonds. Both models are related by a two-fold symmetry axis that is perpendicular to the helix axes. In either case the interface is stabilized by salt bridges which perhaps explain the salt dependence of the GADD45 $\gamma$  dimer stability shown by analytical ultracentrifugation (Fig. 4.27). Therefore, both interfaces show merits of their own and hence both interface models had to be analyzed to probe the correct assembly.

#### 4.2.4 Comparison with mouse interface(s)

Like human GADD45 $\gamma$ , mouse GADD45 $\gamma$  has been reported to be characterized by several potential dimer interfaces [149]. The two most probable dimerization modes in murine GADD45 $\gamma$  (interface 1 and interface 2) involve helix  $\alpha$ 3 and are similar to interface 1 and interface 3 of human GADD45 $\gamma$ . Like human protein, the juxtaposition of two residues, Thr79 and Leu80 in  $\alpha$ 3 of mouse GADD45 $\gamma$  makes them ideal candidates for point mutation to identify

the correct interface because Leu80 is only involved in interaction 2 while Thr79 stabilizes interaction 1 (Fig. 4.28).



**Figure 4.28: Ribbon representation of interface 1 and 2 in mouse GADD45 $\gamma$ .** Chain A is colored cyan and chain B violet. (A) Dimer generated by interaction 1. (B) Dimer generated by interaction 2. (C) Close-up view of the interactions of helix  $\alpha 3$  in dimer 1. (D) Close-up view of the interactions of helix  $\alpha 3$  in dimer 2. (E) Close-up view of helix  $\alpha 2$  showing the interaction of Lys48.

As the  $\alpha 2$ – $\alpha 2$  interaction is involved only in interface 3, mutations in  $\alpha 2$  to disrupt dimerization involving interface 3 were also considered. Furthermore, Lys48 involved in a salt bridge with Glu45 was mutated to glutamate to examine the effects of reversing the charge at this position (Fig. 4.28E).

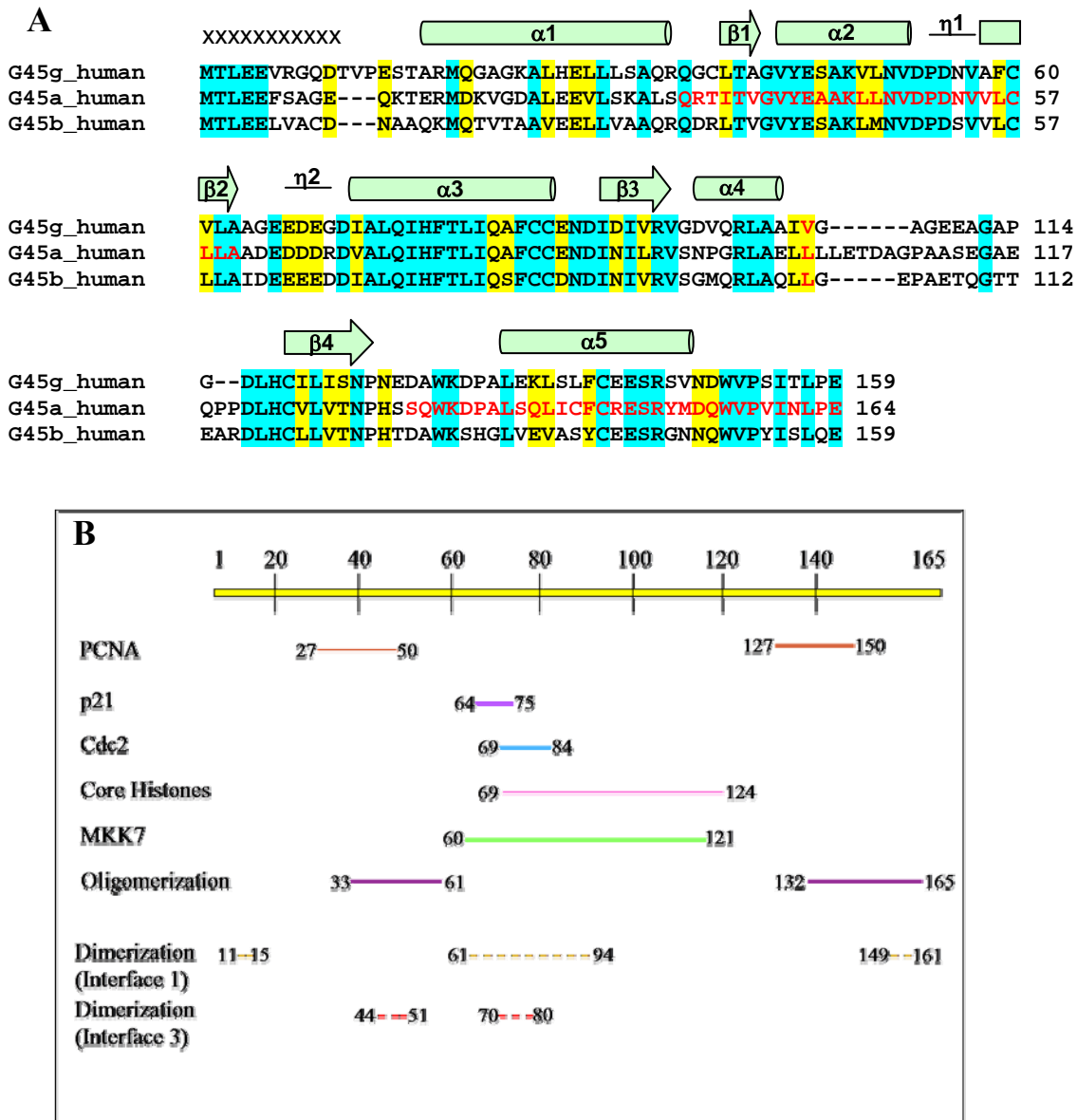
The oligomerization state of each mutant form was assessed by gel filtration and by dynamic light scattering (DLS). The K48E mutation, which would potentially disrupt interaction 3, did not affect the dimerization state of the protein. This can be explained by the observation that most of the interactions between the two  $\alpha 2$  helices in interaction 3 involve backbone and  $C_{\beta}$  interactions. Dimerization of the T79E mutant appeared to be concentration- dependent for which the presence of a dynamic equilibrium between monomeric and dimeric states for this mutant was suggested. True to the assumption, the L80E mutation, which was predicted to disrupt interaction 3, completely abolished the dimerization of mouse GADD45 $\gamma$  therefore strongly supporting interface 3 as the mode of dimerization for GADD45 $\gamma$  in solution.

Human GADD45 $\gamma$  and its mouse homolog share 95% sequence identity. Both structures are highly similar with an r.m.s.d value of 1.2 Å (Fig. 4.26). Therefore it is likely that interface 3 would be the mode of dimerization in case of human GADD45 $\gamma$  as well. However, as described in section 4.1.7, independent mutagenesis studies showed that none of the six single point mutants including L80E had any effect on dimerization of human GADD45 $\gamma$  (Table 4.5). It should be noted that since K47E did not show any effect on dimerization in case of mouse homolog, this mutation was not repeated.

It is possible that human GADD45 $\gamma$  is more stable than its mouse counterpart and therefore single point mutations were insufficient to exert any effect. Human GADD45 $\gamma$  has a melting point ( $T_m$ ) of 54 °C as monitored by thermal shift assay. Though a direct comparison would necessitate measuring  $T_m$  of both human and mouse GADD45 $\gamma$  in a similar experimental set up, it is tempting to mention that a  $T_m$  of 46 °C reported for mouse homolog is substantially lower [149]. Most inter-molecular interactions in interface 1 do not follow the symmetric nature of the dimer which is not unusual for non-crystallographic symmetry (NCS) related protomers. However, this asymmetric nature of interaction may be one of the reasons why single point mutations failed to show any effect on dimerization assembly. Therefore combination of mutants involving both interfaces should be subjected to further studies.

#### **4.2.5 Sequence conservation identifies interface 3 as the likely interface**

The four-helix bundle of interface 3 consists of  $\alpha 2$ - $\alpha 3$  whereas interface 1 contains  $\eta 2$  and  $\alpha 3$  as the major interacting regions as well as the N-terminus (residues 12-18),  $\beta 3$  and the C-terminus (residues 144-157). Each interface contains one of the two putative oligomerization regions suggested for GADD45 $\alpha$  (Fig. 4.29A). The interactions of GADD45 members with Cdc2, p21, p38, MKK7 and histones all have been mapped to the central region of GADD45 ( $\alpha$ ) (Fig. 4.29B). Both interface 1 and interface 3 identified in this study involve this region. Since the residues that play a critical role for dimerization in interface 3 have been shown to be essential for growth inhibition by GADD45 $\gamma$  [149], it is possible that the studies mapping GADD45 binding regions for other cellular proteins have identified interactions, for which dimerization is necessary. Although it is possible that GADD45 $\alpha$  and GADD45 $\gamma$  oligomerize differently, the sequence similarity among GADD45 family members makes it unlikely. Though both interfaces involve residues that are conserved amongst GADD45 protein family,  $\alpha 2$  and  $\alpha 3$  which form the four-helix bundle in interface 3 are most highly conserved (Fig. 4.29A).



**Figure 4.29: A)** Sequence alignment of the three isoforms of human GADD45. Residues predicted to be involved in GADD45α dimerization are colored red. Secondary structural elements observed in GADD45γ are marked above the sequence. Identical residues are highlighted on a magenta background; type-conserved residues are highlighted in yellow. Interface 3 consists of α2-α3 four-helix bundle. Interface 1 is made up of N-η2-α3-β3-C. **B)** Regions of GADD45α implicated in interaction with other cellular proteins and self-association. Oligomerization residues have been reported for GADD45α. Residues corresponding GADD45γ interface 1 and 3 are also shown. Most of the interactions have been mapped to the central region.

Moreover, GADD45 isoforms are known to hetero-oligomerize which has been demonstrated by co-immunoprecipitation [172]. This is consistent with the fact that interface 3 consisting of α2 and α3 involve the most conserved regions in the GADD45 family sequences. An interface

comprising most highly conserved residues appears much more likely to be the correct interface than an interface made up of less conserved regions and loops as in the case of interface 1.

