Structural Basis for Complex Formation between Human IRSp53 and the Translocated Intimin Receptor Tir of Enterohaemorrhagic E. coli

&

Structural Characterisation of Amorfrutins Bound to PPARγ

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina
zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Jens Christian de Groot
aus Wermelskirchen
1. Referent: Honorarprofessor Dr. Dirk W. Heinz
2. Referent: Professor Dr. Michael Steinert
eingereicht am: 12.12.2011
mündliche Prüfung (Disputation) am: 23.04.2012

Druckjahr 2012
Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Publikationen

# Contents

Abbreviations........................................................................................................................................ VI

Zusammenfassung ................................................................................................................................... 10

Summary .............................................................................................................................................. 12

1 Chapter I ............................................................................................................................................. 13

Structural Basis for Complex Formation between Human IRSp53 and the Translocated Intimin Receptor Tir of Enterohaemorrhagic *E. coli* ............. 13

1.1 Introduction ..................................................................................................................................... 13

1.1.1 *Escherichia coli* and the pathogenic deputies ........................................................................... 13

1.1.2 Enterohaemorrhagic *E. coli* (EHEC) and the haemolytic-uremic syndrome (HUS) ......................................................................................................................... 14

1.1.3 The EHEC serotype O157:H7 account for 90% of all HUS cases .............................................. 15

1.1.4 Shiga toxins (Stxs) of EHEC cause HUS manifestation .............................................................. 16

1.1.4.1 Stxs and their causal role in HUS .............................................................................................. 16

1.1.4.2 Stxs are released into the gut by EHEC .................................................................................. 17

1.1.4.3 Stxs trafficking ....................................................................................................................... 18

1.1.4.4 Stxs and the fatal consequences in patients ............................................................................ 19

1.1.4.5 Non-Stx-HUS or atypical-HUS .............................................................................................. 19

1.1.5 The mechanism of EHEC to colonise the human intestine ....................................................... 20

1.1.5.1 The locus of enterocyte effacement (LEE) ........................................................................... 20

1.1.5.2 Attaching-and-effacing (A/E) lesions induce actin pedestal formation .................................. 22

1.1.5.3 Comparison of EHEC and EPEC A/E lesions ....................................................................... 26

1.1.6 The domain organisation of IRSp53 and the homologue proteins IRTKS and FLJ22582 ................ 27

1.1.6.1 The IMD (I-BAR domain) ...................................................................................................... 27

1.1.6.2 The SH3 domain .................................................................................................................... 28

1.1.7 Therapeutic treatment of HUS .................................................................................................. 29
1.2 **Aim of this work** .................................................................................................................. 30
1.3 **Materials** ................................................................................................................................ 31
1.3.1 Molecular weight standards ............................................................................................................. 31
1.3.2 Media, buffers and stock solutions ..................................................................................................... 31
1.3.3 Bacterial strains ..................................................................................................................................... 32
1.3.4 Plasmids for IMD-IRSp53 expression .................................................................................................... 32
1.3.5 IMD constructs ........................................................................................................................................ 32
1.3.6 Peptides ................................................................................................................................................ 33
1.3.7 Crystallisation screens .......................................................................................................................... 33
1.4 **Methods** ............................................................................................................................................. 33
1.4.1 Production and purification of IRSp53-IMD and IRSp53-IMD mutants ................................................. 33
1.4.2 IRSp53-IMD:Tir complex formation ....................................................................................................... 34
1.4.3 Analytical methods ............................................................................................................................... 34
1.4.3.1 Photometric quantification of protein concentrations ........................................................................ 34
1.4.3.2 Discontinuous SDS-PAGE .............................................................................................................. 34
1.4.3.3 Dynamic light scattering (DLS) ......................................................................................................... 35
1.4.3.4 Mass spectroscopy (MS) .................................................................................................................. 35
1.4.3.5 Isothermal titration calorimetry (ITC) ............................................................................................... 36
1.4.4 Crystallisation by vapour diffusion ....................................................................................................... 37
1.4.4.1 Sitting-drop vapour diffusion .......................................................................................................... 37
1.4.4.2 Hanging-drop vapour diffusion ....................................................................................................... 37
1.4.4.3 Microseeding .................................................................................................................................... 38
1.4.5 Data collection, structure determination and refinement ....................................................................... 38
1.4.6 Bioinformatics ...................................................................................................................................... 39
1.4.6.1 Sequence analysis ............................................................................................................................ 39
1.4.6.2 Structural analysis ........................................................................................................................... 39
1.4.6.3 Figure preparation .......................................................................................................................... 39
1.5 **Results** .............................................................................................................................................. 40
1.5.1 Production and purification of IMD-IRSp53 ...................................................................................... 40
1.5.2 Production and purification of IMD-IRSp53 mutants ............................................................................ 41
CONTENTS

1.5.3 Crystallisation of the IMD-IRSp53:Tir complex .................................................42
1.5.4 Crystal structure of the IMD-IRSp53:Tir complex .............................................42
1.5.4.1 Both Tir molecules bound by the IMD point into the same direction ........43
1.5.4.2 A novel peptide binding site of IRSp53 recognises the NPY motif ..........46
1.5.4.3 The Tir binding site is conserved .................................................................50
1.5.4.4 ITC analysis to study the effect of mutations on the IMD-IRSp53:Tir complex formation .................................................................51

1.6 Discussion .................................................................................................................54
1.6.1 The NPY motif of Tir: a ‘hot spot’ for IMD recognition .....................................54
1.6.2 The conserved Tir binding pocket of IMD: the essential residues and their relevance for NPY recognition .................................................................54
1.6.3 The conserved Tir binding pocket of IMD and the preference for NPY motifs...58
1.6.4 Tir is the first unambiguously IMD binder .........................................................58
1.6.5 The affinity of Tir ...............................................................................................58
1.6.6 Model of the intimin:Tir:IRSp53 complex ..........................................................59

1.7 Outlook .....................................................................................................................61
1.7.1 NPY motif containing host proteins may represent physiological ligands for IRSp53/IRTKS .................................................................................................61
1.7.2 The Tir:IMD interaction is likely to compete with membrane binding ..........63
1.7.3 The binding affinity of Tir to IMD .......................................................................63

2 Chapter II .....................................................................................................................64

Structural Characterisation of Amorfrutins Bound to the Peroxisome

Proliferator-activated Receptor gamma .................................................................64

2.1 Introduction ...............................................................................................................64
2.1.1 The nuclear receptor PPARγ ..............................................................................64
2.1.1.1 The activation function 1 (AF1) .....................................................................66
2.1.1.2 The ligand-binding domain (LBD) ...............................................................66
2.1.1.3 Ligand-mediated LBD stabilisation regulates PPARγ’s transcriptional activity .................................................................................................67
2.1.1.3.1 The PPARγ full agonist rosiglitazone .......................................................68
2.1.1.3.2 PPARγ partial and intermediate agonists ........................................69
2.1.1.4 Co-regulator recruitment ........................................................................70
2.1.2 A new class of high affinity PPARγ partial agonists ................................70
2.2 Aim of this work .........................................................................................72
2.3 Materials .....................................................................................................73
2.3.1 Plasmid for hPPARγ LBD expression .......................................................73
2.3.2 Compounds .............................................................................................73
2.4 Methods .......................................................................................................73
2.4.1 Production and purification of PPARγ’s LBD ...........................................73
2.4.2 Analytical methods ..................................................................................74
2.4.2.1 Analytical gel permeation chromatography ........................................74
2.4.2.2 AF4 and DLS .........................................................................................74
2.4.3 Crystallisation by vapour diffusion .........................................................75
2.4.3.1 Sitting-drop vapour diffusion ................................................................75
2.4.3.2 Hanging-drop vapour diffusion ............................................................75
2.4.3.3 Crystal soaking .....................................................................................75
2.4.3.4 Co-crystallisation ................................................................................75
2.4.4 Data collection, structure determination and refinement .........................75
2.4.5 Bioinformatics ........................................................................................76
2.4.5.1 Sequence analysis ................................................................................76
2.5 Results .........................................................................................................77
2.5.1 Production and purification of the PPARγ ligand-binding domain (LBD) ....77
2.5.2 Crystallisation of PPARγ’s LBD with different amorfrutins .....................80
2.5.2.1 Crystal soaking .....................................................................................80
2.5.2.2 Co-crystallisation ................................................................................81
2.5.3 Structural characterisation of different amorfrutins in complex with the LBD of PPARγ .........................................................................................82
2.5.4 Crystallographic contacts of the inactive conformation lead to an artificial orientation and localisation of the amorfrutins ........................................89
2.6  Discussion .................................................................................................................. 91
2.6.1 Structural insights into amorfrutin recognition by PPARγ ........................................ 91
2.6.1.1 Structural comparison with other related PPARγ ligands ........................................ 92
2.6.1.2 Possible mechanism of differential co-regulator recruitment .................................. 93
2.6.1.3 The active and inactive state of the PPARγ’s LBD ..................................................... 94

2.7  Outlook ....................................................................................................................... 95

3  References .................................................................................................................... 97

4  Accession Numbers ..................................................................................................... 110

Danksagung ...................................................................................................................... 111
### Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>9(S)-HODE</td>
<td>(9S)-9-Hydroxyoctadeca-10,12-dienoic acid</td>
</tr>
<tr>
<td>A</td>
<td>Absorption</td>
</tr>
<tr>
<td>Å</td>
<td>Ångström (0.1 nm)</td>
</tr>
<tr>
<td>ABBA</td>
<td>actin-bundling protein with BAIAP2 homology</td>
</tr>
<tr>
<td>Abl</td>
<td>Abelson murine leukemia viral oncogene homologue</td>
</tr>
<tr>
<td>A/E</td>
<td>Attaching-and-effacing</td>
</tr>
<tr>
<td>AF4</td>
<td>Asymmetric flow field flow fractionation</td>
</tr>
<tr>
<td>Arp2/3</td>
<td>Actin-related protein 2/3</td>
</tr>
<tr>
<td>BAIAP2</td>
<td>Brain-specific angiogenesis inhibitor 1-associated protein 2</td>
</tr>
<tr>
<td>BAR</td>
<td>Bin-Amphiphysin-Rvs</td>
</tr>
<tr>
<td>Bcr</td>
<td>breakpoint cluster region</td>
</tr>
<tr>
<td>CCP4</td>
<td>Collaborative Computational Project No. 4</td>
</tr>
<tr>
<td>Cdc42</td>
<td>Cell division control protein 42</td>
</tr>
<tr>
<td>Cif</td>
<td>Cycle inhibiting factor</td>
</tr>
<tr>
<td>CRIB</td>
<td>Cdc42/Rac interactive binding</td>
</tr>
<tr>
<td>Crk</td>
<td>Chicken tumor nr. 10 regulator of kinase</td>
</tr>
<tr>
<td>CRKL</td>
<td>Crk-like protein</td>
</tr>
<tr>
<td>DRLPA</td>
<td>Dentatorubral-pallidoluysian atrophy</td>
</tr>
<tr>
<td>CV</td>
<td>Column volume</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton</td>
</tr>
<tr>
<td>DAEC</td>
<td>Diffusely adherent <em>E. coli</em></td>
</tr>
<tr>
<td>DLS</td>
<td>Dynamic light scattering</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl sulfoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EAEC</td>
<td>Enteroaggregative <em>E. coli</em></td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>ECP</td>
<td>E. coli common pilus</td>
</tr>
<tr>
<td>EH</td>
<td>Eps15 homology</td>
</tr>
<tr>
<td>EHEC</td>
<td>Enterohaemorrhagic E. coli</td>
</tr>
<tr>
<td>EIEC</td>
<td>Enteroinvasive E. coli</td>
</tr>
<tr>
<td>EPEC</td>
<td>Enteropathogenic E. coli</td>
</tr>
<tr>
<td>Eps8</td>
<td>Epidermal growth factor receptor kinase substrate 8</td>
</tr>
<tr>
<td>Erk1</td>
<td>Extracellular signal-regulated kinase 1</td>
</tr>
<tr>
<td>Esp</td>
<td>E. coli secreted proteins</td>
</tr>
<tr>
<td>Gb3</td>
<td>glycosphingolipid globotriaosylceramide</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione S-transferase</td>
</tr>
<tr>
<td>Hck</td>
<td>Hematopoietic cell kinase</td>
</tr>
<tr>
<td>HCP</td>
<td>Haemorrhagic coli pilus</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>hPPARγ</td>
<td>human peroxisome proliferator-activated receptor gamma</td>
</tr>
<tr>
<td>IEC</td>
<td>Ion exchange chromatography</td>
</tr>
<tr>
<td>IMAC</td>
<td>Immobilised metal ion affinity chromatography</td>
</tr>
<tr>
<td>IMD</td>
<td>IRSp53-MIM homology domain</td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl-β-D-thiogalactoside</td>
</tr>
<tr>
<td>IRSp53</td>
<td>Insulin receptor substrate p53</td>
</tr>
<tr>
<td>IRTKS</td>
<td>Insulin receptor tyrosine kinase substrate</td>
</tr>
<tr>
<td>ITC</td>
<td>Isothermal titration calorimetry</td>
</tr>
<tr>
<td>LBD</td>
<td>ligand-binding domain</td>
</tr>
<tr>
<td>M</td>
<td>Molecular mass</td>
</tr>
<tr>
<td>Map</td>
<td>Mitochondrial-associated protein</td>
</tr>
<tr>
<td>mDia</td>
<td>Mammalian homologue of <em>Drosophila diaphanous</em></td>
</tr>
<tr>
<td>Mena/VASP</td>
<td>Mammalian enabled/Vasodilator-stimulated phosphoprotein</td>
</tr>
<tr>
<td>MIM</td>
<td>Missing in metastasis</td>
</tr>
<tr>
<td>MWCO</td>
<td>Molecular weight cut off</td>
</tr>
<tr>
<td>N-CoR</td>
<td>Nuclear receptor co-repressor</td>
</tr>
<tr>
<td>Nef</td>
<td>Negative factor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>NleA</td>
<td>Non-LEE-encoded effector A (also known as Espl)</td>
</tr>
<tr>
<td>nTZDpa</td>
<td>Non-thiazolidinedione selective partial agonist</td>
</tr>
<tr>
<td>NMEC</td>
<td>Neonatal meningitis <em>E. coli</em></td>
</tr>
<tr>
<td>N-WASP</td>
<td>Neural Wiskott-Aldrich syndrome protein</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamid gel electrophoresis</td>
</tr>
<tr>
<td>PDB</td>
<td>Protein Data Bank</td>
</tr>
<tr>
<td>PDZ</td>
<td>Postsynaptic density-95(PSD-95)-discs large-zone occludens-1</td>
</tr>
<tr>
<td>PGC-1α</td>
<td>Peroxisome proliferator-activated receptor gamma co-activator-1 alpha</td>
</tr>
<tr>
<td>PGC-2</td>
<td>Peroxisome proliferator-activated receptor gamma co-activator-2</td>
</tr>
<tr>
<td>15d-PGJ2</td>
<td>15-Deoxy-Delta-12,14-prostaglandin J2</td>
</tr>
<tr>
<td>PI(4,5)P₂</td>
<td>Phosphatidylinositol-4,5-bisphosphate</td>
</tr>
<tr>
<td>PPAR</td>
<td>Peroxisome proliferator-activated receptor</td>
</tr>
<tr>
<td>PTB</td>
<td>Phosphotyrosine binding</td>
</tr>
<tr>
<td>r.m.s.d.</td>
<td>Root mean square deviation</td>
</tr>
<tr>
<td>rpm</td>
<td>Rounds per minute</td>
</tr>
<tr>
<td>RXRα</td>
<td>Retinoid X receptor alpha</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulphate</td>
</tr>
<tr>
<td>SH3</td>
<td>Src homology 3</td>
</tr>
<tr>
<td>Shank-1</td>
<td>SH3 and multiple ankyrin repeat domains 1</td>
</tr>
<tr>
<td>SMRT</td>
<td>Silencing mediator co-repressor for retinoid and thyroid-hormone</td>
</tr>
<tr>
<td>SPPARγM</td>
<td>Selective PPARγ modulators</td>
</tr>
<tr>
<td>Src</td>
<td>Proto-oncogene tyrosine-protein kinase Src (sarcoma)</td>
</tr>
<tr>
<td>SRC1</td>
<td>Steroid receptor co-activator-1</td>
</tr>
<tr>
<td>Stx</td>
<td>Shiga toxin</td>
</tr>
<tr>
<td>T3SS</td>
<td>Type three secretion system</td>
</tr>
<tr>
<td>TCEP</td>
<td>Tris(2-carboxyethyl)phosphine</td>
</tr>
<tr>
<td>Tir</td>
<td>Translocated intimin receptor</td>
</tr>
<tr>
<td>TirC</td>
<td>Carboxy-terminal cytoplasmic domain of Tir</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TJ</td>
<td>Tight junctions</td>
</tr>
<tr>
<td>TLS</td>
<td>Translation/libration/screw</td>
</tr>
<tr>
<td>$T_m$</td>
<td>Melting temperature</td>
</tr>
<tr>
<td>TZD</td>
<td>Thiazolidinedione</td>
</tr>
<tr>
<td>UPEC</td>
<td>Uropathogenic <em>E. coli</em></td>
</tr>
<tr>
<td>WASP</td>
<td>Wiskott-Aldrich syndrome protein</td>
</tr>
<tr>
<td>WAVE</td>
<td>WASP-family verpolin-homologous protein</td>
</tr>
</tbody>
</table>
**Zusammenfassung**

**Strukturelle Grundlagen zur Komplexbildung von humanem IRSp53 mit dem *translocated intimin receptor* Tir enterohämorrhagischer E. coli**


**Strukturelle Charakterisierung der Bindung von Amorfrutinen an den *peroxisome proliferator-activated receptor gamma***

ungiftige Naturstoffe sind, könnten diese nicht nur als alternative Medikation bei Typ-II-Diabetes ohne Nebenwirkungen dienen, sondern wären auch als Nahrungsergänzungsmittel zur Prävention solcher Krankheiten geeignet.
Summary

**Structural Basis for Complex Formation between Human IRSp53 and the Translocated Intimin Receptor Tir of Enterohaemorrhagic *E. coli***

Actin assembly beneath enterohaemorrhagic *E. coli* (EHEC) attached to its host cell is triggered by the intracellular interaction of its translocated effector proteins Tir and EspF with human IRSp53 family proteins and N-WASP. This study reveals the structure of the amino-terminal IMD domain of IRSp53 in complex with a peptide derived from the Tir sequence, in which the homodimeric IMD domain binds two Tir molecules aligned in parallel. This arrangement provides a protein scaffold linking the bacterium to the host cell's actin polymerisation machinery. The structure uncovers a novel, specific peptide binding site on the surface of the IMD, which is conserved between IRSp53 and its homologue IRTKS. The Tir Asn-Pro-Tyr (NPY) motif, which is essential for pedestal formation, is specifically recognised by this binding site. The site was confirmed by mutagenesis and *in vivo* binding assays. It is possible that IRSp53 utilises the NPY binding site for additional interactions with as yet unknown partners within the host cell.

**Structural Characterisation of Amorfrutins Bound to the Peroxisome Proliferator-activated Receptor gamma**

Amorfrutins were recently identified as a non-toxic class of natural products with high affinity to the peroxisome proliferator-activated receptor gamma (PPARγ), a nuclear receptor with regulatory functions in lipid and glucose metabolism. The PPARγ agonist rosiglitazone is currently used in treatment of type II diabetes. It increases insulin sensitivity but has adverse effects such as weight gain. Amorfrutins were also shown to increase insulin sensitivity and to normalise triglyceride metabolism, but do not enhance undesired fat storage. Amorfrutins therefore represent a very promising alternative for treatment or prevention of type II diabetes and the metabolic syndrome in general. This study reveals the structural basis of PPARγ activation by the currently most promising amorfrutins, amorfrutin 1, 2, and B. In contrast to rosiglitazone, the high resolution structures revealed that PPARγ binding and activation by amorfrutins are helix H12-independent. Comparison with complex structures of PPARγ with other ligands confirmed that the amorfrutins belong to the group of partial PPARγ agonists. Few partial agonists, like amorfrutins 1, were already shown to separate adverse effects like adipogenesis from the induced insulin sensitivity *in vivo*. Thus, the amorfrutins have not only the potential to work as alternative drugs to treat diseases like type II diabetes without the induction of adverse effects. These non-toxic naturally compounds may also act as a dietary supplement to prevent these diseases.
Chapter I

Structural Basis for Complex Formation between Human IRSp53 and the Translocated Intimin Receiver Tir of Enterohaemorrhagic E. coli

1.1 Introduction

1.1.1 *Escherichia coli* and the pathogenic deputies

The ubiquitous commensal bacterium *E. coli* requires only a combination of few mobile genetic elements for changing into a highly adapted pathovar, capable of causing a range of diseases from gastroenteritis to extraintestinal infections of the urinary tract, bloodstream, and central nervous system (Croxen and Finlay, 2010). These genetic elements are called pathogenicity islands (PAIs), encoding for an arsenal of virulence factors like toxins, colonization factors, and other virulence factors. PAIs are located on plasmids or integrated into the bacterial chromosome and are flanked by prophages, transposons, integrases, and transposases to facilitate their mobility (Kirsch et al., 2004).

Annually hundreds of millions of people worldwide suffer from pathogenic *E. coli* infections, rationalizing extensive research to characterise the different pathovars. Eight pathovars have been well characterised and can be classified as either diarrhoeagenic or extraintestinal (Kaper et al., 2004). The group of diarrhoeagenic *E. coli* include enterohaemorrhagic *E. coli* (EHEC), enteropathogenic *E. coli* (EPEC), enterotoxigenic *E. coli* (ETEC), enteroinvasive *E. coli* (EIEC), enteroaggregative *E. coli* (EAEC) and diffusely adherent *E. coli* (DAEC). Uropathogenic *E. coli* (UPEC) and neonatal meningitis *E. coli* (NMEC) are classified as extraintestinal *E. coli* (Table 1.1-1), causing urinary tract infections and sepsis/meningitis, respectively. All these pathovars need to adhere for their pathogenicity to the host cells,
which they achieve by characteristic mechanisms of attaching and exploiting the host cells (Croxen and Finlay, 2010). Only EIEC invades the host cell for multiplication and dissemination (Kaper et al., 2004).

Table 1.1-1: Sites of pathogenic *E. coli* colonization and classification into diarrhoeagenic and extraintestinal *E. coli*. EHEC: enterohaemorrhagic *E. coli*, EPEC: enteropathogenic *E. coli*, ETEC: enterotoxigenic *E. coli*, EIEC: enteroinvasive *E. coli*, EAEC: enteroaggregative *E. coli*, DAEC: diffusely adherent *E. coli*, UPEC: uropathogenic *E. coli*, NMEC: neonatal meningitis *E. coli*.

