Biochemical Studies on IGF and IGF-Binding Proteins Interactions & Structural Investigations on the SH3 Domain of Crk-associated Tyrosine Kinase Substrate p130cas (CAS)

Magdalena Wisniewska

2005
Biochemical Studies on IGF and IGF-Binding Proteins Interactions &
Structural Investigations on the SH3 Domain of Crk-associated Tyrosine Kinase Substrate p130cas (CAS)

Magdalena Wisniewska

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Dr. A. Bacher
Prüfer der Dissertation: 1. apl. Prof. Dr. Dr. h. c. R. Huber
                         2. Univ.-Prof. Dr. J. Buchner

Die Dissertation wurde am 08.03.2005 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 10.05.2005 angenommen.
Acknowledgements

The thesis presented here is a result of the studies carried out in the Department of Structural Research at the Max Planck Institute for Biochemistry from June 2001 to December 2004.

I would like to thank Professor Robert Huber for giving me the opportunity to work in his department and for being my Doktorvater.

I am indebted to my supervisor Doctor Rick A. Engh for his scientific contributions, guidance and financial support during all these years.

I am very grateful to Doctor Tad A. Holak, who, although not involved directly in the supervision of my work, gave me the strong sense of membership of his group.

I would like to express my deep gratitude to the people from Roche Diagnostics, Penzberg for getting me involved in a number of very interesting projects. In particular to Doctor Martin Lanzendörfer, Doctor Kurt Lang, Michael Schräml and Birgit Essig for offering a friendly and scientific atmosphere in their laboratories during my stay at the department.

I am indebted to all members of both: Roche and NMR Groups at MPI for their help, advice and pleasant atmosphere. Especially, I am very grateful to Christine and Grzesiek for their inestimable assistance with crystallographic tasks, and Ola for her friendship.

In special case I would like to thank my sister Ania, not only for spending a lot of time on reading my manuscripts, but also for her strong support, which has accompanied me from the beginning of my scientific path.

Last, but not least, I would like to acknowledge Igor thanks to whom I made this thesis in the MPI and my parents for all their love, care and constant encouragement.
Publications


Wisniewska, M., Schräml, M., Siwanowicz, I., Lanzendörfer, M, Ambrosius, D., Huber, R. and Engh, R. A. Application of the open cell-free rapid translation system (RTS) to the expression of the insulin-like growth factor binding protein-4. *manuscript in preparation*


**Abbreviations**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Ångström ($10^{-10}$ m)</td>
</tr>
<tr>
<td>ALS</td>
<td>acid labile subunit</td>
</tr>
<tr>
<td>APS</td>
<td>ammoniumperoxodisulfate</td>
</tr>
<tr>
<td>BirA</td>
<td>protein biotin ligase</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CAS</td>
<td>crk-associated tyrosine kinase substrate</td>
</tr>
<tr>
<td>CECF</td>
<td>continuous-exchange cell free technology</td>
</tr>
<tr>
<td>CIP</td>
<td>calf intestinal alkaline phosphatase</td>
</tr>
<tr>
<td>Crk</td>
<td>adapter protein</td>
</tr>
<tr>
<td>Da</td>
<td>Dalton (g/mol)</td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethylsulfoxide</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>ECM</td>
<td>extracellular matrix</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGF</td>
<td>epidermal growth factor</td>
</tr>
<tr>
<td>EGL</td>
<td>ethylene glycol</td>
</tr>
<tr>
<td>EGTA</td>
<td>ethyleneglycotetraacetic acid</td>
</tr>
<tr>
<td>ELISA</td>
<td>enzyme-linked immunosorbant assay</td>
</tr>
<tr>
<td>ESI-MS</td>
<td>electrospray ionisation mass spectrometry</td>
</tr>
<tr>
<td>Esf/Sin</td>
<td>embryonal Fyn-associated substrate/Src-interacting protein</td>
</tr>
<tr>
<td>FAs</td>
<td>focal adhesions</td>
</tr>
<tr>
<td>FAK</td>
<td>focal adhesion kinase</td>
</tr>
<tr>
<td>FGF</td>
<td>fibroblast growth factor</td>
</tr>
<tr>
<td>FSH</td>
<td>follicle-stimulating hormone</td>
</tr>
<tr>
<td>g</td>
<td>acceleration of gravity ($g = 9.81 \text{ m s}^{-2}$)</td>
</tr>
<tr>
<td>GH</td>
<td>growth hormone</td>
</tr>
<tr>
<td>GSH</td>
<td>reduced glutathione</td>
</tr>
<tr>
<td>GSSG</td>
<td>oxidized glutathione</td>
</tr>
<tr>
<td>GST</td>
<td>glutathione S-transferase</td>
</tr>
<tr>
<td>HBS</td>
<td>HEPES buffered saline</td>
</tr>
<tr>
<td>HEF1</td>
<td>human enhancer of filamentation</td>
</tr>
<tr>
<td>HEPES</td>
<td>N-(2-hydroxyethyl)piperazine-N’-(2-ethanesulfonic acid)</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>IGF</td>
<td>insulin-like growth factor</td>
</tr>
<tr>
<td>IGFBP</td>
<td>insulin-like growth factor binding protein</td>
</tr>
<tr>
<td>IGFBP-rP</td>
<td>IGFBP-related proteins</td>
</tr>
<tr>
<td>IGF-IR</td>
<td>insulin-like growth factor receptor type I</td>
</tr>
<tr>
<td>IGF-IIR</td>
<td>insulin-like growth factor receptor type II</td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropyl-β-thiogalactopyranoside</td>
</tr>
<tr>
<td>ITC</td>
<td>isothermal titration calorimetry</td>
</tr>
<tr>
<td>JKN</td>
<td>Jun N-terminal kinase</td>
</tr>
<tr>
<td>$K_A$</td>
<td>association constant</td>
</tr>
<tr>
<td>$K_D$</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Broth medium</td>
</tr>
<tr>
<td>MALDI</td>
<td>matrix-assisted laser desorption/ionization mass spectrometry</td>
</tr>
<tr>
<td>MAD</td>
<td>multiwavelength anomalous diffraction</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple cloning site</td>
</tr>
<tr>
<td>MES</td>
<td>N-ethylsulfonic acid morpholine</td>
</tr>
<tr>
<td>MIR</td>
<td>multiple isomorphous replacement</td>
</tr>
<tr>
<td>MR</td>
<td>molecular replacement</td>
</tr>
<tr>
<td>MW</td>
<td>molecular weight</td>
</tr>
<tr>
<td>NiNTA</td>
<td>nickel-nitrilotriacetic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate buffered saline</td>
</tr>
<tr>
<td>PDGF</td>
<td>platelet-derived growth factor</td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
</tr>
<tr>
<td>PI3K</td>
<td>phosphatidylinositol 3-kinase</td>
</tr>
<tr>
<td>pIVEX</td>
<td>In Vitro Expression</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonfylfluoride</td>
</tr>
<tr>
<td>PRM</td>
<td>proline-rich motif</td>
</tr>
<tr>
<td>RBS</td>
<td>ribosomal binding site</td>
</tr>
<tr>
<td>RMSD</td>
<td>root mean square deviation</td>
</tr>
<tr>
<td>RTS</td>
<td>rapid translation system</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulphate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Src</td>
<td>sarcoma tyrosine kinase</td>
</tr>
<tr>
<td>SH2</td>
<td>Src homology 2 domain</td>
</tr>
<tr>
<td>SH3</td>
<td>Src homology 3 domain</td>
</tr>
<tr>
<td>SHIP2</td>
<td>SH2-containing inositol 5’-phosphatase 2</td>
</tr>
<tr>
<td>SIR</td>
<td>single isomorphous replacement</td>
</tr>
<tr>
<td>SPR</td>
<td>surface plasmon resonance</td>
</tr>
<tr>
<td>T7P</td>
<td>T7 promoter</td>
</tr>
<tr>
<td>T7T</td>
<td>T7 terminator</td>
</tr>
<tr>
<td>TEMED</td>
<td>N,N,N’,N’- tetramethylethyldiamine</td>
</tr>
<tr>
<td>TBE</td>
<td>Tris-Borate-EDTA</td>
</tr>
<tr>
<td>TGF-β</td>
<td>transforming growth factor-β</td>
</tr>
<tr>
<td>Tris</td>
<td>tris(hydroxymethyl)aminomethane</td>
</tr>
</tbody>
</table>

Amino acids and nucleotides are abbreviated according to either one or three letter IUPAC code.
Table of Contents

1 Biochemical studies on IGF and IGF-binding proteins interactions 1
   1.1 Biological background 1
      1.1.1 Introduction 1
      1.1.2 Insulin-like growth factors 2
      1.1.3 IGF receptors 2
      1.1.4 IGF-binding proteins 4
      1.1.5 IGFs and cancer 9
   1.2 Aim of the project 10
   1.3 Materials and methods 11
      1.3.1 Materials 11
         1.3.1.1 E. coli strains 11
         1.3.1.2 Plasmids 11
         1.3.1.3 Antibiotics 11
         1.3.1.4 Cell growth media and stocks 11
         1.3.1.5 Common buffers 12
         1.3.1.6 Buffers for purification of the proteins under denaturing conditions 13
         1.3.1.7 Buffers for purification of the proteins under native conditions 14
         1.3.1.8 Enzymes, antibodies and other proteins 15
         1.3.1.9 Kits and reagents 15
         1.3.1.10 Protein and nucleic acids markers 16
      1.3.2 General laboratory methods 16
         1.3.2.1 Transformation of E. coli 16
            1.3.2.1.1 Transformation by heat shock 16
            1.3.2.1.2 Transformation by electroporation 16
         1.3.2.2 Preparation of plasmid DNA 16
         1.3.2.3 Digestion with restriction enzymes 17
         1.3.2.4 Purification of DNA fragments 17
         1.3.2.5 DNA Agarose gel electrophoresis 17
         1.3.2.6 Sonication 17
         1.3.2.7 SDS polyacrylamide gel electrophoresis (SDS PAGE) 17
         1.3.2.8 Staining of proteins 19
         1.3.2.9 Determination of protein concentration 19
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3.3 Mutagenesis of the IGF-miniBP5</td>
<td>20</td>
</tr>
<tr>
<td>1.3.4 Protein expression, refolding and purification</td>
<td>22</td>
</tr>
<tr>
<td>1.3.5 1 D-1H NMR</td>
<td>23</td>
</tr>
<tr>
<td>1.3.6 Binding on the column</td>
<td>23</td>
</tr>
<tr>
<td>1.3.7 Kinetic analysis</td>
<td>23</td>
</tr>
<tr>
<td>1.3.7.1 SPR principles</td>
<td>24</td>
</tr>
<tr>
<td>1.3.7.2 BIAcore analysis</td>
<td>24</td>
</tr>
<tr>
<td>1.3.7.3 ITC principles</td>
<td>25</td>
</tr>
<tr>
<td>1.3.7.4 ITC experiments</td>
<td>25</td>
</tr>
<tr>
<td>1.3.8 Crystallization trials</td>
<td>27</td>
</tr>
<tr>
<td>1.3.9 In vitro translation system as a high throughput method for screening of the proteins</td>
<td>27</td>
</tr>
<tr>
<td>1.3.9.1 Principles of rapid translation system (RTS)</td>
<td>28</td>
</tr>
<tr>
<td>1.3.9.2 Mono-biotinylation of the proteins using RTS</td>
<td>29</td>
</tr>
<tr>
<td>1.3.9.3 Expression vectors</td>
<td>30</td>
</tr>
<tr>
<td>1.3.9.4 Cloning of human IGFBP4 into pIVEX vectors</td>
<td>31</td>
</tr>
<tr>
<td>1.3.9.5 In vitro protein expression</td>
<td>34</td>
</tr>
<tr>
<td>1.3.9.6 Western blot analysis of the RTS products</td>
<td>34</td>
</tr>
<tr>
<td>1.3.9.7 Purification of HisTag-BP4</td>
<td>35</td>
</tr>
<tr>
<td>1.3.9.8 Purification of StrepTag-BP4</td>
<td>35</td>
</tr>
<tr>
<td>1.3.9.9 Surface plasmon resonance experiments</td>
<td>36</td>
</tr>
<tr>
<td>1.4 Results – mutagenesis studies</td>
<td>37</td>
</tr>
<tr>
<td>1.4.1 Protein purification</td>
<td>37</td>
</tr>
<tr>
<td>1.4.2 Size exlusion binding test</td>
<td>40</td>
</tr>
<tr>
<td>1.4.3 Characteristics of the binding of IGF-I to the miniBP5 mutants determined by surface plasmon resonance (SPR)</td>
<td>41</td>
</tr>
<tr>
<td>1.4.4 Characteristics of the binding of IGF-I to the miniBP5 mutants determined by isothermal titration calorimetry</td>
<td>43</td>
</tr>
<tr>
<td>1.4.5 Crystallization of miniBP5</td>
<td>45</td>
</tr>
<tr>
<td>1.4.5.1 Optimisation of crystallization conditions</td>
<td>45</td>
</tr>
<tr>
<td>1.4.5.2 Data collection</td>
<td>46</td>
</tr>
<tr>
<td>1.5 Discussion-mutagenesis studies</td>
<td>48</td>
</tr>
<tr>
<td>1.6 Results – RTS</td>
<td>52</td>
</tr>
</tbody>
</table>
# Table of Contents