It is likely that dimerization (homo- or hetero-) through highly conserved residues would leave the less conserved regions and loops open for isoform-specific interactions. In addition, residues involved in interfacial interactions are quite asymmetric in interface 1, for example the monodentate salt bridge. In contrast, interaction 3 represents a perfectly symmetric molecule. Taken together, based on the sequence conservation as well as findings from SAXS and fluorescence spectroscopy and comparison with the structure of mouse GADD45 $\gamma$ , it can be concluded that interface 3 involving helices  $\alpha 2$  and  $\alpha 3$  represent the likely dimerization interface of human GADD45 $\gamma$ .

#### 4.2.6 Significance of dimerization involving smaller interfacial contact area

GADD45 proteins are known to form homo- and heterodimers [152, 153]. They are also known to interact with a large variety of proteins from cellular machinery (Fig. 4.29B). This reactivity would require the protein to be flexible to ensure ready dissociation of the complex when required. The presence of salt bridges in the interface as well as the observed salt dependence of stability suggests that the dimer association of GADD45 $\gamma$  is not very tight. A rather low binding affinity as indicated by a  $K_D$  value of  $\sim 5\mu\text{M}$  also suggests that the dimer might be dissociated by ligands that can bind to the interface region. Moreover, interface 3 is characterized by a smaller contact surface area ( $413\text{ \AA}^2$ ). Therefore, I may be allowed to speculate that the weak association observed in GADD45 $\gamma$  probably assisted by small contact surface area of interface 3 as well as low binding affinity and salt dependence would ensure ready dissociation of complex depending on the type and nature of interaction. It is of course striking that the "true" dimer would be stabilized by an interface which is  $< 1/3$  in buried SSA and has a much lower calculated  $\Delta G$  of stabilization (Table. 4.2) than the competing arrangement. However, there are some examples in literature which exhibit similar behavior. Thornton and Nooren from their analysis on 16 experimentally validated 'weak' transient homodimers have reported that weak, transient homodimers are mostly characterized by low-affinity (high  $K_D$  value) and smaller contact surface area (ranges between  $478\text{--}900\text{ \AA}^2$ , which is 7% to 18% of the protomer surface) and elevated monomer-dimer equilibrium [155]. Most notable example is  $\beta$ -lactoglobulin dimer which has  $K_D \sim 20\mu\text{M}$  at neutral pH and contact surface area of  $529\text{ \AA}^2$  (6.6% of the total surface area in the interface) [156]. Small interfaces are probably easier to control and can accommodate more

readily to changes in the cell environment. To change the oligomeric state in such interfaces, only a limited number of interactions would need to be broken/made and no large structural rearrangements would be required [155].

#### 4.2.6 Implication of LXXLL motifs in GADD45 $\gamma$

The LXXLL sequence is a protein-recognition motif widely found in coactivator proteins that bind to the activation function-2 region of ligand-binding domains (LBDs) of nuclear-receptors. This motif has been reported to be necessary and sufficient for the nuclear receptors and coactivators binding [98]. LXXLL motifs normally form short, amphiphatic  $\alpha$ -helices. The termini of the LXXLL helix are locked in place by a ‘charge clamp’ involving two conserved lysine and glutamate of the LBD. These residues grip the LXXLL helix so that the internal residues can pack into the receptor hydrophobic pocket (Fig. 1.6) [99]. It has been reported that the residues flanking the core motif might play a role in determining receptor selectivity and binding [157]. In most cases, motifs lacking a non-polar residue at position -1 bind with lower affinity [158] (Fig. 4.30).

		-1	
HEB	10-AIGTDKEL	SDLLDFSAMFSP-	29
E2-A	10-PVGTDKEL	SDLLDFSMMFPL-	29
c-Myb	88-KEKRIKEL	ELLLMSTENELK-	107
STAT6	795-LLPPTEDLT	TKLLLEGQGESG-	814
CBP LXD1	63-AASKHKQL	SELLRGGSGSSI-	82
CBP LXD2	351-RKLIQQQL	VLLHAHKCQRR-	370
CBP LXD3	2130-RSISPSAL	QDLLRTLKSPSS-	2149
GADD45g1	19-MQGAGKAL	HELLLSAQRQGC-	38
GADD45g2	94-RVGDVQRL	AAIVGAGEEAGA-	109
GADD45g3	110-EGGAPGDL	LHCILISNPEDT-	129

**Figure 4.30:** Sequence alignment of LXXLL motifs from different coactivators that are not directly implicated in nuclear-receptor binding shows the tendency for a small or polar residue at position -1 (green). LXXLL motifs (complete or incomplete) shown in yellow. GADD45g1, 2 & 3 refers to <sup>26</sup>LHELL<sup>30</sup>, <sup>101</sup>LAAIV<sup>105</sup> and <sup>117</sup>LHCIL<sup>121</sup> motifs present in GADD45 $\gamma$ , respectively.

GADD45 $\gamma$  exhibits small or polar residues at -1 position and low binding affinity with nuclear receptor PPAR $\gamma$ -LBD ( $K_D \sim 2.5 \mu M$ ). However, the mutational analysis of LXXLL motifs performed in the present study questions LXXLL motifs’ requirement for nuclear receptor binding in case of GADD45 $\gamma$ . It remains to be seen if the flanking residue(s) would contribute to the binding as has been reported for PPAR $\gamma$ -LBD binding with its coactivator NCoA1 [99]. However, it is possible that the LXXLL motifs in GADD45 $\gamma$  are non-functional. In this context it is worth noting that non-functional examples of LXXLL motifs can be found in proteins that are involved in transcriptional regulation or other cellular processes [158]. Structural elucidation of GADD45 complexed with nuclear receptor is therefore necessary to know the exact role played by this motif in GADD45 proteins.

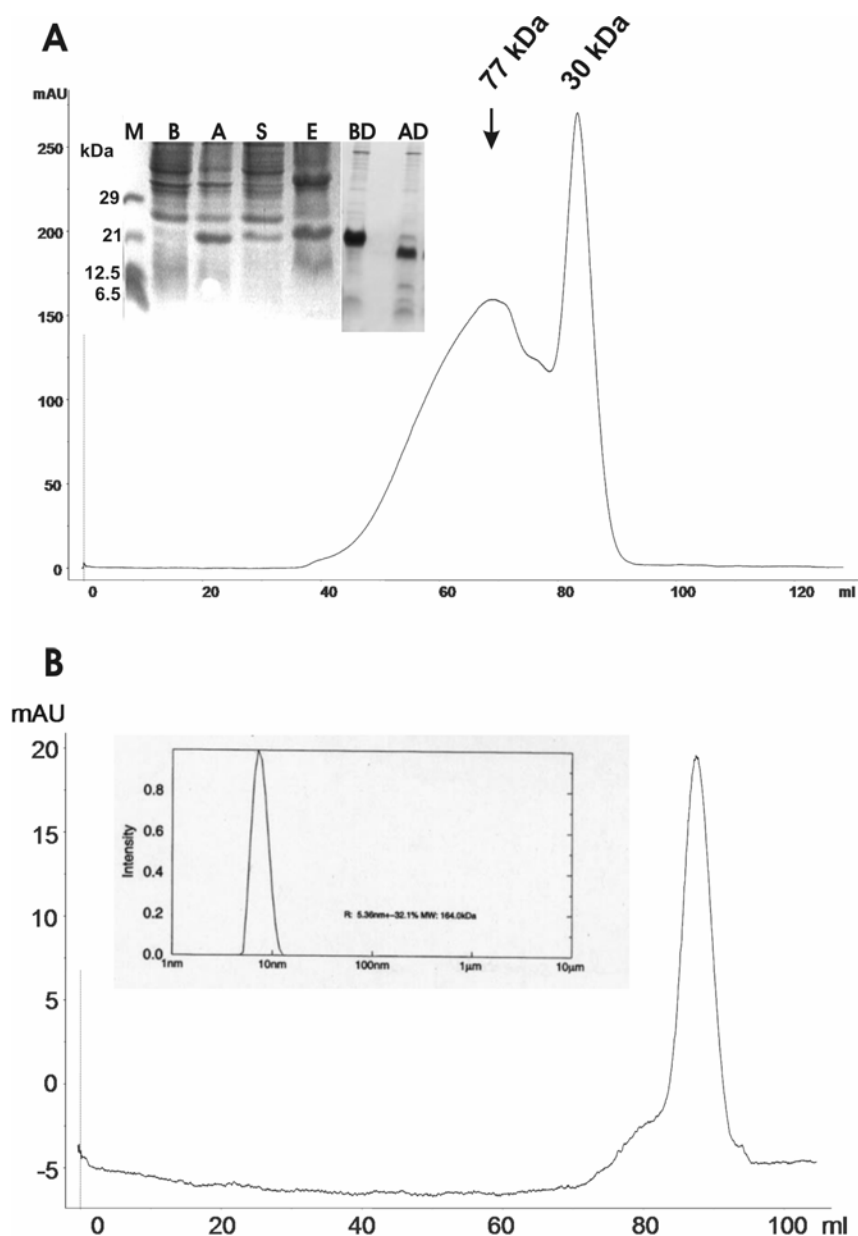
## 5. Characterization of human GADD45 $\alpha$

### 5.1 Results

Success with the structure analysis of GADD45 $\gamma$  automatically led to GADD45 $\alpha$ , since this is the most prominent member of the GADD45 protein family. Although it shares 58% sequence identity with human GADD45 $\gamma$  and therefore is expected to have a closely similar structure, GADD45 $\alpha$  differs in its expression pattern and activation mechanism and exhibits different and distinct biological properties in cells. It was therefore interesting to examine, which structural determinants are responsible for the distinct cellular functions of GADD45 $\alpha$  that are not shared by the other GADD45 proteins. In this chapter, various attempts to crystallize GADD45 $\alpha$  alone and with one of its major interacting proteins are described. There will be some redundancy with regard to Chapters 3 and 4 which have been allowed to simplify the presentation of this work.