<table>
<thead>
<tr>
<th>Sites of Pathogenic <em>E. coli</em> Colonization and Classification</th>
<th>Diarrhoeagenic <em>E. coli</em></th>
<th>Extraintestinal <em>E. coli</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Brain</td>
<td>EHEC</td>
<td>UPEC</td>
</tr>
<tr>
<td>Bloodstream</td>
<td>EPEC</td>
<td>NMEC</td>
</tr>
<tr>
<td>Large bowel</td>
<td>ETEC</td>
<td></td>
</tr>
<tr>
<td>Small bowel</td>
<td>EIEC</td>
<td></td>
</tr>
<tr>
<td>Kidney</td>
<td>EAEC</td>
<td></td>
</tr>
<tr>
<td>Bladder</td>
<td>DAEC</td>
<td></td>
</tr>
</tbody>
</table>

1.1.2 Enterohaemorrhagic *E. coli* (EHEC) and the haemolytic-uremic syndrome (HUS)

Enterohaemorrhagic *E. coli* (EHEC) are food-born human pathogenic bacteria that cause severe bloody diarrhoea. They are capable of triggering the fatal haemolytic uremic syndrome (HUS), the main cause of acute renal failure in childhood and adolescence, accompanied by low platelet count and destruction of red blood cells. The course of disease (Figure 1.1-1) initiates with watery diarrhoea after a mean incubation period of 3 days post infection with EHEC, accompanied by cramping abdominal pain (Karch, 2001). During the next 2 to 3 days, in most patients watery diarrhoea turns bloody. Most patients recover spontaneously within 7 days (Mead and Griffin, 1998). However, about 15% of infected patients younger than 10 years develop HUS (Karch, 2001). The incidence of HUS is about 1.0/100 000 children below 15 years in Germany and Austria, with a mortality of about 3-5% (Mead and Griffin, 1998; Zimmerhackl et al., 2002). Up to 40% of HUS patients exhibit long term sequelae 10 to 15 years after onset of the disease, such as chronic renal failure that leads to a lifetime dependence on dialysis, or other serious complications, such as persisting renal hypertension, proteinuria, diabetes mellitus, or neurological disorders (Karch, 2001; Mead and Griffin, 1998; Zimmerhackl et al., 2002). HUS is the most severe complication of a potentially avoidable food-borne infection, as the bacteria are transmitted through contaminated food in most cases.
1.1.3 The EHEC serotype O157:H7 account for 90% of all HUS cases

The EHEC strains of the O157:H7 serotype, so-named because it expresses the 157th somatic ‘O’ and the 7th flagellar ‘H’ antigen according to the Kauffmann-White scheme (Institut Pasteur, 2007; Orskov and Orskov, 1992), are the best studied EHEC strains, because they are clinically the most relevant pathogenic E. coli in Europe, North America, and Japan (Nataro and Kaper, 1998). Strains of the O157:H7 serotype were the first strains denoted as enterohaemorrhagic E. coli, which are believed to account for over 90% of all cases of HUS in industrial countries (Siegler, 1995), rationalising for instance the routinely examination of all diarrhoeal stools for E. coli O157 in the UK (Mead and Griffin, 1998).

![Figure 1.1-1: History of Infection with EHEC O157:H7.](image)

Screening for EHEC O157:H7 infections relies on the inability of this serotype to ferment sorbitol rapidly, a phenotype distinct from most other E. coli, which leads to the formation of colourless colonies on sorbitol-containing MacConkey agar (Mead and Griffin, 1998). However, due to an increasing incidence of a sorbitol-fermenting serotype of EHEC O157:H7, as in the outbreak in Southern Germany of 2002 (Robert Koch-Institut, 2003), additional detection methods were introduced such as the use of antisera against the O157
antigen or assays for Shiga toxins, which are produced by EHEC (O’Brien et al., 1984) but for instance not by the related EPEC (see also section 1.1.4).

The EHEC serotype O157:H7 possesses a highly conserved 92 kb plasmid, which encodes for different potential virulence factors. For instance, it encodes for EHEC haemolysin (Schmidt et al., 1995) and a type II secretion system (Schmidt et al., 1997). Although the plasmid encodes for several proteins which were expressed during the course of HUS, data supporting a role in the pathogenic process for each protein are still missing. Other virulence factors clearly play essential roles during infection, mainly the prophage encoded Shiga toxins that trigger the HUS (see section 1.1.4), but also other PAI encoded virulence factors, important for adhesion and dissemination of EHEC (see section 1.1.5.1).

1.1.4 Shiga toxins (Stxs) of EHEC cause HUS manifestation

Shiga toxins, or verotoxins, are the main virulence factors of EHEC and are the defining characteristics of Shiga toxin-producing E. coli (STEC), which were absent for instance in the EHEC related EPEC (Hayward et al., 2006). STEC produce two major toxins with a sequence identity of 56%, Stx1 and Stx2 (Jackson et al., 1987). Two variants of Stx1 (Stx1 and Stx1c) as well as several variants of Stx2 (Stx2, Stx2c, Stx2d, Stx2e and Stx2f) are known, with a sequence identity of 84-99% of the homologous to Stx2 (Johannes and Romer, 2010). The Shiga toxins are prophage encoded multi-subunit protein complexes that bind to the glycosphingolipid receptor Gb3 (CD77) on the surface of certain eukaryotic cell types, leading, depending on the target cell type, to apoptosis or the induction of an inflammatory response (Hayward et al., 2006; Obrig, 2010). During the course of HUS in humans, Stx2 and the variant Stx2c are the two mainly produced Shiga toxins (Bockemühl et al., 1997). HUS caused by these toxins typically affects the kidney and the gastrointestinal tract (Johannes and Romer, 2010).

1.1.4.1 Stxs and their causal role in HUS

The causal role of Stxs in HUS, summarized by Karch in 2010, has been inferred from: (1) Stx production is common to all HUS-associated E. coli isolate regardless of the serotype (Karmali et al., 1985); (2) in vivo produced Stxs can be detected in faeces of HUS patients (Karmali et al., 1985); (3) HUS patients develop immune responses to Stxs produced by the infecting strain (Karmali et al., 1985); (4) intravenous injection of purified Stx causes severe disease in laboratory animals, which is prevented by Stx neutralizing antibodies (Richardson et al., 1992); (5) in non-immune animals, circulating Stx binds rapidly to the vascular
endothelium of target organs (Richardson et al., 1992); and (6) HUS-like disease has been described in racing greyhounds orally infected with *E. coli* O157:H7 (Hertzke et al., 1995) and in baboons following injection of purified Stx (Taylor et al., 1999).

### 1.1.4.2 Stxs are released into the gut by EHEC

Stx2 and its variant Stx2c seems to be more relevant than Stx1 during the course of HUS development and Stx2, in contrast to the almost completely cell associated Stx1, is found in the extracellular fraction (Shimizu et al., 2009). Therefore, huge effort has been invested to examine especially the mechanism of the Stx2 release into the gut. To date, there are two mechanisms suggested to be involved in the release of Stx2 into the gut: (1) a specific induction of the Stx2-encoding prophage and (2) a specific Stx2 secretion mechanism (Shimizu et al., 2009).

The release of Stx2 by EHEC into the gut as a result of specific induction of the Stx2-encoding prophage is mediated by the SOS response, a repair mechanism of bacteria induced by DNA damage according to different kind of triggers, like antibiotics or DNase colicins (Toshima et al., 2007). The SOS response is regulated by LexA and the recombinase A (RecA) and involves the expression of more than 40 genes (Croxen and Finlay, 2010). Under normal condition, RecA is transcriptionally repressed by LexA. However, DNA damage activates RecA by the interaction with single stranded DNA, catalysing the cleavage of its repressor LexA, but also the cleavage of the protein CI, a repressor of the lysogenic lambda phage. This leads to an increased production of the prophage encoded Shiga toxins, replication of the lambda phage and finally the induction of the lambdoid phaged-mediated lysis of the host cell and thus the release of the toxins into the gut. Therefore, EHEC infection treatment with antibiotics is often counterproductive as it leads to the release of the intracellular toxins and even triggers their expression, as e.g. shown for mitomycin C that greatly induce Stx2 expression (Shimizu et al., 2009; Wong et al., 2000; Zhang et al., 2000).

Along with the SOS response-mediated Stxs release by lysis of the host cell, a Stx2 specific secretion mechanism is also thought to contribute to the translocation of this toxin into the gut (Shimizu et al., 2007; Shimizu et al., 2009). Ser31 of the Stx’s B subunit (see also section 1.1.4.3) seems to be recognised by a Stx2 secretion system in EHEC, as a S31N mutant results in an inhibited extracellular but an increasing cell-associated localisation of Stx2 (Shimizu et al., 2007; Shimizu et al., 2009). This suggested mechanism causes a continuously release of basally expressed Stx2, in contrast to the cell-associated Stx1
(Shimizu et al., 2009). However, the detailed mechanism of Stx2 release into the gut remains still elusive.

1.1.4.3 Stxs trafficking

X-ray crystallography revealed a hexameric structure with an AB$_5$ molecular configuration of the Shiga toxin, with a molecular mass of 32 kDa of the A and 7.7 kDa of each B subunit, respectively (Fraser et al., 1994; Stein et al., 1992). The B subunits, encircling the carboxy-terminus of the A subunit, dictate the localisation of the Shiga toxins within the bacterial cells and thus are responsible for the almost complete cell associated localisation of Stx1 in EHEC and the generally extracellular localisation of Stx2 (Shimizu et al., 2007; Weinstein et al., 1989). Beside its crucial role regarding bacterial localisation, the B subunit also dictates the toxins internalization by the host cells. The B subunit exhibit two high affinity and one low affinity binding site for three Gb3 molecules, leading to a binding constant in the nanomolar range (Bast et al., 1999; Fuchs et al., 1986; Soltyk et al., 2002). The interaction of the B subunit with Gb3 on the surface of paneth cells (Schuller et al., 2007), located at the bottom of the crypt, induces endocytosis of the toxins. The toxins are then transported in a retrograde manner via the golgi to the endoplasmic reticulum (Johannes and Romer, 2010). Thereafter the A subunit is activated by an enzymatically cleavage, released into the cytoplasm and inactivates the eukaryotic ribosome by the removal of a single adenine base from the 28S rRNA within the large 60S ribosomal subunit (Endo et al., 1988). This inhibits protein synthesis which may result in apoptosis (Obrig et al., 1988).

However, it should be noted that Stx is shown to be also present in human Gb3-negative intestinal cells, possibly after being taken up by macropinocytosis (Malyukova et al., 2009). Macropinocytosis is a non-selective uptake mechanism based on actin cytoskeleton rearrangement at the plasma membrane (Lim and Gleeson, 2011), as it occurs during the course of EHEC infection (see section 1.1.5.2). However, inside these cells, Stx does not prevent protein synthesis or induce apoptosis (Schüler et al., 2004), but may inhibit chemokine expression and that in turn would inhibit inflammation (Gobert et al., 2007). Furthermore, Stx2 increases the expression of the cell surface protein nucleolin in epithelial cells, shown to interact with the cell surface protein intimin of EHEC and thus may contribute to the bacterial adhesion (see also section 1.1.5.2) (Robinson et al., 2006; Sinclair and O'Brien, 2002).
1.1.4.4 Stxs and the fatal consequences in patients

Stxs provoke local damage of the microcirculation, causing vasculitis, which exacerbates mucosal infraction and leads to bleeding into the bowel and thus bloody diarrhoea (Johannes and Romer, 2010), the primary clinical symptom of typical diarrhoea-associated HUS. Cattle are the main reservoir of STEC, but surprisingly without any clinical symptoms of infection, which may be explained by the lack of the Gb3 receptor (Pruimboom-Brees et al., 2000).

The Stxs are also able to travel by the blood stream to other organs with fatal consequences. Peripheral blood monocytes, shown to be Stxs resistant despite their Gb3 receptor, respond to Stx interaction with releasing pro-inflammatory cytokines (Harrison et al., 2004; Ramegowda and Tesh, 1996) that in turn stimulate Gb3 expression by endothelial cells of different organs (Harrison et al., 2004; Johannes and Romer, 2010). Thereby, organs like the lung, the central nervous system, but most notably the kidney, were sensitized to Stx intoxication. The primary severe damage of the kidney may be explained by its most abundant amount of expressed Gb3, but also by the high volume of blood flow and the high blood filtration rate and thus a more excessive interaction with the toxin, compared with other organs (Obrig, 2010). As mentioned above, the most affected patients are younger than 10 years which may be ascribed to the fact that only pediatric kidneys express Gb3 in the glomeruli (Lingwood, 1994), cells located at the blood-urine barrier. The microenvironment of the Gb3 receptor seems to be also critical for the sensitivity of the receptor to the toxins, as Gb3 was shown to interact with membrane components like cholesterol, other glycolipids, fatty acids, or other proteins (Lingwood et al., 2010a; Lingwood et al., 2010b; Mahfoud et al., 2010; Nutikka and Lingwood, 2004).

Furthermore, Stxs trigger a thrombotic mechanism. The toxin induces platelet-leukocyte aggregation and expression of tissue factor in the kidney, the brain (Sugatani et al., 2000), and blood cells (Stahl et al., 2009). Tissue factor containing microparticles were released, enter the circulation and initiate the conversion of fibrinogen to fibrin, resulting in the formation of an insoluble fibrin clot (Bach, 1988) and thus increasing the risk of vascular congestion.

Even if Shiga toxins are now commonly accepted as main virulence factors of pathogenic STEC (Karch, 2001; O’Brien et al., 1992; Obrig, 2010), their role during the pathogenesis of the typical diarrhoea-associated HUS is not yet completely understood.

1.1.4.5 Non-Stx-HUS or atypical-HUS

The Non-Stx-HUS is less common than the typical Stx mediated and diarrhoea-associated HUS, is not diarrhoea-associated, accounts for only 5 to 10% of all cases of the disease and,
in contrast to the Stx mediated HUS, may manifest at all ages, but is more frequent in adults (Noris and Remuzzi, 2005). As the name already implies, the non-Stx-HUS is not caused by Stx, but by non-Stx producing bacteria (*Streptococcus pneumoniae*), or e.g. by viruses such as HIV, different drugs, or systemic diseases like lupus (Noris and Remuzzi, 2005). Therefore, this type of HUS is not EHEC mediated and only mentioned in this work for the sake of completeness.

### 1.1.5 The mechanism of EHEC to colonise the human intestine

After ingestion of the pathogenic bacteria, EHEC have to compete against the bacteria of the healthy intestine flora to colonise the gut. To accomplish this task, bacteria take advantage of different evolutionary evolved genetic modifications (see section 1.1.5.1), leading to expression of different effector proteins that induce host cell’s actin cytoskeleton rearrangement and promote their tight adhesion and dissemination. Furthermore, they induce diarrhoea, which flushes out less adherent bacteria and thus helps EHEC to colonise the intestine (see section 1.1.5.2).

#### 1.1.5.1 The locus of enterocyte effacement (LEE)

EPEC and most EHEC strains, including EHEC serotype O157:H7, contain almost identical copies of the locus of enterocyte effacement (LEE) (Perna *et al.*, 1998). The LEE is a chromosomal pathogenicity island that encodes for additional virulence factors aside from the prophage encoded Shiga toxins. These factors are essential for EHEC-mediated rearrangement of the host cell’s actin cytoskeleton and facilitate adhesion and dissemination of the pathogens. The LEE is organized into five major operons, LEE 1 to 5. The LEE, the most important PAI, has evolved over the last 4.5 million years (Reid *et al.*, 2000) to a tightly packed sequence which contains only the genes of capital importance during the course of infection (Kirsch *et al.*, 2004). It comprises 41 open reading frames, encoding transcriptional regulators, a type III secretion system (T3SS), chaperones, translocators, the adhesin intimin, and effector proteins like the translocated intimin receptor (Tir) (Figure 1.1-2) (Garmendia *et al.*, 2005).
Figure 1.1-2: Localisation of the major genes involved in EHEC O157:H7 pathogenicity. Shown are the locus of enterocyte effacement (LEE), the EspF_U and Shiga toxins encoding prophage-U, and the plasmid pO157, encoding for proteins like haemolysin (HlyA). The chromosomal pathogenicity island LEE of *E. coli* O157:H7 is shown according to Garmendia et al. (2005). The five putative polycistronic operons (LEE1 to LEE5) of EHEC LEE are indicated. EHEC image: Manfred Rohde, HZI.

EPEC as well as EHEC infections lead to so-called attaching-and-effacing (A/E) lesions, defined by the attachment of the bacteria to the epithelial cells of the intestinal mucosa and the subsequent loss (effacement) of their microvilli. EHEC require the LEE-encoded proteins intimin, the corresponding receptor Tir, and EspF_U to induce A/E lesion formation in the host cell. The prophage-U encoded effector protein EspF_U, named according to the similarity to the LEE encoded effector protein EspF (Campellone et al., 2004), and Tir need to be translocated by the T3SS into the host cell. These three proteins are sufficient to induce A/E lesions (see section 1.1.5.2) by a pathway related to EPEC infections (see section 1.1.5.3).

A/E lesions seem to play a crucial role during EHEC O157:H7 infection. However, LEE is not essential for the pathogenesis of all STEC, as some cases of severe disease, including HUS, are also caused by LEE-negative STEC strains (Paton et al., 2001). But, these strains are thought to use other mechanisms for adhesion, like the LEE-negative O113:H12 STEC strain, shown to utilise the adhesin Saa (Paton et al., 2001).
1.1.5.2 Attaching-and-effacing (A/E) lesions induce actin pedestal formation

A/E lesions induced by EHEC lead next to the loss of microvilli to localised actin polymerisation and to the formation of pseudopod-like “actin pedestals” beneath the attached bacteria (Figure 1.1-2). *E. coli* typically possess fimbiae, enabling the bacteria to adhere to host cells. The Type 1 fimbiae were described first and are the most common adhesins in *E. coli* (Duguid *et al*., 1955; Welinder-Olsson and Kaijser, 2005) which mediate adherence to mannose-containing glycoproteins found on the surface of many eukaryotic cells (Johnson, 1991). However, the Stx producing EHEC strain O157 is unable to express type 1 fimbiae (Enami *et al*., 1999), but adhere through their common pilus (ECP) (Croten and Finlay, 2010) as well as through their type IV pilus, termed haemorrhagic coli pilus (HCP) (Figure 1.1-3, lower panel) (Xicohtencatl-Cortes *et al*., 2007).

Attachment can also be enhanced by interactions of the bacterial outer membrane protein intimin with nucleolin, a surface-localised intimin receptor, whose expression is increased by Stx2 (see also section 1.1.4.3). The biggest impact on adherence, however, is mediated by the interaction of intimin to the effector protein Tir. Tir and a variety of other LEE and phage encoded effector proteins are translocated into the host cell via the T3SS (Figure 1.1-3, lower panel) (Hayward *et al*., 2006). Therefore, a deletion of the *tir* gene results in a profound reduction in EHEC colonisation (~10,000-fold), as shown for the infant rabbit intestine (Ritchie and Waldor, 2005). Upon secretion, Tir is integrated into the plasma membrane in a hairpin-loop conformation and its extracellular domain serves as another receptor for intimin, thereby anchoring the bacterium to the host cell (Campellone and Leong, 2003). In this arrangement, both termini of Tir project into the host cell cytoplasm providing the recruitment sites for the host’s actin polymerisation machinery (Figure 1.1-3). The amino-terminal cytoplasmic domain is not strictly required for pedestal formation but modulates pedestal length (Campellone *et al*., 2006). Its exact link to the actin cytoskeleton is still largely unclear. In contrast, the carboxy-terminal cytoplasmic domain (TirC) is essential for pedestal formation (Campellone *et al*., 2006).
Figure 1.1-3: Schematic representation of A/E lesions during the course of EHEC infection according to Croxen et al. (2010) and Weiss et al. (2009), respectively. EHEC adhere to colonocytes in the large bowel through the *E. coli* common pilus (ECP) and the haemorrhagic coli pilus (HCP). The EHEC secrete a large arsenal of effector proteins into the host cell by a type III secretion system (T3SS). The effector protein Tir integrates into the plasma membrane and serves as a receptor for the EHEC cell surface protein intimin. The autoinhibited host protein IRSp53 (or its paralogue IRTKS) is activated by interacting with active Cdc42 and is thereby recruited to the membrane. Active IRSp53 interacts with Tir and EspFu, another secreted effector protein. EspFu recruits and activates N-WASP, leading to ARP2/3 dependent actin polymerisation. Thereby, EHEC are linked to the actin cytoskeleton of the host cell, resulting in actin pedestal formation. Shiga toxin (Stx) released into the gut is internalized, possibly by macropinocytosis, into colonocytes, resulting in inhibition of the innate immune response and increased nucleolin expression, which enhances adhesion via additional intimin interactions. Stx is internalized into paneth cells via Gb3 receptor-mediated endocytosis. Upon enzymatic activation, the toxin prevents protein synthesis of the host cell, which can lead to apoptosis. After absorption, Stx can travel with the blood stream to other organs, causing severe damage especially to the kidneys. The effector proteins EspG, EspF and Map trigger a disrupted ion balance and inhibit water absorption, leading to watery diarrhoea. Inset: Schematic representation of the
domain interactions linking TirEHEC to Arp2/3 complex-mediated actin polymerisation during pedestal formation. The IMD and SH3 domains of IRSp53 or IRTKS interact with the NPY-motif of TirEHEC and the proline-rich repeats of EspF_U, respectively. The proline-rich repeats of EspF_U are capable of interacting with N-WASPs, bound to the Arp2/3 complex which drives actin polymerisation, leading to a direct link of the extracellular pathogenic bacterium to the host cell’s actin cytoskeleton. The LEE and prophage encoded effector proteins are shown in red and green, respectively.

EspF_U, as mentioned before, is also transported via the T3SS into the host cell. EspF_U comprises a variable number of almost identical, proline-rich 47 amino acid repeats (Garmendia et al., 2004; Sallee et al., 2008) (Figure 1.1-3, inset), each including a hydrophobic helix that binds to and activates the autoinhibited N-WASP (Campellone et al., 2004; Cheng et al., 2008), leading to Arp2/3 dependant actin polymerisation. Furthermore, the proline-rich repeats are also capable of binding to the Src homology 3 (SH3) domain of either the host cell protein IRSp53 (insulin receptor substrate p53) (Weiss et al., 2009) or its paralogue IRTKS (insulin receptor tyrosine kinase substrate) (Vingadassalom et al., 2009). Due to multiple proline-rich repeats, EspF_U was shown to simultaneously bind and activate multiple N-WASP proteins (Sallee et al., 2008) and multiple SH3 domains. This allows the focal clustering of IRSp53 and Arp2/3 complexes and drives focal actin assembly (Weiss et al., 2009), leading to pedestal formation.

IRSp53 has an SH3 domain, a central Cdc42/Rac interactive binding (CRIB) motif and an amino-terminal IRSp53-MIM homology domain (IMD) (see also section 1.1.6.1). The CRIB motif seems to be important for the activation of IRSp53, as IRSp53 exists in an autoinhibited state in the cytoplasm but becomes activated after the interaction of the CRIB domain with the membrane associated active form of Cdc42 (Ahmed et al., 2009). The CRIB domain of IRSp53 is atypical, as it comprises less conserved amino acids compared to the typical CRIB of e.g. WASP (Krugmann et al., 2001) and thus is also called “partial CRIB”. Despite this, the CRIB of IRSp53 binds specifically to active Cdc42 but not to the activated Rac (Krugmann et al., 2001; Scita et al., 2008). The activation of the small Rho GTPase Cdc42 is mediated by the T3SS secreted effector protein Map (Figure 1.1-3) (Huang et al., 2009). Map was shown to induce Cdc42 dependant cell surface filopodia (Kenny et al., 2002) and to work as a guanine nucleotide exchange factor (GEF) that specifically activates Cdc42 (Huang et al., 2009), recruiting the activated IRSp53 to the membrane (Ahmed et al., 2009). Furthermore, Map acts as a multifunctional protein that also disrupts the mitochondrial structure and function (Ma et al., 2006).