1.6.1 Optimisation of RTS conditions 52
   1.6.1.1 Time 52
   1.6.1.2 Temperature 54
   1.6.1.3 Proteases inhibitors and chaperones 57
1.6.2 Optimisation of RTS conditions – Summary 58
1.6.3 Purification of IGFBP-4 59
1.6.4 BIACore analysis 60
1.7 Discussion – RTS 62

2 Structural studies of SH3 domain of recombinant human Crk-associated tyrosine kinase substrate p130cas (CAS) 67
   2.1 Biological background 67
      2.1.1 p130cas (CAS) 67
      2.1.2 SH3 domains as potential anti-tumor targets 71
   2.2 Aim of the project 72
   2.3 X-ray crystallography as the most powerful method for structural studies 73
      2.3.1 General background 73
      2.3.2 Crystal space groups and symmetries 73
      2.3.3 Crystal growth 73
      2.3.4 X-ray diffraction by crystals 75
      2.3.5 How to solve the phase problem? 75
   2.4 Expression and purification of CAS SH3 domain 80
   2.5 Crystallization of the CAS SH3 domain 80
      2.5.1 Co-crystallization trials 80
         2.5.1.1 Compounds-putative inhibitors for CAS SH3 80
            2.5.1.1.1 Soaking with compounds 82
            2.5.1.1.2 Cross-linking trials 82
         2.5.1.2 Peptide inhibitors 82
      2.5.2 Crystallization of the CAS SH3 domain expressed in the RTS 82
   2.6 Results and discussion 83
      2.6.1 Data collection and structure determination 83
         2.6.1.1 Data collection-crystals of the protein produced in the RTS 85
      2.6.2 Molecular structure 86
      2.6.3 Water molecules 89
      2.6.4 Crystal packing effects 90
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.6.5 Comparing CAS SH3 structures from the three different packing</td>
<td>90</td>
</tr>
<tr>
<td>environments</td>
<td></td>
</tr>
<tr>
<td>2.6.6 Implications for protein-ligand interactions</td>
<td>94</td>
</tr>
<tr>
<td>2.6.7 General Conclusions</td>
<td>99</td>
</tr>
<tr>
<td>3 Summary</td>
<td>101</td>
</tr>
<tr>
<td>4 Zusammenfassung</td>
<td>103</td>
</tr>
<tr>
<td>5 References</td>
<td>105</td>
</tr>
<tr>
<td>6 Appendix</td>
<td>123</td>
</tr>
<tr>
<td>6.1 Amino acid sequenceses of the constructs used in a work</td>
<td>123</td>
</tr>
<tr>
<td>6.2 List of chemicals</td>
<td>124</td>
</tr>
<tr>
<td>6.3 Laboratory equipment</td>
<td>126</td>
</tr>
<tr>
<td>6.4 Computer programs</td>
<td>128</td>
</tr>
</tbody>
</table>
1 Biochemical studies on IGF and IGF-binding proteins interactions

1.1 Biological background

1.1.1 Introduction

The insulin-like growth factors (IGFs) family comprises the polypeptide ligands IGF-I and IGF-II, two types of cell membrane receptors (IGF-IR and IGF-IIR), six binding proteins (IGFBP-1 through IGFBP-6), and an emerging group of IGFBP-related proteins (IGFBP-rP 1-9). In addition, a large group of IGFBP proteases hydrolyzes, resulting in the release of bound IGFs that then resume their ability to interact with IGF-IR. Thus, as far as IGF action is concerned, IGFBP proteases may also be regarded as part of the IGF family because they indirectly regulate the action of IGFs. The scheme of IGF system is presented in figure 1.1.

Figure 1.1. The IGF system. IGFBPs bind IGFs with high affinity, regulating the availability of free IGFs. As indicated, specific proteases cleave IGFBPs and thus regulate levels of free IGF as well as IGFBPs. The IGF-I receptor binds both IGFs and insulin, whereas IGF-II receptor interacts with IGFs and proteins containing mannose-6-phosphate but not insulin. IGFBPs are involved in IGF-dependent and IGF-independent actions (Wetterau et al., 1999).
1.1.2 Insulin-like growth factors

The insulin-like growth factors (IGF-I and IGF-II) are mitogens that play a pivotal role in regulating cell proliferation, differentiation, apoptosis, and transformation (Jones and Clemmons, 1995; Wetterau et al., 1999). IGF-I and IGF-II are 67% identical single polypeptide chains of 70 and 67 amino acids, respectively, sharing with insulin ~40% sequence identity and presumed structural homology. The first 29 residues of IGFs are homologous to the B-chain of insulin (B region, 1–29), followed by 12 residues that are analogous to the C-peptide of proinsulin (C region, 30–41), and a 21-residue region that is homologous to the A-chain of insulin (A region, 42–62). The C-terminal octapeptide (D region, 63–70) has no counterpart in insulins and proinsulins (Baxter et al., 1992; Murray-Rust et al., 1992). The IGFs are the only members of the insulin superfamily, in which the C region is not removed proteolytically after translation. Most of the IGF-I and IGF-II molecules in serum are found in a 150 kDa ternary complex formed by an IGF, IGFBP-3, and a glycoprotein known as the acid labile subunit (ALS) (Baxter, 1994). To a smaller degree, IGFBP-5 also forms such a complex (Twigg et al., 1998). A small proportion of IGFs is carried by other IGFBPs, and less than 1% of IGFs circulate in the free forms (Baxter, 1994).

Expression of the IGF-I gene is regulated primarily by growth hormone (GH). GH, however, has no regulatory effect on IGF-II expression, and the primary regulator of IGF-II gene transcription remains unclear (Jones and Clemmons, 1995). Expression of IGFs is also influenced by various hormones, including estrogens, adrenocorticotropic hormone, thyrotropin, luteinizing hormone, follicle-stimulating hormone (FSH), human chorionic gonadotropin, as well as by other growth factors, such as platelet-derived growth factor (PDGF), epidermal growth factor (EGF), and fibroblast growth factor (FGF). Diet and nutrition affect circulating IGF-I level, suggesting a possible impact on IGF-I production (Sara et al., 1990).

1.1.3 IGF receptors

There are two known receptors that specifically recognize the IGFs. The IGF-IR receptor (also known as the type 1 IGF receptor) has a high degree of homology with the insulin receptor, which is mostly involved in the transduction of metabolic signals. The IGF-II receptor (also known as the type 2 IGF receptor) is identical to the cation-independent mannose 6-phosphate receptor. The major biological roles of the IGF-II receptor appear to be recycling of lysosomal enzymes containing Man-6-P residues and clearance of IGF-II.
Both IGF-IR and IGF-IIR are glycoproteins and are located on the cell membrane. The two receptors, however, differ completely in structure and function (Le Roith et al., 1995). IGF-IR is a tetramer of two identical α-subunits and two identical β-subunits. The α- and β-subunits are linked by disulfide bonds to form αβ-half receptor, which, in turn, is subsequently linked to another αβ-half receptor (by disulfide bonds between the α subunits) to form the mature α2β2 holoreceptor (Figure 1.2). Ligand binding specificity is conferred by the cysteine-rich regions of the α-subunit extracellular domain, while tyrosine kinase activity resides in the cytoplasmic β-domain. Just as there is structural homology between insulin and the IGFs, the IGF-I receptor and the insulin receptor are both α2β2-heterotetramers with subunits of almost identical size and have regions where there is a high degree of sequence homology (~50-60% overall sequence identity and 84% in the tyrosine kinase domain) (Czech, 1989).

![Figure 1.2. Structural characteristics of the IGF-IR receptor (based on Jones and Clemmons, 1995).](image)

One result of the close homology between IGF-I receptors and insulin receptors is the formation of a hybrid insulin/IGF receptor. This receptor is composed of one α-subunit and one β-subunit of IGF-IR and one α-subunit and one β-subunit of the insulin receptor (Le Roith et al., 1995). The amount of the insulin/IGF-I hybrid receptor varies substantially from tissue to tissue. Studies with isolated hybrid receptors show that they bind both IGF-I and insulin.
with affinities comparable to the IGF-I receptors (15- to 50-fold lower affinity for insulin than for IGF-I) (Jones and Clemmons, 1995).

The IGF-II/cation-independent mannose-6-phosphate (IGF-II/Man-6-P) receptor is monomeric. Three ligand-binding regions are found in the extracellular domain of the receptor, one for IGF-II binding and two for proteins containing mannose-6-phosphate, such as rennin, proliferin, thyroglobulin, and the latent form of transforming growth factor TGF-β. IGF-II/Man-6-P receptors bind IGF-II with high affinity (reported \( k_d \) range 0.017-0.7 nM) but the affinity for IGF-I is more than 500-fold lower, and the receptor does not bind insulin (Jones and Clemmons, 1995). IGF-IIR has no tyrosine kinase activity. Since binding of IGF-IIR to IGF-II results in degradation of IGF-II, IGF-IIR acts like an antagonist to IGF-II, reducing its biological activity (Oates et al., 1998). Because of this effect, IGF-IIR has been considered to be a potential tumor suppressor molecule. A unique feature of IGF-IIR may contribute to its ability to act as a scavenger for circulating IGF-II. Upon proteolytic cleavage, the extracellular domain of the receptor is disassociated from the cell membrane as a soluble fragment that circulates in the blood with the ability to bind to IGF-II and to facilitate its degradation (Costello et al., 1999).