#### 5.1.1 Protein expression and purification

The gene encoding the 165 amino-acid long GADD45 $\alpha$  (molecular mass of 18.3 kDa and pI of 4.3) was cloned into the pET28a vector (Novagen) with an N-terminal thrombin-cleavable 6-His tag and expressed in *E. coli* BL 21 DE3 (Novagen). Cells were grown in LB medium at 37 °C to an OD<sub>600</sub> of 0.5 and induced with 1 mM IPTG for 3 h. The cells were harvested and lysed by passage through a French Press in His-lysis buffer. The lysate was centrifuged, and the supernatant was applied to HisTrap HP columns and then washed with 5 column volumes of His-washing buffer. Finally, bound protein was eluted by washing with 100% His-elution buffer. Thrombin cleavage of the His-tag and buffer exchange was simultaneously carried out by dialysis against 20 mM Tris (pH 7.5) and 50 mM NaCl. The second step in the purification involved a Mono-Q anion-exchange column (Pharmacia), where a linear gradient was run from 50 to 1000 mM NaCl over 20 ml. Protein eluted at about 500 mM NaCl. Final polishing was done on a HiLoad 16/60 Superdex 200 16/60 (Pharmacia) gel-filtration column equilibrated with 50 mM NaCl and 20 mM Tris pH 7.6. GADD45 $\alpha$  is detected in two peaks eluting at 77 kDa and 30 kDa [Fig. 5.1].



**Figure 5.1: Expression of human *GADD45α* and purification of its gene product.**

**A) Inset** - SDS PAGE showing the expression and affinity purification (M: marker, B: lysate before induction, A: after induction (expressed protein marked by red arrow), Sup: supernatant after cell lysis and centrifugation, E: Elute from the  $\text{Ni}^{2+}$ -NTA His-Trap column. BD: pooled fractions after  $\text{Ni}^{2+}$ -NTA step for before thrombin digestion, AD: after digestion. **A)** - *GADD45α* runs as a mixture of monomers and dimers and higher oligomers in gel filtration. **B)** The second peak from the first gel filtration runs as a single species in a second gel filtration. **Inset** - but shows large hydrodynamic radius in DLS, thereby suggesting the presence of aggregates.

The first peak did not arise from a homogeneous protein sample as found by DLS measurement. In DLS, the observed large hydrodynamic radius indicated aggregation of this sample which rendered it unfit for crystallization. The second peak, although monodisperse was not



homogeneous either in DLS. This second, smaller peak fraction was pooled separately, concentrated and applied to an S200 16/60 column which resulted in a single peak whose retention time suggested a mixture of GADD45 $\alpha$  monomer and dimer again. DLS measurements confirmed the presence of a single species, but the presence of non-specific aggregates as evidenced by large hydrodynamic radius could not be ruled out (Fig. 5.1B Inset). There seemed to be concentration-dependent monomer-dimer equilibrium, since the dimer fraction increased when concentrating the monomeric sample as monitored by DLS, which clearly indicated the unhomogeneity of the protein sample and, hence, gave a poor prognosis for crystallization experiments.

### 5.1.2 Attempts to crystalize GADD45 $\alpha$

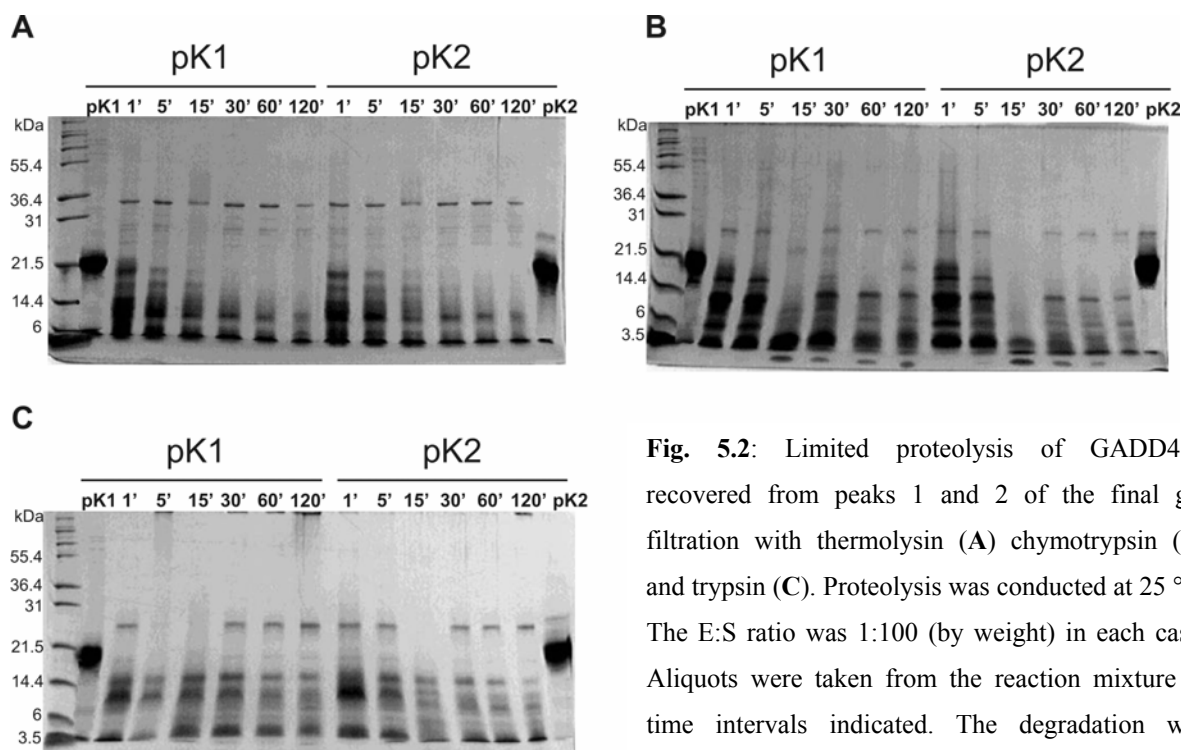
In order to obtain a homogeneous preparation of GADD45 $\alpha$  suitable for structure analysis, several different parameters were modified. Modification of the culture conditions involving different media, temperature or IPTG concentrations proved inefficient. A screening of different buffer pH values, salt concentrations and additives were performed in a Thermofluor analysis (Pantoliano et al, 2001) with the aim to find the best condition for stability and homogeneity. However, protein stability showed little dependence on pH or salt concentration (up to 500 mM). Modulating physico-chemical parameters proved to be inefficient, as crystals could not be obtained (Table 5.1).

Parameters	Expression	Solubility	Crystals
Culture medium	yes	yes	no
IPTG (0.1 – 1mM)	yes	yes	no
Temp. (18 °C - 37 °C)	yes	yes	no
pH (4.5 – 10)	–	yes	no
NaCl (50 - 500mM)	–	yes	no
Glycerol (5-10%)	–	yes	no
Reducing agent	–	yes	no
Ultracentrifugation (100,000g for 30 min)	–	yes	no

**Table 5.1:** Summary of different physico-chemical parameters and their effect on GADD45 $\alpha$  crystallization.

### 5.1.3 In search of stable core fragment

To identify a stable core fragment, limited proteolytic cleavage of GADD45 $\alpha$  recovered from peak 1 and peak 2 from the gel filtration was performed using thermolysin, trypsin and chymotrypsin at different time intervals. Subsequently, proteolysis was quenched by the addition of phenylmethylsulfonyl fluoride (PMSF). The sample was boiled immediately in SDS sample buffer for 5 min to stop proteolysis and analysed on SDS-PAGE. All three proteases showed similar degradation patterns, i.e numerous fragments were generated after time-dependent cleavage, but a single, stable fragment of GADD45 $\alpha$  was not obtained (Fig. 5.2).

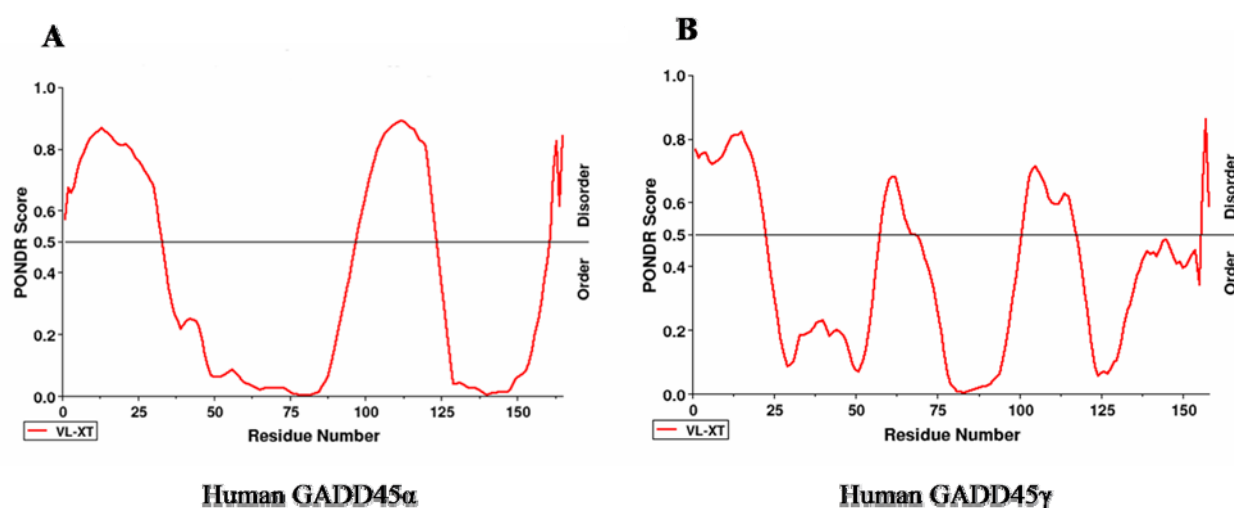


**Fig. 5.2:** Limited proteolysis of GADD45 $\alpha$  recovered from peaks 1 and 2 of the final gel filtration with thermolysin (A) chymotrypsin (B) and trypsin (C). Proteolysis was conducted at 25 °C. The E:S ratio was 1:100 (by weight) in each case. Aliquots were taken from the reaction mixture at time intervals indicated. The degradation was monitored by SDS-PAGE. Pk1: peak 1, pk2: peak 2.

### 5.1.4 Utility of *in silico* tools

Two web-based servers were used to design strategies rendering GADD45 $\alpha$  amenable to crystallization. First, Surface Entropy Reduction (SER) [159] was analyzed using a server maintained by the NIH MBI Laboratory at UCLA [160]. This server suggested mutating surface-exposed, high-entropy amino-acid residues such as lysines and glutamates with residues that have small, low-entropy side chains such as alanine. These mutations are likely to enhance a protein's crystallizeability by the generation of crystal contacts [161]. The proposed mutations were clustered in three categories: (1) E10A, Q11A, K12A; (2) E114A, E117A, Q118A; and (3) E63A.