The IMD of the active IRSp53 interacts with TirC of EHEC, leading to a direct connection of the extracellular EHEC with the actin polymerisation machinery of the host cell (Vingadassalom et al., 2009; Weiss et al., 2009) and the recruitment of the N-WASP/Arp2/3 complex to the plasma membrane. The interaction surface within the IMD has been mapped
down to a stretch of nine amino acids with a central NPY motif (Campellone et al., 2006). Substitutions with alanine demonstrated that each position of the NPY motif but none of the neighbouring positions are critical for actin pedestal formation. Interestingly, a conservative substitution of the tyrosine by phenylalanine attenuated pedestal formation but could not entirely abrogate it (Brady et al., 2007). Despite the related structures of the Tir binding proteins IRSp53 and IRTKS, ectopic expression of IRSp53 mediates the formation of filopodia (Lim et al., 2008), while cytoskeletal changes induced by IRTKS are more subtle (Millard et al., 2005). In any case, these proteins link changes of plasma membrane morphology to actin cytoskeleton dynamics (Millard et al., 2005; Scita et al., 2008; Suetsugu et al., 2006).

As mentioned above, along with Tir and EspF, a lot of other effector proteins are injected into the host cell, like Nle (also known as EspI) that reduces protein trafficking and also disrupts tight junctions, or the cycle inhibiting factor (Cif) that inhibits the cell cycle progression. Some effector proteins account for the primary clinical symptom of an EHEC infection, diarrhoea. Diarrhoea provides the EHEC with an advantage over competing flora of the intestine, since the intimately bound EHEC remain attached to the host's intestine and other less adherent bacteria are flushed away. The diarrhoea is mediated by the loss of absorptive surface due to the loss of microvilli, disrupted ion balance, and water absorption. This endothelial dysfunction is triggered by the T3SS translocated effector proteins like EspF and EspG, leading to a decreased activity of the Na⁺/H⁺ exchanger 3 (NHE3) (Hodges et al., 2008) and the Cl⁻/OH⁻ exchanger (Gill et al., 2007), respectively (Figure 1.1-3, lower panel). The affected ion absorption leads to diarrhoea, as water moves according to the osmotic effect to areas of higher salt concentrations. Furthermore, both proteins induce mislocalisation of aquaporin water channels from the membrane into the cytoplasm (Guttman et al., 2007). EspF, intimin, Map and Tir also cause the inhibition of the sodium-D-glucose co-transporter (SGLT1), a major water pump that accounts for approximately 4 to 6 litre water uptake from the intestine per day (Dean et al., 2006; Meinild et al., 1998). The inhibition of the intestinal serotonin transporter SERT (Esmali et al., 2009) and the upregulation of connexin 43 (Cx43) (Guttman et al., 2011) were also shown to contribute to diarrhoea due to A/E bacterial infection (not shown in the figure). SERT is involved in the uptake of serotonin, a hormone regulating fluid and electrolyte secretion and absorption (Croxen and Finlay, 2010). Connexin 43 (Cx43) forms open connexon hemichannels in apical membranes of infected cells (Guttman et al., 2011). However, up to now no T3SS effector proteins are identified to be associated in the SERT inhibition or Cx43 upregulation. But these dramatic effects of the secreted virulence factors on the epithelial cells, that also include the loss of tight junctions, lead unavoidably to severe diarrhoea. The initially watery
diarrhoea turns to bloody diarrhoea within the first week (Figure 1.1-1), due to the mucosal infraction in response to the EHEC produced Shiga toxin (see section 1.1.4.4).

1.1.5.3 Comparison of EHEC and EPEC A/E lesions

Although EHEC and EPEC are closely related and have an almost identical LEE, EHEC injects around twice as many effectors into host cells as EPEC, even if most of these are redundant (Tobe et al., 2006). These pathovars exhibit distinct adhesion mechanisms and pathways upon attachment that are exploited for inducing actin assembly, as shown for EHEC O157:H7 and EPEC O127:H6 or by other, more recently identified EHEC and EPEC strains (Frankel and Phillips, 2008). However, there are some correlations in particular between the prototypical strains EPEC 1 and the typical non-sorbitol fermenting EHEC O157, as shown in Figure 1.1-4 and discussed in the following. Both strains translocate Tir by the T3SS into the host. The Tir molecule is well conserved between EPEC, EHEC, and also other related bacteria such as Citrobacter rodentium. Only a short stretch of TirC around Tyrosine 474 is specific for EPEC-Tir. Phosphorylation of EPEC Y474 by host cell kinases (Phillips et al., 2004; Swimm et al., 2004) triggers binding of the SH2-domain of Nck proteins (Frese et al., 2006; Gruenheid et al., 2001), resulting in recruitment and activation of N-WASP and stimulation of Arp2/3 dependant actin polymerisation (reviewed in Hayward et al., 2006). Thus, in contrast to EHEC, EPEC exploit Nck as part of their infection strategy (Frese et al., 2006) and links the bacterium IRSp53 and EspF_U independant to the hosts actin cytoskeleton.

Figure 1.1-4: Comparison of the signalling cascade of the prototypical strains EHEC O157 (left) and EPEC 1 (right). While EHEC require IRSp53 and EspF_U to link the bacterium to the host cell’s actin cytoskeleton, EPEC exploit Nck as a linker to the actin cytoskeleton, which implicates the phosphorylation of Tir’s Tyr474. Figure according to (Croxen and Finlay, 2010).
However, even though IRSp53 is present in EPEC but dispensable for EPEC pedestal formation, introduction of EspF₀₁ into EPEC leads to Nck-independent pedestal formation by using an ₄₅₂NP₄₅₄ motif of Tir similar to EHEC Tir (Brady et al., 2007). Thus the canonical EPEC is capable of interacting with IRSp53, whereas no information are currently available about whether Tyr₄₅₄ is phosphorylated or not when bound to IRSp53.

1.1.6 The domain organisation of IRSp53 and the homologue proteins IRTKS and FLJ22582

The proteins most closely related to IRSp53 are IRTKS and the poorly characterised FLJ22582. These proteins share the two main domains: IMD (see section 1.1.6.1) and SH3 (see section 1.1.6.2), but also a WASP homology 2 (WH2)-like domain (Figure 1.1-5), involved in mediating binding to G-actin (Scita et al., 2008). The partial CRIB domain is a unique feature of IRSp53. FLJ22582 lacks also the partial CRIB but in addition the WW domain-binding motif (PPXY).

![Figure 1.1-5: Domain organization of IRSp53 and the homologue proteins IRTKS and FLJ22582.](image)

The different domains of the proteins are indicated.

1.1.6.1 The IMD (I-BAR domain)

The alpha-helical IMDs are distant members of the Bin-Amphiphysin-Rvs (BAR) domain superfamily. Whereas canonical banana-shaped BAR domains bind membranes with their concave surface and induce invagination, IMDs are cigar-shaped and possess a convex surface that interacts with the cell membrane, leading to outward protrusions (Figure 1.1-6) like filopodia and are therefore referred to as inverse BAR or I-BAR domains (Frost et al., 2009; Mattila et al., 2007). Interestingly, IMDs also directly bind to and bundle F-actin (Disanza et al., 2006; Millard et al., 2005; Suetsugu et al., 2006). However, two recent studies suggested that the isolated IMD itself does not crosslink actin filaments under physiological conditions (Lee et al., 2007; Mattila et al., 2007). The actin binding sites overlap, at least in part, with those that target lipid membranes (Millard et al., 2005; Suetsugu...
IMDs deform PI(4,5)P$_2$-rich membranes (Mattila et al., 2007; Saarikangas et al., 2009) whereas the IMDs of IRSp53 or IRTKS interact mainly through electrostatic interactions with phospholipids of the membrane (Saarikangas et al., 2009). In the resulting tubules, the IMDs remain dynamically associated with the inner leaflet and are thought to form an inner helical scaffold (Figure 1.1-6B) (Saarikangas et al., 2009). It should be noted though, that the basic cluster of amino acids that confers lipid and potentially actin binding to IMDs (Millard et al., 2005; Suetsugu et al., 2006) is not required for formation of pedestals by EHEC bacteria (Weiss et al., 2009).

Figure 1.1-6: Hypothesis of BAR domain mediated membrane curvature. (A) Structural comparison of the three different types of BAR domains. Dashed lines indicate the assumed membrane binding site. (B) BAR domains lead according to their structure to the formation of invaginations (left) or protrusions (right). The BAR domains are hypothesized to exhibit a helical scaffold on the cytosolic side of the tubules. Figure (B): (Scita et al., 2008).

1.1.6.2 The SH3 domain

The SRC homology 3 domain (SH3) plays a critical role in a wide variety of biological processes like increasing the local concentration of distinct proteins, altering the subcellular localisation of components of signalling pathways or mediating the assembly of large multiprotein complexes (Mayer, 2001). SH3 domains interact with proline-rich motifs of different proteins. The domain forms a typically β-barrel structure, with anti-parallel β-strands that are connected by variable loops. The small interaction surface of about 400Å$^2$ results in usually weak selectivity and affinities in the low micromolar range to the ligand (Kuriyan and Cowburn, 1997). However, affinity and specificity can be greatly enhanced by additional contacts between SH3 loop regions and residues of the ligand outside the proline-rich motif (Mayer and Saksela, 2004).

It is not surprisingly that SH3 domains exhibit also a high impact for the IRSp53 signalling, as they mediate the interactions to a huge variety of different ligands like WAVE2 (Miki et al.,
2000), Mena/VASP (Krugmann et al., 2001), Eps8 (Disanza et al., 2006), mDia (Fujiwara et al., 2000), espin (Sekerkova et al., 2003), DRLPA (Okamura-Oho et al., 1999), Shank-1 (Bockmann et al., 2002), and synaptopodin (Yanagida-Asanuma et al., 2007). As the SH3 of IRSp53 and IRTKS also interacts with EspFU (Figure 1.1-3), it displays an essential component to physically link Tir and thus EHEC to the actin cytoskeleton of the host cell.

1.1.7 Therapeutic treatment of HUS

The treatment of Shiga toxin-producing bacteria with conventional anti-microbial therapies was often reported to be counterproductive since it may stimulate toxin production (Shimizu et al., 2009; Wong et al., 2000; Zhang et al., 2000). Due to the worst ever outbreak of the very aggressive Stx2 producing HUSECO41 (O104:H4) in Germany 2011, there is a huge demand for therapeutic strategies to treat STEC infections and in particular the associated HUS. Currently no other therapeutic treatment other than fluid volume control, apheresis and dialysis are available to reduce or prevent renal failure. However, there are some promising approaches like the synthetic Shiga toxin binder “DAISY” (Mulvey et al., 2003) or “SUPER TWIG” (Nishikawa et al., 2002) which acts as a Gb3 analogue neutralizing the toxin. Probiotic bacteria with Gb3 mimics on their surface (Paton et al., 2000) or monoclonal antibodies against Stx1 and Stx2 were also tested (Mukherjee et al., 2002). Low purity of the commercially available Stxs are thought to be the reason for diverging results from related experiments, as contaminating substances like lipopolysaccharides are much more potent pro-inflammatory agents than Stxs (Obrig, 2010).

Treatment of HUSECO41 patients with the monoclonal antibody Eculizumab (Soaris; Alexion) to prevent the destruction of red blood cells, dramatically demonstrates the need for new therapeutic strategies. The antibody is developed to treat paroxysmal nocturnal haemoglobinuria, preventing integration of a pore forming terminal complement complex into erythrocytes which would cause haemolysis (Parker et al., 2007; Rother et al., 2007). But unfortunately so far only limited data is available which describes treatment of human HUS patients using this antibody (Gruppo and Rother, 2009; Lapeyraque et al., 2011).
1.2 Aim of this work

IRSp53 and IRTKS were recently discovered to link the human pathogenic bacterium EHEC to the host cell actin polymerisation machinery by interactions of their IMD and SH3 domains to the translocated effector proteins Tir and EspF, respectively.

The aim of this study was to solve the crystal structure of the amino-terminal human IMD of IRSp53 in complex with a synthetic peptide derived from the EHEC effector Tir, shown to interact with IMD (Weiss et al., 2009). The structure should uncover the unknown binding site on the surface of IMD for the NPY motif of the carboxy-terminus of Tir which was already proven to be essential for binding (Brady et al., 2007).

Isothermal titration calorimetry should be used to give information about the thermodynamics and the stoichiometry of this interaction. Close inspection of the structure should facilitate design of IMD point mutants, predicted to be impaired in Tir binding, to verify the structural data. Furthermore, the mutants should identify the most essential residues of the IMD:Tir interaction and give information about their impact.

Hence, this study should reveal a detailed description of a new binding pocket on the surface of IRSp53’s IMD, responsible for Tir recognition. The study should unveil how the NPY motif provides a central molecular scaffold, linking the bacterium and the host cells actin polymerisation machinery. Furthermore, the structure should reveal a novel binding site for NPY motifs, which is probably also implicated in interactions with other so far unknown interaction partners, effecting pathogenesis and/or actin cytoskeletal regulation.
1.3 Materials

All chemicals were purchased from the following companies: ApliChem, Gold Biotechnology, Fluka, Roth, Roche, Hampton Research, Merck, Millipore, Sigma-Aldrich and QIAGEN. The quality standard was “pro analysis” (p.a.).

1.3.1 Molecular weight standards

Precision Plus Protein Unstained Standards and Precision Plus Protein All Blue Standards (Biorad) were used for SDS-PAGE analysis.

1.3.2 Media, buffers and stock solutions

Table 1.3-1: Media, buffer and stock solutions used in this study.

<table>
<thead>
<tr>
<th>Media</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luria Bertani (LB)</td>
<td>10 g/L Bactotryptone, 5 g/l NaCl, 5 g/L yeast extract</td>
</tr>
<tr>
<td>Super broth (SB)</td>
<td>900 ml of solution 1 (12 g/L Bactotryptone, 24 g/L yeast extract, 0.4% (v/v) glycerol) were mixed with 100 ml of solution 2 (0.17 M KH₂PO₄ and 0.72M K₂HPO₄). The solutions were sterilized separately.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Buffers</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>8x SDS-sample-buffer</td>
<td>7% SDS (w/v), 20% (v/v) glycerol, 0.5 M β-mercaptoethanol, 0.25 mg/ml Coomassie Blue G-250, 10 mM Tris/HCL pH: 6.8</td>
</tr>
<tr>
<td>SDS-PAGE lower buffer (4x)</td>
<td>1.5 M Tris/HCl pH 8.8</td>
</tr>
<tr>
<td>SDS-PAGE upper buffer (4x)</td>
<td>0.5 M Tris/HCl pH 6.8, 0.4% (w/v) SDS</td>
</tr>
<tr>
<td>SDS-PAGE running buffer</td>
<td>25 mM Tris/HCl, 192 mM glycine, 0.1% (w/v) SDS</td>
</tr>
<tr>
<td>Protease inhibitors</td>
<td>Solution 1 (0.88 g PMSF, 10 mg Pepstatin A, 10 mg Chymostatin, dissolved in 10 ml DMSO) and solution 2 (1.57 g Benzamidine, 5 mg Leupeptin, 5 mg Bestatin, 5 mg Aprotinin, dissolved in 10 ml ddH2O) were sterilized by filtration, stored at -20°C and separately added to the protein solution in a ratio of 1:1000.</td>
</tr>
</tbody>
</table>
1.3.3  Bacterial strains

Table 1.3-2: Bacterial strains used in this study.

<table>
<thead>
<tr>
<th>Bacterial strains</th>
<th>Genotype</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL21(DE3)</td>
<td>F^-ompT hsdS_Bgal dcm(DE3)</td>
<td>Novagen</td>
</tr>
<tr>
<td>SCS1</td>
<td>recA1 endA1 gyrA96 thi-1 hsdR17(rBmB) supE44 relA1</td>
<td>Stratagene</td>
</tr>
<tr>
<td>Rosetta2(DE3)</td>
<td>F^-ompT hsdS_Bgal dcm(DE3) pRARE2(Cam)</td>
<td>Novagen</td>
</tr>
</tbody>
</table>

1.3.4  Plasmids for IMD-IRSp53 expression

pQTEV: A PCR fragment encoding amino acids 1-250 (MSLS…VASN) of human IRSp53 (UniProt BAIP2_HUMAN), cloned with BamH1 and NotI into a pQTEV (GenBank AY243506), was used for expression, resulting in IMD_{IRSp53} fused to an amino-terminal His<sub>7</sub>-tag and TEV cleavage site with the sequence GSTMSLS…VASN upon cleavage.

pRARE: The pRARE plasmid (Novagen) encodes tRNA genes to improve the codon usage in E. coli of Arg, Ile, Gly, Leu and Pro, except for Arg CCA/CGG (Novy et al., 2001). In this study, the pRARE plasmid was used to enhance protein expression of IMD-IRSp53 in the SCS1 cells.

1.3.5  IMD constructs

Table 1.3-3: IMD constructs used in this study.

<table>
<thead>
<tr>
<th>Construct</th>
<th>plasmids</th>
<th>Bacterial strain</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMD_{IRSp53}wt</td>
<td>1-250</td>
<td>pQTEV/pRARE</td>
</tr>
<tr>
<td>IMD_{IRSp53}L28E*</td>
<td>as above</td>
<td>as above</td>
</tr>
<tr>
<td>IMD_{IRSp53}K108A*</td>
<td>as above</td>
<td>as above</td>
</tr>
<tr>
<td>IMD_{IRSp53}R193S*</td>
<td>as above</td>
<td>as above</td>
</tr>
<tr>
<td>IMD_{IRSp53}F196A*</td>
<td>as above</td>
<td>as above</td>
</tr>
<tr>
<td>IMD_{IRSp53}K108A/R193S*</td>
<td>as above</td>
<td>as above</td>
</tr>
</tbody>
</table>

*The IMD mutants used in this study were kindly provided by Prof. Dr. Theresia Stradal (Institute for Molecular Cell Biology, University of Münster, Germany).
1.3.6 Peptides

Table 1.3-4: Peptides used for ITC.

<table>
<thead>
<tr>
<th>Peptides</th>
<th>Sequenze</th>
<th>SwissProt</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHEC-Tir-peptide</td>
<td>452GTVQNPYADVKT^463</td>
<td>Q7DB77</td>
</tr>
<tr>
<td>EHEC-Tir-peptide-Y458F</td>
<td>452GTVQNPFADVKT^463</td>
<td>as above</td>
</tr>
<tr>
<td>EPEC-Tir-peptide</td>
<td>448SEVVNPYAEVGG^459</td>
<td>B7UM99</td>
</tr>
<tr>
<td>NPxY-peptide of the human insulin receptor</td>
<td>995YSNPEYLSASDV^1005</td>
<td>P06213</td>
</tr>
</tbody>
</table>

1.3.7 Crystallisation screens

Initial crystallisation conditions were identified by using the following screens from QIAGEN: JCSG+, JCSG Core I, JCSG Core II, JCSG Core III, JCSG Core IV, The Classics, The PEGs, The PEGs II, The MPD, Cryos and The AmSO₄.

1.4 Methods

1.4.1 Production and purification of IRSp53-IMD and IRSp53-IMD mutants

_E. coli_ SCS1 or Rosetta2 (Table 1.3-3) cells bearing the IMD-IRSp53 expression plasmid and pRARE were grown in SB medium supplemented with 100 μg/ml ampicillin and 34 μg/ml chloramphenicol at 37°C to an OD of 1.5. Protein expression was induced with 1 mM IPTG at 37°C for 4 hr. Harvested cells were disrupted with a high pressure cell disrupter (Constant Systems) in 20 mM Hepes, pH 7.5, 100 mM NaCl, 10 mM imidazole, 1 mM TCEP and protease inhibitors. After centrifugation (35,000×g, 45 min), the soluble protein was purified by nickel affinity chromatography (5 ml HiTrap Chelating HP column, GE Healthcare) and eluted with an imidazole gradient (10-500 mM) over 20 column volumes. The protein was cleavage with his-tagged TEV protease. Uncleaved material and the protease were removed by nickel rechromatography, followed by gel filtration (HiLoad 16/60 Superdex 75 prep grade column, GE Healthcare) in 20 mM HEPES-NaOH, 100 mM NaCl, 0.1 mM EDTA and 1 mM DTT, pH 7.5. The protein purification for isothermal titration calorimetry (section 1.4.3.5) was carried out with exception of the His-tag truncation.
1.4.2 IRSp53-IMD:Tir complex formation

A peptide with the sequence $452^{GTVQNPYADVKT}463$, derived from EHEC Tir (see section 1.3.6), was synthesized and HPLC-purified in-house by Dr. Werner Tegge. The peptide contains a carboxy-terminal amide and was used for complex formation and crystallisation (see section 1.4.4). The peptide buffer was exchanged against 20 mM HEPES-NaOH, 100 mM NaCl, 0.1 mM EDTA and 1 mM DTT, pH 7.5. The complex was prepared by adding a 2-fold molar excess of peptide to purified IMD with a final concentration of 11 mg/ml.

1.4.3 Analytical methods

1.4.3.1 Photometric quantification of protein concentrations

The protein concentrations were determined at 280 nm by using a NanoDrop ND-1000 photometer (peqlab) according to the Lambert-Beer law (Equation 1.4-1), with the extinction (E), the intensity of the transmitted beam ($I_t$), the intensity of the incident beam ($I_0$), the extinction coefficient ($\varepsilon_\lambda$), the protein concentration ($c$), and the path length ($d$). The extinction coefficients were calculated using the software ProtParam (www.ExPASy.org).

\[ E_\lambda = -\log \left( \frac{I_t}{I_0} \right) = \varepsilon_\lambda \cdot c \cdot d \]  

Equation 1.4-1

1.4.3.2 Discontinuous SDS-PAGE

SDS polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse protein expression and quality (Laemmli, 1970). Protein samples were denatured in 8x SDS sample buffer for 5 min at 95 °C and applied next to a molecular standard to an SDS-gel (Table 1.4-1). The electrophoresis was initially performed at 120 V until the samples were concentrated to a thin line in the stacking gel. The electrophoretic separation in the running gel was performed at 160 V. The gels were briefly washed in dH$_2$O and subsequently stained with InstantBlue (Biozol).
Table 1.4-1: SDS gel composition. Specifications according six gels.

<table>
<thead>
<tr>
<th>Solutions</th>
<th>Running gel</th>
<th>Stacking gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acrylamide/bisacrylamide 30 % (w/v)</td>
<td>12 ml</td>
<td>1.5 ml</td>
</tr>
<tr>
<td>4x lower buffer</td>
<td>7.6 ml</td>
<td></td>
</tr>
<tr>
<td>4x upper buffer</td>
<td></td>
<td>2.5 ml</td>
</tr>
<tr>
<td>10% SDS</td>
<td>0.3 ml</td>
<td></td>
</tr>
<tr>
<td>H₂O</td>
<td>10 ml</td>
<td>5.9 ml</td>
</tr>
<tr>
<td>TEMED</td>
<td>40 µl</td>
<td>30 µl</td>
</tr>
<tr>
<td>25% APS</td>
<td>60 µl</td>
<td>30 µl</td>
</tr>
</tbody>
</table>

1.4.3.3 Dynamic light scattering (DLS)

Dynamic light scattering (DLS) is a tool to determine size distribution profiles of particles or proteins in solution, providing information about the monodispersity or polydispersity and thus about complex formation, protein aggregation or degradation. The size distribution profile gives also allusion to the crystallisability, whereas proteins with a monodispersity of 70-80% were shown to have a high probability of producing some kind of crystals (Bergfors, 1999). DLS determines the scattering intensities of particles exposed to a focused laser beam as a function of time. The resulting fluctuation of the intensities, due to the Brownian Motion, depend on the particle size. Increasing particle size leads to decreased Brownian Motion and increased half-width of the peak maxima. The diffusion constant can be calculated by applying an autocorrelation function, leading to the possibility to calculate the hydrodynamic radius of the particles, giving information about the size distribution profile of particles in the solution.