1.1.4 IGF-binding proteins

Structural features
The IGF-binding proteins family comprises six proteins (IGFBP-1 to 6), ranging in length from 216 to 289 amino acids, that bind to IGFs with high affinity and an emerging group of IGFBP-related proteins (IGFBP-rP 1-9), which bind IGFs with low affinity. It should be noticed that N- and O-linked glycosylation can increase the size of proteins considerably (in case of IGFBP-3, from 28.7 kDa to ~ 45 kDa for the fully glycosylated protein). IGFBPs, originally identified as serum transport proteins for the IGFs, are ubiquitously produced in virtually all tissues, with each tissue having specific levels of certain IGFBPs. Figure 1.3 presents the amino acid alignment of the six human IGFBPs.
Table 1.3. Sequence alignment of human IGFBP-1 to -6. miniBP-5 that was studied is in magenta and bold. Conserved residues are indicated by cyan shading. Residues that were mutated are highlighted in yellow. Linker region (L-domain) is underlined.
All IGFBPs share a common domain organization. The highest conservation is found in the N- (residues 1 to 100) and C- (from residue 170) terminal cysteine rich regions. Twelve conserved cysteines are found in the N-terminal domain (exception is IGFBP-6 with only 10 conserved cysteines) and six in the C-terminal domain. IGFBP-4 is unique among the IGFBPs in having two extra cysteine residues in the variable L-domain. IGFBPs contain a GCGCCXXC motif and a CWCV sequence within their N- and C-terminus, respectively. The central, weakly conserved part (L-domain) contains most of the cleavage sites for specific proteases (Chernausek et al., 1995). It should be underlined that the limited proteolysis plays a key role in modulating levels and actions of free IGFs and IGFBPs (Wetterau et al., 1999). In general, proteolyzed forms of IGFBPs tend to have greatly reduced binding activity compared to the intact proteins, as demonstrated by the ability of protease resistant forms of IGFBP-4 and -5 to retain IGF-inhibitory activity under conditions where protease-sensitive IGFBPs are inactive (Imai et al., 1997; Rees et al., 1998).

Both the N- and C-domains participate in the binding to IGFs (Baxter et al., 1992; Baxter, 2000; Buckway et al., 2001; Clemmons, 2001; Firth and Baxter, 2002; Firth et al., 2001; Payet et al., 2003; Shand et al., 2003) although specific roles of each of these domains in IGF binding have not been decisively established. In the absence of the complete tertiary structures of any IGFBP neither alone nor in complex with IGF, inferences about important structural domains of the proteins have been based on their natural variants (e.g. partially proteolyzed protein forms), as well as, on mutagenesis and expression studies. Huhtala et al. (1986) described a naturally occurring N-terminal 21 kDa fragment of human IGFBP-1 isolated from placenta, which had IGF-I-binding activity and his results were confirmed thanks to mutagenesis studies performed by Brinkman et al. (1991). During last years, a number of publications reporting the structural characterization of the small domains of the IGFBPs, which are known to mediate IGFs binding, has rapidly risen (Kalus et al., 1998; Zeslawski et al., 2001; Headey et al., 2003; Siwanowicz et al., 2005a), giving a hope for true understanding of the IGF-IGFBPs interactions.

The various post-translational modifications of IGFBPs (except proteolysis there are phosphorylation and glycosylation) can affect their binding to IGFs, although the effects differ for different proteins, and perhaps also different species of origin. For example, phosphorylated forms of IGFBP-1 from human serum and other sources have up to 10-fold higher affinity for IGF-I than the fully dephosphorylated form, although the affinities of the two forms for IGF-II do not differ so distinctly. In case of glycosylation, there is no evidence that this modification has strong influence on protein ability to bind IGFs.
IGFBP-1, IGFBP-2, IGFBP-3, IGFBP-4 and IGFBP-5 demonstrate high affinity for both IGF-I and IGF-II and share at least 50% homology among themselves, and 80% homology between different species (Lamson et al., 1991).

Regulation of IGFBPs gene transcription is complex and tissue specific. A number of hormones (estrogens, glucocorticoids, FSH, GH, thyroid hormone, insulin, vitamin D and cortisone), as well as growth factors (FGF, EGF, TGF-β, PDGF, and IGFs themselves) have been found to regulate the expression of these proteins (Yu and Rohan, 2000).

**Functional features of the IGFBPs**

IGFBPs have multiple and complex functions, which can be either IGF-dependent or IGF-independent. With respect to IGF-dependent function, IGFBPs are able to inhibit or enhance the action of IGFs, resulting in suppression or stimulation of cell proliferation. Several IGFBPs (e.g. IGFBP-4) inhibit IGF actions by binding to IGFs and preventing the binding of IGFs to the IGF receptors (Qin et al., 1998). Because the affinity of IGFBPs for IGFs is an order of magnitude greater than the affinity of IGF receptors towards IGFs, very little IGF binds to receptors in the presence of an equimolar concentration of receptor and binding protein. On the other hand, it is known that a number of IGFBPs (IGFBP-1, -3 and -5) stimulate IGF actions in a variety of cell types. It should be emphasized that the same protein could stimulate or inhibit IGF actions, depending on several variables, including culture conditions, cell type, IGFBP dose and posttranslational modifications (Clemmons, 1998; Ferry et al., 1999; Hwa et al., 1999; Baxter, 2000; Collettt-Solberg and Cohen, 2000; Clemmons, 2001). In this regard, non-phosphorylated IGFBP-1 has been shown to increase the IGF-I actions, whereas its phosphorylated form decreases IGF-I actions in cultured smooth muscle cells or fibroblasts (Busby et al., 1988).

IGF-independent actions of IGFBPs include effects on cell growth, migration and apoptosis, which involve putative BP receptors (reviewed in Firth and Baxter, 2002). IGFBP-1 interacts with α5β1 integrin, influencing cell adhesion and migration. IGFBP-2, -3, -5, and -6 have heparin-binding domains and can bind glycosaminoglycans. IGFBP-3 and -5 have carboxyl-terminal basic motifs incorporating heparin-binding and additional basic residues that interact with the cell surface and matrix, the nuclear transporter importin-β, and other proteins. Serine/threonine kinase receptors have been proposed for IGFBP-3 and -5, but their signaling functions are poorly understood. Other cell surface IGFBP-interacting proteins have been described but not identified as functional receptors. However, IGFBP-3 binds and modulates the retinoid X receptor-α, interacts with TGF-β signaling through Smad proteins,
and influences other signaling pathways (Fanayan et al., 2000). These interactions can modulate cell cycle and apoptosis (Firth and Baxter, 2002; Rajah et al., 2002). Because IGFBPs regulate cell functions by diverse mechanisms, manipulation of IGFBP-regulated pathways may offer therapeutic opportunities in cancer, ageing and other diseases (for recent reviews see: Bach, 1999; Cohen et al., 2000; Khandwala et al., 2000; Firth and Baxter, 2002; Ali et al., 2003; LeRoith and Helman, 2004; Mazerbourg et al., 2004; Nemoto and Finkel, 2004).

The main functions of the insulin-like growth factor binding proteins are presented below and summarized in Table 1.1

Functions of the IGFBPs:

- limit bioavailability of free IGFs to bind IGF receptors
- regulate transport of IGFs between intra- and extravascular spaces
- prolong the half-life of IGFs in circulation
- enhance actions of IGFs by forming a slow-releasing pool of IGFs
- affect cellular proliferation/death via IGFBP receptors
- nuclear actions
- prevent IGF-induced hypoglycaemia

<table>
<thead>
<tr>
<th></th>
<th>IGFBP-1</th>
<th>IGFBP-2</th>
<th>IGFBP-3</th>
<th>IGFBP-4</th>
<th>IGFBP-5</th>
<th>IGFBP-6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phosphorylation</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>I</td>
</tr>
<tr>
<td>Proteolysis</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>I</td>
</tr>
<tr>
<td>Glycosylation</td>
<td>-</td>
<td>-</td>
<td>N</td>
<td>N</td>
<td>O</td>
<td>O</td>
</tr>
<tr>
<td>Heparin binding</td>
<td>I</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>IGF-I/IGF-II preference</td>
<td>≈ IGF-II</td>
<td>≈ IGF-II</td>
<td>≈ IGF-II</td>
<td>IGF-II</td>
<td>IGF-II</td>
<td></td>
</tr>
<tr>
<td>IGF-dependent actions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>suppression</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>enhancement</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>IGF-independent actions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>migration (CHO cells)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>induce apoptosis (mouse fibroblasts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>increase cell differentiation (human osteoblasts)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 1.1. IGFBPs characteristics. I indicates that data regarding this characteristics are incomplete or inconclusive. N, N-linked glycosylation; O, O-linked glycosylation (Jones and Clemmons, 1995; modified).

1.1.5 IGFs and cancer

The IGF-I and IGF-II, with potent mitogenic and antiapoptotic effects, have been widely studied for their role in cancer. Both IGFs are strong mitogens for a wide variety of cancer cell lines such as sarcoma, leukaemia, and cancers of the prostate, breast, lung, colon, stomach, esophagus, liver, pancreas, kidney, thyroid, brain, ovary and uterus (LeRoith et al., 1995; Yu and Rohan, 2000; LeRoith and Roberts, 2003). In vitro studies have shown consistently that members of the IGF family not only regulate the growth of various cancer cells but also interact with other cancer-related molecules. In general, IGFs interact synergistically with other mitogenic growth factors and steroids and antagonize the effects of antiproliferative molecules in cancer cells (Yu and Rohan, 2000).

The effects of IGFs on cancer cells are mediated through the IGF-I receptor. Eliminating IGF-IR from the cell membrane, blocking the interaction of IGFs with IGF-IR, or interrupting the signal transduction pathway of IGF-IR can abolish the mitogenic action of IGFs (LeRoith et al., 1995; Yu and Rohan, 2000). IGF-IR also plays a critical role in cell transformation that is induced by tumor-virus proteins and oncogene products. Moreover, this receptor is involved not only in the induction of cell transformation but also in the maintenance of the transformed phenotype (LeRoith et al., 1995). A number of studies strongly suggest that the IGF-binding proteins family plays a crucial role in the regulation of human cancer. In most situations, the binding proteins suppress the mitogenic action of IGFs and promote apoptosis (Yu and Rohan, 2000). IGFBP-3 has been found to inhibit cell growth and induce apoptosis in many experimental systems also via IGF-independent manner (Gill et al., 1997; Rajah et al., 1997). It should be emphasized that because of the presence of IGFBP proteases, some in vitro studies have shown that BPs are able to stimulate the growth of cancer cells (Camacho-Hubner et al., 1991; Chen et al., 1994).