Next, the GADD45 $\alpha$  sequence was analyzed with the Predictor of Naturally Disordered Regions (PONDR) [162] accessible from Molecular Kinetics, Inc. [163]. Fig 5.4A shows the output of the PONDR analysis for GADD45 $\alpha$ . The PONDR algorithm predicts natively unfolded (disordered) protein regions based on the amino-acid sequence and an analysis of folded protein structures in the Protein Data Bank ([www.pdb.org](http://www.pdb.org)) [164]. The PONDR neural network is trained on a specific set of ordered and disordered sequences, and then returns a value for order/disorder probability of the central amino acid in the window. If a residue value exceeds a threshold of 0.5, this residue is predicted as disordered. Proteins containing natively unfolded peptide regions are less likely to crystallize than completely structured polypeptides.



**Fig. 5.3:** A) PONDR prediction of structured and unstructured regions of GADD45 $\alpha$ . Residues with a disorder probability greater 0.5 are predicted to reside in an unstructured protein region. The N-terminus, middle and C-terminus are predicted to be disordered. B) GADD45 $\gamma$  prediction is shown for comparison..

PONDR prediction for GADD45 $\alpha$ (corresponding residues in GADD45 $\gamma$ )	PONDR prediction for GADD45 $\gamma$ (residues found disordered in structure)
Residues 1 – 32 (1 -36)	Residues 1 – 22 (1-10)
X	Residues 58 - 67
Residues 97 - 123 (100-118)	Residues 101–117 (105-118) (123-32)
Residues 161 - 165 (156-159)	Residues 156 – 158 (157-158)

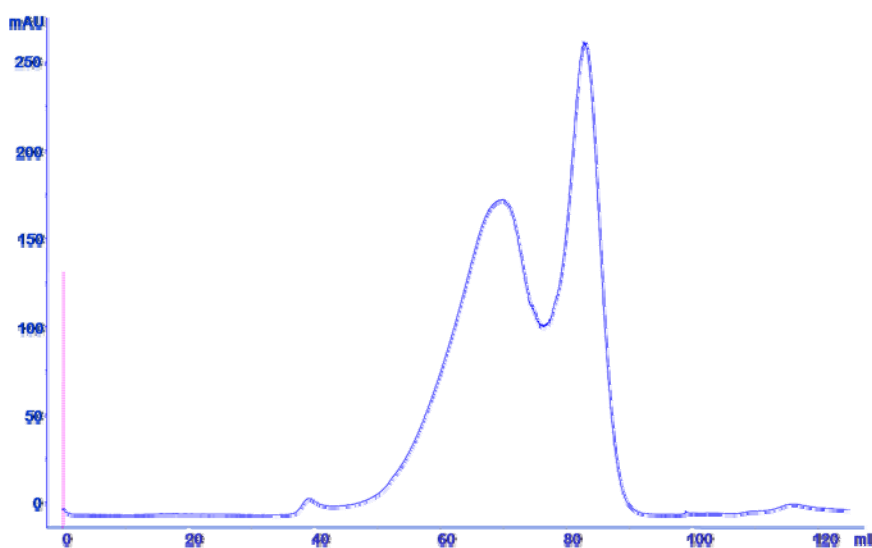
**Table 5.1:** Summary of the prediction for the disordered residues in GADD45 $\alpha$  and GADD45 $\gamma$  by PONDR

PONDR predicted the N-terminal residues 1-32, a central region (97-123) and C-terminus residues 161-165 of GADD45 $\alpha$  to be disordered. The predictions from SER and PONDR suggested internal and terminal modifications, respectively, for enhancing the crystallizability of GADD45 $\alpha$ . The PONDR prediction of disordered segments for GADD45 $\alpha$  is not very different from GADD45 $\gamma$  (Table 5.1). In accordance with the prediction, N-terminus was found disordered in GADD45 $\gamma$  where 10 residues have no electron density.

The long L1 loop of GADD45 $\gamma$  involving residues 101 – 117 was also found disordered in one of the two GADD45 $\gamma$  molecules in the asymmetric unit of the crystal as predicted by PONDR. Both PONDR and SER suggested modification of the GADD45 $\alpha$  N-terminus. Based on the predictions from PONDR and SER and the structural alignment with GADD45 $\gamma$  it was therefore decided to truncate the N-terminal 10 amino acids of GADD45 $\alpha$  to increase the probability of crystallization.

### 5.1.5 Generation and crystallization of N-terminally truncated GADD45 $\alpha$

Using Quick-Change mutagenesis, a shorter fragment of GADD45 $\alpha$  was generated by deleting the 10 N-terminal residues. This protein was purified following the protocol which was used for the purification of full-length GADD45 $\alpha$ . The N-terminal truncation seemed to have no effect on self association as the protein again eluted from gel filtration in two peaks representing protein fractions with higher oligomers and monomer-dimer equilibrium (Fig. 5.4). In DLS the N-terminally truncated GADD45 $\alpha$  showed a similar concentration-dependent aggregation tendency as the full-length protein.



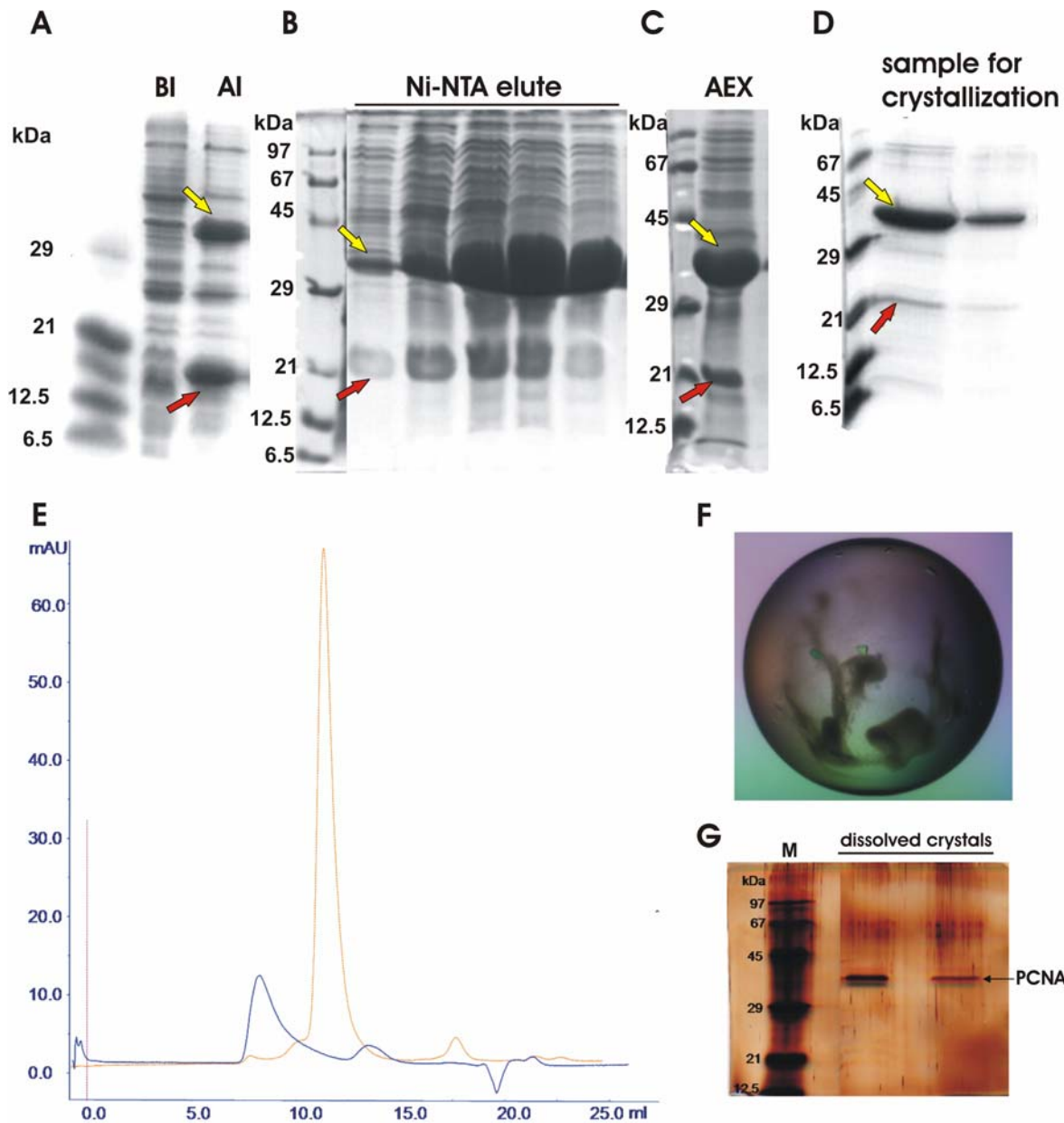
**Figure 5.4: Gel-filtration elution profile of N-terminally truncated GADD45 $\alpha$ .** The profile is similar to that observed with the full-length protein

### 5.1.6 Co-crystallization with PCNA

It has been observed occasionally that disordered protein segments acquire structure when bound to a suitable partner protein. One of the best documented interactions of GADD45 $\alpha$  is that with PCNA. Several studies have mapped the interacting regions of both proteins and suggested the stoichiometry to be 2:1 GADD45 $\alpha$ :PCNA [53]. However, to date there is no structural information of this complex available. In an attempt to obtain structural information on GADD45 $\alpha$ , co-crystallization experiments with PCNA were carried out. Human PCNA which is 261 amino acids long and has a molecular mass of about 32.6 kDa was co-expressed by cloning its coding region into the MCS-I of the pCDFDuet-1 vector, encoding a protein with an N-terminal thrombin-cleavable 6x-His tag. Full-length GADD45 $\alpha$  was cloned into the MCS-II of the same vector without any tag. *E. coli* cells were grown following the expression protocol that allowed the purification of full-length GADD45 $\alpha$ . Complex formation was evident as both the proteins could be co-purified following initial affinity purification using a Ni-NTA His-Trap column (Fig. 5.6).