For DLS the protein samples with a concentration of 0.5 – 1 mg/ml (leading to approximately 10% scattering of the laser beam) were filtrated with a 0.2 µm filter to avoid the interference of dust. DLS measurements were carried out at room temperature by using a DynaPro Titan (Wyatt Technologies) and a laser power of at least 1,000,000 counts/sec. Protein solutions with a polydispersity <30% were used for crystallisation.

1.4.3.4 Mass spectroscopy (MS)

Protein samples from SDS-PAGE were analysed in-house by Dr. Manfred Nimtz and Undine Felgenträger using MALDI-TOF-MS.
1.4.3.5 Isothermal titration calorimetry (ITC)

Isothermal titration calorimetry (ITC) is a technique to study biomolecular interactions by the direct measurement of protein binding energetics (Leavitt and Freire, 2001). By recording the heat of reaction (binding enthalpy, $\Delta H$), directly proportional to complex formation, characteristic parameters can be calculated. Numerical approximation to a curve of cumulated $\Delta H$ ($\Delta H_{\text{cum}}$) versus the total ligand concentration ($B_t$) according to Equation 1.4-2 facilitates the determination of the enthalpy of formation ($\Delta H_f$) and the association constant ($K_a$). The dissociation constant ($K_d$) corresponds to the reciprocal association constant according to Equation 1.4-3. The Gibbs energy changes ($\Delta G$) and entropy changes ($\Delta S$) are calculated according to the Van’t Hoff and Gibbs-Helmholtz equation (Equation 1.4-4 and Equation 1.4-5).

$$\Delta H_{\text{cum}} = \frac{V \cdot \Delta H_f}{2 \cdot K_a} \cdot [K_a \cdot (C + Bt) + 1 - (K_a^2 \cdot [C - Bt]^2 + 2 \cdot K_a \cdot [C + Bt] + 1)^{0.5}]$$  \hspace{1cm} \text{Equation 1.4-2}

$$K_d = \frac{1}{K_a}$$ \hspace{1cm} \text{Equation 1.4-3}

$$\Delta G = -R \cdot T \cdot \ln(K_a)$$ \hspace{1cm} \text{Equation 1.4-4}

$$\Delta G = \Delta H - T \cdot \Delta S$$ \hspace{1cm} \text{Equation 1.4-5}

To determine these binding parameters by ITC, a MCS-ITC calorimeter (Microcal, Northhampton, MA, USA) was used. Peptides and IMD were dialyzed against 20 mM HEPES-NaOH, 100 mM NaCl, 0.1 mM EDTA and 0.8 mM tris(2-carboxyethyl)phosphine (TCEP), pH 7.5. During titration, 2 mM peptide was injected into 0.1 mM of His-tagged IMD protomers at 24°C in 10 μl steps up to a threefold molar excess. The mixing heat of the peptide, measured by injecting peptide into buffer, was subtracted from the signals obtained from the binding reactions. ITC data of three independant measurements were analysed using Origin V7.0 with Microcal ITC add-on. Peptides $^{452}\text{GTVQNPYADVKT}^{463}$, derived from EHEC Tir, $^{448}\text{GTVPNPYADVKT}^{459}$, derived from EPEC Tir (SwissProt B7UM99), and a sequence of the human insulin receptor $^{995}\text{YSNPYEYLSADY}^{1005}$ (SwissProt P06213) were synthesized in-house with a carboxy-terminal amide and HPLC-purified by Dr. Werner Tegge.
1.4.4 Crystallisation by vapour diffusion

The vapour diffusion technique is the most commonly used method for crystallisation. This method based on the presence of a reservoir with precipitant containing mother liquor, absorbing water from a crystallisation drop in a closed system. Water vapour diffuses from the crystallisation drop, a mixture of protein solution and mother liquor, into the higher concentrated reservoir, until equilibrium is reached. The diffusion leads to a gradual supersaturation in the crystallisation drop and thus facilitates entropic driven crystallisation (Rupp, 2010).

Prior crystallisation, the concentrated an purified protein of the IRSp53-IMD:Tir complex was centrifuged (20 min., 18,000g, 4°C) to separate possible precipitates from the protein solution. Crystallisation was performed at 19°C using the vapour phase diffusion method, either on a sitting drop for initial screening or on a hanging drop for optimisation.

1.4.4.1 Sitting-drop vapour diffusion

The sitting-drop vapour phase diffusion method was used for initial screening in 96-well format, facilitating the use of robotic systems to mix small amounts of pure protein with screen conditions (section 1.3.7).

200 nl of the purified IMD-IRSp53:Tir protein was mixed with an equal amount of the screening conditions in a droplet by using a pipetting robot (Mosquito robot, TTP or HoneyBee 961, DIGILAB Genomic Solutions Inc.). The crystallisation was carried out in 3 x 96 well plates (Greiner), whereas two wells were loaded with different protein concentrations (11 mg/ml and 5.5 mg/ml) and the third was loaded with a buffer/precipitant control, to check for salt crystal formation. The 96-well plates were sealed using a Manco™ Crystal Clear tape (Jena Bioscience) and incubated at 19°C.

1.4.4.2 Hanging-drop vapour diffusion

The hanging drop vapour phase diffusion method was used for optimisation in a 24-well format, by varying the pH, the protein, or precipitant concentration. Protein solution was mixed on a cover slip with an equal amount of mother liquor and placed on a well with a greased rim upside down over 500 µl of mother liquor.
The IMD-IRSp53:Tir complex was crystallised by mixing 2-3 µl of concentrated protein (11mg/ml) and 2-3 µl of mother liquor (16% (w/v) PEG 3350, 0.3 M (NH₄)₂SO₄), incubated at 19°C.

1.4.4.3 Microseeding

The Seeding technique was used to prevent microcrystal showers and to optimise crystal growth and quality. Microcrystal showers are caused or at least favoured by a crystallisation setup in a region of high supersaturation and spontaneous nucleation or by an experiment that proceeds far into the nucleation zone. This kind of setup often results in the formation of many microcrystals, consuming the protein in the crystallisation drop and preventing the formation of only a few big crystals. Lowering the concentration may prevent these microcrystals, but increase the time for nucleation. Thus it is sometimes worthwhile to identify the amount of seeds and the crystallisation conditions, often slightly reduced conditions, to optimise crystal quality by microseeding. The strategy of microseeding is to place microseeds, e.g. from a crashed protein crystal, into a low supersaturated protein-precipitant solution, where spontaneous nucleation is improbable but nucleation growth of the seeds will occur, leading in theory to only a few big crystals.

5-10 IMD-IRSp53:Tir crystals, resulting from the hanging-drop method, were crushed in 100 µl of mother liquor (16% (w/v) PEG 3350, 0.3 M (NH₄)₂SO₄) and were diluted ten times 1:10. IMD-IRSp53:Tir solution was mixed with an equal amount of each dilution and incubated according to the hanging-drop vapour diffusion method.

1.4.5 Data collection, structure determination and refinement

To prevent or minimise crystal damage by radiation and thermal vibrations, the crystals were transferred into reservoir solution with additional 20% (v/v) PEG 400 before flash freezing in liquid nitrogen and data collection at 100 K. The data sets were collected at beamline 14.2 of the synchrotron storage ring BESSY II in Berlin. Images were indexed and processed with the XDS program package (Kabsch, 1993) and the structure was solved by molecular replacement using CCP4 MOLREP (Vagin and Teplyakov, 1997) with PDB entry 1Y2O as the search model. REFMAC5 (Murshudov et al., 1997) was used for refinement. TLS (Translation/libration/screw) (Painter and Merritt, 2006) parameters were included to allow domain movements to be refined and Coot (Emsley and Cowtan, 2004) was used for manual model building. The validation of the structure was performed using MolProbity (Chen et al., 2010; Davis et al., 2007).
1.4.6 Bioinformatics

1.4.6.1 Sequence analysis

Protein sequences were aligned with EMBOSS emma/ClustalW2 (Larkin et al., 2007) and displayed with ESPript 2.2 (Gouet et al., 2003). Disordered protein regions were predicted by DISOPRED2 (Ward et al., 2004). The proportion of human protein sequences in SwissProt containing the NPY motif was determined with Scansite (Obenauer et al., 2003).

1.4.6.2 Structural analysis

Schematic diagrams of atomic interactions were calculated with LIGPLOT (Wallace et al., 1995). Buried surface areas were calculated with PISA (Krissinel and Henrick, 2007). Root mean square deviations (r.m.s.d.) between common Cα-positions were calculated with ProFit using the McLachlan algorithm (McLachlan, 1982).

1.4.6.3 Figure preparation

Figures were prepared with PyMOL (http://pymol.org), LIGPLOT (Wallace et al., 1995) and CorelDRAW X (Corel, Ottawa, Canada).
1.5 Results
1.5.1 Production and purification of IMD-IRSp53

The IMD of IRSp53 was expressed and purified according to section 1.4.1, using immobilised metal ion affinity chromatography (IMAC) for initial purification. Elution of the His$_7$-tag bound IMD by a gradient of imidazole, leads already to a high purity of the 31.7 kDa protein due to a late elution of the IMD at approximately 450 mM imidazole (Figure 1.5-1). The yield of IMAC purified IMD was approximately 40 to 80 mg per litre culture.

![Figure 1.5-1: IMAC purification of IMD-IRSp53. Bacterial lysate containing the IMD-IRSp53 was loaded on a 5 ml HisTrap HP column and eluted with a gradient of 10 to 500 mM imidazole (A). The grey and green line depicts the fractions analysed by SDS-PAGE and Coomassie Blue staining (B). The fractions marked green were pooled for further purification. M: protein standard.](image)

The His$_7$-tag was truncated from IMD using His$_6$-tagged TEV protease. The truncation avoids steric hindrance of crystallisation by the tag and minimises disordered regions, reducing unfavourable loss of entropy of the IMD during crystallisation. Nickel rechromatography was used to separate the His$_7$-tag and protease from the IMD after cleavage, and gel permeation chromatography was used for buffer exchange and final purification. Due to the high purity of the protein after the first purification step, the gel filtration column was loaded with the maximum of the recommended sample volume of 3% column volume, resulting in an expanded protein peak, followed by a second peak which can be attribute to nucleic acids (Figure 1.5-2A). SDS-PAGE was used to verify the purity and removal of the ~3 kDa His$_7$-tag (Figure 1.5-2B). Prior crystallisation trials, the pooled protein fractions of the gel filtration were analysed by MALDI-TOF-MS (section 1.4.3.4) and by DLS (section 1.4.3.3), to ensure the identity of the 28.8 kDa IMD and its monodispersity, respectively.
Figure 1.5-2: Gel filtration. The IMD elutes at about 62 ml from a Superdex 75 HR 16/60 column (A). The peak fraction (50-70 ml) was analysed by SDS-PAGE and Coomassie Blue staining (B), with His$_7$-tagged IMD protein of the IMAC as a control to verify the successful truncation of the approximately 3 kDa His$_7$-tag. M: protein standard.

1.5.2 Production and purification of IMD-IRSp53 mutants

The IMD-IRSp53 mutants used in this study (Table 1.3-3) were expressed and purified according to section 1.4.1 with yields and purification profiles during the IMAC and gel filtration similar to the wild type IMD (section 1.5.1). SDS-PAGE of purified IMD mutants is shown in Figure 1.5-3.

Figure 1.5-3: SDS-PAGE of the gel filtration purified His$_7$-tagged IMD mutants. The different mutants are indicated. M: protein standard.
1.5.3 Crystallisation of the IMD-IRSp53:Tir complex

Crystallisation trials were performed with either unbound IMD or IMD in complex with an EHEC or EPEC Tir derived synthetic dodecapeptide. The complexes were formed according to section 1.4.2. Crystallisation conditions were initially screened in 96-well format by using the sitting-drop vapour phase diffusion method (section 1.4.4.1). Using the JCSG screens, multiple crystallisation conditions for IMD, working as a control, and the IMD in complex with either an EHEC or EPEC peptide were identified (Table 1.5-1). However, only the IMD in complex with the EHEC derived Tir dodecapeptide could be optimised to well diffracting crystals. Even though, crystals appear already after one day using the optimised condition of 0.3 M ammonium sulphate and 16% PEG 3350 as mother liquor, and microseeding to optimise crystal growth, well diffracting crystals needed to grow up to one month at 19°C to reach a proper size (Figure 1.5-4) and quality for diffraction up to 2.1 Å (measured at the beamline 14.2 of the synchrotron storage ring BESSY II in Berlin).

Figure 1.5-4: Optimisation of the crystallisation condition of IMD in complex with the EHEC derived Tir dodecapeptide. Initial optimisation of the crystallisation conditions by the hanging-drop vapour diffusion technique leads to conjoined needles (A). Further optimisation by using the microseeding technique leads to single occurring needles (B). Well diffracting crystal need to grow up to one month at the optimised condition and a temperature of 19°C.

1.5.4 Crystal structure of the IMD-IRSp53:Tir complex

The crystal structure of the complex of the IMD of IRSp53 with a dodecapeptide derived from EHEC Tir (GTVQNPYADVKT, residues 452-463) was solved by molecular replacement and refined at a resolution of 2.1 Å. The crystallographic data are summarized in Table 1.5-2. IMD forms a 170 Å long, cigar-shaped homodimer formed by a bundle of parallel coiled coils of kinked α-helices (Figure 1.5-5). The two protomers of the dimer measure about 120 Å in length and 30 Å in diameter and bury a surface area of 2,600 Å². The IMD structure is in accord with previously published IMD structures (Lee et al., 2007; Millard et al., 2005;
Suetsugu et al., 2006), with a calculated r.m.s. deviation of 0.95 Å to Protein Data Bank (PDB) entry 1Y2O (218 common Cα-positions aligned).

Table 1.5-1: Crystallisation conditions for IMD, IMD in complex with an EHEC and EPEC Tir derived peptide. Optimised crystallisation condition used for IMD-Tir\textsubscript{EHEC} crystallisation is highlighted in blue.

<table>
<thead>
<tr>
<th>Crystallisation conditions</th>
<th>Crystal form</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMD</td>
<td>needles</td>
</tr>
<tr>
<td>20% (w/v) PEG 3350, 0.2 M lithium nitrate</td>
<td>spheres</td>
</tr>
<tr>
<td>20% (w/v) PEG 3350, 0.2 M calcium acetate</td>
<td>spheres</td>
</tr>
<tr>
<td>IMD+Tir\textsubscript{EHEC}</td>
<td>spheres</td>
</tr>
<tr>
<td>0.1 M hepes pH 7.5, 40% (v/v) PEG 400, 0.2 M calcium acetate</td>
<td>spheres</td>
</tr>
<tr>
<td>0.1 M sodium cacodylate pH 6.5, 40% (v/v) PEG 600, 0.2 M calcium acetate</td>
<td>spheres</td>
</tr>
<tr>
<td>0.1 M MES, 1.6 M ammonium sulphate, final pH 6.0</td>
<td>spheres</td>
</tr>
<tr>
<td>0.1 M citric acid, 30% (w/v) PEG 6000, final pH 4.0</td>
<td>spheres</td>
</tr>
<tr>
<td>0.2 M ammonium sulphate, 20% (w/v) PEG 3350</td>
<td>needles</td>
</tr>
<tr>
<td>IMD+Tir\textsubscript{EPEC}</td>
<td>spheres</td>
</tr>
<tr>
<td>0.1 M sodium cacodylate pH 6.5, 40% (v/v) MPD, 5% (w/v) PEG 8000</td>
<td>spheres</td>
</tr>
<tr>
<td>0.1 M sodium phosphate-citrate pH 4.2, 40% (w/v) Ethanol, 5% (w/v) PEG 1000</td>
<td>needles</td>
</tr>
<tr>
<td>0.2 M calcium acetate, 0.1 M imidazole pH 8.0, 20% PEG 1000</td>
<td>spheres</td>
</tr>
<tr>
<td>0.2 M lithium sulphate, 0.1 M sodium acetate pH 4.5, 50% PEG 400</td>
<td>spheres</td>
</tr>
</tbody>
</table>

1.5.4.1 Both Tir molecules bound by the IMD point into the same direction

The electron density map clearly reveals the position of the bound Tir-derived peptide (Figure 1.5-6A). The IMD homodimer binds two Tir-derived peptides, one at the centre of each protomer (Figure 1.5-5). The peptides point into the same direction, almost parallel to the two-fold rotation axis of the dimer and perpendicular to the IMD cigar shape. The carboxy-termini of the peptides point to the site of the IMD that is assumed to face the cytoplasm in the membrane bound state. Peptides are located far apart with 91 Å and 61 Å distances between their carboxy- and amino-termini, respectively. Each peptide buries 564 Å\textsuperscript{2} of the IMD surface, which is close to the average of buried surface areas in peptide-protein interactions (London et al., 2010). The whole length of the Tir-derived peptide is in close proximity to the domain, with the two Tir valine side-chains (V454, V461) oriented inwards, thus maximising solvent entropy upon binding. All amino acid residues of the peptide and the domain are well defined in the electron density map, except for the loop between helices H2 and H3 and residues 234-250 encoding helix H4.
Table 1.5-2: Summary of crystallographic analysis.

<table>
<thead>
<tr>
<th>Data collection</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C222₁</td>
<td></td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a, b, c (Å)</td>
<td>88.35, 187.36, 36.81</td>
</tr>
<tr>
<td>X-ray source</td>
<td>Beamline 14.2, BESSY II, Berlin</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.9184</td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>9.44-2.11</td>
<td></td>
</tr>
<tr>
<td>Last shell (Å)</td>
<td>2.16-2.11</td>
<td></td>
</tr>
<tr>
<td>(R_{\text{merge}}) (%)</td>
<td>7.7 (53.2)</td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>231684 (9729)</td>
<td></td>
</tr>
<tr>
<td>Unique reflections</td>
<td>18190 (1343)</td>
<td></td>
</tr>
<tr>
<td>Mean (I)/s.d. (I)</td>
<td>20.13 (3.13)</td>
<td></td>
</tr>
<tr>
<td>Completeness</td>
<td>99.8 (100)</td>
<td></td>
</tr>
<tr>
<td>Multiplicity</td>
<td>12.7 (7.2)</td>
<td></td>
</tr>
</tbody>
</table>

| Structure refinement |          |          |
| Resolution range (Å) | 9.44-2.11 (2.16-2.11) |          |
| \(R_{\text{work}}\) (%) | 23.3 |          |
| \(R_{\text{free}}\) (%) | 27.6 |          |

| Total number of |          |          |
| Non-hydrogen atoms | 1991 |          |
| Protein atoms | 1837 |          |
| Peptide atoms | 103 |          |
| Water molecules | 46 |          |
| Sulfate molecules | 1 |          |

| r.m.s.d. |          |          |
| Bond length (Å) | 0.022 |          |
| Bond angle (deg) | 1.90 |          |
| Main-chain B-factors (Å³) | 39.4 |          |
| Side-chain B-factors (Å³) | 43.5 |          |
| Wilson B-factor (Å²) | 41.8 |          |
| Average B-factor protein atoms (Å²) | 41.5 |          |
| Average B-factor peptide atoms (Å²) | 27.0 |          |
| Average B-factor solvent atoms (Å²) | 40.8 |          |

| Ramachandran statistics |          |          |
| Most favoured regions (%) | 99.1 |          |
| Allowed regions (%) | 0.9 |          |
| Disallowed regions (%) | 0 |          |

*Values in parentheses are for the highest resolution shell.*
Figure 1.5-5: Complex of IMD_{IRSp53} with the EHEC Tir-derived peptide. IMD_{IRSp53} binds a 12-residue peptide derived from the EHEC virulence factor Tir, incorporating the NPY motif. Ribbon diagrams of the two subunits of the IMD homodimer are drawn in cyan and wheat. Two identical peptide molecules are coloured blue and yellow, respectively. Peptide amino-termini are labelled with an encircled N.
1.5.4.2 A novel peptide binding site of IRSp53 recognises the NPY motif

In the structure of the IMD:Tir complex, the NPY motif of Tir interacts with a large hydrophobic binding pocket in the central region of the IMD<sub>IRSp53</sub> protomer (Figure 1.5-6A). This pocket is also present in the free IMD structure reported by Suetsugu <em>et al.</em> (PDB entry 1WDZ). On the contrary, in the structure reported by Millard <em>et al.</em> (PDB entry 1Y2O), the binding pocket of one of the two protomers is partly occluded by the side-chain of Lys108 (Figure 1.5-6C,D). The reason for this difference remains unclear.

All three side-chains of the NPY motif (456-458) are specifically bound by IMD<sub>IRSp53</sub>. The conformation of the peptide backbone is defined by two overlapping type I β-turns (456-459, 458-461), which are stabilised by three intramolecular hydrogen bonds and an ordered water molecule (Figure 1.5-6B, Figure 1.5-7). The pyrrolidine ring of Pro457 of the NPY motif further rigidifies the backbone and is accommodated in a shallow hydrophobic groove on the IMD surface (Figure 1.5-6A,D).

The side-chain of Tyr458 of the NPY motif is accommodated in a deep hydrophobic pocket and is sandwiched between the hydrophobic portions of the IMD Arg193 and Lys108 side-chains (Figure 1.5-6A,B). Two ordered water molecules are bound in the interior of the pocket and link the Tyr458 hydroxyl group to helices H1, H2, and H3 of IMD (Figure 1.5-7, Figure 1.5-6B). Substitution of Tyr458 by phenylalanine would compromise these water-mediated interactions, which rationalises the reduced pedestal formation of the EHEC-Tir Y458F mutant (Brady <em>et al.</em>, 2007).

Recognition of Asn456 of the NPY motif is based on the unique hydrogen bonding potential of its side-chain (Figure 1.5-7). The side-chain's carbonyl oxygen is hydrogen bonded to a nitrogen of the guanidinium group of IMD Arg193 and accepts an additional, intrapeptide hydrogen bond from the backbone nitrogen of Tir Tyr458, an i to i+2 interaction that often stabilises Asn-Pro-containing β-turns (Wilson and Finlay, 1997). The side-chain amide nitrogen of Asn456 is hydrogen bonded to IMD Glu189. Hydrophobic contacts of the Asn456 side-chain to Tyr115 further strengthen the interaction. Clearly, asparagine is essential at this position, as no other residue could fully satisfy all of the observed bonding constraints.

Arg193 of IMD<sub>IRSp53</sub>, part of the signature sequence <sup>183</sup>EE<sup>193</sup>RR<sup>193</sup> of the IMD family (Yamagishi <em>et al.</em>, 2004), appears to be the most critical residue for the recognition of the NPY motif. In the free domain, its side-chain is already oriented in the correct conformation for binding to Tir Asn456 and Tyr458 by a salt bridge between its guanidinium group and Asp112. This is in contrast to Lys108, which has a variable side-chain conformation in the absence of the peptide.
Figure 1.5-6: Atomic details of the IMD_{IRSp53}:Tir complex. (A) Profile of the binding pocket. The Tir-derived peptide \textsuperscript{452}GTVQNPYADVKT\textsuperscript{463} is drawn in yellow with its electron density map in blue. Dashed lines indicate hydrogen bonds. Tyr458 of the NPY motif is centred by two water molecules (red) in the interior of the binding pocket. (B) Stereo view of the interactions of the Tir-derived peptide comprising the NPY motif (456-458). Residues involved in the complex interface and in β-turn
formation are labelled. The green line represents the backbone of the peptide. Hydrogen bonds involved in β-turn formation are shown as red dashed lines. (C and D) Superposition of the free IMD structure of Millard et al. (green, PDB 1Y2O) and the peptide-bound IMD (wheat). The side-chain of Lys108 blocks the binding pocket in the free structure. The conformation of the side-chain of Lys108 in the free IMD structure of Suetsugu et al. is similar to the complex (1WDZ, not shown).