In summary, the IGF system plays a central role in many aspects of tumorigenesis. A better understanding of this complex system will facilitate the development of novel approaches to diagnose and treat the various human cancers.
1.2 Aim of the project

The main goal of the project was to find a mutant of truncated form of IGFBP-5, which would have a higher affinity for IGF-I than the native protein. Based on the knowledge of the mini-IGFBP-5 structure, several mutants have been designed, purified and subsequently analyzed in context of the IGF-mini-IGFBP-5 interactions. This project was also aimed at investigations of the *in vitro* protein production by means of the rapid translation system (RTS). In this study, IGFBP-4 was expressed with the different RTS, and the further attempts were focused on the optimisation of the reaction conditions in order to obtain protein, which would be soluble and active. The results were verified by the standard biochemical methods (SDS PAGE electrophoresis, immunoblotting) as well as by more advanced techniques (isothermal titration calorimetry, surface plasmon resonance, NMR).
1.3 Materials and methods

1.3.1 Materials

1.3.1.1 *E. coli* strains

Cloning strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>XL1-Blue</td>
<td>Stratagene (USA)</td>
</tr>
<tr>
<td>One Shot® TOP10</td>
<td>Invitrogen (Holland)</td>
</tr>
<tr>
<td>DH5α</td>
<td>Novagen (Canada)</td>
</tr>
</tbody>
</table>

Protein expression strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>One Shot® BL21 Star™ (DE3)</td>
<td>Invitrogen (Holland)</td>
</tr>
<tr>
<td>One Shot® BL21 Star™ (DE3) pLysS</td>
<td>Invitrogen (Holland)</td>
</tr>
</tbody>
</table>

1.3.1.2 Plasmids

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>pET-28a(+)</td>
<td>Novagen (Canada)</td>
</tr>
<tr>
<td>pIVEX 2.4 bNde</td>
<td>Roche (Germany)</td>
</tr>
<tr>
<td>pIVEX 2.2 D</td>
<td>Roche (Germany)</td>
</tr>
<tr>
<td>pIVEX 2.3-MSC</td>
<td>Roche (Germany)</td>
</tr>
<tr>
<td>pIVEX-GST</td>
<td>Roche (Germany)</td>
</tr>
</tbody>
</table>

1.3.1.3 Antibiotics

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin, sodium salt</td>
<td>Sigma (USA)</td>
</tr>
<tr>
<td>Kanamycin, monosulfate</td>
<td>Sigma (USA)</td>
</tr>
</tbody>
</table>

1.3.1.4 Cell growth media and stocks

LB medium

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto tryptone</td>
<td>10 g/l</td>
</tr>
<tr>
<td>Bacto yeast extract</td>
<td>5 g/l</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>10 g/l</td>
</tr>
</tbody>
</table>

pH was adjusted to 7.0. For the preparation of agar plates the medium was supplemented with 15 g agar.
TB medium

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacto tryptone</td>
<td>12 g/l</td>
</tr>
<tr>
<td>Bacto yeast extract</td>
<td>24 g/l</td>
</tr>
<tr>
<td>Sodium chloride</td>
<td>10 g/l</td>
</tr>
<tr>
<td>Glycerol</td>
<td>4 ml</td>
</tr>
<tr>
<td>Deionized water</td>
<td>900 ml</td>
</tr>
</tbody>
</table>

The medium was autoclaved, cooled to 60°C, 100 ml sterile K phosphate and glucose were added. The final concentration of glucose was 1 %

K phosphate

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>KH$_2$PO$_4$</td>
<td>23.1 g/l</td>
</tr>
<tr>
<td>K$_2$HPO$_4$</td>
<td>125.4 g/l</td>
</tr>
</tbody>
</table>

Stock solution of glucose

20 g glucose were dissolved in distilled water to the final volume of 100ml and autoclaved.

Stock solution of ampicillin

100 mg/ml of ampicillin were dissolved in distilled water. The stock solution was sterile filtrated and stored in aliquots at -20 until used.

Stock solution of kanamycin

50 mg/ml of kanamycin were dissolved in distilled water. The stock solution was sterile filtrated and stored in aliquots at -20 until used.

Stock solution of IPTG

A sterile 1 M stock of IPTG in distilled water was prepared and stored in aliquots at -20 until used.

1.3.1.5 Common buffers

**Buffer P(0)**

- 8 mM KH$_2$PO$_4$
- 16 mM Na$_2$HPO$_4$
- 0.05% NaN$_3$
- pH 7.0

**Buffer P(1000)**

- 8 mM KH$_2$PO$_4$
- 16 mM Na$_2$HPO$_4$
- 1 M NaCl
- 0.05% NaN$_3$
- pH 7.0
### Buffer T

(thrombin cleavage buffer)

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>60 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>60 mM KCl</td>
<td></td>
</tr>
<tr>
<td>2.5 mM CaCl₂</td>
<td></td>
</tr>
<tr>
<td>50 mM Tris</td>
<td></td>
</tr>
<tr>
<td>0.05% NaN₃</td>
<td></td>
</tr>
<tr>
<td>pH 8.0</td>
<td></td>
</tr>
</tbody>
</table>

### PBS

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>140 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>2.7 mM KCl</td>
<td></td>
</tr>
<tr>
<td>10 mM Na₂HPO₄</td>
<td></td>
</tr>
<tr>
<td>1.8 mM KH₂PO₄</td>
<td></td>
</tr>
<tr>
<td>0.05% NaN₃</td>
<td></td>
</tr>
<tr>
<td>pH 7.3</td>
<td></td>
</tr>
</tbody>
</table>

### HBS-EP

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>10 mM Hepes</td>
<td></td>
</tr>
<tr>
<td>150 mM NaCl</td>
<td></td>
</tr>
<tr>
<td>3 mM EDTA, pH 7.4</td>
<td></td>
</tr>
<tr>
<td>0.005% surfactant P20</td>
<td></td>
</tr>
</tbody>
</table>

**Surfactant P20**

Polyoxyethylenesorbitan, a nonionic surfactant recommended for inclusion in the buffers used in BIAcore systems.

### 1.3.1.6 Buffers for purification of the proteins under denaturing conditions

#### Buffer A

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 M guanidinium chloride</td>
<td></td>
</tr>
<tr>
<td>100 mM NaH₂PO₄ x H₂O</td>
<td></td>
</tr>
<tr>
<td>10 mM Tris</td>
<td></td>
</tr>
<tr>
<td>10 mM β-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td>pH 8.0</td>
<td></td>
</tr>
</tbody>
</table>

#### Buffer B

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>6 M guanidinium chloride</td>
<td></td>
</tr>
<tr>
<td>100 mM NaH₂PO₄ x H₂O</td>
<td></td>
</tr>
<tr>
<td>10 mM Tris</td>
<td></td>
</tr>
<tr>
<td>10 mM β-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td>pH 6.5</td>
<td></td>
</tr>
</tbody>
</table>
Buffer C
6 M guanidinium chloride
100 mM NaAc x 3H₂O
10 mM β-mercaptoethanol
pH 4.0

Buffer D
6 M guanidinium chloride
pH 3.0

Buffer E
200 mM arginine
1 mM EDTA
100 mM Tris
2 mM red GSH
2 mM ox GSH
0.05% NaN₃
pH 8.4

1.3.1.7 Buffers for purification of the proteins under native conditions

Binding buffer
50 mM NaH₂PO₄
300 mM NaCl
10 mM imidazole
pH 8.0

Wash buffer
50 mM NaH₂PO₄
300 mM NaCl
20 mM imidazole
pH 8.0

Elution buffer
50 mM NaH₂PO₄
300 mM NaCl
250 mM imidazole
pH 8.0
1.3.1.8 Enzymes, antibodies and other proteins

BSA       New England BioLabs (USA)
Pfu turbo DNA Polymerase     Stratagene (USA)
Pwo DNA Polymerase     Roche (Germany)
BamHI       New England BioLabs (USA)
KspI       New England BioLabs (USA)
NdeI       New England BioLabs (USA)
StuI       New England BioLabs (USA)
DpnI       Stratagene (USA)
Xa factor       Sigma (USA)
Thrombin       Sigma (USA)
Thrombin     Roche (Germany)
BirA       Roche (Germany)
Ab-IGFBP-4     Roche-Penzberg
hIGF-I receptor grade     GroPep (Australia)

1.3.1.9 Kits and reagents

QIAquick PCR Purification Kit     Qiagen (Germany)
QIAprep Spin Miniprep Kit     Qiagen (Germany)
QIAGEN Plasmid Midi Kit     Qiagen (Germany)
QIAGEN Plasmid Maxi Kit     Qiagen (Germany)
QuikChange Site-Directed Mutagenesis Kit     Stratagene (USA)
Rapid Ligation Kit     Roche (Germany)
BM chemiluminescence Western Blotting Kit (Mouse/Rabbit)     Roche (Germany)
RTS100HY     Roche (Germany)
RTS500     Roche (Germany)
RTS500HY     Roche (Germany)
RTS9000     Roche (Germany)
RTS GroE Supplement     Roche (Germany)
Complete Protease Inhibitor Cocktail     Roche (Germany)
NuPAGE Bis-Tris pre-cast gels     Invitrogen (Holland)
Chapter 1  
Biochemical studies on IGF and IGF-binding proteins interactions

NuPAGE MES SDS Buffer Kit  
Invitrogen (Holland)

1.3.1.10 Protein and nucleic acids markers

Mark12 Standard for SDS-PAGE  
Invitrogen (Holland)

Prestained Protein Marker, Broad Range (6-175 kDa)  
New England BioLabs (USA)

1 kb DNA-Leiter  
Peqlab (Germany)

1.3.2 General laboratory methods

1.3.2.1 Transformation of *E. coli*

1.3.2.1.1 Transformation by heat shock

2 µl of a ligation mix or 1 µl of plasmid DNA were added to 50 µl of chemically competent cells. The mixture was incubated on ice for 30 min followed by a heat shock of 30 s at 42°C, short cooling on ice, and addition of 250 µl SOC medium. After 1 h of incubation at 37°C, 20-50 µl of the mixture were spread out on LB agar plates including selective antibiotic and incubated overnight at 37°C.

1.3.2.1.2 Transformation by electroporation

1 µl of plasmid DNA was added to 50 µl of electrocompetent cells, the mixture was pipetted into a 2 mm electroporation cuvette. The electroporation was performed in an electroporation vessel (Gene pulser) at 1650 V. Then the suspension was transferred into an Eppendorf tube and mixed with 1 ml LB medium. After 1 h of incubation at 37°C, 20-50 µl of the mixture were spread out on LB agar plates including selective antibiotic and incubated overnight at 37°C.

1.3.2.2 Preparation of plasmid DNA

The isolation of plasmid DNA from *E. coli* was done with assistance of plasmid kits offered by Qiagen company. The preparation of plasmid DNA in a small scale (up to 20 µg) was performed to check the successful cloning. Larger amounts of DNA (up to 500 µg) were needed to RTS. Both types of preparation were carried out following the instructions of the manufacturer (Qiagen Plasmid Mini Kit and Plasmid Maxi Kit protocol, respectively).
1.3.2.3 Digestion with restriction enzymes

Usually, 1-2 units of restriction enzyme were employed per µg DNA to be digested. The digestion was performed in the buffer specified by the manufacturer at the optimal temperature (37°C) overnight. The fragment ends that occurred after digestion were cohesive or blunt ends.

1.3.2.4 Purification of DNA fragments

DNA obtained from restriction digestion, phosphatase treatment or PCR was purified from primers, nucleotides, enzymes, mineral oil, salts, agarose, ethidium bromide, and other impurities using silica-gel column (QIAquick PCR Purification Kit, Qiagen). The QIAquick system uses a simple bindwash-elute procedure. Binding buffer was added directly to the PCR sample or other enzymatic reaction, and the mixture was applied to the spin column. Nucleic acids adsorbed to the silica-gel membrane in the high-salt conditions provided by the buffer. Impurities were washed away and pure DNA was eluted with a small volume of low-salt buffer provided or water, ready to use in all subsequent applications.

1.3.2.5 DNA agarose gel electrophoresis

To verify the DNA samples, agarose gel electrophoresis was performed. For this purpose 1% agarose in TBE buffer plus ethidium bromide was prepared. The solution was poured into a horizontal gel chamber to cool down. The DNA samples were mixed with sample buffer and loaded into the gel’s pockets. Electrophoresis was carried out at 50-100 V. Results evaluation was done by UV illuminator.

1.3.2.6 Sonication

Sonication is a simple method used for the disruption cells by ultrasounds. The bacterial suspension was filled into the pre-cooled glass and the suitable pulse of ultrasounds was applied (output control 7.5, 50%). To avoid overheat of the sample sonication was carried out in two steps of 5 min each with 5 min intervals between steps and on ice.