The complex was further purified on a Mono-Q anion-exchange column (Pharmacia) and a final gel filtration using a Superdex200 16/60 size-exclusion column. Throughout the purification process, the complex remained intact, albeit a gradual decrease in the affinity of GADD45 $\alpha$  was observed.

Crystals appeared within a week in 25% PEG 3350 0.2 M Li<sub>2</sub>SO<sub>4</sub>, 0.1 M Bis-Tris pH 6.5. Analysis of the crystals by SDS-PAGE revealed the presence of PCNA only. It seemed that the complex was not stable enough and fell apart during crystallization. Apparently, the interactions between PCNA molecules that stabilize a PCNA lattice are energetically favored over interactions between GADD45 $\gamma$  and PCNA.



**Figure 5.6: Purification and crystallization of a PCNA: GADD45 $\alpha$  complex.** **A)** SDS-PAGE showing expression pattern. M: marker proteins in kDa, Bi: lysate before induction, Ai: after induction. **B)** Affinity purification using His-Trap. Elute fractions are shown. **C)** Pool fraction after anion-exchange purification. **D)** Purified complex after gel filtration for crystallization. PCNA and GADD45 are marked by yellow and red arrows, respectively. **E)** Elution profile of PCNA (dotted orange line) and PCNA: GADD45 $\alpha$  (solid blue line) suggests complex formation as indicated by shift in gel filtration. **F)** Crystals grown from solution containing GADD45 $\gamma$  and PCNS in 2:1 molar ratio. **G)** Silver-stained PAGE shows that the crystal contains only PCNA.

## 5.2 Discussion

### 5.2.1 Self-association driven by hydrophobic interactions

The hydrodynamic method (gel filtration) strongly suggests that recombinant GADD45 $\alpha$  is capable of self-assembly. It has been reported that recombinant GADD45 $\alpha$  can form dimeric, trimeric and tetrameric species *in vitro* with dimer being the prevalent form [152]. The calculated molecular mass of GADD45 $\alpha$  is 18.4 kDa. Under the conditions of chromatography performed in this study GADD45 $\alpha$  exists mostly in monomer-dimer equilibrium with clear presence of aggregating species as well. DLS provides an estimation of particle size based on diffusion properties. The assembly of GADD45 $\alpha$  subunits into oligomers is accompanied by an increase in the mean apparent hydrodynamic radius of the molecule which indicates that big particles are present.

There are several inter-molecular reaction types which can lead to aggregation. It can stem from interactions between solvent-exposed hydrophobic surfaces or patches, inter-molecular disulfide bond formation between cysteine residues, cross-beta polymerization etc. Human GADD45 $\alpha$  has 5 cysteines. However, an increase in DTT concentration had no effect on the extent and pattern of oligomerization. Notably, GADD45 $\gamma$  is also characterized by the absence of disulfide bonds despite the presence of 6 cysteines. The absence of a DTT effect argues strongly against disulfide-formation-mediated nonspecific aggregation of the protein. Furthermore, the increase in salt (NaCl) concentration from 50 to 300 mM did not alter the fraction of higher-order oligomeric species in gel filtration, suggesting the hydrophobic nature of oligomerization. Moreover, according to the published report [57] GADD45 $\alpha$  is strongly retained on a hydrophobic matrix, phenyl-Sepharose. Therefore, it might not be wholly unreasonable to assume that there are more hydrophobic patches on the surface which will lead to increased oligomerization and/or aggregation.

<b>G45g_human</b>	MTLEEVRGQDTVPESTARMQGAGKALHELLLSAQRQGCLTAGVYESAKVLNVDPDNVAF	60
<b>G45a_human</b>	MTLEEFSSAGE---QKTERMDKVGDALEEVLSKALSQRTITVGVVYEAARKLLNVDPDNVVL	57
<b>G45g_human</b>	VLAAGEEDEGDIALQIHFTLIQAFCCENDIDIVRVGDVQRLAAIVG-----AGEEAGAP	114
<b>G45a_human</b>	LLAAEDDDRDVALQIHFTLIQAFCCENDINILRVSNPGRLLAEELLLETDAAGPAASEGAE	117
<b>G45g_human</b>	G--DLHCILISNPNEADWDPALEKLSLFCEESRSVNDWVPSITLPE	159
<b>G45a_human</b>	QPPDLHCVLVTNPHSSQWKDPALEKLSLFCEESRSYMDQWVPVINLPE	164

Figure 5.7: Alignment of GADD45 $\alpha$  and GADD45 $\gamma$  showing hydrophobic clusters exposed on surface (in green and red for GADD45 $\gamma$  and GADD45 $\alpha$ , respectively) as predicted by iPDA server [164].

Indeed it has been predicted by iPDA server - <http://biominer.cse.yzu.edu.tw/ipda/> [165] that GADD45 $\alpha$  is more hydrophobic than GADD45 $\gamma$ , as there are more and/or larger hydrophobic clusters present in GADD45 $\alpha$  than GADD45 $\gamma$  (Fig. 5.7). This tendency for aggregation is perhaps what renders the protein non-homogenous at any given concentration and therefore makes it less likely to crystallize.

PONDR predicts both N- and C-termini as well as central region to be disordered for GADD45 $\alpha$  and SER also suggested mutation involving residues from these regions for both proteins. Deletion of N-terminal 10 residues of GADD45 $\alpha$  was found to have no effect on non-specific aggregation (Fig. 5.5) which suggests that the extreme N-terminus is not involved in oligomerization and highlights the likelihood of its being unstructured, as predicted. PONDR and SER analyses for GADD45 $\alpha$  and GADD45 $\gamma$  are similar. True to the prediction, both human and mouse GADD45 $\gamma$  were indeed found to be characterized by their disordered N- & C-termini and flexible loop region in the middle. Despite the presence of disordered regions, both full length human and mouse GADD45 $\gamma$  could be crystallized. On the other hand GADD45 $\alpha$ , although sharing 58% sequence identity with GADD45 $\gamma$  is not amenable to crystallization. The non-crystallizability therefore could stem out from some other reasons. Presence of larger hydrophobic clusters on the surface of GADD45 $\alpha$  as predicted and the resulting self-association could be a reason. It would be, therefore, interesting to check the veracity of predictions by other structured techniques such as NMR.

### **5.2.3 Potential for crystallization of a GADD45 $\alpha$ :PCNA complex**

The interaction of GADD45 $\alpha$  and PCNA is well documented, and the stoichiometry of this interaction has been reported to be 2:1 GADD45 $\alpha$ :PCNA. This study shows that both proteins can indeed form a complex as evidenced by co-purification of this complex. However, the gradual loss of one interacting partner and finally dissociation of the complex could happen out of two reasons: Either this interaction is transient, occurring only at cellular level and cannot be reproduced recombinantly, or, the inadequate presence and contribution of second partner (untagged GADD45 $\alpha$ ) was ineffective to the cause of stable complex generation. The latter seems more likely in the present scenario as the complex formation was clearly indicated by gel filtration. Therefore engineering a second affinity tag to the currently untagged GADD45 $\alpha$  might help to increase its contribution to the generation of a stable complex which might lead to crystallization eventually.



## REFERENCES

1. Friedberg, E. C., DNA damage and repair. *Nature*, 2003. 421: 436-40.
2. Rassool, F. V., DNA double strand breaks (DSB) and non-homologous end joining (NHEJ) pathways in human leukemia. *Cancer Lett*, 2003. 193: 1-9.
3. Peltomaki, P., Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J Clin Oncol*, 2003. 21: 1174-9.
4. Jiricny, J. and Marra, G., DNA repair defects in colon cancer. *Curr Opin Genet Dev*, 2003. 13: 61-9.
5. Fisher, D. E., Apoptosis in cancer therapy: crossing the threshold. *Cell*, 1994. 78: 539-42.
6. Muschel, R. J., Soto, D. E., McKenna, W. G., and Bernhard, E. J., Radiosensitization and apoptosis. *Oncogene*, 1998. 17: 3359-63.
7. Holbrook, N. J. and Fornace, A. J., Jr., Response to adversity: molecular control of gene activation following genotoxic stress. *New Biol*, 1991. 3: 825-33.
8. Friedberg, E. W., GC; Siede, W; Wood, RD; Schultz, RA; Ellenberger, T, ed. *DNA Repair & Mutagenesis* 2ed. 2006, ASM Press: Washington D.C. 1-58.
9. Saikumar, P., Dong, Z., Weinberg, J. M., and Venkatachalam, M. A., Mechanisms of cell death in hypoxia/reoxygenation injury. *Oncogene*, 1998. 17: 3341-9.
10. Pierce, A. J., Johnson, R. D., Thompson, L. H., and Jasin, M., XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev*, 1999. 13: 2633-8.
11. Fornace, A. J., Jr., Mammalian genes induced by radiation; activation of genes associated with growth control. *Annu Rev Genet*, 1992. 26: 507-26.
12. O'Connor, P. M., Mammalian G1 and G2 phase checkpoints. *Cancer Surv*, 1997. 29: 151-82.
13. Elledge, S. J., Cell cycle checkpoints: preventing an identity crisis. *Science*, 1996. 274: 1664-72.
14. Hartwell, L. H. and Weinert, T. A., Checkpoints: controls that ensure the order of cell cycle events. *Science*, 1989. 246: 629-34.
15. Liebermann, D. A., Hoffman, B., and Steinman, R. A., Molecular controls of growth arrest and apoptosis: p53-dependent and independent pathways. *Oncogene*, 1995. 11: 199-210.
16. Canman, C. E. and Lim, D. S., The role of ATM in DNA damage responses and cancer. *Oncogene*, 1998. 17: 3301-8.
17. May, P. and May, E., Twenty years of p53 research: structural and functional aspects of the p53 protein. *Oncogene*, 1999. 18: 7621-36.
18. Lakin, N. D. and Jackson, S. P., Regulation of p53 in response to DNA damage. *Oncogene*, 1999. 18: 7644-55.
19. Ichijo, H., From receptors to stress-activated MAP kinases. *Oncogene*, 1999. 18: 6087-93.
20. Basu, S. and Kolesnick, R., Stress signals for apoptosis: ceramide and c-Jun kinase. *Oncogene*, 1998. 17: 3277-85.
21. Holbrook, N. J., Liu, Y., and Fornace, A. J., Jr., Signaling events controlling the molecular response to genotoxic stress. *Exs*, 1996. 77: 273-88.
22. Bulavin, D. V., Amundson, S. A., and Fornace, A. J., p38 and Chk1 kinases: different conductors for the G(2)/M checkpoint symphony. *Curr Opin Genet Dev*, 2002. 12: 92-7.
23. Carrier, F., *et al.*, Characterization of human Gadd45, a p53-regulated protein. *J Biol Chem*, 1994. 269: 32672-7.
24. Zhan, Q., *et al.*, The gadd and MyD genes define a novel set of mammalian genes encoding acidic proteins that synergistically suppress cell growth. *Mol Cell Biol*, 1994. 14: 2361-71.