Figure 1.5-7: Schematic diagram of atomic interactions. EHEC Tir and the IMD’s binding pocket are drawn with black and red labels and bonds, respectively. Cyan spheres represent water and dashed lines hydrogen bonds. Hydrophobic interactions are indicated by spoked arcs and atoms with spokes. An ordered water molecule and three intra-molecular hydrogen bonds (red) stabilise the peptide’s two overlapping β-turns. Oxygen atoms are coloured red, nitrogen atoms blue and carbon atoms black.
In the structure of the free IMD homodimer of Millard et al. (PDB 1Y2O), Arg191 binds to Asp232 of the loop linking the opposing protomer’s helices H3 and H4 (Figure 1.5-8A). In complex to Tir, the Arg191 guanidinium group contacts the oxygens of the peptide’s terminal carboxylic amide group and of Ala229 of the opposing protomer (Figure 1.5-8B). This induced an altered conformation of the loop linking helix H3 and the last helix, H4, which was in turn not observed in the structure. These differences may be causative of the entirely different crystal packing of the complex (Figure 1.5-8C,D).

**Figure 1.5-8: Differences in crystal packing of free and peptide loaded IMD IRSp53 facilitates a conformational change of H4.** (A) In the free structure (PDB 1Y2O), Arg191 of one monomer (green) binds to the loop linking helix H3 and helix H4 of the other monomer (orange) and thereby stabilises helix H4 adjacent to the dimer. In the structure of the complex (B), the peptide’s terminal amide group (yellow) binds to the guanidinium group of Arg191 of one IMD monomer (wheat), leading to a conformational change of helix H4 and the preceding loop of the other monomer (blue). (C) Crystal packing according to Millard et al. of the free IMD (PDB 1Y2O). IMDs are densely packed. The two protomers of each dimer are in cyan and wheat, respectively. (D) Crystal packing of the peptide loaded IMD displays a more loose array, explaining the undetectability of the loop between H2 and H3 and of entire H4. Tir peptides are shown in red. Space groups are indicated.
1.5.4.3 The Tir binding site is conserved

The paralogues IRSp53 and IRTKS interact with both EHEC effectors Tir and EspF via their IMD and SH3 domains, respectively. In contrast, the third family member in mammals, FLJ22582, only binds EspF (Weiss et al., 2009). The sequence identity for the IMDs of IRSp53 and IRTKS is 47% and all residues contacting the NPY motif are conserved (Figure 1.5-9).

Figure 1.5-9: IMD sequence conservation. Sequence alignment of the known human IMD domains of IRSp53, IRTKS, FLJ22582, MIM and ABB-1 and IRSp53 orthologs from Xenopus tropicalis (claw frog), Danio rerio (zebrafish) and Drosophila melanogaster (fruit fly) (GenBank NP_059344, NP_061330, NP_079321, NP_055566, NP_612392, NP_001121436, NP_001035335, NP_729679). Blue boxes indicate IRSp53 residues in contact to the Tir NPY motif. Red and yellow boxes highlight identical and similar residues, respectively. Positions that were mutated in this study are marked by asterisks.
The IMD of FLJ22582, however, has only a sequence identity of 28% to IRSp53, and the residues forming the binding pocket differ markedly. The most prominent difference is FLJ22582 Asn108, which replaces Lys108 of IRSp53 (Figure 1.5-10). Lys108 is one of most important residues in the binding pocket, stabilising Tyr458 of the NPY motif (Figure 1.5-6B). The lack of Tir binding in FLJ22582 may thus be explained by its absence. The Tir binding site is highly conserved within the group of IRSp53 and IRTKS orthologues of evolutionary distant vertebrates including claw frog and zebrafish, but not in invertebrate orthologues. The site is also not conserved in the human IMD proteins MIM and ABBA-1 (Figure 1.5-9).

Figure 1.5-10: Mapped IMD sequence conservation. Sequence conservation between the IRSp53 residues, constituting the hydrophobic binding pocket, and IRTKS (left) or FLJ22582 (right). Deep green indicates identity, while purple represents differing residues.

### 1.5.4.4 ITC analysis to study the effect of mutations on the IMD-IRSp53:Tir complex formation

The complex formation between the IMD<sub>IRSp53</sub> and the Tir-derived peptide was studied by isothermal titration calorimetry (ITC). The stoichiometry of two peptide molecules per IMD homodimer was confirmed, and binding was apparently non-cooperative. A dissociation constant of 59 ± 10 µM was determined (Figure 1.5-11), which indicates a weak interaction, comparable to SH3-domain interactions with proline-rich motifs (Ball et al., 2005; Kuriyan and Cowburn, 1997; Mayer and Saksela, 2004). A corresponding 12-mer peptide was derived from EPEC-Tir (SEVVNPYAEVGG) and a similar dissociation constant of 37 ± 14 µM was obtained.
Binding of the IMD$_{IRSp53}$ to additional peptides with motifs related to NPY was studied by ITC, including the $^{456}{\text{NPY}}^{458} \rightarrow \text{NPF}$ mutant, described in section 1.3.6, and a peptide derived from the insulin receptor comprising the related NPXY motif. No binding was detectable by ITC for these NPF and NPXY motif containing peptides (data not shown), demonstrating preference of IMD for the NPY motif. To interpret these results, it should be noted that the protein concentration used for ITC (100 µM) was close to the dissociation constant of the wild type domain and the Tir-derived peptide (60 µM), and it would not have been possible to detect binding at lower affinities.

**Figure 1.5-11**: Isothermal titration calorimetry of the binding reaction of IMD$_{IRSp53}$ and four IMD point mutants with peptides derived from EHEC and EPEC Tir, comprising the NPY motif. The upper panel shows considerable heat release upon titration of the wild type IMD$_{IRSp53}$ (black) with the EHEC Tir-derived peptide, in contrast to the R193S mutant (red). The heat data, corrected for dilution, is plotted against the molar ratio of peptide molecules and IMD protomers in the lower panel. Sigmoidal curves typical of exothermic binding reactions were obtained for the wild type IMD$_{IRSp53}$ and the mutant F196A, but not for the R193S, K108A and L28E mutants.
Close inspection of the binding site allows the prediction of IMD point mutants, impaired in Tir binding without disturbing the overall IMD structure (Figure 1.5-9). These point mutants were generated in-house by the group of Theresia Stradal. Substitution of Leu28 by glutamic acid (L28E) was predicted to fill the hydrophobic pocket. The residue Lys108 forms a hydrogen bond and hydrophobic contacts with the peptide. It is changed to an asparagine in FLJ22582, the IRSp53 paralogue that does not recognise Tir. The conformation of the Lys108 side-chain is stabilised by Phe196. Arg193 may be critically important for recognizing Tir, since it is the residue with most interactions with the peptide (Figure 1.5-7) and since the salt bridge to Asp112 appears to stabilise the binding pocket. No binding of the Tir-derived peptide was detected by ITC when Leu28, Lys108, or Arg193 were substituted in IMD$_{IRSp53}$ (L28E, K108A, R193S in Figure 1.5-11). Binding of the Phe196 point mutant F196A to the peptide was weaker than for the wild type domain but still clearly detectable.
1.6 Discussion

1.6.1 The NPY motif of Tir: a ‘hot spot’ for IMD recognition

The crystal structure presented here reveals the molecular details of the interaction between EHEC Tir and the IMD of human IRSp53. It elucidates why the highly conserved NPY motif of EHEC Tir is critical for interaction with the IMD of IRSp53 and IRTKS, whereas neighbouring residues can be substituted with alanine without any effect (Brady et al., 2007; Weiss et al., 2009). The specificity of recognition of Tir by the IMD is determined by the characteristic features of the NPY motif’s three residues, 456-458. The Asn456 side chain is bound by specific hydrogen bonds. The Tyr458 phenyl ring projects into a hydrophobic pocket, forming extensive hydrophobic interactions to the IMD, while the phenolic hydroxyl group is bound by two water molecules at the bottom of the pocket. Correspondingly, substitution of Tyr458 with alanine virtually abolished pedestal formation by EHEC bacteria, while substitution with phenylalanine led to a milder defect (Brady et al., 2007). The pyrrolidine ring of Pro457 of the NPY motif fits into a shallow hydrophobic cavity on the IMD surface. It rigidifies the backbone and supports the formation of two overlapping type I β-turns in the Tir polypeptide chain.

The structure described here adds the IMD to the list of known peptide binding domains. The binding energy of protein-protein and peptide-protein interactions generally depends on only a small number of ‘hot spot’ residues (London et al., 2010), obviously represented here by the NPY motif, including the tightly bound Tyr458. Interestingly, tyrosine ranks on third position of the most frequent amino acid residues identified in peptide-protein complexes (London et al., 2010). Specific recognition of phosphorylated Tyr474 by Nck proteins is essential for pedestal formation by EPEC, as mentioned above. Other tyrosines of EPEC Tir at positions 483 and 511 also become phosphorylated and are essential for recruiting SHIP2, a host inositol phosphatase controlling pedestal formation by generating a phosphatidylinositol 3,4-bisphosphate-enriched lipid platform (Smith et al., 2010). It appears that tyrosine-mediated interactions of Tir with specific peptide recognition domains of the host cell are critical for pedestal formation of both EHEC and EPEC.

1.6.2 The conserved Tir binding pocket of IMD: the essential residues and their relevance for NPY recognition

The crucial role of certain residues within the IMD NPY binding site revealed by the structure was confirmed by site-directed mutagenesis and isothermal titration calorimetry (section 1.5.4.4). Further binding studies as pull-down assays from lysates and interaction assays in living cells, using the Protein Interaction Platform (PIP) assay in yeast (Schmitz et al., 2009),
were carried out in cooperation with the group of Prof. Dr. Theresia Stradal (Institute for Molecular Cell Biology, University of Münster, Germany), Dr. John M. Leong (Department of Molecular Genetics and Microbiology, University of Massachusetts Medical School, Worcester, MA, USA) and Cammie F. Lesser (Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA, USA), respectively.

### Table 1.6-1: IMD mutants used to study the interaction with Tir.

ITC and pull-down assays were carried out using the IMD of IRSp53, whereas the IMD of IRTKS was used in the PIP assay. The sequence of IMD$_{IRTKS}$ aligns the sequence of IMD$_{IRSp53}$ with a shift of -1. X indicates the use of the domain in the corresponding assay.

<table>
<thead>
<tr>
<th>IMD$_{IRSp53}$</th>
<th>IMD$_{IRTKS}$</th>
<th>ITC</th>
<th>Pull-down assay</th>
<th>PIP assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wt</td>
<td>Wt</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>L28E</td>
<td>L27E</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>K108A</td>
<td>K107A</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>R193S</td>
<td>R192S</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>F196A</td>
<td>F195A</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>K108A/R193S</td>
<td>K107A/R192S</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td>K107A/F195A</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>R193S/F196A</td>
<td>R192S/F195A</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

The pull-down assays were used to probe the ability of the described IMD mutants (section 1.5.4.4 and Table 1.6-1) to interact with the Tir-derived peptides or, in contrast to the isothermal titration calorimetry, with the entire cytoplasmic carboxy-terminal tail of EPEC- and EHEC-Tir (TirC). The pull-down assays were performed from lysates of COS7 cells, ectopically expressing the IMD variants as amino-terminal GFP fusions, using biotinylated synthetic Tir-derived peptides immobilised on streptavidin-agarose beads as bait. Under these conditions, wild type EPEC and EHEC-derived peptides showed significant binding to the IMD of IRSp53, and could effectively deplete GFP-tagged IMDs from the lysate and enrich in the precipitate (Figure 1.6-1). In contrast, neither the IMD mutants K108A, F196A (Figure 1.6-1), nor L28E, R193S, or the IMD double mutants K108A/R193S and K108A/F196A (not shown) could be detected in the precipitates. Vice versa experiments were performed by using recombinant His-tagged wild type and mutant IMDs as baits in pull-down assays from lysates of COS7 cells, ectopically expressing the entire TirC. The recombinant wild type IMD effectively bound to TirC from both EPEC and EHEC. In accordance with the data obtained by ITC, weak but reproducible binding of TirC was also...
observed for the F196A mutant of IMD_{IRSp53}, while all other IMD mutants were abrogated for binding (Figure 1.5-11 and Figure 1.6-1B).

**Figure 1.6-1**: Pull-down assays of Tir with wild type and mutant IMD of IRSp53. (A) Pull-down assays of Tir-derived peptides from lysates of mammalian cells ectopically expressing GFP-tagged IMD variants of human IRSp53 as indicated on the right. Shown are cell lysates before (load) and after (supernatant) pull-downs and precipitates (pull) of three different peptides as indicated above. EHEC Tir-derived peptides that bind to Nck instead of IRSp53 (Y474) as well as streptavidin beads were employed as bait-control and GFP alone in the cell lysate was the prey-control. (B) Pull-down assays of recombinant His-tagged IMD from lysates of COS7 cells ectopically expressing the cytoplasmic carboxy-termini of EPEC or EHEC Tir as indicated on the right. Shown is the cell lysate before the pull-down (load) and the precipitates (His-IMD pull-down) of wild type IMD (wt) and six variants as indicated above. Note the weak binding to mutant F196A in accordance with the ITC measurements.

PIP-assays were used to investigate whether the mutations analysed above within the IMD of IRSp53 would also affect the interaction of Tir with IRTKS. Given the strong conservation of the Tir binding site, IRSp53 and IRTKS most likely bind to the Tir receptor in the same fashion. EHEC TirC and a $^{456}$NPY$^{458}$→AAA substitution mutant (TirC AAA) (Vingadassalom et al., 2009) were fused to the scaffolding protein µNS that forms large focal inclusions ('platforms') in yeast cells. Co-expression of GFP-tagged IMD of IRTKS with the TirC fusion protein resulted in complex assembly and formation of fluorescent foci in almost all cells (Figure 1.6-2A). The TirC AAA mutant was not able to alter the diffuse intracellular localisation of IMD_{IRTKS}, as expected. Substitution of IRTKS Lys107 or Phe195
(corresponding to IRSp53 Lys108 and Phe196) partially inhibited the interaction with TirC (K107A and F195A, Figure 1.6-2A,B). This was exacerbated by combining both mutations (K107A/F195A). Strong inhibition was also observed for the substitution of Leu27 by glutamic acid (L27E). Finally, substitution of Arg192 in IRTKS completely abolished the Tir:IMD<sub>IRTKS</sub> interaction (R192S), as the double mutations R192S/F195A and K107A/R192S did (Figure 1.6-2A,B).

Figure 1.6-2: Protein Interaction Platform assay of Tir with wild type and mutant IMD of IRTKS <em>in vivo</em>. (A) Yeast cells co-expressing µNS-TirC and wild type or mutant GFP-tagged IMD of IRTKS were visualized by fluorescence microscopy 4-5 hours post-induction of fusion protein expression. Exemplary GFP inclusions are marked by arrows. The percentage of cells containing GFP inclusions was determined visually and is indicated at the bottom right of each image and (B) visualized in the histogram.

The interaction of the IMD of IRSp53 and IRTKS with Tir <em>in vitro</em> and in living cells was thus impaired by substitution of residues constituting the peptide binding site (Figure 1.5-11, Figure 1.6-1 and Figure 1.6-2). The results were conclusive, also concerning residual binding that was always found to be strongest for the F196A or F195A mutants, respectively. The tight interaction of Arg193 with Tir in the crystal structure was reflected by the complete loss of binding upon substitution of the corresponding residues in IRSp53 and IRTKS. Substitution of Lys107 or Lys108 in respectively IRTKS or IRSp53 considerably reduced or abolished binding. The homologous FLJ22582 has an asparagine at this position, which is probably the reason for its lack of binding to Tir (Weiss <em>et al.</em>, 2009).

Full-length IRSp53 was shown to equally well precipitate with both EPEC and EHEC Tir-derived peptides, containing this NPY motif (Weiss <em>et al.</em>, 2009). Moreover, both GFP-tagged IRSp53 and endogenous IRSp53 can be readily detected in both EPEC and EHEC pedestals
(Weiss et al., 2009). However, in contrast to EHEC, that essentially require IRSp53, it is dispensable for EPEC pedestal formation as exemplified by using IRSp53 knockout cells (Weiss et al., 2009). Notwithstanding this, introduction of EspF_U into EPEC drastically increases Nck-independent pedestal formation (Brady et al., 2007). Thus, IRSp53 effectively binds to EPEC Tir but no information is available to date on the state of EPEC Tir Y454 phosphorylation, when bound to IRSp53.

In Summary, our current data confirm a virtually complete conservation of the binding characteristics between IRSp53 and IRTKS to the carboxy-termini of EPEC and EHEC Tir.

1.6.3 The conserved Tir binding pocket of IMD and the preference for NPY motifs

The mode of binding of the NPY motif is strikingly similar to the recognition of NPXY and NPF motifs by phosphotyrosine binding (PTB) and Eps15 homology (EH) domains, respectively. The NPXY and NPF motifs also form stabilised Asn-Pro type I β-turns, their asparagine side-chains are specifically hydrogen bonded and their aromatic side chains enter hydrophobic pockets in the respective domains (de Beer et al., 2000; Eck et al., 1996). Binding of peptides containing an NPXY or NPF motif by IMD was not detected by ITC, indicating preference for NPY over related motifs.

1.6.4 Tir is the first unambiguously IMD binder

The striking specificity of the recognition of the NPY motif by the IMD suggests that the domain has evolved to recognise this motif, a feature that just becomes usurped by pathogenic E. coli. Thus it is tempting to speculate that this binding site may have cellular ligands that harbour NPY motifs, which occur in up to 3% of human proteins. However, up until now, no binding partners for the IMD domains of IRSp53 and IRTKS other than Tir have been pinned down unambiguously because the in vivo relevance of its binding to both Rac1 (Disanza et al., 2006; Krugmann et al., 2001; Miki et al., 2000) and actin (Mattila et al., 2007; Millard et al., 2005) is still the subject of debate.

1.6.5 The affinity of Tir

The almost parallel orientation and the distance of the two Tir-derived peptides in the complex suggest that one IRSp53 homodimer can link two Tir molecules in vivo. An affinity of
60 µM was found for the binding reaction under \textit{in vitro} conditions, which is comparable to the relatively weak interactions between SH3-domains and proline-rich motifs (Ball \textit{et al.}, 2005; Kuriyan and Cowburn, 1997). However, the affinity and specificity of ligand binding by SH3 can be greatly enhanced by additional contacts between the SH3 loop regions and residues of the ligand outside the proline-rich motif (Arold \textit{et al.}, 1998; Mayer and Saksela, 2004). This is greatly illustrated by the SH3 blocker of Kardinal and colleagues, optimizing a high affinity SH3 blocker peptide with a K$_D$ in the nanomolar range that competes for the SH3 of CRKL with Bcr-Abl and hence reduces the proliferation of primary blast cells from chronic myeloid leukemia patients (Kardinal \textit{et al.}, 2000). Thus, regions outside the Tir-derived peptide might likewise contribute to the binding to IRSp53. This notion is supported by the fact that weak binding of the IMD mutant F196A to a longer carboxy-terminal Tir fragment was observed in pull down experiments (Figure 1.6-1B), while binding to short Tir-derived peptides was not detectable (Figure 1.6-1A). Furthermore, Tir forms homodimers and is clustered beneath the attached bacteria upon interaction with intimin (Luo \textit{et al.}, 2000; Thierry \textit{et al.}, 2004). Together with the presence of two binding sites on the IMD homodimer, this is expected to lead to a higher apparent affinity due to the avidity effect and, consequently, efficient recruitment of IRSp53 at physiological concentrations.

1.6.6 Model of the intimin:Tir:IRSp53 complex

The EHEC bacterium utilises a large protein scaffold including bacterial and host proteins that immobilises the bacterium and leads to N-WASP mediated Arp2/3-complex dependant actin polymerisation (Weiss \textit{et al.}, 2009). Figure 1.6-3 shows a model of this scaffold consisting of Tir, intimin and IRSp53. An alternative model with one IRSp53 homodimer linking two distinct Tir homodimers appears also possible. IMDs recognise and even induce convex deformation of lipid membranes (Scita \textit{et al.}, 2008; Suetsugu \textit{et al.}, 2006). Membrane binding by human IRSp53 is short-lived and highly dynamic (Saarikangas \textit{et al.}, 2009) and it is dispensable for pedestal formation (Weiss \textit{et al.}, 2009). The orientation of the Tir-derived peptides on the IMD suggest that the flat IMD surface, which is assumed to contact the plasma membrane, is oriented towards the membrane in the complex, thus supporting the possibility of Tir-bound IRSp53 contacting the plasma membrane.
Figure 1.6-3: Model of the intimin:Tir:IRSp53 complex. A model of the scaffold formed by EHEC Tir, intimin, and human IRSp53 was drawn to scale based on the available structures. Whether the Tir-bound IMD interacts with the plasma membrane directly is currently unknown. The extracellular Tir and intimin domains are represented by the structure of the complex of the homologues EPEC proteins by Luo et al. (PDB entry 1FO2). The model excludes the amino-terminus of Tir, which modulates pedestal length (Campellone et al., 2006). The IRSp53 SH3 domain is represented by the similar PDB entry 2KXC. Regions of unknown structure are represented by ropes with end-to-end distances corresponding to root mean square end-to-end distances of random coils of the respective chain length (Tanford et al., 1966).
1.7 Outlook

1.7.1 NPY motif containing host proteins may represent physiological ligands for IRSp53/IRTKS

The emergency of the critically increasing incidence of antibiotic resistance due to the widespread and, to some extent, careless use of antibiotics requires a more detailed understanding of the molecular basis of host pathogen interactions to generate alternative methods to treat infectious diseases. Moreover, diseases like the HUS, described to be triggered by the use of antimicrobial substances (see section 1.1.4.2), currently lack a well-established method of treatment and rely on the development of such alternative strategies. Structural insights into the key mechanisms of infection of human pathogenic E. coli can pave the way to design inhibitors to disrupt key interactions during infection and enable the treatment of infectious diseases without the use of antibiotics.

Several studies already aim at the development of strategies to neutralise Shiga toxins or to disrupt the mechanism of adhesion of human pathogenic bacteria like blocking the interaction of Tir:intimin of intimin positive EHEC. However, the Tir:IRSp53 interaction currently seems to be no drug target to treat EHEC infections as this interaction seems to trigger only actin polymerisation. Inhibition of actin polymerisation and thus pedestal formation after EHEC infection might, to some extent, probably improve absorption and reduce intestinal tissue damage and the dissemination of EHEC, which is suggested to be enhanced by the pedestals which move along the cell surface, the so-called ‘pathogen surfing’. But, according to the current knowledge, inhibited pedestal formation will probably not affect adhesion or toxin production of EHEC, the central issue of EHEC infections. Nevertheless, unveiling the molecular mechanism of pedestal formation and the investigation of the bacterial benefit of this feature will contribute to the substantial understanding of EHEC infections and hence the development of strategies to treat this disease.