1.3.2.7 SDS polyacrylamide gel electrophoresis (SDS PAGE)

To verify the protein samples, SDS polyacrylamide gel electrophoresis was done. Because small proteins were analyzed, tricine gels were chosen to use (Schagger and von Jagow, 1987).
Chapter 1: Biochemical studies on IGF and IGF-binding proteins interactions

**Reagents:**

Anode Buffer (+): 200 mM Tris pH 8.9

Cathode Buffer (-):
- 100 mM Tris pH 8.25
- 100 mM tricine
- 0.1% SDS

Separation Buffer:
- 1 M Tris pH 8.8
- 0.3% SDS

Stacking Buffer:
- 1 M Tris pH 6.8
- 0.3% SDS

Separation acrylamide:
- 48% acrylamide
- 1.5% bis-acrylamide

Stacking acrylamide:
- 30% acrylamide
- 0.8% bis-acrylamide

**Pouring gels**

Separation gel:
- 1.675 ml H₂O
- 2.5 ml separation buffer
- 2.5 ml separation acrylamide
- 0.8 ml glycerol
- 25 µl APS
- 2.5 µl TEMED

Intermediate gel:
- 1.725 ml H₂O
- 1.25 ml separation buffer
- 0.75 ml separation acrylamide
- 12.5 µl APS
- 1.25 µl TEMED
Stacking gel:  
- 2.575 ml H₂O
- 0.475 ml stacking buffer
- 0.625 ml stacking acrylamide
- 12.5 µl 0.5 M EDTA, pH 8.0
- 37.5 µl APS
- 1.9 µl TEMED

The guanidinium HCl-free protein samples were prepared by mixing 20 µl of protein solution with 5 µl of sample buffer (SB) followed by 3 min incubation at 100 °C. Due to the rapid precipitation of SDS in contact with guanidine, the samples to be examined by PAGE after Ni-NTA chromatography under denaturing conditions had to be prepared in a following fashion: 20 µl of the protein solution in a denaturing buffer was diluted with 400 µl 20% trichloroacetic acid (TCA). The sample was incubated for 5 min at room temperature followed by centrifugation for 5 min at 20 000 x g. Supernatant was discarded by suction, precipitated protein pellet was washed once by vortexing with 400 µl ethanol. After centrifugation and ethanol removal, protein pellet was resuspended in 20 µl of 2x SB and the sample was boiled for 3 min.

1.3.2.8 Staining of proteins
Staining of proteins with Coomassie-Blue solution was performed as described in Sambrook and Russell (2001).

Coomassie-Blue solution:  
- 0.025% Coomassie Brillant Blue R250
- 45% ethanol
- 10% acetic acid

Destaining solution:  
- 5% ethanol
- 10% acetic acid

1.3.2.9 Determination of protein concentration
The concentration of proteins in solution was determined with assistance of the Bradford reagent (BioRad). For this purpose, 5 µl of the protein sample was added to 1 ml (10 x diluted) Bradford reagent into a plastic cuvette. After gentle mixing, A₅₉₅ was measured and converted to the protein concentration on the basis of a BSA calibration curve.
To estimate the concentration of proteins more precisely, $A_{285}$ was measured and converted to the protein concentration on the basis of extinction coefficients. It has been demonstrated (Gill and von Hippel, 1989) that it is possible to estimate the molar extinction coefficient of a protein from knowledge of its amino acid composition. From the molar extinction coefficient of tyrosine, tryptophan and cystine (cysteine residues do not absorb appreciably at wavelengths $>260$ nm, while cystine does) at a given wavelength the extinction coefficient of a denaturated protein can be computed using the equation:

$$E(Prot) = \text{Numb(Tyr)} \times \text{Ext(Tyr)} + \text{Numb(Trp)} \times \text{Ext(Trp)} + \text{Numb(Cystine)} \times \text{Ext(Cystine)}$$

The absorbance (optical density) can be calculated using the following formula:

$$\text{Absorb(Prot)} = \frac{E(Prot)}{\text{Molecular weight}}$$

The conditions at which these equations are valid are: 6.0 M guanidium hydrochloride, 0.02 M phosphate buffer, pH 6.5.

### 1.3.3 Mutagenesis of the IGF-miniBP5

Mutagenesis procedure was performed by means of QuikChange Site-Directed Mutagenesis Kit (Stratagene). The mutagenic oligonucleotide primers were designed according to suggestions provided by manufacturer: they possessed 30-32 bases in length, their melting temperatures were greater than or equal to 78°C, they have a minimum GC content of 40% and terminated in one or more C or G bases. The desired mutation was in the middle of the primer with ~10-15 bases of correct sequence on both sides.

Mini-IGFBP-5 (residues 40-92 of IGFBP-5) subcloned into pET 28a vector was used as a DNA template. Amino acid sequence of the construct is presented below.

```
```

Amino acids of the His-T7-tag, including the thrombin-cleavage sites (underlined), are in italics. Mini-IGFBP-5 after cleavage by thrombin comprises the amino acid sequence denoted in capital letters.
Introduced mutation and primers:

V49 => I  GTC => ATT, ATG, ATA
V49IF  5’- GGG CAG TCG TGC GGC ATC TAC ACC GAG CGC - 3’
V49IR  5’- GCG CTC GGT GTA GAT GCC GCA CGA CTG CCC - 3’

Y50 => K  TAC => AAA, AAG
Y50KF  5’- GCA GTC GTG CGG CGT CAA AAC CGA GCG CTG CG - 3’
Y50KR  5’- CGC AGC GCT CGG TCT TGA CGC CGC ACT GC - 3’

Y50 => A  TAC => GCT, GCC, GCA, GCG
Y50AF  5’- GCA GTC GTG CGG CGT CAC CGA GCG CTG CG - 3’
Y50AR  5’- CGC AGC GCT CGG TGC TGC GCC CGC ACT GC - 3’

R53 => Y  CGC => TAT, TAC
R53YF  5’- GGC GTC TAC ACC GAG TAC TGC GCC CAG GGG - 3’
R53YR  5’- CCC CTG GGC GCA GTA CTC GGT GTA GAC GCC - 3’

L73 => W  CTG => TGG
L73WF  5’- GCC GCT GCA CGC CTC GCT GCA CGG CCG CGG - 3’
L73WR  5’- CCG CGG CCG TGC AGC CAG GCC TGC AGC GGC - 3’

Mutant strand synthesis reactions were carried out using PCR (Mullis and Faloora, 1987) as described below:

**Sample reaction:**

5 µl of 10x reaction buffer
50 ng of dsDNA template
125 ng of primer F
125 ng of primer R
1 µl of dNTP mix
dd H₂O to a final volume of 50 µl
then 1 µl of Pfu Turbo DNA polymerase was added
### PCR conditions:

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>95°C</td>
<td>1 min</td>
</tr>
<tr>
<td>95°C</td>
<td>30 s</td>
</tr>
<tr>
<td>55°C</td>
<td>1 min</td>
</tr>
<tr>
<td>68°C</td>
<td>12 min</td>
</tr>
</tbody>
</table>

After PCR, each amplification reaction was incubated with Dpn I restriction enzyme (37°C, for 2 h) to digest the parental supercoiled dsDNA. Plasmid DNA was isolated using QIAprep Spin Miniprep Kit (Qiagen) and was subjected to verification by automated sequencing.

### 1.3.4 Protein expression, refolding and purification

One Shot® BL21 Star™ (DE3) chemically competent cells were transformed with the plasmid DNA (mini-IGFBP-5 mutant in pET 28a) according to the protocol described in section 1.3.2.1.1. Growth media contained kanamycin as a selective antibiotic. Typically, the cells were induced at OD$_{600}$ = 0.8-0.9 by addition of IPTG (1 mM final concentration) and protein expression was performed for next 3 hours with vigorous shaking at 37°C. The culture was harvested by centrifugation (5500 x g, 20 min, 4°C) subsequently bacterial pellets were resuspended in buffer A and left for solubilisation at room temperature overnight. Note: compositions of all used buffers were described in section 1.3.1.6. The cell suspension was sonicated 2 x 6 minutes using macrotip (output control 7.5, 50%) and centrifuged at 60000 x g for 1 h at 20°C. Obtained supernatant was incubated with Ni-NTA slurry (Qiagen), equilibrated previously in buffer A, for 1 h at room temperature with gentle agitation. Next, the mixture was loaded onto an empty column and washed with buffer A and B. The protein was eluted with buffer C. The fractions containing desirable protein were pooled, concentrated and dialysed against buffer D to reduce all disulfide bridges and get rid of reducing agents without causing reoxidation of cysteines. Subsequently, refolding of the protein was performed. For this purpose, protein sample was diluted in buffer E (1:50) in ~200 µl aliquots, with 10-20 minutes intervals between aliquots. Thanks to the gradual dilution, the accumulation of high concentration of folding intermediates was avoided, which could otherwise lead to aggregation. The refolding mixture was left with stirring at 4°C. After 3 days, the mixture was concentrated and dialysed against buffer P(0) and then was loaded onto cation exchange column (MonoS). The protein of interest was eluted with a gradient by buffer P(1000), verified by SDS PAGE.
Chapter 1  
Biochemical studies on IGF and IGF-binding proteins interactions

pooled and dialysed against thrombin cleavage buffer. Thrombin cleavage was carried out as presented below.

- 1 µl enzyme (~25 units) per 1 mg protein was needed to cut the His-T7-tag
- sample was left at RT with gentle agitation for 2 days
- digestion efficiency was investigated by SDS PAGE

Finally, proteins of interest were separated from the HisTag by gel filtration on the Superdex S75 prep grade 26/60XR column. The buffer used contained 140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 0.05% NaN₃, pH 7.3. To determine the purity of the samples, SDS-PAGE was performed. Molecular weights of miniBP5 mutants were subjected to verification by mass spectrometry (MALDI).

1.3.5 1D-¹H NMR

1D-proton NMR experiment was performed in order to verify the correct folding of the miniBP5 mutants. All spectra were recorded at 300K on the Bruker DRX600 spectrometer. The protein samples were concentrated and dialyzed against PBS buffer and subsequently centrifuged to clear out potential aggregates. Subsequently, 450 µl of the protein solution was mixed with 50 µl of D₂O and transferred to an NMR tube.

1.3.6 Binding on the column

Primary binding tests, which verified the interactions between IGF-I and IGF-mini BP5 mutants were performed by means of size exclusion chromatography. For this purpose equimolar amounts of the components were mixed and loaded on the Superdex S75 column. The buffer used contained 140 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 0.05% NaN₃, pH 7.3.

1.3.7 Kinetic analysis

Measurement of the binding interaction between IGF and IGF-miniBP5 mutants was performed using two methods: surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC).
1.3.7.1 SPR principles

In a basic SPR biosensor experiment, one reactant, referred to as the ligand (B) is attached to the sensor surface. The other reactant, referred to as the analyte (A), flows past this surface in solution. When the analyte and the ligand interact to form a complex (AB), a response is generated. Figure below shows schematically the interactions that occur at the sensor surface. Because binding responses are recorder in real time, it is possible to interpret kinetic information about the interaction (Myszka, 1997).

Figure 1.4. Schematic diagram shows basic interaction at the biosensor surface. The ligand (B) is attached to the surface via a flexible linker. During the association phase, analyte (A) is flowed past the surface. $k_m$ is the mass transport coefficient used to describe the diffusion of analyte through the unstirred solvent layer. $k_a$ and $k_d$ are the reaction rate constants, association and dissociation respectively, which describe the formation of the complex (AB) (Myszka, 1997).

1.3.7.2 BIAcore analysis

In IGF-miniBP5 mutants studies BIAcore 3000 instrument (BIAcore Inc., Melbourne, Australia) was employed. All experiments were performed at 25°C and HBS-EP was used as a running buffer and for dilution of analytes. Coupling of hIGF-I (receptor grade) to CM5 BIAsensor chip was achieved using standard coupling procedure (Lofas and Johnsson, 1990). Subsequently, kinetic studies with different mini-IGFBP-5 constructs were carried out.
Additionally, a range of analyte concentrations (0.78-100 nM) was tested. After the dissociation phase, the hIGF-I-coated biosensor chip was successfully regenerated with 100 mM HCl and washed with HBS-EP buffer.