25. Fornace, A. J., Jr., Alamo, I., Jr., and Hollander, M. C., DNA damage-inducible transcripts in mammalian cells. *Proc Natl Acad Sci U S A*, 1988. 85: 8800-4.
26. Abdollahi, A., Lord, K. A., Hoffman-Liebermann, B., and Liebermann, D. A., Sequence and expression of a cDNA encoding MyD118: a novel myeloid differentiation primary response gene induced by multiple cytokines. *Oncogene*, 1991. 6: 165-7.
27. Liebermann, D. A. and Hoffman, B., MyD genes in negative growth control. *Oncogene*, 1998. 17: 3319-29.
28. Beadling, C., Johnson, K. W., and Smith, K. A., Isolation of interleukin 2-induced immediate-early genes. *Proc Natl Acad Sci U S A*, 1993. 90: 2719-23.
29. Hollander, M. C., *et al.*, Genomic instability in Gadd45a-deficient mice. *Nat Genet*, 1999. 23: 176-84.
30. Gupta, M., Gupta, S. K., Balliet, A. G., Hollander, M. C., Fornace, A. J., Hoffman, B., and Liebermann, D. A., Hematopoietic cells from Gadd45a- and Gadd45b-deficient mice are sensitized to genotoxic-stress-induced apoptosis. *Oncogene*, 2005. 24: 7170-9.
31. Zhang, W., *et al.*, CR6: A third member in the MyD118 and Gadd45 gene family which functions in negative growth control. *Oncogene*, 1999. 18: 4899-907.
32. Takekawa, M. and Saito, H., A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. *Cell*, 1998. 95: 521-30.
33. Wang, X., Gorospe, M., and Holbrook, N. J., gadd45 is not required for activation of c-Jun N-terminal kinase or p38 during acute stress. *J Biol Chem*, 1999. 274: 29599-602.
34. Shaulian, E. and Karin, M., Stress-induced JNK activation is independent of Gadd45 induction. *J Biol Chem*, 1999. 274: 29595-8.
35. Zhang, W., Hoffman, B., and Liebermann, D. A., Ectopic expression of MyD118/Gadd45/CR6 (Gadd45beta/alpha/gamma) sensitizes neoplastic cells to genotoxic stress-induced apoptosis. *Int J Oncol*, 2001. 18: 749-57.
36. Gupta, S. K., Gupta, M., Hoffman, B., and Liebermann, D. A., Hematopoietic cells from gadd45a-deficient and gadd45b-deficient mice exhibit impaired stress responses to acute stimulation with cytokines, myeloablation and inflammation. *Oncogene*, 2006. 25: 5537-46.
37. Selvakumaran, M., Lin, H. K., Sjin, R. T., Reed, J. C., Liebermann, D. A., and Hoffman, B., The novel primary response gene MyD118 and the proto-oncogenes myb, myc, and bcl-2 modulate transforming growth factor beta 1-induced apoptosis of myeloid leukemia cells. *Mol Cell Biol*, 1994. 14: 2352-60.
38. Yoo, J., *et al.*, Transforming growth factor-beta-induced apoptosis is mediated by Smad-dependent expression of GADD45b through p38 activation. *J Biol Chem*, 2003. 278: 43001-7.
39. Papathanasiou, M. A., *et al.*, Induction by ionizing radiation of the gadd45 gene in cultured human cells: lack of mediation by protein kinase C. *Mol Cell Biol*, 1991. 11: 1009-16.
40. Kastan, M. B., *et al.*, A mammalian cell cycle checkpoint pathway utilizing p53 and GADD45 is defective in ataxia-telangiectasia. *Cell*, 1992. 71: 587-97.
41. Zerbini, L. F., Wang, Y., Correa, R. G., Cho, J. Y., and Liebermann, T. A., Blockage of NF-kappaB induces serine 15 phosphorylation of mutant p53 by JNK kinase in prostate cancer cells. *Cell Cycle*, 2005. 4: 1247-53.
42. Canman, C. E., Gilmer, T. M., Coutts, S. B., and Kastan, M. B., Growth factor modulation of p53-mediated growth arrest versus apoptosis. *Genes Dev*, 1995. 9: 600-11.
43. Vairapandi, M., Balliet, A. G., Fornace, A. J., Jr., Hoffman, B., and Liebermann, D. A., The differentiation primary response gene MyD118, related to GADD45, encodes for a nuclear protein which interacts with PCNA and p21WAF1/CIP1. *Oncogene*, 1996. 12: 2579-94.
44. Yu, Y., Kovacevic, Z., and Richardson, D. R., Tuning cell cycle regulation with an iron key. *Cell Cycle*, 2007. 6: 1982-94.

45. Tront, J. S., Hoffman, B., and Liebermann, D. A., Gadd45a suppresses Ras-driven mammary tumorigenesis by activation of c-Jun NH2-terminal kinase and p38 stress signaling resulting in apoptosis and senescence. *Cancer Res*, 2006. 66: 8448-54.
46. Sancar, A., Mechanisms of DNA excision repair. *Science*, 1994. 266: 1954-6.
47. Jonsson, Z. O. and Hubscher, U., Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. *Bioessays*, 1997. 19: 967-75.
48. Kelman, Z. and Hurwitz, J., Protein-PCNA interactions: a DNA-scanning mechanism? *Trends Biochem Sci*, 1998. 23: 236-8.
49. Smith, M. L., *et al.*, Interaction of the p53-regulated protein Gadd45 with proliferating cell nuclear antigen. *Science*, 1994. 266: 1376-80.
50. Smith, M. L., *et al.*, p53-mediated DNA repair responses to UV radiation: studies of mouse cells lacking p53, p21, and/or gadd45 genes. *Mol Cell Biol*, 2000. 20: 3705-14.
51. Vairapandi, M., Azam, N., Balliet, A. G., Hoffman, B., and Liebermann, D. A., Characterization of MyD118, Gadd45, and proliferating cell nuclear antigen (PCNA) interacting domains. PCNA impedes MyD118 AND Gadd45-mediated negative growth control. *J Biol Chem*, 2000. 275: 16810-9.
52. Azam, N., Vairapandi, M., Zhang, W., Hoffman, B., and Liebermann, D. A., Interaction of CR6 (GADD45gamma) with proliferating cell nuclear antigen impedes negative growth control. *J Biol Chem*, 2001. 276: 2766-74.
53. Hall, P. A., Kearsey, J. M., Coates, P. J., Norman, D. G., Warbrick, E., and Cox, L. S., Characterisation of the interaction between PCNA and Gadd45. *Oncogene*, 1995. 10: 2427-33.
54. Kearsey, J. M., Coates, P. J., Prescott, A. R., Warbrick, E., and Hall, P. A., Gadd45 is a nuclear cell cycle regulated protein which interacts with p21Cip1. *Oncogene*, 1995. 11: 1675-83.
55. Zhan, Q., Antinore, M. J., Wang, X. W., Carrier, F., Smith, M. L., Harris, C. C., and Fornace, A. J., Jr., Association with Cdc2 and inhibition of Cdc2/Cyclin B1 kinase activity by the p53-regulated protein Gadd45. *Oncogene*, 1999. 18: 2892-900.
56. Yang, Q., Manicone, A., Coursen, J. D., Linke, S. P., Nagashima, M., Forgues, M., and Wang, X. W., Identification of a functional domain in a GADD45-mediated G2/M checkpoint. *J Biol Chem*, 2000. 275: 36892-8.
57. Jin, S., *et al.*, The GADD45 inhibition of Cdc2 kinase correlates with GADD45-mediated growth suppression. *J Biol Chem*, 2000. 275: 16602-8.
58. Vairapandi, M., Balliet, A. G., Hoffman, B., and Liebermann, D. A., GADD45b and GADD45g are cdc2/cyclinB1 kinase inhibitors with a role in S and G2/M cell cycle checkpoints induced by genotoxic stress. *J Cell Physiol*, 2002. 192: 327-38.
59. Lu, B., Yu, H., Chow, C., Li, B., Zheng, W., Davis, R. J., and Flavell, R. A., GADD45gamma mediates the activation of the p38 and JNK MAP kinase pathways and cytokine production in effector TH1 cells. *Immunity*, 2001. 14: 583-90.
60. Carrier, F., *et al.*, Gadd45, a p53-responsive stress protein, modifies DNA accessibility on damaged chromatin. *Mol Cell Biol*, 1999. 19: 1673-85.
61. Jin, S. G., Guo, C., and Pfeifer, G. P., GADD45A does not promote DNA demethylation. *PLoS Genet*, 2008. 4: e1000013.
62. Hollander, M. C., Kovalsky, O., Salvador, J. M., Kim, K. E., Patterson, A. D., Haines, D. C., and Fornace, A. J., Jr., Dimethylbenzanthracene carcinogenesis in Gadd45a-null mice is associated with decreased DNA repair and increased mutation frequency. *Cancer Res*, 2001. 61: 2487-91.
63. Hartwell, L. H. and Kastan, M. B., Cell cycle control and cancer. *Science*, 1994. 266: 1821-8.
64. Zhan, Q., Bae, I., Kastan, M. B., and Fornace, A. J., Jr., The p53-dependent gamma-ray response of GADD45. *Cancer Res*, 1994. 54: 2755-60.