The data of this study characterise the NPY motif of Tir as a ‘hot spot’ for the interaction with the host’s IMD. The domain seems to be evolutionary evolved to recognise this motif, a feature that becomes usurped by pathogenic E. coli, and it is tempting to speculate that this binding site may have cellular ligands that harbour NPY motifs, which occur in up to 3% of human proteins. The huge variety of possible ligands to this binding site of the host’s IMD illustrate again, that even if the IMD:Tir interaction represented a pharmacological relevant drug target, which is currently not the case, the Tir binding pocket of IMD seems to be inappropriate due to the presumably adverse effects by blocking also diverse cellular ligands. But interestingly, among these NPY motif containing proteins, numerous candidates can be found that are related to cellular processes that may involve IRSp53 or IRTKS (see Table 1.7-1). Thus the discovery and characterisation of this new binding pocket may lead to the
identification of new interaction partners, possibly involved in the pathogenesis and cytoskeletal regulation. Furthermore, these studies will also contribute to a better understanding of the cellular role of IRSp53. It will be an exciting task to probe these interactions and analyse the recently described IRSp53 knockout mouse (Sawallisch et al., 2009) for potential phenotypes that may relate to defective IMD:NPY interactions.

Table 1.7-1: Potential NPY-containing interactors of IMD proteins. The SwissProt database comprises 20,329 human proteins, of which 749 contain the NPY motif as detected using Scansite (http://scansite.mit.edu/dbsequence_one.html). Candidates that potentially might interact with the IMD of IRSp53 or IRTKS were selected according to the following criteria: (i) nuclear, intraluminal or extracellular motifs are excluded; (ii) the motif is conserved at least between human and mouse orthologs and (iii) conservation between paralogues is preferred. The selection shown is further confined based on involvement in IRSp53-signalling (e.g. Shank), in cytoskeleton reorganisation (e.g. LIMK), in Rho-GTPase signalling (Intersectin), and for those that were implicated in filopodium (or similar like stereocilia/microvilli, e.g. Myosin X) formation before (de Groot et al., 2011).

<table>
<thead>
<tr>
<th>Short name</th>
<th>Full name</th>
<th>NPY position</th>
<th>NPY location</th>
<th>Protein class</th>
</tr>
</thead>
<tbody>
<tr>
<td>PK3CB</td>
<td>Pi3K, p110b, Phosphotyidylinositol-3 kinase, catalytic subunit</td>
<td>482-84</td>
<td>C2 domain</td>
<td>phospholipid kinase</td>
</tr>
<tr>
<td>ROCK2</td>
<td>Rho-associated protein kinase 2</td>
<td>1195-97</td>
<td>very C-terminus</td>
<td>protein kinases</td>
</tr>
<tr>
<td>LIMK1 &amp; -2</td>
<td>LIM domain kinase 1</td>
<td>512-14</td>
<td>kinase domain</td>
<td>regulators of Rho-GTPase activity</td>
</tr>
<tr>
<td>ITSN1 &amp; -2</td>
<td>Intersectin 1 (Cdc42 GEF)</td>
<td>1618-20</td>
<td>C2 domain</td>
<td></td>
</tr>
<tr>
<td>ARHGAP10</td>
<td>GRAF2 GAP10 associated with FAK (RhoA, Cdc42)</td>
<td>625-27</td>
<td>between Rho-GAP and SH3</td>
<td></td>
</tr>
<tr>
<td>Myo IA</td>
<td>Brush border myosin 1 /BBM-I</td>
<td>49-51</td>
<td>head domain</td>
<td>myosins</td>
</tr>
<tr>
<td>Myo X</td>
<td>Myosin 10 (filopodia)</td>
<td>104-06</td>
<td>head domain</td>
<td></td>
</tr>
<tr>
<td>Myo XV</td>
<td>Myosin 15 (stereocilia)</td>
<td>423-25</td>
<td>head domain NT</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1263-65</td>
<td>head domain CT</td>
<td></td>
</tr>
<tr>
<td>CLAP1 &amp; -2</td>
<td>CLIP115 associated protein1</td>
<td>1265-67</td>
<td>CLIP-binding site</td>
<td>microtubule tip proteins</td>
</tr>
<tr>
<td>SHANK2</td>
<td>SH3 and multiple ankyrin repeat domains protein 2</td>
<td>608-10</td>
<td>central part</td>
<td>partners of IRSp53</td>
</tr>
<tr>
<td>BAI2</td>
<td>Brain-specific inhibitor of angiogenesis 2 (GPCR)</td>
<td>1417-19</td>
<td>cytoplasmic tail</td>
<td></td>
</tr>
</tbody>
</table>
1.7.2 The Tir:IMD interaction is likely to compete with membrane binding

The cigar shaped IMD dimer possesses positively charged amino acids exclusively concentrated at the ends of IMD, interacting with high affinity to PI(4,5)P₂, and localise the IMD to the plasma membrane (Mattila et al., 2007; Saarikangas et al., 2009). However, it is also shown that membrane binding is not required for the formation of pedestals by EHEC bacteria (Weiss et al., 2009). It would be interesting to examine whether the carboxy-terminal region of Tir competes with the membrane for IRSp53 binding or if membrane bound IRSp53 is also capable to interact simultaneously with this Tir fragment. This could be examined by e.g. vesicle cosedimentation assays (Saarikangas et al., 2009), where the ability of labeled Tir to cosediment with membrane bound IMD would be examined.

1.7.3 The binding affinity of Tir to IMD

The quite low dissociation constant of 60 µM of the EHEC Tir:IMD interaction, measured by ITC, may be due to the use of a short peptide instead of the whole carboxy-terminal cytoplasmic Tir. As already discussed (section 1.6.5), the SH3 domain of IRSp53 as well as the avidity effect may contribute to this interaction and possibly dramatically increase the binding affinity. Thus, it would be exciting to test whether other regions of Tir and IRSp53 also contribute to this interaction. Furthermore, it would be of interest to test the effects of IRSp53 mutants in physiological relevant cell culture models, as the pull down assays were performed using cell lysates and the PIP assays were performed in yeast.
2 Chapter II

Structural Characterisation of Amorfrutins Bound to the Peroxisome Proliferator-activated Receptor gamma

2.1 Introduction

2.1.1 The nuclear receptor PPAR\(\gamma\)

The peroxisome proliferator-activated receptor gamma (PPAR\(\gamma\)) belongs to the so-called adopted orphan receptors of the nuclear receptor superfamily. The name is derived from the ability of the first identified member of this family, PPAR\(\alpha\), responding, in contrast to the other family members, to various compounds that induce peroxisome proliferation (Tontonoz and Spiegelman, 2008). The PPARs exhibit distinct tissue distributions with different activities. PPAR\(\alpha\) is expressed in liver, kidney, heart, and muscle where it regulates fatty acid catabolism. PPAR\(\delta\) is ubiquitously expressed with less defined functions, but is implicated in e.g. wound healing and keratinocyte differentiation (Michalik et al., 2001). PPAR\(\gamma\), the target protein of this study, is mainly expressed in adipocytes and macrophages. PPAR\(\gamma\) consists of an amino-terminal, ligand-independent activation function 1 (AF1), a DNA binding domain which is highly conserved among the superfamily, followed by a hinge region and a carboxy-terminal conserved ligand-binding domain (LBD), containing the activation function 2 (AF2) (Figure 2.1-1A) (Evans, 1988; Glass and Ogawa, 2006). PPAR\(\gamma\) acts in the nucleus as a heterodimer with the retinoid X receptor RXR\(\alpha\) and is a ligand-activated transcription factor (Berger and Moller, 2002), interacting with food ingredients like fatty acids and their metabolites (Figure 2.1-1B) (Kliewer et al., 1997; Tontonoz and Spiegelman, 2008). PPAR\(\gamma\) is a well-known drug target of e.g. type II diabetes treatment as it works as a dominant sensor and regulator in adipose cell differentiation and glucose and lipid metabolism, but
plays also a key role in inflammation, atherosclerosis and cancer, and has strong anti-growth properties (Murphy and Holder, 2000; Tontonoz and Spiegelman, 2008).

Figure 2.1-1: Domain organisation of PPARγ and the structural characteristics of its LBD. (A) PPARγ consists of an amino-terminal activation function domain 1 (AF1), a central DNA-binding domain, and a carboxy-terminal ligand-binding domain (LBD) (Glass and Ogawa, 2006). (B) PPARγ and the mechanism of DNA binding according to Glass and Ogawa (2006). PPARγ heterodimerises with RXRα and binds to DNA in a ligand-dependent manner. (C-E) The protein surface and ribbon representation of PPARγ’s LBD reveal a huge ligand entry site in the lower portion of the protein, located between helix H3 and H2’. (D) The receptor dynamic is illustrated by plotting the temperature factors (B-factors) on the structure (PDB 1PRG), revealing the omega loop and helix H12 as the most mobile segments of the LBD. The phosphorylation site at Ser245 (Ser273 in PPARγ2), and helix H12 are highlighted. (E) The huge ligand binding cavity of PPARγ is illustrated by white surfaces (Nolte et al., 1998).
2.1.1.1 The activation function 1 (AF1)

The amino-terminal AF1 of the nuclear receptor is different in two isoforms of PPARγ. In contrast to PPARγ1, PPARγ2 has 28 additional amino acids at the amino-terminus. It is expressed in a more adipose-selective manner and about ten times more abundant than PPARγ1 (Tontonoz et al., 1994). Furthermore, PPARγ2 appears to be more potent to induce adipogenesis (Rosen et al., 2002). MAP kinase (specifically Erk1, also known as p44) mediated phosphorylation of AF1’s Ser112 reduces ligand-binding affinity of the receptor and negatively regulates the transcriptional activity, resulting in the inhibition of its biological action (Adams et al., 1997; Hu et al., 1996). The phosphorylation was suggested to induce a conformational change of the unbound receptor, in which the amino-terminal AF1 inhibits, via intramolecular interactions, ligand binding to the LBD (Shao et al., 1998). The AF1 was also shown to bind to PPARγ co-activator 2 (PGC-2), not affected by Ser112 phosphorylation, which itself exhibits pro adipogenic action (Castillo et al., 1999). However, the exact mechanism of AF1 mediated regulation of PPARγ’s adipogenic action remains still elusive and requires further investigation.

2.1.1.2 The ligand-binding domain (LBD)

The LBD has several regulator functions. It determines the subcellular localisation, initiates heterodimerisation with RXRα and regulates the transcriptional activation or repression in a ligand-dependant manner. The LBD consists of 13 α-helices and a small four-stranded β-sheet region (Figure 2.1-1D). The AF2, the receptors surface facilitating co-activator interaction, is formed by the LBD’s helix 3-4 loop and the carbox-terminal ends of helix H11 and H12. The LBD has a phosphorylation site at Ser245 (Ser273 in PPARγ2) (Choi et al., 2010) and possesses a cavity of about 1300 Å³ in the lower portion of the LBD (Nolte et al., 1998), located between helix H3 and the β-sheet region (Figure 2.1-1C-E). Ligand binding to the LBD of the nuclear receptor leads to co-activator recruitment. Co-activators like the Steroid receptor co-activator-1 (SRC1) or the Peroxisome proliferator-activated receptor gamma co-activator-1 alpha (PGC-1α) possess an LXXLL motif (where L is leucine and X is any amino acid) that dock with their hydrophobic sides, formed by the leucine residues, into a hydrophobic groove on the surface of the receptor (Figure 2.1-2) (Heery et al., 1997; Li et al., 2008; Nolte et al., 1998). A positively charged lysine (K301) and a negatively charged glutamic acid (E471) residue at each side of the hydrophobic groove contribute to the selectivity by clamping the LXXLL motif by hydrogen bonds with the motif’s backbone at both, the amino- and carboxy-terminal ends (Darimont et al., 1998; Nolte et al., 1998; Shiau et al., 1998). This kind of interactions of the hydrophobic groove with the co-activator motif leads to the name ‘charge clamp’. In the unbound state, nuclear receptors are
frequently blocked by co-repressors like the silencing mediator co-repressor for retinoid and thyroid-hormone (SMRT) or the nuclear receptor co-repressor (N-CoR) (Chen and Evans, 1995; Horlein et al., 1995) containing an LXX[I/H]XXX[I/L] motif (Hu and Lazar, 1999; Nagy et al., 1999; Perissi et al., 1999). The ligand dependant distinction between co-repressor and co-activator was proposed to be caused by the different length of the motifs that have to be accommodated in the binding pocket. Ligand binding leads to a conformational change in which the amphipathic helix H12 is stabilised tightly packed against the LBD (see also section 2.1.1.3) and shortens the hydrophobic groove, such that it only suits the smaller co-activator motif (Perissi et al., 1999).

Figure 2.1-2: Complex of PPARγ’s LBD with an SRC1 peptide. The SRC1 peptide (red) interacts with the charge clamp of the LBD (PDB 2PRG) in response to a ligand like rosiglitazone. The accessibility of the charge clamp is regulated by helix H12 (yellow) in a ligand dependant manner. Residues defining the charge clamp are indicated. Residues in parentheses correspond to residues in PPARγ2.

2.1.1.3 Ligand-mediated LBD stabilisation regulates PPARγ’s transcriptional activity

Crystallographic temperature factors (B-factors) are linearly related to the mean square of an atom’s displacement and indicate atomic flexibility. Crystal structures of unbound PPARγ have high temperature factors in their lower portion (Figure 2.1-1D), which thus seems to be less rigid than the upper portion. Helix H12 is the most mobile helix and the omega-loop, linking helix H2’ and helix H3, is the most mobile loop (Nagy et al., 1999; Nolte et al., 1998). Both elements are located in the lower portion. Ligand binding to PPARγ reduces atomic thermal vibrations and stabilises the receptor. This stabilisation induces a conformational
change of helix H12, leading to a tight association of this helix to the LBD and results in a more compact and rigid structure (Choi et al., 2010; Nagy and Schwabe, 2004). This affects, as described above, co-regulator recruitment. The ligand dependant degree of LBD stabilisation is believed to be directly proportional to the receptors transcriptional activity (Nagy and Schwabe, 2004) and may cause differential co-regulator recruitment (Choi et al., 2010). Stabilisation can be achieved by direct contacts of the ligand to helix H12, as shown for full agonists like rosiglitazone (see Figure 2.1-4 and section 2.1.1.3.1), or by a helix H12 independant stabilisation, as shown for partial and intermediate agonists like the non-thiazolidinedione selective partial agonist (nTZDpa) and BVT.13 (see Figure 2.1-3 and section 2.1.1.3.2). However, both mechanisms of stabilisation result in a similar position of helix H12.

![Figure 2.1-3: The PPARγ partial agonist nTZDpa and the intermediate agonist BVT.13.](image)

2.1.1.3.1 The PPARγ full agonist rosiglitazone

The synthetic compound rosiglitazone belongs to the group of thiazolidinediones (TZDs) and is a potent, high-affinity PPARγ activator, which is used as insulin sensitizer to treat type II diabetes. The small, lipophilic rosiglitazone molecule targets adipose tissue where PPARγ is mainly expressed and stabilises preferentially helix H12 by direct contacts (Figure 2.1-4A). This kind of stabilisation leads to co-activator recruitment and is suggested to ‘freeze’ the receptor in a rigid conformation. Obesity-induced Cdk5-mediated phosphorylation of the receptor LBD’s Ser245 is thereby inhibited (Choi et al., 2010). The inhibition of the phosphorylation is apparently closely associated with the anti-diabetic effects of rosiglitazone, as it alters the expression of a subset of genes with regulatory functions in metabolism, like an enhanced transcription of the insulin-sensitising adipokine, adiponectin (Choi et al., 2010). Rosiglitazone is currently used to treat type II diabetes, but causes adverse effects including weight gain, oedema, and congestive heart failure (Kahn and
CHAPTER II - INTRODUCTION

McGraw, 2010; Lebovitz, 2002). These adverse effects rationalise research to explore new agonists for PPARγ which also induce insulin sensitivity but lack these adverse effects.

2.1.1.3.2 PPARγ partial and intermediate agonists

Full, intermediate, and partial agonists are defined by transcriptional activation potentials of >80%, 50-80%, or <50%, relative to the full agonist rosiglitazone. In contrast to full agonists, intermediate and partial agonists like BVT.13 and nTZDpa or MRL-24 (see Figure 2.1-4) do not stabilise helix H12 by direct contacts but rather globally stabilise the LBD by contacts to helix H3 and the β-sheet region of the binding pocket in the lower portion of the domain (Figure 2.1-4) (Bruning et al., 2007). The mode of stabilisation of the receptor by partial agonists leads to a more effective inhibition of Ser245 phosphorylation and to a more specific and in general reduced gene activation profile, compared with the full agonist rosiglitazone. These effects were already described for the partial agonist MRL-24 (Choi et al., 2010).

Partial agonists, also known as selective PPARγ modulators (SPPARγMs) are able to separate adverse effects like adipogenesis from inducing insulin sensitivity in vivo (Rangwala and Lazar, 2002).

![Figure 2.1-4: Ligand dependant stabilisation of the LBD.](image)
LBD stabilisation by the full PPARγ agonist rosiglitazone (A) and the partial agonist MRL-24 (B) according to the hydrogen/deuterium exchange data of Choi et al. 2010. The LBD’s stabilised regions are coloured blue. Helix H12, H3, and the phosphorylation site are indicated. Note the differentially stabilisation of helix H12 and the omega loop, containing the phosphorylation site.
Co-activators like SRC-1 are known to bind to the charge clamp at helix H12 in response to ligands like the full agonist rosiglitazone. But, as there are no significant conformational differences observed between the LBD bound to full, intermediate or partial agonists, it was suggested that the differentially stabilisation of the receptor and in particular the diminished stabilisation of helix H12 by partial agonists is reasonable for the reduced and more specific gene activation profile. Furthermore, the phosphorylation of Ser245 on the opposite site of the domain may constitute a second ligand-gated co-regulator binding site near this residue (Choi et al., 2010). The differentially regulation of this phosphorylation site by partial or intermediate agonists may thus also affect co-regulator recruitment. Agonists with the potential to induces insulin sensitivity with milder or even without the induction of adverse effects may pave the way to more specific PPARγ activation for the clinically use as therapeutic drugs to treat diseases like type II diabetes, while maintaining the patients quality of life.

2.1.1.4 Co-regulator recruitment

It was suggested that distinct ligands generate by differential stabilisation unique co-activator-binding surfaces on the LBD, resulting in a differential recruitment of co-regulators and thus a different transcriptional activation by PPARγ (Choi et al., 2010; Kodera et al., 2000; Wu et al., 2003). Furthermore, sequences flanking the LXXLL motif of the co-activator also affect the recognition by the receptor (Klein et al., 2005; Wu et al., 2003). These suggestions are further supported by the findings that PGC-1α is able to act as a co-activator of PPARγ in the presence or absence of rosiglitazone (Li et al., 2008; Wu et al., 2003). While the rosiglitazone-mediated co-activation is independant of the charge clamp mutation E471A, the ligand-independant activation is abolished by this mutation. This indicates a unique, ligand-dependant structural feature of the LBD that associates with PGC-1α (Wu et al., 2003). On the contrary, the rosiglitazone mediated activation is inhibited by the PGC-1α mutant S(-2)A, with the -2 position relative to the first leucine of the 144LXXLL148 motif (Wu et al., 2003). Thus both, the ligand but also the co-activator configurations are critical for co-regulator recognition and nuclear receptor activity.

2.1.2 A new class of high affinity PPARγ partial agonists

A new class of high affinity PPARγ partial agonists, the amorfrutins (see Figure 2.1-5), was identified by our cooperation partner, the group of Dr. Sascha Sauer, Max Planck Institute for Molecular Genetics, Berlin. They have screened a library of about 8000 pure natural products using an electrospray ionisation mass spectrometry based high-throughput
screening method and identified the amorfrutins as structurally new and affine PPARγ agonists. Amorfrutins are a group of natural products with the characteristics of SPPARγMs (Weidner et al., submitted). They are non-toxic ingredients of edible plant roots of liquorice, *Glycyrhiza foetida*, and fruits of a related legume, *Amorpha fruticosa*, also known as desert false indigo or bastard indigobush. The name ‘amorfrutin’ is derived from *Amorpha fruticosa*, in which the molecules were originally identified. Amorfrutins are selective and high affinity PPARγ agonists with dissociation constants for amorfrutin 1, 2, and B of about 20-300 nM, determined by time resolved FRET assays. The small, lipophilic amorfrutin molecules consist of a 2-hydroxy benzoic acid core structure with diverse phenyl and isoprenyl moieties. Amorfrutin 1 was shown to increase insulin sensitivity without the induction of adipogenesis and even lead to weight loss in vivo (Weidner et al., submitted). Amorfrutins have great potential for treatment or prophylaxis of type II diabetes and the metabolic syndrome in general.

![Figure 2.1-5: Structures of amorfrutin 1, 2, and B, used in this study.](image-url)
2.2 Aim of this work

The aim of this work was to solve the crystal structures of amorfrutin 1, 2, and B in complex with the ligand-binding domain (LBD) of human PPARγ for structural characterisation of these novel natural products at the atomic level. Recent research regarding ligand specific selectivity, affinity, and transcriptional activation (Weidner et al., submitted) led to the assumption that different structural features of amorfrutins stabilise the transcriptional receptor. Structural comparison, also with related complex structures, should help to identify these features, elucidate the mechanism of activation, and classify, whether amorfrutins are full, intermediate, or partial agonists. Furthermore, the study should explain how the structural data will fit into the current working model of PPARγ activation. The results of this study should contribute to the understanding of the regulatory mechanism of PPARγ in insulin sensitisation and fat metabolism.
2.3 Materials

If not stated otherwise, the materials used in this chapter are similar to the materials in chapter I.

2.3.1 Plasmid for hPPARγ LBD expression

A PCR fragment encoding amino acids 223-505 (AEIS...KDLY) of the LBD of hPPARγ (SwissProt PPARG_HUMAN P37231), cloned with NdeI and BamH1 into a pET-28a vector (Novagene), was used for expression, resulting in hPPARγ LBD fused to an amino-terminal His7-tag and thrombin cleavage site, with the sequence GSHMAEI...KDLY upon cleavage.

2.3.2 Compounds

The compounds amorfrutin 1 (NP-003520), amorfrutin 2 (NP-003521), and amorfrutin B were synthesized by Frank C. Schröder (Boyce Thompson Institute and Department of Chemistry and Chemical Biology, Cornell University, Ithaca, NY, USA) and obtained dissolved in DMSO at 100 mM. Higher concentrations of the amorfrutins couldn’t be achieved due to their solubility limits.

2.4 Methods

If not stated otherwise, the methods used in this chapter are similar to the methods in chapter I.

2.4.1 Production and purification of PPARγ’s LBD

The PPARγ LBD was expressed with a previously described expression plasmid kindly provided by Krister Bamberg (Cronet et al., 2001). Transformed BL21(DE3) cells were induced with 0.1 mM IPTG at 18°C for 20 h. Harvested cells were disrupted with a high pressure cell disrupter (Constant Systems) in 20 mM Tris-HCl, 150 mM NaCl, 10% (v/v) glycerol, 1 mM tris(2-carboxyethyl)phosphine HCl (TCEP, Gold Biotechnology), 10 mM imidazole, pH 8.0 in the presence of protease inhibitors (Complete EDTA-free, Roche) and centrifuged (35,000xg, 45 min). The supernatant was loaded on a 5 ml HisTrap HP column (GE Healthcare) and eluted with an imidazole gradient (10-300 mM imidazole) over 20
column volumes. The protein was diluted to a final concentration of 20 mM NaCl, immediately loaded onto a MonoQ HR 10/10 column (GE Healthcare) and eluted with a NaCl gradient (20-500 mM NaCl), followed by incubation with thrombin protease (1 U/mg) at 4°C for 20 h to cleave off the His-tag. His-tag peptides and uncleaved material were removed by rechromatography with Ni-NTA agarose (Qiagen), followed by gel filtration (HR 10/30 Superdex 75 prep grade column, GE Healthcare) in 20 mM Tris-HCl, 100 mM NaCl, 0.5 mM DTT, and 2 mM EDTA, pH 8.0.