1.3.7.3 ITC principles

Isothermal titration calorimetry (ITC) is a thermodynamic technique for monitoring any chemical reaction initiated by the addition of a binding component, and has become the method of choice for characterizing biomolecular interactions. When substances bind, heat is either generated or absorbed. Measurement of this heat allows accurate determination of binding constants ($K_B$), reaction stoichiometry ($n$), enthalpy ($\Delta H$) and entropy ($\Delta S$), thereby providing a complete thermodynamic profile of the molecular interaction in a single experiment. In ITC experiment, a syringe containing a “ligand” solution is titrated into a cell containing a solution of the “macromolecule” at constant temperature. When ligand is injected into the cell, the two materials interact, and heat is released or absorbed in direct proportion to the amount of binding. As the macromolecule in the cell becomes saturated with ligand, the heat signal diminishes until only background heat of dilution is observed.

1.3.7.4 ITC experiments

In IGFBP’s studies, the VP-ITC system (MicroCal) was used. It is very sensitive isothermal titration calorimeter that uses a cell feedback network (CFB) to differentially measure heat produced or absorbed between the sample and reference cell. Twin coin-shaped cells are mounted in a cylindrical adiabatic environment, and connect to the outside through narrow access tubes (Figure 1.5). A thermoelectric device measures the temperature difference between the two cells and a second device measures the temperature difference between the cells and the jacket. As chemical reactions occur in the sample cell, heat is generated or absorbed. The temperature difference between the sample and reference cells ($\Delta T_1$) is kept at zero by the addition of heat to the sample or reference cell, as appropriate, using the CFB system. The integral of the power required to maintain $\Delta T_1 = 0$ over time is a measure of total heat resulting from the process being studied.
All experiments were carried out according to references provided by the manufacturer. Proteins were used at 0.2 mM in PBS (140 mM NaCl, 2.7 mM KCl 10 mM Na₂HPO₄, 1.8 mM KH₂PO₄, 0.05% NaN₃, pH 7.3) and titrated from a 300 µl syringe into a sample chamber holding 1.43 ml of 0.018 mM respective binding partner (IGF-I). All solutions were degassed prior to measurements. Heat generated by protein dilution was determined in separate experiments by injecting protein solution into PBS filled sample chamber. All data were corrected for the heat of protein dilution. The details of the experimental and injection parameters are described below.

**Experimental parameters:**
- total number of injections: 58
- cell temperature [°C]: 20
- initial delay [sec]: 60
- reference power [µCal/sec]: 15
- syringe concentration [mM]: 0.2096
- cell concentration [mM]: 0.018
- stirring speed: 270

**Injection parameters:**
- volume [µl]: 5
- duration [sec]: 10
- spacing [sec]: 450
- filter period [sec]: 2

All steps of the data evaluation were performed using ORIGIN(V5.0) software provided by the manufacturer.
1.3.8 Crystallization trials

Protein samples (miniBP5 wild type and miniBP5 Y50K mutant) were prepared for crystallization trials in the following manner. Both proteins were expressed and purified as described before. In the last step of purification procedure desirable proteins were separated from any excess of either protein by gel filtration on the Superdex S75 prep grade 26/60XR column. The buffer used contained 10 mM Tris, pH 7.3, 50 mM NaCl and 0.1% NaN3. Collected fractions were pooled and concentrated until the expected protein concentration was achieved (5-30 mg/ml). Subsequently, the samples were centrifuged to clear out small particles and precipitate. The crystallization procedure was performed as quickly as possible (usually directly after concentration of the proteins). Remaining protein samples were stored at -80°C. To determine the preliminary crystallization conditions the Hampton Research Crystal Screens were employed. Crystallization of the miniBP5 WT and Y50K was carried out with the sitting drop vapor diffusion method. 0.5 ml of crystallization buffer was pipeted into each reservoir of the Cryschem 24 well plate (Hampton Research) then 1 µl of protein sample and 1 µl of crystallization buffer were pipeted into the post of each reservoir. The drops were not mixed. Crystallization was carried out at 4°C.

1.3.9 In vitro translation system as a high throughput method for screening of the proteins

For the last decade, bacterial in vitro systems for high-throughput protein expression have been considered as an alternative for the standard cell-based expression of proteins. However, the idea that cell lysates could be used to produce protein in vitro appeared much earlier, and was presented by Nirenberg and Matthaei in 1961. That system was mostly used for the analysis of the genetic codon and was modified by Zubay (1973) to facilitate the direct expression of an exogenous gene. Two different types of cell-free expression systems, namely static and continuous, are used presently (Luke, 2004). In the static system, transcription and translation take place in a single reaction vessel until the substrates and energy components essential for the transcription/translation machinery are available. When the reagents are used, the reaction stops. On the contrary, in the continuous system, all needed reagents for sustained protein expression are continuously supplied from the so-called feeding solution across a semi-
permeable membrane. Rapid translation system (RTS) used in these studies was developed by Spirin et al. (1988) and belongs to the continuous cell-free system.

**1.3.9.1 Principles of Rapid Translation System (RTS)**

RTS is a scalable in vitro transcription/translation protein expression system that produces sufficient amounts of protein for further characterization studies, functional assays, or structural analysis. The main advantage of the RTS is using the Continuous-Exchange Cell Free technology (CECF), which eliminates the need for laborious up- and down-stream steps (e.g., host cell transformation, culturing, or lysis) typically associated with cell-based expression systems. Moreover, in contrast to standard coupled transcription/translation reactions in which by-products ultimately inhibit the reaction, CECF technology by employing a semi-permeable membrane, provides continuous removal of by-products and a steady supply of substrates necessary to sustain the reaction, resulting in high levels of protein synthesis. Coupled transcription/translation occurs in the reaction chamber, where the membrane retains high-molecular-weight species, such as the DNA template, T7 RNA polymerase, ribosomes, tRNAs and the protein product. At the same time, substrates - nucleoside triphosphates, amino acids and energy substrates - continuously diffuse from the feeding compartment into the reaction compartment. Likewise, reaction by-products, such as phosphate, nucleoside mono- and diphosphates and other low-molecular-weight species, diffuse into the larger feeding compartment (Roche, RTS manual)

![Figure 1.6. Schematic illustration of CECF protein expression (Roche, RTS500 manual)](image-url)
The main advantage of this cell-free system is its open character allowing direct manipulation of the reaction conditions and applications that are impossible in the standard \textit{in vivo} systems. First of all the cell-free protein synthesis enables addition of detergents, chaperones, and appropriate ligands to the reaction chamber, that may help in proper folding of the protein and shows promise for obtaining soluble and functional proteins. Moreover, the protein production can be monitored during the RTS by taking samples from the reaction chamber without the risk of aggravation of protein expression level. Such a safe monitoring can help in optimisation of the reaction conditions. Other properties of the cell-free system including the fast expression of the protein (Betton, 2003; Luke, 2004; Murthy \textit{et al.}, 2004), the ability to express toxic proteins that cannot be produced \textit{in vivo} in bacterial and insect cells (Renesto and Raoult, 2003) and the ability to incorporate non-natural amino acids (i.e. selenomethionine for X-ray analysis, and labeled amino acids for NMR studies (Kigawa \textit{et al.}, 1995; Hirao \textit{et al.}, 2002; Gite \textit{et al.}, 2003; Traverso \textit{et al.}, 2003), cause the systems based on bacterial lysate to become the method of choice for many investigators. During last years a number of publications concerning this subject have rapidly risen, suggesting a significant increase of interest in these systems.

\textbf{1.3.9.2 Mono-biotinylation of the proteins using RTS}

Labelling proteins with biotin is a common method to obtain a system for detection, purification and immobilization with streptavidin (SA) -based reagents. As alternative for chemical labelling (e.g., with biotin-NHS-esters) enzymatic labelling offers the possibility to introduce single biotin residues at defined positions in a protein. This can be accomplished by introducing defined peptide tags, which contain the specific recognition and biotinylation site for protein biotin ligase (BirA, EC 6.3.4.15). In an ATP-dependent process, this enzyme activates biotin to form biotinyl 5' adenylate and transfers biotin to the $\varepsilon$-amino group of a lysine residue within the specific peptide tag called AviTag (Avidity LLC, Denver, USA). The RTS system thanks to its open character employs this tag in combination with the \textit{E. coli} biotin protein ligase, biotin and ATP to achieve this.
Figure 1.7. AviTag principles. The 15 amino acid AviTag-peptide sequence is introduced into the protein by fusing the coding sequence to the gene of interest by conventional or PCR-based cloning methods. Addition of the template to a RTS reaction will lead to the translation of the protein with the AviTag fused to the C- or N-terminus. If the BirA enzyme, biotin and ATP are added to the expression reaction, biotinylation of this sequence takes place during the expression reaction (Roche, RTS manual).

1.3.9.3 Expression vectors

According to recommendation provided by manufacturer, the pIVEX vector family was employed in RTS experiments. Any vector from this family includes the following elements and structural features:

- target gene under control of T7 promoter located downstream of a RBS sequence
- distance between T7 promoter and start ATG should not exceed 100 bp
- distance between the RBS sequence and start ATG should not be longer than 5-8 bp
- T7 terminator sequence at the 3’ end of the gene

Additionally, each pIVEX vector contains at least one fusion tag sequence to allow further easy detection and purification of the expressed protein (Figure 1.8)
Figure 1.8. Scheme of the pIVEX-GST vector. T7P, T7 promoter; RBS, ribosomal binding site; His₆-Tag, hexa histidine tag; GST, glutathione S-transferase tag; XA, factor Xa protease cleavage site; MCS, multiple cloning site for the insertion of the target gene; T7T, T7 terminator (Roche, RTS manual)

1.3.9.4 Cloning of human IGFBP4 into pIVEX vectors

DNA sequence encoding full-length IGFBP-4 (residues 1-237) was cloned into four differed in tags pIVEX vectors: pIVEX 2.4b Nde (HisTag), pIVEX 2.2D (StrepTag), pIVEX 2.3-MSC (AviTag) and pIVEX-GST. Restriction sites for the following enzymes were introduced into IGFBP-4 gene by means of the standard PCR procedure.

KspI and BamHI =\> constructs: pIVEX_His-IGFBP-4
pIVEX_Strep-IGFBP-4

StuI and BamHI =\> construct: pIVEX_GST-IGFBP-4

PCR reaction mix:
1 µl of DNA template
2 µl of dNTP mix (10 mM)
125 ng of reverse primer (R)*
125 ng of foreword primer (F)*
10 µl of reaction buffer
ddH₂O to a final volume of 50 µl
then 1µl of Pwo DNA polymerase was added

*- full list of primers used in the work is present in Table 1.2.
Reaction conditions:

94°C  2 min
94°C  1 min
52°C/48°C  2 min       30 cycles
72°C  3 min

Restriction sites for NdeI and BamHI were introduced into IGFBP-4 gene together with the factor Xa cleavage site and Avi-Tag by means of two-step PCR procedure (Figure 1.9). In the first PCR (performed as described above) cleavage sites for factor Xa and BamHI were introduced. DNA amplified in this reaction was used as a template for a second PCR in which four different primers were employed:

F1: Avi-Tag-coding, containing NdeI restriction site
F2: Avi-Tag-coding template
R1: Reverse primer for gene amplification
R2: Avi-Tag-coding template

Vectors were cut with appropriate restriction enzymes and then were dephosphorylated with CIP enzyme. Vectors as well as inserts (gene of interest) were purified using QIAquick PCR purification kit (Qiagen). Subsequently, ligation of fragments to be cloned and vectors was performed using Rapid Ligation Kit (Roche Diagnostics GmbH). Protocol for the ligation reaction is presented below.
Biochemical studies on IGF and IGF-binding proteins interactions

vector (the molar ratio of insert to vector should be 1+3)
insert
2 µl of DNA dilution buffer
H₂O (to a final volume of 10 µl)
10 µl of T4 DNA ligation buffer
1 µl of T4 DNA ligase
The mixture was incubated for 5 min at room temperature.