65. el-Deiry, W. S., *et al.*, WAF1, a potential mediator of p53 tumor suppression. *Cell*, 1993. 75: 817-25.
66. Nurse, P., Ordering S phase and M phase in the cell cycle. *Cell*, 1994. 79: 547-50.
67. Wang, X. W., *et al.*, GADD45 induction of a G2/M cell cycle checkpoint. *Proc Natl Acad Sci U S A*, 1999. 96: 3706-11.
68. Zhao, H., *et al.*, The central region of Gadd45 is required for its interaction with p21/WAF1. *Exp Cell Res*, 2000. 258: 92-100.
69. O'Connor, P. M. and Fan, S., DNA damage checkpoints: implications for cancer therapy. *Prog Cell Cycle Res*, 1996. 2: 165-73.
70. Hoffman, B. and Liebermann, D. A., Role of gadd45 in myeloid cells in response to hematopoietic stress. *Blood Cells Mol Dis*, 2007. 39: 344-7.
71. Maga, G. and Hubscher, U., Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci*, 2003. 116: 3051-60.
72. Hunter, T. and Pines, J., Cyclins and cancer. II: Cyclin D and CDK inhibitors come of age. *Cell*, 1994. 79: 573-82.
73. Xiong, Y., Zhang, H., and Beach, D., D type cyclins associate with multiple protein kinases and the DNA replication and repair factor PCNA. *Cell*, 1992. 71: 505-14.
74. Xiong, Y., Zhang, H., and Beach, D., Subunit rearrangement of the cyclin-dependent kinases is associated with cellular transformation. *Genes Dev*, 1993. 7: 1572-83.
75. Grana, X. and Reddy, E. P., Cell cycle control in mammalian cells: role of cyclins, cyclin dependent kinases (CDKs), growth suppressor genes and cyclin-dependent kinase inhibitors (CKIs). *Oncogene*, 1995. 11: 211-9.
76. Sherr, C. J., G1 phase progression: cycling on cue. *Cell*, 1994. 79: 551-5.
77. Prelich, G., Tan, C. K., Kostura, M., Mathews, M. B., So, A. G., Downey, K. M., and Stillman, B., Functional identity of proliferating cell nuclear antigen and a DNA polymerase-delta auxiliary protein. *Nature*, 1987. 326: 517-20.
78. Shivji, K. K., Kenny, M. K., and Wood, R. D., Proliferating cell nuclear antigen is required for DNA excision repair. *Cell*, 1992. 69: 367-74.
79. Gulbis, J. M., Kelman, Z., Hurwitz, J., O'Donnell, M., and Kuriyan, J., Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell*, 1996. 87: 297-306.
80. Chen, I. T., Smith, M. L., O'Connor, P. M., and Fornace, A. J., Jr., Direct interaction of Gadd45 with PCNA and evidence for competitive interaction of Gadd45 and p21Waf1/Cip1 with PCNA. *Oncogene*, 1995. 11: 1931-7.
81. Warbrick, E., PCNA binding through a conserved motif. *Bioessays*, 1998. 20: 195-9.
82. Kyriakis, J. M. and Avruch, J., Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol Rev*, 2001. 81: 807-69.
83. Kyriakis, J. M. and Avruch, J., Sounding the alarm: protein kinase cascades activated by stress and inflammation. *J Biol Chem*, 1996. 271: 24313-6.
84. Ip, Y. T. and Davis, R. J., Signal transduction by the c-Jun N-terminal kinase (JNK)--from inflammation to development. *Curr Opin Cell Biol*, 1998. 10: 205-19.
85. Davis, R. J., Signal transduction by the JNK group of MAP kinases. *Cell*, 2000. 103: 239-52.
86. Chi, H., Lu, B., Takekawa, M., Davis, R. J., and Flavell, R. A., GADD45beta/GADD45gamma and MEKK4 comprise a genetic pathway mediating STAT4-independent IFNgamma production in T cells. *Embo J*, 2004. 23: 1576-86.
87. Ribeiro, R. C., Kushner, P. J., and Baxter, J. D., The nuclear hormone receptor gene superfamily. *Annu Rev Med*, 1995. 46: 443-53.

88. Michalik, L. and Wahli, W., Peroxisome proliferator-activated receptors: three isotypes for a multitude of functions. *Curr Opin Biotechnol*, 1999. 10: 564-70.
89. Kliewer, S. A., *et al.*, Differential expression and activation of a family of murine peroxisome proliferator-activated receptors. *Proc Natl Acad Sci U S A*, 1994. 91: 7355-9.
90. Forman, B. M., Chen, J., and Evans, R. M., Hypolipidemic drugs, polyunsaturated fatty acids, and eicosanoids are ligands for peroxisome proliferator-activated receptors alpha and delta. *Proc Natl Acad Sci U S A*, 1997. 94: 4312-7.
91. Chawla, A., Schwarz, E. J., Dimaculangan, D. D., and Lazar, M. A., Peroxisome proliferator-activated receptor (PPAR) gamma: adipose-predominant expression and induction early in adipocyte differentiation. *Endocrinology*, 1994. 135: 798-800.
92. Tontonoz, P., Nagy, L., Alvarez, J. G., Thomazy, V. A., and Evans, R. M., PPARgamma promotes monocyte/macrophage differentiation and uptake of oxidized LDL. *Cell*, 1998. 93: 241-52.
93. Brun, R. P., Tontonoz, P., Forman, B. M., Ellis, R., Chen, J., Evans, R. M., and Spiegelman, B. M., Differential activation of adipogenesis by multiple PPAR isoforms. *Genes Dev*, 1996. 10: 974-84.
94. Umesono, K., Murakami, K. K., Thompson, C. C., and Evans, R. M., Direct repeats as selective response elements for the thyroid hormone, retinoic acid, and vitamin D3 receptors. *Cell*, 1991. 65: 1255-66.
95. Horwitz, K. B., Jackson, T. A., Bain, D. L., Richer, J. K., Takimoto, G. S., and Tung, L., Nuclear receptor coactivators and corepressors. *Mol Endocrinol*, 1996. 10: 1167-77.
96. Glass, C. K., Rose, D. W., and Rosenfeld, M. G., Nuclear receptor coactivators. *Curr Opin Cell Biol*, 1997. 9: 222-32.
97. Voegel, J. J., Heine, M. J., Tini, M., Vivat, V., Chambon, P., and Gronemeyer, H., The coactivator TIF2 contains three nuclear receptor-binding motifs and mediates transactivation through CBP binding-dependent and -independent pathways. *Embo J*, 1998. 17: 507-19.
98. Heery, D. M., Kalkhoven, E., Hoare, S., and Parker, M. G., A signature motif in transcriptional co-activators mediates binding to nuclear receptors. *Nature*, 1997. 387: 733-6.
99. Nolte, R. T., *et al.*, Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-gamma. *Nature*, 1998. 395: 137-43.
100. Yi, Y. W., Kim, D., Jung, N., Hong, S. S., Lee, H. S., and Bae, I., Gadd45 family proteins are coactivators of nuclear hormone receptors. *Biochem Biophys Res Commun*, 2000. 272: 193-8.
101. Amundson, S. A., Zhan, Q., Penn, L. Z., and Fornace, A. J., Jr., Myc suppresses induction of the growth arrest genes gadd34, gadd45, and gadd153 by DNA-damaging agents. *Oncogene*, 1998. 17: 2149-54.
102. Zerbini, L. F., *et al.*, NF-kappa B-mediated repression of growth arrest- and DNA-damage-inducible proteins 45alpha and gamma is essential for cancer cell survival. *Proc Natl Acad Sci U S A*, 2004. 101: 13618-23.
103. Zhang, X., Sun, H., Danila, D. C., Johnson, S. R., Zhou, Y., Swearingen, B., and Klibanski, A., Loss of expression o
104. Chung, H. K., *et al.*, Gadd45gamma expression is reduced in anaplastic thyroid cancer and its reexpression results in apoptosis. *J Clin Endocrinol Metab*, 2003. 88: 3913-20.
105. Reddy, S. P., *et al.*, Novel glioblastoma markers with diagnostic and prognostic value identified through transcriptome analysis. *Clin Cancer Res*, 2008. 14: 2978-87.
106. Blundell, T. L. J., L.N., Protein Crystallography, Academic Press (1976).
107. Ewald, P., Das reziproke Gitter in der Strukturtheorie. *Z. Kristallogr.*, 1921. 56.
108. Hendrickson, W. A., Horton, J. R., and LeMaster, D. M., Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *Embo J*, 1990. 9: 1665-72.

109. Terwilliger, T. C. and Berendzen, J., Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr*, 1999. 55: 849-61.
110. Miller, R., Gallo, S. M., Khalak, H. G. a., and Weeks, C. M., *SnB*: crystal structure determination via *Shake-and-Bake*. *J. Appl. Crystallogr.*, 1994. 27: 613-621.
111. Sheldrick, G. M., Phase annealing in SHELX-90: Direct methods for larger structures. *Acta Crystallogr., A*, 1990. 46.
112. Terwilliger, T. C., Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr*, 2000. 56: 965-72.
113. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M., Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, 1991. 47 ( Pt 2): 110-9.
114. Winn, M. D., Isupov, M. N., and Murshudov, G. N., Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Cryst.*, 2001. D57: 122-133.
115. Winn, M. D., Isupov, M. N., and Murshudov, G. N., Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr D Biol Crystallogr*, 2001. 57: 122-33.
116. Sambrook, J. and Russell, D. W., Molecular Cloning. *Cold Spring Harbor, New York*, 2001.
117. Mullis, K. B. and Faloona, F. A., Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol*, 1987. 155: 335-50.
118. Hanahan, D., Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol*, 1983. 166: 557-80.
119. Wang, W. and Malcolm, B. A., Two-stage PCR protocol allowing introduction of multiple mutations, deletions and insertions using QuikChange Site-Directed Mutagenesis. *Biotechniques*, 1999. 26: 680-2.
120. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R., High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur J Biochem*, 1995. 230: 788-96.
121. Scheich, C., Niesen, F. H., Seckler, R., and Bussow, K., An automated in vitro protein folding screen applied to a human dynactin subunit. *Protein Sci*, 2004. 13: 370-80.
122. Bussow, K., *et al.*, Structural genomics of human proteins--target selection and generation of a public catalogue of expression clones. *Microb Cell Fact*, 2005. 4: 21.
123. Nesterenko, M. V., Tilley, M., and Upton, S. J., A simple modification of Blum's silver stain method allows for 30 minute detection of proteins in polyacrylamide gels. *J Biochem Biophys Methods*, 1994. 28: 239-42.
124. Thiede, B., Hohenwarter, W., Krah, A., Mattow, J., Schmid, M., Schmidt, F., and Jungblut, P. R., Peptide mass fingerprinting. *Methods*, 2005. 35: 237-47.
125. Laue, T. M. and Stafford, W. F., 3rd, Modern applications of analytical ultracentrifugation. *Annu Rev Biophys Biomol Struct*, 1999. 28: 75-100.
126. Kelly, S. M. and Price, N. C., The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci*, 2000. 1: 349-84.
127. Royer, C. A., Probing protein folding and conformational transitions with fluorescence. *Chem Rev*, 2006. 106: 1769-84.
128. Svergun, D. I. and Koch, M. H., Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol*, 2002. 12: 654-60.
129. Glatter, O., Kratky, O., Small-angle X-ray scattering, Eds., Academic: London, 1982.