2.4.2 Analytical methods

Photometric quantification of protein concentrations, SDS-PAGE, DLS, and MS were carried out in accord with the description in chapter I (see section 1.4.3).

2.4.2.1 Analytical gel permeation chromatography

Analytical gel permeation chromatography was performed in 20 mM Tris-HCl, 100 mM NaCl, 0.5 mM DTT, and 2 mM EDTA, pH 8.0 using a HiLoad 16/60 Superdex 75 column (GE Healthcare). 1 ml buffer containing a protein mixture of 20 mg/ml ribonuclease A (14.6 kDa), 20 mg/ml ovalbumin (46.7 kDa), and 10 mg/ml thyroglobulin (699 kDa) was applied to the column and used as molecular weight standard. The gel-phase distribution coefficient ($K_{av}$) was calculated according to Equation 2.4-1, where $V_0$ is column void volume, $V_e$ is elution volume, and $V_c$ is geometric column volume, calculated from column dimensions. $V_0$ was approximated from the elution volume of thyroglobulin. $K_{av}$ was plotted versus the logarithm of the molecular weight of the different standard proteins. The equation of the linear fit of the relationship between $K_{av}$ and the different molecular weights of the standard proteins was calculated and used to estimate the molecular weight of LBD.

$$K_{av} = \frac{V_e - V_0}{V_c - V_0} \quad \text{Equation 2.4-1}$$

2.4.2.2 AF4 and DLS

Protein samples of the LBD were analysed by Asymmetric Flow Field-Flow Fractionation (AF4), coupled to a dynamic light scattering (DLS) device, in-house by Dr. Thorsten Lührs.
2.4.3 Crystallisation by vapour diffusion

2.4.3.1 Sitting-drop vapour diffusion

The sitting-drop vapour diffusion method was used for initial screening in 96-well format as described in chapter I (see section 1.4.4.1).

2.4.3.2 Hanging-drop vapour diffusion

The protein was concentrated to 11 mg/ml and crystallised using hanging drop-vapour phase diffusion by mixing equal volumes of concentrated protein and reservoir solution (0.8 M tri-sodium citrate and 0.1 M imidazole, pH 8.0). Crystals appear after 2-5 days at 19°C.

2.4.3.3 Crystal soaking

Soaking crystals with the different amorfrutins was carried out using crystals grown under conditions described in section 2.4.3.2. The crystals were transferred into soaking buffer (10 mM Tris-HCl, 50 mM NaCl, 0.25 mM DTT, 1 mM EDTA, 0.8 M tri-sodium citrate, 0.1 M imidazole, final pH 8.0) with a stepwise increase in reagent concentration (0.5 mM, 1 mM, and 1.5 mM ligand and a maximum of 1.5% DMSO) with varying incubation times.

2.4.3.4 Co-crystallisation

The complexes of the PPARγ LBD bound to the different amorfrutins were prepared by adding a 10-fold molar excess of each amorfrutin to the purified LBD at a final concentration of 1 mg/ml. DMSO and unbound ligand were removed by using a HR 10/30 Superdex 75 prep grade column (GE Healthcare) in 20 mM Tris-HCl, 100 mM NaCl, 0.5 mM DTT, and 2 mM EDTA, pH 8.0. The complexes were crystallised using the hanging-drop vapour diffusion method according to section 2.4.3.2, but had to be optimised by microseeding (see section 1.4.4.3).

2.4.4 Data collection, structure determination and refinement

To prevent or minimise crystal damage by radiation, the crystals were transferred into a mild hypertonic cryo-protectant solution (0.84 M tri-sodium citrate, 25% (v/v) glycerol, and 0.1 M imidazole, pH 8.0) and immediately thereafter flash frozen in liquid nitrogen before data collection at 100 K. The data sets were collected in-house on a Saturn 944+ detector.
(Rigaku) using Cu Kα X-rays from a Rigaku MicroMax-007 HF rotating-anode X-ray generator with a VariMax Optic (Rigaku). Images were indexed and processed with XDS (Kabsch, 1993) and the structures were solved by molecular replacement using CCP4 MOLREP (Vagin and Teplyakov, 1997) with PDB entry 1PRG (Nolte et al., 1998) as the search model. REFMAC5 (Murshudov et al., 1997) and Phenix.refine (Adams et al., 2010) were used for refinement. TLS (translation/libration/screw) parameters (Painter and Merritt, 2006) were included to allow domain movements to be refined and Coot (Emsley and Cowtan, 2004) was used for manual model building. Figures were prepared with PyMOL. Root mean square deviations (r.m.s.d.) between common Cα-positions were calculated with ProFit using the McLachlan algorithm (McLachlan, 1982). The validation of the structure was performed using MolProbity (Chen et al., 2010; Davis et al., 2007).

2.4.5 Bioinformatics

Structural analysis and figure preparations were carried out according to chapter I (see section 1.4.6).

2.4.5.1 Sequence analysis

Protein sequences were aligned with ClustalW2 (Larkin et al., 2007) and displayed with ESPript 2.2 (Gouet et al., 2003).
2.5 Results

2.5.1 Production and purification of the PPARγ ligand-binding domain (LBD)

The PPARγ LBD was expressed and purified according to section 2.4.1, using IMAC as the initial purification step. The LBD was eluted from the matrix using a gradient up to 300 mM imidazole. The protein was already completely eluted after 130 mM imidazole (Figure 2.5-1). The yield of IMAC purified PPARγ LBD was about 70 mg/L culture.

**Figure 2.5-1:** IMAC purification of the PPARγ LBD. Bacterial lysate containing the LBD was loaded on a 5 ml HisTrap HP column and eluted with a gradient of 10 to 300 mM imidazole (A). The grey and green line depicts the fractions analysed by SDS-PAGE and Coomassie Blue staining (B). The fractions marked green were pooled for further purification. M: protein standard.

**Figure 2.5-2:** Anion exchange purification. The PPARγ LBD purified by IMAC was loaded on an anion exchange MonoQ HR 10/10 column and eluted with a gradient of 20 to 500 mM NaCl (A). The green line marks the fractions analysed by SDS-PAGE and Coomassie Blue staining (B). The marked fractions were pooled for further processing. M: protein standard.
The LBD was further purified and concentrated by anion exchange chromatography. It was eluted early in a gradient of 20 mM to 500 mM NaCl (Figure 2.5-2). The ion exchange chromatography leads to a high purity of the protein, which was subsequently digested using thrombin protease to cleave off the His-tag (Figure 2.5-3). IMAC Rechromatography with Ni-NTA superflow was used to remove the His-tag peptide and uncleaved material. Gel filtration was used for final purification and buffer exchange (Figure 2.5-4).

**Figure 2.5-3**: SDS-PAGE verifying successful removal of the approximately 3 kDa His-tag. M: protein standard.

**Figure 2.5-4**: Gel filtration. The LBD of PPARγ elutes at 71 ml from a Superdex 75 HR 16/60 column (A). The grey and green lines depict the fractions analysed by SDS-PAGE (B), whereas the fractions marked green were pooled. Arrows indicated artefacts of the gel electrophoresis. M: protein standard.

SDS-PAGE of the gel filtration fractions resulted in two additional bands at 75 kDa and 250 kDa next to the expected 32.7 kDa LBD band. Mass spectrometry, analytical gel filtration as well as AF4, coupled to DLS, were carried out to identify these additional bands. Mass
spectrometry, performed in-house by Dr. Manfred Nimtz and Undine Felgenträger, showed that all three bands consisted of hPPARγ LBD. Furthermore, analytical gel filtration and AF4-DLS both revealed a monodisperse protein solution (Figure 2.5-4A and Figure 2.5-5). Analytical gel filtration, using thyroglobulin to approximate the void volume of the column and ovalbumin as well as ribonuclease A as molecular standards, resulted in a single peak with a calculated molecular weight of about 40 kDa (Figure 2.5-4A and Figure 2.5-5A,B). The AF4-DLS, as the much more precise method, was carried out in-house by Dr. Thorsten Lührs and revealed unambiguously a monodisperse protein solution and the exact molecular weight of 32.6 kDa for the LBD (Figure 2.5-5C). These data verify the desired purity and monodispersity of the monomeric protein in solution and confirm an artificial oligomerisation of the protein during the SDS-PAGE, due to the formation of stable multimers under the condition of the SDS-sample-buffer used for the SDS-PAGE.

Figure 2.5-5: Monodispersity of the purified LBD of PPARγ. Elution profile of the analytical gel filtration (A) and the derived standard line (B). The standard line was calculated from ovalbumin and ribonuclease A elution volumes. Thyroglobulin was used to approximate the column’s void volume. Using the standard line, a molecular weight of about 40 kDa was derived for the LBD. (C) AF4-DLS profile of the LBD. Shown are UV absorbance (continuous lines) and the molar mass (dashed lines) as a function of time. The analyses were performed using different protein amounts: 10 µg (red lines), 20 µg (blue lines), and 50 µg (green lines), and reveal the exact molecular mass of the LBD (32.6 kDa).
CHAPTER II - RESULTS

2.5.2 Crystallisation of PPARγ’s LBD with different amorfrutins

Protein crystals of the LBD in complex with the different amorfrutins were obtained by crystal soaking (section 2.5.2.1) and co-crystallisation (section 2.5.2.2), respectively.

2.5.2.1 Crystal soaking

Crystal soaking is fast and more convenient for obtaining crystal structures of one protein with different ligands, as it requires only one crystallisation condition. Thus, crystallisation conditions for ligand-free PPARγ LBD were screened based on the previously published crystallisation conditions of this protein (Nolte et al., 1998) using the hanging-drop vapour phase diffusion method (section 1.4.4.2), but also using the sitting-drop vapour phase diffusion method in 96-well format and the JCSG screens (section 1.4.4.1). Protein crystals appeared only with sodium citrate as the precipitant. Crystallisation conditions were optimised to 0.8 M tri-sodium citrate and 0.1 M imidazole, pH 8.0 (section 2.4.3.2). Protein crystals appeared already after two days at 19°C. Within two weeks, some of the crystals grew up to large monoclinic crystals with edge lengths of about 400 µm (Figure 2.5-6). The protein crystals diffracted X-rays up to 2.0 Å resolution.

![Figure 2.5-6: Crystal of the ligand-free LBD of PPARγ. The scale bar corresponds to 200 µm.](image)

The amorfrutins were obtained dissolved in DMSO at a maximal concentration of 100 mM, due to their solubility limits. To avoid DMSO mediated crystal damage and to determine the maximal amorfrutin concentrations that could be used during the soaking, the effect of the organic solvent on crystal integrity was tested prior to the soaking of the compounds.

The crystals were transferred into soaking buffer with different DMSO concentrations. After 24 h incubation, damage of the crystals became visible at DMSO concentrations above 1.5%, like rounded edges or crystal cracks (Figure 2.5-7A). These effects did not exacerbate even after three days of incubation. Thus, a maximum of 1.5% DMSO and hence a maximum
concentration of the amorfrutins of 1.5 mM was used during the stepwise soaking of the LBD crystals. However, crystal soaking with the different amorfrutins resulted in crystal damage even at 1.5% DMSO (Figure 2.5-7B) and a decreased diffraction from up to 2.0 Å to about 2.5 Å, even at lower concentrations and varying incubation times. The resulting electron densities turned out to be too badly defined to identify the precise position and molecular interactions of the amorfrutins in the LBD.

![Figure 2.5-7](image_url)

**Figure 2.5-7: The effects of soaking on crystal integrity.** (A) DMSO lead to rounded edges and crystal cracks at concentrations above 1.5% and an incubation time of 24 h. (B) The amorfrutins, dissolved in DMSO, lead to affected crystal integrity even at DMSO concentrations below 1.5% and varying incubation times. Crystal damages are indicated by arrows.

### 2.5.2.2 Co-crystallisation

In comparison to crystal soaking, much better data were obtained by co-crystallisation of amorfrutins with PPARγ. The complexes of the LBD bound to different amorfrutins were formed according to section 2.4.3.4. The crystallisation conditions were similar to the crystallisation of the ligand free protein, but resulted in adnate, distorted, and cracked crystals. However, the microseeding technique was used to optimise crystal quality, leading to diffraction up to 2.0 Å and a well-defined electron density map that allowed the precise definition of the amorfrutins in the ligand binding pocket of PPARγ and the description of all molecular interactions.
2.5.3 Structural characterisation of different amorfrutins in complex with the LBD of PPARγ

The crystal structures of the LBD in complex with three different amorfrutins were solved by molecular replacement and were all refined at a resolution of 2.0 Å (Table 2.5-1). The electron density maps reveal clearly the canonical three layer α-helical sandwich composed of 13 α-helices and a small, four-stranded β-sheet (Figure 2.5-8). The structures are in accord with previously published PPARγ LBD structures (Bruning et al., 2007; Cronet et al., 2001; Nolte et al., 1998) with a calculated r.m.s. deviation of 0.87 ± 0.09 Å to Protein Data Bank (PDB) entry 1PRG (258 common Cα-positions of chain A aligned).

Figure 2.5-8: Structural comparison of the three amorfrutins bound to the LBD of PPARγ. (A) The aligned overall structures of the LBDs bound to the three amorfrutins are almost indistinguishable. Only amorfrutin B (orange sticks) is shown. The amorfrutins bind between the lower portion of helix H3 and the β-sheet region. Helices (blue) and β-strands (green) are labelled H1 to H12 and S1 to S4, respectively. (B) Superimposition of the three amorfrutins in the binding pocket, generated by alignment of the three respective protein structures. Amorfrutin 1 (white), 2 (yellow), and B (orange) are shown as sticks. Only the amorfrutin B-bound domain is displayed as ribbons. (C) 2F₀-Fₑ electron density omit map calculated around amorfrutin 1, 2, and B, contoured at 1.0 σ. The amorfrutins dynamic is illustrated by plotting B-factors on the structures.
Table 2.5-1: Summary of crystallographic analysis.

<table>
<thead>
<tr>
<th><strong>Data collection</strong></th>
<th><strong>amorfrutin 1</strong></th>
<th><strong>amorfrutin 2</strong></th>
<th><strong>amorfrutin B</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C121</td>
<td>C121</td>
<td>C121</td>
</tr>
<tr>
<td>Cell dimensions a, b, c (Å)</td>
<td>92.81, 61.10, 118.22</td>
<td>92.29, 60.97, 117.9</td>
<td>92.14, 60.86, 117.6</td>
</tr>
<tr>
<td>Unique angle β (°)</td>
<td>102.78</td>
<td>102.65</td>
<td>102.58</td>
</tr>
<tr>
<td>X-ray source</td>
<td>Rigaku MicroMax-007 HF</td>
<td>1.5419</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>27.2-0.0</td>
<td>27.2-0.0</td>
<td>29.2-0.0</td>
</tr>
<tr>
<td>Last shell (Å)</td>
<td>2.15-2.00</td>
<td>2.10-2.00</td>
<td>2.10-2.00</td>
</tr>
<tr>
<td>R$_{merge}$ (%)</td>
<td>10.7 (64.4)</td>
<td>10.9 (56.3)</td>
<td>8.7 (45.4)</td>
</tr>
<tr>
<td>Observations</td>
<td>293322 (51030)</td>
<td>149072 (18878)</td>
<td>148018 (18473)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>43633 (8488)</td>
<td>43005 (5857)</td>
<td>42829 (5697)</td>
</tr>
<tr>
<td>Mean (I)/s.d. (I)</td>
<td>15.8 (3.1)</td>
<td>13.71 (2.82)</td>
<td>17.68 (3.41)</td>
</tr>
<tr>
<td>Completeness</td>
<td>99.5 (99.8)</td>
<td>99.0 (99.5)</td>
<td>99.2 (97.7)</td>
</tr>
<tr>
<td>Multiplicity</td>
<td>6.7 (6.0)</td>
<td>3.4 (3.2)</td>
<td>3.5 (3.2)</td>
</tr>
</tbody>
</table>

**Structure refinement**

| Resolution range (Å) | 27.2-0.0         | 27.2-0.0         | 29.2-0.0         |
| R$_{work}$ (%)       | 20.4             | 20.89            | 20.3             |
| R$_{free}$ (%)       | 25.3             | 25.9             | 23.3             |

**Total number of**

| Non-hydrogen atoms | 4554             | 4671             | 4670             |
| Protein atoms      | 4299             | 4380             | 4339             |
| Ligand atoms       | 75               | 44               | 60               |
| Water molecules    | 180              | 247              | 271              |

**r.m.s.d.**

| Bond length (Å)     | 0.011            | 0.010            | 0.010            |
| Bond angle (°)      | 1.240            | 1.219            | 1.186            |
| Main chain B-factors (Å$^2$) | 39.1            | 37.9             | 39.0             |
| Side chain B-factors (Å$^2$) | 42.2            | 41.3             | 42.3             |
| Average B-factor protein atoms (Å$^2$) | 40.7            | 39.7             | 40.7             |
| Average B-factor ligand atoms (Å$^2$) | 43.1            | 47.6             | 43.6             |
| Average B-factor solvent atoms (Å$^2$) | 44.3            | 44.9             | 44.3             |

**Ramachandran statistics**

| Most favoured regions (%) | 98.8            | 97.3             | 98.3             |
| Allowed regions (%)       | 1.2             | 2.7              | 1.7              |
| Disallowed regions (%)    | 0               | 0                | 0                |

Values in parentheses are for the highest resolution shell.
All three complex structures had an asymmetric unit containing two amorfrutin molecules bound to one LBD homodimer of almost identical monomers. The structures of the two monomers, denoted chain A and B, differ due to crystal contacts. The two different conformations of the LBD chains, termed active and inactive, are commonly observed in PPARγ LBD structures (Cronet et al., 2001; Nolte et al., 1998). The inactive conformation (chain B) is caused by helix H12 of chain B forming crystal contacts with a neighbouring molecule that probably prevent it from moving into the active conformation like chain A (see also section 2.5.4). The active conformation of chain A is similar to the complexed receptor with the full agonist rosiglitazone and the co-activator SRC1 (Nolte et al., 1998) and is commonly regarded as a suitable model for PPARγ LBD activation. The following sections will therefore refer only to the complex in the active conformation.

The electron density maps of each amorfrutin is well defined in the complex structures (Figure 2.5-8C) and clearly reveal their position in the cavity of the LBD. Corresponding to their related structures, the three amorfrutins are bound with almost identical localisation and orientation (Figure 2.5-8A and B). Whereas full agonists like rosiglitazone or MRL-20 stabilise especially helix H12 of the LBD (Bruning et al., 2007; Nolte et al., 1998), the amorfrutins stabilise the receptor between helix H3 and the β-sheet region, close to the ligand entry site. This recognition of the amorfrutins by PPARγ is similar to the partial agonists nTZDpa and MRL-24, and also to the intermediate agonist BVT.13 (Bruning et al., 2007). These ligands all stabilise helix H3 and the β-sheet with hydrogen bonds to Ser342 and Arg288 and form similar extensive van der Waals contacts, most prominently to Ile341 of the β-sheet and Cys285 of helix H3 (Figure 2.5-9, Figure 2.5-10 and Figure 2.5-11A). The mechanism of stabilisation and the low transcriptional activation (<50%) compared to the full agonist rosiglitazone (Weidner et al., submitted) (Table 2.5-2), imply that these amorfrutins belong to the group of partial agonists.

Despite their similar structures, the amorfrutins interact differently with the LBD. Arg288 of the amorfrutin 1-bound PPARγ adopts two alternative conformations (Figure 2.5-9A), which are also observed in nTZDpa-bound PPARγ, but not in the unbound structure (PDB entry 1PRG), where Arg288 is hydrogen bonded to Ser289 (Figure 2.5-10). In the amorfrutin 1 complex, Arg288 either interacts with Glu343 of the β-sheet and the ligand’s carboxyl group, or with Glu295 of helix H3 via hydrogen bonds. Arg288 and Ser342 thus form a bond network upon binding of amorfrutin 1 that, supported by extensive van der Waals contacts of the phenyl and isoprenoid moieties, stabilises helix H3 and the β-sheet and explains the high binding affinity of 236 nM (Figure 2.5-9A and Table 2.5-2).
Figure 2.5-9: Structural details of amorfrutins bound to the LBD of PPARγ. (A) Amorfrutin 1 exerts a hydrogen-bonding network directly to Ser342 of the β-sheet region and Arg288 of H3. Note the alternative conformations of Arg288 (red arrows), stabilised by Glu291 and Glu343. (B, C) Amorfrutin 2 and B coordinate water molecules to form hydrogen-bonding networks between Arg288
of H3 and the β-sheet region, but also directly interact with Ser342. Additional hydrophobic interactions of amorfrutin B’s longer geranyl side chain to Arg288 are the likely cause of this ligand’s higher binding affinity and also prevent the formation of an alternative conformation of Arg288, as observed in the complex with amorfrutin 1. In the schematic diagrams of atomic interactions on the right, calculated using LIGPLOT (Wallace et al., 1995), the ligands and the residues of the binding pocket are drawn with black and orange bonds, respectively. Hydrogen bonds are depicted by dashed lines and van der Waals contacts are indicated by spoked arcs and atoms with spokes. Oxygen atoms are coloured red, nitrogen atoms blue and carbon atoms black. S: sheet, H: helix, L: loop.

Figure 2.5-10: Comparison of previously published PPARγ structures. Shown are the unbound LBD of PPARγ (PDB 1PRG) with Arg288 hydrogen bonded to Ser289 (Nolte et al., 1998), the LBD in complex with the intermediate agonist BVT.13 (PDB 2Q6S), and the LBD in complex with the partial agonists nTZDpa (PDB 2Q5S) and MRL-24 (PDB 2Q5P) (Bruning et al., 2007).
Amorfrutin 2 and B also form direct hydrogen bonds to Ser342 of the β-sheet but fail to form a salt bridge to Arg288 and do not trigger the alternative conformation of this residue. The lack of this interaction results in a lower degree of stabilisation of the 2-hydroxy benzoic acid core of amorfrutin 2 and B, depicted by increased B-factors (see Figure 2.5-8). However, both ligands use the same residues of the LBD to form a water network that stabilises Arg288 of helix H3 (Figure 2.5-9B, C and Figure 2.5-11A), similar to the structure with BVT.13 of Bruning et al. (2007) (Figure 2.5-10). Amorfrutin 2 has a binding constant of 287 nM, similar to amorfrutin 1 with 236 nM (Table 2.5-2), whereas binding of amorfrutin B is markedly stronger with a binding constant of 19 nM.

Amorfrutin B’s long meta-geranyl side chain at the 2-hydroxy benzoic acid core leads to additional hydrophobic contacts especially to Arg288 of helix H3 and to helix H4-5, compared with amorfrutin 1 and 2 (Figure 2.5-9 and Figure 2.5-11A). These additional contacts of amorfrutin B lead to reduced B-factor values of the meta-geranyl side chain (see Figure 2.5-8) and likely prevent that Arg288 adopts the alternative conformation, required for direct contacts to the ligand’s carboxyl group. The additional hydrophobic contacts also explain the almost 12-fold higher binding affinity of amorfrutin B compared to amorfrutin 1 (Table 2.5-2).