To avoid the vector religation during ligation reaction, dephosphorylation at the 5’-end of plasmid DNA with alkaline phosphatase (CIP) was performed. Incubation was carried out for 1 h at 37°C. To verify the existence of inserts of the proper size, control PCR was done. For this purpose, plasmid DNA was isolated by mini-prep from randomly picked colonies. Agarose gel electrophoresis analysis (for details see 1.3.2.5) revealed presence of the insert in all tested colonies. Additionally, each construct was subjected to verification by automated sequencing (ABI PRISM). The plasmid DNA was amplified in a 100 ml LB culture of DH5α E. coli cells and isolated with a plasmid maxi kit (Qiagen). The final concentration of the purified DNA was 1-3 µg/µl.

Table 1.2. List of primers used in the work

<table>
<thead>
<tr>
<th>Name</th>
<th>Construct</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>KspI_F</td>
<td>pIVEX_His-IGFBP-4</td>
<td>CCGGCGCGCGGCACGACGAAGCCATCCACTGC</td>
</tr>
<tr>
<td></td>
<td>pIVEX_Strep-IGFBP-4</td>
<td></td>
</tr>
<tr>
<td>StuI_F</td>
<td>pIVEX_GST-IGFBP-4</td>
<td>GACGAAGCCATCCACTGC</td>
</tr>
<tr>
<td>BamHI_R</td>
<td>pIVEX_His-IGFBP-4</td>
<td>GCCAAGCTTGATCCAGGTCGAC</td>
</tr>
<tr>
<td></td>
<td>pIVEX_Strep-IGFBP-4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>pIVEX_GST-IGFBP-4</td>
<td></td>
</tr>
<tr>
<td>Fxa_F</td>
<td>pIVEX_Avi-IGFBP-4</td>
<td>CATCGAAGGCCGACGAAGCCATCCACTGCC</td>
</tr>
<tr>
<td>BamHI_R2</td>
<td>pIVEX_Avi-IGFBP-4</td>
<td>ATCGGATCCTCATATTACTCTCGAAAGCTGTGAC</td>
</tr>
<tr>
<td>F1</td>
<td>pIVEX_Avi-IGFBP-4</td>
<td>GAAGGCATATGGGTCTGAACG</td>
</tr>
<tr>
<td>F2</td>
<td>pIVEX_Avi-IGFBP-4</td>
<td>CTCAGAAAAATCGAATGGCAGAAAGTCGACATC</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GAAGGCCGCGACGAAGC</td>
</tr>
</tbody>
</table>
### 1.3.9.5 In vitro protein expression

Four different RTS kits (RTS100 HY, RTS500, RTS500 HY and RTS9000) were used to examine the production and solubility of IGFBP-4 constructs. General procedure was carried out as described in the instruction manuals supplied by the manufacturer (Roche). Reaction conditions such as temperature and time have been optimised and these results are comprised in section 1.4. Amounts of the plasmid DNA used in the experiments ranged from 10 to 15 µg. Avi-IGFBP-4 construct was biotinylated during expression in RTS500 and RTS500 HY by adding 2 µl of birA enzyme into reaction mix and 48 µl of biotin: 24 µl into feeding mix and 24 µl into reaction mix.

### 1.3.9.6 Western blot analysis of the RTS products

After in vitro expression the reaction solutions were separated in supernatant and pellet fractions by centrifugation for 15 min at 14000 rpm and then analyzed with the assistance of Western blot. In first step, proteins produced in RTS were transferred from NuPAGE Bis-Tris gels (Invitrogen) onto nitrocellulose membrane (Amersham) as describe below.

3 x 3 Whatman papers cut to the size of the gel were soaked in transfer buffers (1, 2, and 3), additionally the gel and membrane were soaked in buffer 2. Next, the “sandwich” was created with the following layers: filter papers soaked in buffer 3, filter papers soaked in buffer 2, membrane, gel and filter papers soaked in buffer 1. The transfer was carried out 1.10 h at current of 150 mA. Afterwards, the membrane was washed with PBS-tween buffer, blocked for 30 min with 5% milk powder in PBS-tween and then incubated with the primary antibody (polyclonal Ab-BP4) for 1 h with gentle agitation. After incubation the membrane was washed 3 x 10 min with PBS-tween buffer and subsequently, the incubation with POD-labelled secondary antibody (1:10000 in PBS-tween with 2.5% milk powder, for 1.5 h) was performed. In case of Avi-IGFBP-4 construct streptavidin-POD conjugate was used (15 µl in 100 µl of PBS-tween with 2.5 % milk powder, for 1 h) instead of Ab-BP4. After washing (5 x
(5 min) chemiluminescence detection was performed. This step was carried out according to instructions of manufacturer (Roche Diagnostics GmbH).

<table>
<thead>
<tr>
<th>Transfer buffer 1</th>
<th>25 mM Tris pH 9.4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40 mM aminohexanoic acid</td>
</tr>
<tr>
<td></td>
<td>20% isopropanol</td>
</tr>
<tr>
<td>Transfer buffer 2</td>
<td>25 mM Tris pH 10.4</td>
</tr>
<tr>
<td></td>
<td>20% isopropanol</td>
</tr>
<tr>
<td>Transfer buffer 3</td>
<td>0.3 M Tris pH 10.4</td>
</tr>
<tr>
<td></td>
<td>20% isopropanol</td>
</tr>
<tr>
<td>PBS-tween</td>
<td>Standard PBS plus tween tablets (Gibco)</td>
</tr>
</tbody>
</table>

Proteins which have been expressed in RTS and contained HisTag or StrepTag, were further purified using affinity chromatography method: NiNTA column and Strep-Tactin column, respectively.

1.3.9.7 Purification of His-IGFBP-4

His-tagged IGFBP-4 was purified under native conditions directly after RTS reaction, according to protocol provided by company Qiagen. Reaction solution was loaded onto pre-equilibrated NiNTA column. Non-binding proteins were washed away with buffer A and B. Protein of interest was eluted with buffer C and was investigated by gel electrophoresis (NuPage, Invitrogen) as well as by Western blot.

1.3.9.8 Purification of Strep-IGFBP-4

Purification of the IGFBP-4-StrepTag fusion protein was performed with the assistance of Strep-Tactin Sepharose column (IBA GmbH). Directly after RTS reaction, solution consisting desirable protein, was loaded onto pre-equilibrated column. Unspecific proteins were rapidly washed away with small amounts of physiological buffer (100 mM Tris, pH 8, 150mM NaCl). Subsequently, streptavidin-tagged BP4 was eluted due to addition of the specific
competitor "desthiobiotin". The fractions were analyzed by gel electrophoresis as well by Western blot.

1.3.9.9 Surface plasmon resonance experiments

To verify the activity of tested IGFBP-4 constructs, BIAcore 3000 instrument was employed. All experiments were performed at 25°C and HBS-EP was used as a running buffer. Human IGF-I (receptor grade) was immobilized on the surface of a CM5 BIAsensor chip by conventional coupling procedure. IGFBP-4 samples were injected at 20 µl/min for 5 min with a dissociation time of 2 min. Protein solutions directly after RTS reactions as well as purified proteins were tested. Regeneration of the hIGF-I-coated chip surface was achieved by injection of 100 mM HCl. Subsequently kinetic studies with different IGFBP-4 constructs were carried out.
1.4 Results– mutagenesis studies

1.4.1 Protein purification

miniBP5 wild type and miniBP5 mutants (amino acids 40-92 of human IGFBP-5) were purified successfully by three steps purification procedure as presented in details in section 1.3.4.

**STEP 1 Affinity chromatography- Ni-NTA column**

The interaction of the His-tagged proteins with Ni-NTA matrix was used in order to eliminate the variety of *E.coli* proteins, which were naturally expressed during bacterial growth. Figure 1.10 shows the elution profile of proteins, which bind to the matrix. As it turned out (Figure 1.11), there are not only the miniBP5 but also some bacterial proteins.

![Figure 1.10](image)

**Figure 1.10.** A chromatogram of the miniBP5 mutant purified on Ni-NTA column. Peak corresponds to His-tagged miniBP5 mutant.
**Chapter 1  Biochemical studies on IGF and IGF-binding proteins interactions**

**Figure 1.11.** SDS-PAGE analysis of fractions eluted from Ni-NTA column. Bands corresponding to miniBP5 mutant are indicated with arrow mark.

**STEP 2 Ion exchange chromatography - MonoS column**

Proteins, initially purified on Ni-NTA column, were further refolded and purified by means of cation exchange chromatography. Figure 1.12 shows the elution profile of proteins bound to MonoS column. Main peak corresponds to His-tagged miniBP5 mutants, however as shown in Figure 1.13, the remaining two peaks also indicate the presence of miniBP5.

**Figure 1.12.** A chromatogram of the miniBP5 mutant purified on MonoS column. Main peak corresponds to His-tagged miniBP5 mutant is indicated by cyan arrows.
SDS-PAGE analysis of MonoS eluted protein samples (Figure 1.13) demonstrated that they are highly homogenous and the band migrated at ~10kDa corresponds to the theoretical mass of the His-tagged miniBP5 mutant.

**Figure 1.13.** SDS-PAGE analysis of fractions eluted from MonoS column. Bands corresponding to miniBP5 mutant are indicated with arrow mark.

**STEP 3 Gel filtration - Superdex S75 column**

MonoS purified samples, after thrombin digestion, were passed through the Superdex S75 column in order to eliminate uncut protein and other impurities. Figure 1.14 shows the elution profile of the purified miniBP5 mutant. Single, sharp peak indicates that the protein of interest is pure. SDS-PAGE electrophoresis (Figure 1.15) and mass spectrometry (data not shown) confirmed very high purity of miniBP5 mutants that allowed using them in further studies (NMR experiments, kinetic analysis and crystallization trials).
1.4.2 Size exclusion binding test

To verify the forming of binary complex of miniBP5Y50A and IGF-I as well as miniBP5Y50K and IGF-I, the gel filtration column was employed. Protein complexes eluted from the column indicated appropriate molecular weight (~12kDa).
1.4.3 Characteristics of the binding of IGF-I to the miniBP5 mutants determined by surface plasmon resonance (SPR)

IGF-I binding properties of different miniBP5 mutants were analyzed by BIAcore biosensor measurements. Figure 1.17 shows the sensogram profiles of miniBP5 mutants binding to human IGF-I, while Figure 1.18 presents the binding curves for miniBP5 mutants as a function of their concentration.
Figure 1.17. Biacore analysis of the different miniBP5 mutants bind to human IGF-I. Sensograms show the binding of miniBP5WT (violet), miniBP5V49I (magenta) miniBP5L61Y (dark blue), miniBP5L74M (blue), miniBP5L73W (green), miniBP5R53Y (red) and miniBP5Y50K (dark green) to a sensor chip coated with the human IGF-I. The binding of all miniBPs was measured at a concentration of 25 nM at a flow rate of 20 µl/min.

Figure 1.18. Curves showing the binding of different miniBP5 mutants to the human IGF-I as a function of their concentration.
Although all curves indicated the kinetic conditions, there was no fitting to a 1:1 binding model (Langmuir model). The best fitting was observed with two-state reaction that suggests the conformation changes of investigated complex. Shortly, analyte A (miniBP5 mutant) binds to ligand B (hIGF-I) making the complex AB. This complex is changed to AB* which cannot dissociate directly to A and B unlike observed in case of Langmuir model. Because of this, values of the dissociation constant ($k_d$) could not be determined.