130. Feigin LA, S. D., Structure Analysis By Small-angle X-ray and Neutron Scattering. New York: Plenum Press, 1987.
131. Petoukhov, M. V. and Svergun, D. I., Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr Opin Struct Biol*, 2007. 17: 562-71.
132. Mueller, U., *et al.*, Development of a technology for automation and miniaturization of protein crystallization. *J Biotechnol*, 2001. 85: 7-14.
133. Heinemann, U., Bussow, K., Mueller, U., and Umbach, P., Facilities and methods for the high-throughput crystal structural analysis of human proteins. *Acc Chem Res*, 2003. 36: 157-63.
134. Schneider, T. R. and Sheldrick, G. M., Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr*, 2002. 58: 1772-9.
135. Emsley, P. and Cowtan, K., Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*, 2004. 60: 2126-32.
136. Murshudov, G. N., Vagin, A. A., and Dodson, E. J., Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, 1997. 53: 240-55.
137. Brunger, A. T., *et al.*, Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, 1998. 54: 905-21.
138. Perrakis, A., Harkiolaki, M., Wilson, K. S., and Lamzin, V. S., ARP/wARP and molecular replacement. *Acta Crystallogr D Biol Crystallogr*, 2001. 57: 1445-50.
139. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, 1994. 50: 760-3.
140. Kleywegt, G. J. and Jones, T. A., xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets. *Acta Crystallogr D Biol Crystallogr*, 1996. 52: 826-8.
141. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, 1991. 47 ( Pt 2): 110-9.
142. Laskowski, R. A., Moss, D. S., and Thornton, J. M., Main-chain bond lengths and bond angles in protein structures. *J Mol Biol*, 1993. 231: 1049-67.
143. Hoof, R. W., Vriend, G., Sander, C., and Abola, E. E., Errors in protein structures. *Nature*, 1996. 381: 272.
144. DeLano, W. L., The PyMOL Molecular Graphics System. 2003: DeLano Scientific LLC, San Carlos, CA, USA. <http://www.pymol.org>.
145. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A., Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, 2001. 98: 10037-41.
146. Orengo, C. A., Michie, A. D., Jones, S., Jones, D. T., Swindells, M. B., and Thornton, J. M., CATH--a hierarchic classification of protein domain structures. *Structure*, 1997. 5: 1093-108.
147. Pearl, F. M., *et al.*, The CATH database: an extended protein family resource for structural and functional genomics. *Nucleic Acids Res*, 2003. 31: 452-5.
148. Krissinel, E. and Henrick, K., Inference of macromolecular assemblies from crystalline state. *J Mol Biol*, 2007. 372: 774-97.
149. Schrag, J. D., Jiralerspong, S., Banville, M., Jaramillo, M. L., and O'Connor-McCourt, M. D., The crystal structure and dimerization interface of GADD45gamma. *Proc Natl Acad Sci U S A*, 2008. 105: 6566-71.
150. Caputo, G. A. and London, E., Cumulative effects of amino acid substitutions and hydrophobic mismatch upon the transmembrane stability and conformation of hydrophobic alpha-helices. *Biochemistry*, 2003. 42: 3275-85.

151. Cronet, P., *et al.*, Structure of the PPAR $\alpha$  and - $\gamma$  ligand binding domain in complex with AZ 242; ligand selectivity and agonist activation in the PPAR family. *Structure*, 2001. 9: 699-706.
152. Kovalsky, O., Lung, F. D., Roller, P. P., and Fornace, A. J., Jr., Oligomerization of human Gadd45a protein. *J Biol Chem*, 2001. 276: 39330-9.
153. Liebermann, D. A. and Hoffman, B., Gadd45 in the response of hematopoietic cells to genotoxic stress. *Blood Cells Mol Dis*, 2007. 39: 329-35.
154. Tornatore, L., *et al.*, Gadd45 beta forms a homodimeric complex that binds tightly to MKK7. *J Mol Biol*, 2008. 378: 97-111.
155. Nooren, I. M. and Thornton, J. M., Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol*, 2003. 325: 991-1018.
156. Uhrinova, S., Smith, M. H., Jameson, G. B., Uhrin, D., Sawyer, L., and Barlow, P. N., Structural changes accompanying pH-induced dissociation of the beta-lactoglobulin dimer. *Biochemistry*, 2000. 39: 3565-74.
157. Chang, C., *et al.*, Dissection of the LXXLL nuclear receptor-coactivator interaction motif using combinatorial peptide libraries: discovery of peptide antagonists of estrogen receptors alpha and beta. *Mol Cell Biol*, 1999. 19: 8226-39.
158. Plevin, M. J., Mills, M. M., and Ikura, M., The LxxLL motif: a multifunctional binding sequence in transcriptional regulation. *Trends Biochem Sci*, 2005. 30: 66-9.
159. Goldschmidt, L., Cooper, D. R., Derewenda, Z. S., and Eisenberg, D., Toward rational protein crystallization: A Web server for the design of crystallizable protein variants. *Protein Sci*, 2007. 16: 1569-76.
160. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J., Phaser crystallographic software. *Journal of Applied Crystallography*, 2007. 40: 658-674.
161. Derewenda, Z. S., Rational protein crystallization by mutational surface engineering. *Structure*, 2004. 12: 529-35.
162. Radivojac, P., Obradovic, Z., Brown, C. J., and Dunker, A. K., Prediction of boundaries between intrinsically ordered and disordered protein regions. *Pac Symp Biocomput*, 2003: 216-27.
163. Brunger, A. T., *et al.*, Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallographica Section D*, 1998. 54: 905-921.
164. Berman, H. M., *et al.*, The Protein Data Bank. *Nucleic Acids Res*, 2000. 28: 235-42.
165. Su, C. T., Chen, C. Y., and Ou, Y. Y., Protein disorder prediction by condensed PSSM considering propensity for order or disorder. *BMC Bioinformatics*, 2006. 7: 319.



**ABBREVIATIONS**

$\alpha$	alpha
°C	degree centigrade (Celsius)
$\mu$	micro ( $10^{-6}$ )
A	adenine or adenosine
APS	ammonium persulfate
$\beta$	beta
BESSY	Berliner Elektronenspeicherring-Gesellschaft für Synchrotronstrahlung
bp	base pair(s)
BSA	bovine serum albumin
C	cytosine or cytidine
C-ter	carboxy terminal end of polypeptide / protein
Da	molecular weight unit, Dalton (g/mol)
DNA	deoxyribonucleic acid
ECL	enhanced chemiluminescence
EDTA	ethylene diamine tetraacetate
FPLC	fast performance liquid chromatography
$\gamma$	gamma
G	guanine or guanosine
GADD	growth arrest and DNA damage
HABA	hydroxy-azophenyl-benzoic acid
HEPES	2-[4-(2-hydroxyethyl)-1-piperazino]-ethanesulfonic acid
IL	Interleukin
IPTG	isopropyl thiogalactoside
k	kilo, $10^3$
kb	kilo base(s)
kDa	kilo Dalton
l	volume unit, liter
LB	Luria-Bertani medium
m	milli, $10^{-3}$
M	concentration unit, molar, moles per liter
MAD	multiwavelength anomalous diffraction
MAPK	mitogen activated protein kinase
MCS	multiple cloning site
min	time unit, minute
MIR	multiple isomorphous replacement
MMS	methyl methanesulfonate
mol	molecular unit, mole, equals $6.023 \times 10^{23}$ particles of a substance
MS	mass spectrometry
N	amino terminal end of polypeptide / protein
NCS	non-crystallographic symmetry
NTA	nitrilotriacetic acid
N-ter	amino terminal end of polypeptide / protein
OD	optical density

PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PBS	phosphate-buffered saline
PDB	Protein Data Bank
PEG	polyethylene glycol
PMSF	phenylmethanesulfonyl fluoride
pI	isoelectric point
psi	pressure unit, pounds per square inch
PSPF	Protein Sample Production Facility
PVDF	polyvinylidene fluoride
QCM	QuikChange mutagenesis
rmsd	root-mean-square deviation
s	time unit, second
SDS	sodium dodecylsulfate
SEC	size-exclusion chromatography
SU	subunit
T	thymine or thymidine
T <sub>m</sub>	melting temperature
TBE	TRIS-borate-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
TGF $\beta$	transforming growth factor $\beta$
TNF $\alpha$	tumor necrosis factor $\alpha$
TRIS	Tris (hydroxymethyl)-aminomethane
UV	ultraviolet radiation
V	voltage
YTH	Yeast two hybrid
wt	wild-type

Three-letter and single-letter codes for amino acids

Ala/A	Alanine	Leu/L	Leucine
Arg/R	Arginine	Lys/K	Lysine
Asn/N	Asparagine	Met/M	Methionine
Asp/D	Aspartate	Phe/F	Phenylalanine
Cys/C	Cysteine	Pro/P	Proline
Gly/G	Glycine	Ser/S	Serine
Gln/Q	Glutamine	Thr/T	Threonine
Glu/E	Glutamate	Trp/W	Tryptophan
His/H	Histidine	Tyr/Y	Tyrosine
Ile/I	Isoleucine	Val/V	Valine

## APPENDIX

### Publications in journals

**Sarbani Bhattacharya**, Jürgen J. Müller, Yvette Roske, Andrew P. Turnbull, Claudia Quedenau, Frank Götz, Konrad Büssow and Udo Heinemann. The Crystal Structure of Human GADD45 $\gamma$ . (manuscript in preparation).