Table 2.5-2: Affinity constants (Ki) and transcriptional activation relative to rosiglitazone according to (¹) Weidner et al. (submitted) and (²) Bruning et al., 2007.

<table>
<thead>
<tr>
<th></th>
<th>PPARα Ki</th>
<th>PPARδ Ki</th>
<th>PPARγ Ki</th>
<th>Relative transcriptional activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amorfrutin 1</td>
<td>27 µM¹</td>
<td>27 µM¹</td>
<td>0.236 µM¹</td>
<td>39%¹</td>
</tr>
<tr>
<td>Amorfrutin 2</td>
<td>25 µM¹</td>
<td>17 µM¹</td>
<td>0.287 µM¹</td>
<td>30%¹</td>
</tr>
<tr>
<td>Amorfrutin B</td>
<td>2.6 µM¹</td>
<td>1.8 µM¹</td>
<td>0.019 µM¹</td>
<td>n.a.</td>
</tr>
<tr>
<td>nTZDpa</td>
<td>n.a.</td>
<td>n.a.</td>
<td>0.029 µM¹</td>
<td>&gt;50%²</td>
</tr>
<tr>
<td>MRL-24</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>&gt;50%²</td>
</tr>
<tr>
<td>BVT.13</td>
<td>n.a.</td>
<td>n.a.</td>
<td>n.a.</td>
<td>50-80%²</td>
</tr>
<tr>
<td>Rosiglitazone</td>
<td>7 nM¹</td>
<td>n.a.</td>
<td>7 nM¹</td>
<td>100%¹</td>
</tr>
</tbody>
</table>

n.a., not available

Amorfrutins exhibit high selectivity for PPARγ, as shown for the three amorfrutins with a selectivity factor of 60 to 140 in comparison to the other PPAR subtypes. Even if the conservation of the residues of the lower portion of helix H3 and the β-sheet is quite low (Figure 2.5-11) most of the interactions to the amorfrutins should also be fulfilled by the residues of PPARα or PPARδ. However, the guanidinium group of the unconserved Arg288 (Figure 2.5-11) seems to play a key role in amorfrutin binding and this feature cannot be satisfied by any other residue. Thus the missing conservation of Arg288 is most likely the reason of the high selectivity of the three amorfrutins to the PPARγ subtype.
Figure 2.5-11: PPAR sequence conservation. (A) Sequence alignment of the three human PPARs, PPARα, PPARγ, and PPARδ (SwissProt Q07869, P37231, Q03181). The sequence identity to PPARγ is 69% and 72% for PPARα and PPARδ, respectively. Red and white boxes highlight identical and similar residues, respectively. Hydrophobic contacts of amorfrutin 1, 2, and B to PPARγ are indicated by blue, red, and green lines, respectively, beneath the sequences. Hydrogen bonds to amorfrutin 1 are indicated by black asterisks; hydrogen bonds to amorfrutin 2 and B are indicated by cyan asterisks. Ruler is given according to PPARγ PDB entry 1PRG. (B) Sequence conservation between the three different PPARs. Blue areas indicate identity, white areas similarity, and yellow areas represent differing residues. Similarity score is projected on the complex structure of the LBD with amorfrutin 1.
2.5.4 Crystallographic contacts of the inactive conformation lead to an artificial orientation and localisation of the amorfrutins

The orientation and localisation of the amorfrutin complex structures seem to be unaffected by crystallographic contacts in the active conformation. The inactive conformation (chain B), however, is obviously altered due to several contacts to a crystallographic related molecule. Helix H12 of chain B forms crystal contacts with chain A’s charge clamp of a neighbouring molecule and protrudes from the molecule, as it was observed in other PPARγ structures before (Cronet et al., 2001; Nolte et al., 1998). Helix H12 of chain B thereby occupies the charge clamp into which co-activator peptide binds, thus mimicking the two-turn, amphipathic co-activator α-helix (Figure 2.5-12) (Pochetti et al., 2007). Hence, chain B shows an artificial conformation. Chain A corresponds to the complex structures with co-activator peptides and therefore acts commonly as a model for PPARγ LBD activation. But also the omega loop linking H2’ and H3, the most thermally mobile loop of PPARγ’s LBD, forms crystal contacts with a related molecule. This loop is stabilised close to the ligand entry side by a hydrogen bond between the carbonyl of Lys265 and the nitrogen of Leu311 of the crystallographic related molecule (Figure 2.5-13A). Thereby the loop occludes parts of the ligand entry site of the inactive form, with His266 and Phe264 pointing into the pocket and to the amorfrutins, leading to an energetically unfavourable position of the ligands. Hence, the amorfrutins delocalise and rotate by almost 180° in the cavity, compared to the active state (Figure 2.5-13B), except amorfrutin 1 that exhibits an alternative conformation in the inactive state.

Figure 2.5-12: Crystallographic contacts mimic a co-activator peptide. Cartoon representation of PPARγ’s active chain A (blue), interacting with chain B of a symmetry related molecule (orange). Helix H12 of chain B occupies the charge clamp of chain A, thus mimicking a co-activator peptide like SRC1 (red). SRC1 peptide is shown as it lies in the active conformation of the ternary complex with rosiglitazone (PDB 2PRG), generated by superimposition. Helix H12 of active chain A is drawn in yellow. Ribbon diagram is only shown for the complex structure with amorfrutin B.
Figure 2.5-13: Crystallographic contacts in the inactive conformation lead to an artificial orientation and localisation of the amorfrutins. (A) Surface and ribbon representation of PPARγ’s ligand entry site of the inactive conformation with amorfrutin B (orange) as it lies in the active conformation, generated by superimposition. Interactions of a crystallographically related LBD molecule in a different homodimer (green) stabilise the loop linking H2’ and H3 (yellow) in front of the ligand entry side of the inactive state with His266 and Phe264 pointing into the pocket and to the amorfrutins as they lie in the active conformation. Residues involved are labelled. Oxygen atoms are coloured red and nitrogen atoms blue. Hydrogen bonds are shown as black dashed lines. (B) Localisation of amorfrutin 1 (white), 2 (yellow), and B (orange) in the active and inactive conformation of the LBD homodimer. Note the rotation of the amorfrutins in the inactive conformation by almost 180°, except for amorfrutin 1 (white), showing an alternative conformation. Ribbon diagram is only shown for the amorfrutin B bound form.
2.6 Discussion

2.6.1 Structural insights into amorfrutin recognition by PPARγ

The crystal structures presented here structurally characterise amorfrutin 1, 2, and B, members of a novel class of natural products derived from non-toxic, edible plant roots that possess the ability of a selective PPARγ modulator to separate insulin-sensitisation from adipogenesis (Weidner et al., submitted). In cooperation with the group of Dr. Sascha Sauer (Max Planck Institute for Molecular Genetics, Otto Warburg Laboratory, Berlin, Germany) we identify the three amorfrutins as partial agonists as they show a transcriptional activation of PPARγ below 50% compared to the full agonist rosiglitazone (Weidner et al., submitted) (Table 2.5-2) and interact with the nuclear receptor via a helix H12 independent mechanism by stabilising helix H3 and the β-sheet region, similar to other partial or intermediate agonists of PPARγ (Bruning et al., 2007).

This study shows that the amorfrutin 1, 2, and B bind, due to their structural similarity, with almost identical localisation and orientation close to the ligand entry site of the LBD. All three amorfrutins form extensive van der Waals contacts with the LBD via their isoprenoid, phenyl and aliphatic moieties, most prominently to Ile341 of the β-sheet S3 and Cys285 of helix H3. All three ligands engage a hydrogen bond network including Arg288 of helix H3 and Ser342 of the β-sheet region. Amorfrutin 1 utilises only its carboxyl group for a direct hydrogen bond to Ser342 and to an alternative conformation of Arg288. Amorfrutin 2 and B utilise also their ortho-hydroxyl group to contact Arg288 and Ser342. However, amorfrutin 2 and B do not directly interact with Arg288 and thus do not induce an alternative conformation of this residue’s side chain, but utilise a water-mediated hydrogen bond to Arg288 instead (Figure 2.5-9).

The carboxyl group of the amorfrutins, interacting with the main chain nitrogen of Ser342 and, as shown for amorfrutin 1, the guanidinium group of Arg288, seems to be reasonable for their high affinity. This is supported by a dramatically increased binding constant from 590 nM to 23 µM upon esterification of the acidic head group of the structurally related amorfrutin 5 (Figure 2.6-1) (Weidner et al., submitted).

Regarding the selectivity of amorfrutin 1, 2, and B for PPARγ (60- to 140-fold) over PPARα or PPARδ, Arg288, and in particular its guanidinium group that interacts with the ligands carboxyl or hydroxyl, seems to play the most important role. Arg288 is not conserved among the PPARs. PPARα and PPARδ have a threonine at this position, which cannot interact with the amorfrutins in a similar way to arginine. This suggestion is further supported by a reduced affinity of a naturally occurring PPARγ mutant (R288H) to naturally ligands like the 15-Deoxy-Delta-12,14-prostaglandin J2 (15d-PGJ2) or the fatty acid (9S)-9-Hydroxyoctadeca-10,12-dienoic acid (9(S)-HODE) in comparison to rosiglitazone (Sarraf et al., 1999). Both, 15d-PGJ2 and 9(S)-HODE are shown to use Arg288 for excessive
hydrophobic interactions or hydrogen bondings, respectively (Itoh et al., 2008; Waku et al., 2009), in contrast to the synthetic agonist rosiglitazone (Nolte et al., 1998). Furthermore, hydrogen bonds of the three amorfrutins to the β-sheet region, like the interaction to Ser342, are only dependent on main-chain atoms. Thus these interactions will probably be accomplished by other amino acids, like by the alanine of PPARα and PPARδ at this position. And even Glu343 which is an asparagine or tyrosine in PPARα and PPARδ, respectively, does not seem critical for the recognition, as the side-chain protrudes away from the ligand.

Figure 2.6-1: Esterification of amorfrutin 5. Amorfrutin 5 (A) has a 40-fold higher affinity constant to PPARγ compared to its methylester (B) (Weidner et al., submitted).

2.6.1.1 Structural comparison with other related PPARγ ligands

Comparison of amorfrutin 1, 2, and B with the previously published partial PPARγ agonists nTZDpa and MRL-24 as well as with the intermediate agonist BVT.13 reveals a closely related interaction mode. All these ligands bind at almost the same site and similarly stabilise helix H3 and the β-sheet. Furthermore, all these ligands contain a carboxyl group that directly interacts with the Ser342 of the β-sheet and, with the exception of amorfrutin 1, implicate water molecules for further stabilisation.

The nTZDpa interacts, similar to amorfrutin 1, via its carboxyl group with both, Ser342 and Arg288 in the alternative conformation (Figure 2.5-9A and Figure 2.5-10). Both ligands were described to be capable of avoiding adverse effects like adipogenesis while improving insulin sensitivity in vivo (Berger et al., 2003)(Weidner et al., submitted) and thus have a high potential as drugs to treat type II diabetes without inducing fat storage.
### 2.6.1.2 Possible mechanism of differential co-regulator recruitment

The structure of the ternary complex of PPARγ, RXRα and DNA revealed three distinct interfaces (Chandra et al., 2008). The DNA binding and ligand-binding domains of the two proteins interact with each other. In addition, the LBD of PPARγ faces with the β-sheet region the DNA binding domain of RXRα (Figure 2.6-2). This β-sheet region is important for DNA binding, as shown by mutating Phe347 of the β-sheet S4 of PPARγ’s LBD into alanine, resulting in weakened ability to bind DNA (Chandra et al., 2008). However, rosiglitazone and BVT.13, latter interacting similar to amorfrutins away from helix H12 but involve helix H3 and the LBD’s β-sheet region, induce only minor variations of the overall conformation of the ternary complex (Chandra et al., 2008). Thus the effects of amorfrutins 1, 2, and B, as well as the effects of other related ligands on gene transcription are unlikely to be caused by structural changes affecting the DBD of RXRα and its DNA affinity.

![Diagram](image.png)

**Figure 2.6-2:** Overall structure of the PPARγ-RXRα complex on DNA (PDB 3DZY) according to Chandra et al. (2008). The expressed LBD of PPARγ for the experimental work in this study is drawn in blue, the PPARγ DBD in light blue, and RXRα in red. Co-activator peptides and rosiglitazone are drawn in black and yellow sticks, respectively. The three different interfaces are indicated by grey planes. The most important residues for amorfrutin interactions, Arg288 and Ser342, as well as the phosphorylation site at Ser245 and the residue Phe347, important for DNA interaction of the heterodimer, are labelled.
PPARγ contains a phosphorylation site at Ser245 (Ser273 in PPARγ2), residing between helix H2 and the β-sheet S1 (Choi et al., 2010), close to RXRα’s DBD. High fat feeding induces obesity in mice and activates the protein kinase Cdk5 (cyclin-dependant kinase 5) in adipose tissue, resulting in phosphorylation of this residue (Choi et al., 2010). The phosphorylation has no influence on DNA binding or general transcriptional activity, but alters the expression of a subset of genes with regulatory functions in metabolism, including a reduced transcription of the insulin-sensitising adipokine, adiponectin (Choi et al., 2010). Amorfrutin 1 and rosiglitazone were shown to block the phosphorylation of Ser245, which is apparently closely associated with their anti-diabetic effects (Choi et al., 2010) (Weidner et al., submitted). It is hypothesized that ligand-mediated interactions between helix H3 and the β-sheet ‘freezes’ this region in a conformation less favourable to Cdk5 phosphorylation (Choi et al., 2010). This leads to the suggestion of differentially regulated co-regulator recruitment in a phosphorylation-dependant manner (Choi et al., 2010) by a novel co-activator binding surface close to Ser245. Thus the inhibited phosphorylation by rosiglitazone or amorfrutin 1 may initiate the recruitment of additional co-regulators, reasonable for dysregulated gene activation, like the enhanced transcription of the insulin-sensitising adiponectin, and the increased insulin sensitivity. Furthermore, regions outside the co-activators LXXLL motifs may also be involved and affect co-activator recognition and regulation (Klein et al., 2005).

2.6.1.3 The active and inactive state of the PPARγ’s LBD

The data demonstrate the unreliability of the inactive conformation even for the helix H12 independant PPARγ activation. Crystallographic contacts to the loop linking H2’ and H3, called the omega-loop, lead to rotation and relocation of the amorfrutins in the cavity (section 2.5.4). The omega-loop seems to be responsible for modulation of PPARγ activity by endogenous fatty acids like 15d-PGJ2 (Waku et al., 2009), illustrating the importance to distinguish between active and inactive form. Furthermore, crystallographic contacts to the omega-loop occlude parts of the ligand entry site and may thus exacerbate soakings of different ligands.
2.7 Outlook

Due to the reported insulin-sensitising effects, the reduced fat storage and the improvement of metabolic and inflammatory parameters, which were ascribed to amorfrutin 1, it would be an exciting task to study the different effects of amorfrutin B, which exhibits a 12-fold higher binding affinity to PPARγ and to test if this compound exhibits improved, similar, or probably reduced effects. Especially the effect on the Cdk5 mediated phosphorylation of PPARγ’s LBD at Ser245 would be of special interest, as, to the current knowledge, it seems to be strongly associated with the anti-diabetic effect of substances like rosiglitazone, MRL-24, or amorfrutin 1. The higher binding affinity of amorfrutin B may contribute to a better stabilisation of the receptor and probably rigidify it in a higher degree as amorfrutin 1 or rosiglitazone do and hence may protect Ser245 more efficiently from Cdk5 mediated phosphorylation. It would be also possible that a structural related ligand with a higher binding affinity may reduce the dosing for the desired effect, which is currently with 10 mg/kg/d amorfrutin 1 (Weidner et al., submitted) quite high. On the contrary, the more lipophilic structure and a possibly resulting reduced bioavailability in the organism may lead to reduced effects and a reduced insulin sensitising action. These effects remain elusive and have to be discovered in the future.

Furthermore, as the PPARγ ligand 15d-PGJ2, a naturally occurring prostaglandin, was described to modify the receptors transcriptional activity by inducing structural changes to the omega-loop (Waku et al., 2009), residing between helix H2’ and H3, it would be also an exciting task to test if amorfrutins also induce a comparable and transcriptional relevant change to this region. As this region exhibits a high flexibility and thus was not visible in the present crystal structures, biophysical techniques appear to be more suitable to address this question. Method of choice may be the measurement of tryptophan fluorescence spectra of a K265W mutant of PPARγ (Waku et al., 2009). The mutation seems to be uncritical for ligand binding as K265 protrudes away from the LBD. The residue resides in the omega-loop and hence environmental changes around this area should become visible by changes of the spectra. Furthermore, the mutants I267A and F287A, residues of the omega-loop and helix H3, respectively, were described to dramatically decrease the receptors transcriptional activity in response to 15d-PGJ2, as these residues are responsible for the induced structural changes of the omega-loop without disturbing 15d-PGJ2 affinity (Waku et al., 2009). The amorfrutins interact close to the ligand entry side and the omega-loop and thus probably also affect the omega-loop conformation. Thus it would be of interest to analyse the amorfrutin mediated cell based transcriptional activity of these mutants in comparison to the wilde-type.
Additional mutants like R288T, R288A and R288H may be also of interest to quantify the contribution of the LBD’s Arg288 to amorfrutin binding, transcriptional activity, and to the selectivity of PPARγ to the amorfrutins in comparison to the other PPAR subtypes. R288T should mimic the residue of PPARα and PPARδ at this position and thus may give insights in PPARγ’s amorfrutin selectivity, compared with the wide-type and the R288A mutant. The naturally occurring PPARγ mutant R288H is suggested to be associated with human colorectal cancer and was described to exhibit reduced affinity to naturally occurring ligands like 15d-PGJ2 or 9(S)-HODE (Section 2.6.1), shown to excessively interact with Arg288 (Sarraf et al., 1999) (Section 2.6.1).

However, one of the most important questions regarding PPARγ mediated transcriptional activation or dysregulation is addressed to probably additional co-regulator recruitment. The omega-loop region and the region around the phosphorylation site at Ser245 appear currently as the most suitable regions for such co-regulator recruitment. Thus it would be an exciting task to identify possible new co-regulators of PPARγ and discover new aspects of the mechanism of nuclear receptor regulation.

Anyway, the insulin sensitising effect and the reduced fat storage upon ingestion of the non-toxic amorfrutin 1 makes this naturally compound already very attractive for the possible usage as a drug to treat diseases like type II diabetes and/or obesity. Furthermore, as the amorfrutins are natural and non-toxic compounds, the usage of amorfrutins as dietary supplements to prevent type II diabetes and/or obesity may be also possible. Thus these compounds need to be tested by a human study to unveil possible human specific adverse effects and to give more detailed information about the metabolic action of these substances.
3 References


REFERENCES


REFERENCES


REFERENCES


REFERENCES


REFERENCES


ligand-binding domain in the complex with two enantiomeric ligands. J Biol Chem 282, 17314-17324.


REFERENCES


REFERENCES


4 Accession Numbers

Atomic coordinates have been deposited in the Protein Data Bank [http://www.pdb.org] with the PDB IDs given in Table 2.7-1.

Table 2.7-1: Accession Numbers.

<table>
<thead>
<tr>
<th>Complex</th>
<th>Title of the entry</th>
<th>PDB ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>IMD_{IRSp53}:Tir</td>
<td>Crystal structure of the I-BAR domain of IRSp53 (BAIAP2) in complex with an EHEC derived Tir peptide</td>
<td>2YKT</td>
</tr>
<tr>
<td>hPPARγ-LBD:amorfrutin_1</td>
<td>Ligand-binding domain of human PPAR gamma in complex with amorfrutin 1</td>
<td>2YFE</td>
</tr>
<tr>
<td>hPPARγ-LBD:amorfrutin_2</td>
<td>Ligand-binding domain of human PPAR gamma in complex with amorfrutin 2</td>
<td>4A4V</td>
</tr>
<tr>
<td>hPPARγ-LBD:amorfrutin_B</td>
<td>Ligand-binding domain of human PPAR gamma in complex with amorfrutin B</td>
<td>4A4W</td>
</tr>
</tbody>
</table>
DANKSAGUNG

Danksagung


Prof. Dr. Dirk Heinz danke ich vor allem für die Möglichkeit, meine Doktorarbeit in seiner Abteilung anfertigen zu dürfen. Die hervorragenden Arbeitsbedingungen, exzellentes Equipment sowie sein Bestreben, Kooperationen effektiv mit meinen Projekten zu verknüpfen, haben deutlich zur Qualität und Abschluss dieser Arbeit beigetragen. Zudem möchte ich ihm sehr für die Unterstützung bei unserer Veröffentlichung danken, sowie für die Übernahme der Betreuung meines Promotionsvorhabens als Mentor.

Meinem Arbeitsgruppenleiter Dr. Joop van den Heuvel möchte ich ebenfalls für seine Diskussionsbereitschaft, sein großes Interesse am Gelingen dieser Arbeit, seine uneingeschränkte Unterstützung und für viele Tipps und Tricks im Laboralltag danken.

Prof. Dr. Michael Steinert möchte ich herzlich für die bereitwillige Übernahme des Zweitgutachtens danken.

Ebenso möchte ich mich herzlich bei Prof. Dr. Ralf-Rainer Mendel für die Übernahme des Prüfungsvorsitzes bedanken.

Bei Prof. Dr. Theresia Stradal und Kai Schlüter, sowie der restlichen AG Stradal der Uni Münster, möchte ich mich für die erfolgreiche Kooperation, Diskussionen und Hilfestellungen bedanken.

Bei Dr. Sascha Sauer, Dr. Christopher Weidner und Claudia Quedenau der Gruppe für Nutrigenomics and Gene Regulation des Max-Planck-Instituts in Berlin möchte ich mich ebenfalls herzlich für eine hervorragende Kooperation bedanken.

Bei Dr. Ronald Frank bedanke ich mich für die bereitwillige Teilnahme an meinem Thesis-Komitee und anregenden Diskussionen.

Dr. Victor Wray danke ich für die sehr hilfreichen Tipps, Anregungen und tatkräftige Unterstützung bei der Formulierung und Gestaltung unserer Publikation.
DANKSAGUNG

Bei Dr. Thorsten Lührs bedanke ich mich für die Durchführung der AF4 Experimente.

Bei Dr. Manfred Nimtz, Undine Felgenträger und Anja Meier möchte ich mich für eine Vielzahl von massenspektroskopischen Analysen bedanken.


Bei unserem Hiwi Adrian Zander möchte ich mich für seine Mitarbeit und sein Engagement im Labor bedanken.

Ich danke Ute Widow, Daniela Gebauer, Nadine Konisch und Sarah Tokarski für die Einarbeitung, eine sehr herzliche Arbeitsatmosphäre und für alles, wofür der Platz hier einfach nicht ausreichen würde.

Ich möchte mich bei allen Mitgliedern der RPEX und der gesamten SB für ein perfektes Arbeitsklima, eine gute Zusammenarbeit und eine sehr schöne und lehrreiche Zeit bedanken.

Ein besonderer Dank gilt zudem Alexander Eberth, Slava Ziegler und Katharina Lecher für ihre sehr hilfreichen Unterstützungen.

Frau Dr. Willeke, Herrn Dr. Weiland und Frau Dr. Härtha-Thoma danke ich für schnelle und richtige Entscheidungen im richtigen Moment, die mich wieder auf die Beine gebracht haben und u. a. auch den Abschluss dieser Arbeit erst ermöglichten.

Abschließend möchte ich mich bei meiner Familie und meiner Freundin Jana bedanken, die mich in jeglicher Hinsicht und zu jeder Zeit unterstützt haben.