**1.4.4 Characteristics of the binding of IGF-I to the miniBP5 mutants determined by isothermal titration calorimetry**

Results of isothermal titration calorimetry (ITC) measurements of the miniBP5 wild type and miniBP5 mutants are graphically presented in Figure 1.19 and summarized in Table 1.3.
Chapter 1  
Biochemical studies on IGF and IGF-binding proteins interactions

![Graph C](image)

![Graph D](image)

![Graph E](image)

![Graph F](image)
**Figure 1.19.** Isothermal titration calorimetry (ITC) measurements of the miniBP5 wild type (A) and miniBP5 mutants: V49I (B), L73W (C), R53Y (D), Y50K (E) and Y50A (F), respectively. The raw data of the experiments are presented on the top panel. The area underneath each injection peak is equal to the total heat released for that injection. When this integrated heat is plotted against the molar ratio of titrant added to the protein solution in the cell, a complete binding isotherm for the interaction is obtained (bottom panel). The one site model was used to fit the data. The solid red line is the calculated curve using the best-fit parameters. The values of binding constant are shown in Table 1.3.

<table>
<thead>
<tr>
<th>reservoir</th>
<th>titrant</th>
<th>K_D [nM]</th>
</tr>
</thead>
<tbody>
<tr>
<td>IGF-I</td>
<td>miniBP5 wild type</td>
<td>126</td>
</tr>
<tr>
<td>IGF-I</td>
<td>miniBP5V49I</td>
<td>359</td>
</tr>
<tr>
<td>IGF-I</td>
<td>miniBP5L73W</td>
<td>366</td>
</tr>
<tr>
<td>IGF-I</td>
<td>miniBP5R53Y</td>
<td>472</td>
</tr>
<tr>
<td>IGF-I</td>
<td>miniBP5Y50K</td>
<td>2.75x10^3</td>
</tr>
<tr>
<td>IGF-I</td>
<td>miniBBP5Y50A</td>
<td>17.3x10^3</td>
</tr>
</tbody>
</table>

Summary of the kinetic data of IGF-I binding to miniBP5 wild type and miniBP5 mutants.

**1.4.5 Crystallization of miniBP5**

**1.4.5.1 Optimisation of crystallization conditions**

In case of miniBP5Y50K mutant, crystallization was successful with 0.2 M zinc acetate dihydrate, 0.1 M sodium cacodylate, pH 6.5, 18% PEG 8000. Crystals of the plates form (Figure 1.20) appeared within 2 weeks, but only in drops containing very high concentrated protein (~30 mg/ml). The drops in which the concentration of the protein was lower (5-20 mg/ml) remained completely clear (precipitate has not even occurred). On the contrary, during crystallization of miniBP5 WT, only precipitate was observed. Moreover, it appeared even if the protein concentration was not high (~5 mg/ml). Condition for the growing of
crystals of miniBP5Y50K mutant was further optimized in order to achieve the crystals of better quality. For this purpose, crystallization was carried out with buffers in which the concentration of PEG 8000 and pH of sodium cacodylate were varied (8%-22% for PEG and 6.2-6.8 for sodium cacodylate). Additionally, the micro-seeding technique was employed. The crystal seeds were introduced into equilibrated clear drops using a rabbit’s whisker as previously described (Ducruix and Giege). The new crystals grew along the streak line within next day.

![Figure 1.20](image)

**Figure 1.20.** Crystal plates of miniBP5Y50K mutant growing in 0.2 M zinc acetate dihydrate, 0.1 M sodium cacodylate, pH 6.5, 18% PEG 8000 at 4°C.

### 1.4.5.2 Data collection

X-ray data sets were collected at the Swiss Light Source (SLS), Villigen, Switzerland and at the MPG/GBF beamline BW6 at DESY, Hamburg. For protection, all measured crystals were soaked for 1-2 min in a solution containing 80% (v/v) mother liquor and 20% (v/v) ethyleneglycol and frozen in liquid nitrogen. The quality of collected data sets was mostly not acceptable due to an insufficient number of the reflections or/and their low intensities. Figure 1.21 presents one of the best diffraction pattern of miniBP5Y50K crystal, measured at DESY, Hamburg.
Figure 1.21. Representative diffraction pattern of miniBP5Y50K crystal. The edge of image plate corresponds to 1.6 Å.

Data processing was done with MOSFLM (Leslie, 1991) or XDS software package (Kabsch, 1993). In case of the data presented above, the most possible space group seems to be P2 with unit cell dimensions $a = 38 \, \text{Å}$, $b = 46.9 \, \text{Å}$, $c = 67 \, \text{Å}$, $\alpha = \gamma = 90^\circ$, $\beta = 100.8^\circ$ and the crystal contains 4 molecules per asymmetric unit. Unfortunately, further structural analysis was not possible because the molecular replacement was not successful.
1.5 Discussion-mutagenesis studies

It is commonly accepted that the insulin-like growth factor binding proteins (IGFBPs) have high affinity for both IGF-I and –II, and in most cases, the affinity of the intact proteins is greater than that of IGF-I receptor (Martin and Baxter, 1999). Therefore, IGFBPs are capable of regulating the equilibrium distribution of IGF-I and -II between that bound to the receptor and that bound to the binding proteins in circulation or in extracellular matrix and on cell surfaces. As natural antagonists of the IGFs, IGFBPs logically could be significant therapeutic agents for inhibiting IGF actions in a variety of human diseases, including cancer (LeRoith et al., 1995; Yu and Rohan, 2000), diabetic renal, and vascular, and eye disease (Smith et al., 1999).

The knowledge of the structure and mechanism of the IGFBPs binding to IGFs is crucial for a true understanding of their function. However, the complete tertiary structure of IGFBPs neither alone nor in complex with IGFs is determined to date. Several groups of investigators have focused on the small domains of the IGFBPs, which are known to mediate IGFs binding (Kalus et al., 1998; Zeslawski et al., 2001; Headey et al., 2003; Siwanowicz et al., 2005a). N-terminal fragments of IGFBP-3, -4 and -5 have been shown to have detectable binding affinity for IGFs, although generally their affinities are reduced between 10- and 1000-fold compared with the native proteins (Andress et al., 1993; Cheung et al., 1994; Hashimoto et al., 1997; Kalus et al., 1998; Standker et al., 2000; Carrick et al., 2001; Galanis et al., 2001; Siwanowicz et al., 2005a). The role of the C-terminal domains in binding to IGFs is not so clear. Several studies have shown that C-terminal fragment of IGFBP-2 binds IGFs only ~10-fold weaker than full-length protein (Wang et al., 1988; Ho and Baxter 1997; Carrick et al., 2001) whereas his counterpart in IGFBP-3 binds IGFs with 3 to 1000-fold lower affinity than full length IGFBP-3 (Galanis et al., 2001; Vorwerk et al., 2002; Payet et al., 2003). On the other hand, there is some evidence that isolated C-terminal fragments of IGFBP-3 (Lalou et al., 1996), IGFBP-4 (Qin et al., 1998) and IGFBP-5 (Kalus et al., 1998) do not bind IGF-I and IGF-II. Our last ITC and NMR studies (Siwanowicz et al., 2005a) also have shown no binding between C-terminal domain of IGFBP-4 and IGF-I. In particular, the results obtained from NMR are valuable on account of very high sensitivity of this method. In comparison to the standard molecular biology methods used in protein investigation, such as ELISA, RIA, or BIAcore, which fail to detect ligands with weak (i.e. millimolar) affinities, NMR technique allows detecting ligands that bind only very weakly to target proteins (Rehm et al., 2001). When all of these data are considered, it appears that neither the N- nor C-terminus separately embodies the full, physiologically critical IGF binding domain. Rather, both regions contribute
to high affinity IGF binding as independently organized segments that are spatially coordinated by the linker domain into the binding site (Rosenzweig, 2004).

The fragment analysis studies have distinctly demonstrated that the high affinity binding site for IGFs is located in N-terminal domain. In the article published by Zeslawski et al. (2001) the importance of this domain (named there mini-IGFBP-5) has been confirmed. The principal IGF-I–mini-IGFBP-5 interaction (Figure 1.22) is a hydrophobic sandwich that consists of interlaced protruding side chains of IGF-I and solvent-exposed hydrophobic side chains of the mini-IGFBP-5. The side chains of IGF-I Phe16, Leu54 and also Glu3 are inserted deeply into a cleft on the mini-IGFBP-5. This cleft is formed by side chains of Arg53 and Arg59 residues on the solvent-exposed side of the molecule and by Val49, Leu70 and Leu74 on the opposite inner side, with a base formed by residues Cys60 and Leu61. Phe16 makes direct contacts with the backbone and side chain of Val49, and with Cys60 of mini-IGFBP-5. The hydrophobic cluster is closed on the solvent side by side chains of Glu3 and Glu9 of IGF-I, and His71 and Tyr50 of mini-IGFBP-5. These residues form a network of hydrogen bonds. Mutagenesis studies for IGF-I have indicated that residues Glu3, Thr4, Gln15 and Phe16 are important for binding to IGFBPs, particularly to IGFBP-3 (Baxter et al., 1992). NMR study, performed previously by Kalus et al. (1998), have shown that the hydrophobic residues Val49, Leu70 and Leu73 of IGFBP-5 are crucial for binding to IGFs, which was further confirmed thanks to the structure provided by Zeslawski et al. (2001).

![Figure 1.22. Ribbon plot of IGF (blue) – mini-IGFBP-5 (green) complex. IGF-I residues involved in binding are highlighted (cyan). Residues that were mutated are indicated in green.](image)
Based on the knowledge of the mini-IGFBP-5 structure, several mutants have been
designed in order to modulate and preferably enhance the IGF binding to the mini-IGFBP-5.
Because the previous NMR and crystallographic studies had shown that hydrophobic residues
between Val49 and Leu73 of IGFBP-5 formed the core of the IGF binding site, four residues
in that region (Val49, Tyr50, Arg53, and Leu73) were mutated to Ile, Lys or Ala, Tyr and
Trp, respectively. Generally, these changes purposed increase the volume of side chains of
amino acids (Val49-Ile and Leu73-Trp) as well as preserve hydrophobic character of the
interaction area (Arg53-Tyr). Two remaining mutations were different. In particular, mutation
the tyrosine to lysine was risky because in place of an aromatic and hydrophobic residue a
positively charged amino acid was introduced. Tyrosine to alanine substitution was chosen
because alanine is present at this position in the IGFBP-1 (Figure 1.3). It should be noted that
every mentioned substitution referred to very high conserved residues.

The kinetic analysis, using two methods: surface plasmon resonance (SPR) and
isothermal titration calorimetry (ITC) have shown that mini-IGFBP-5 mutants have lower
binding affinities for IGF-I than mini-IGFBP-5 wild type. Isothermal titration calorimetry
measurements allowed determination of the binding constant values. Native mini-IGFBP-5 has
a $K_D$ of 126 nM, whereas the all mutants bind the IGF-I with 3 to 140-fold lower affinity than
mini-IGFBP-5 wild type. Substitutions for Tyr50 seem to be essential for the binding affinity.
Both mutants, related to this residue, Tyr50Lys and Tyr50Ala, indicate the lowest values of
binding constant ($K_D$ of 2.75 µM and 17.3 µM, respectively) and the result for Tyr50Lys
mutant is in agreement with data obtained from BIAcore experiments (Tyr50Ala mutant was
not measured by this method). Such an exceptionally high $K_D$ value for Tyr50Ala mutant
indicates 140-fold weaker binding for IGF-I, which is somewhat unexpected, because alanine
residue is present at this position in IGFBP-1. Val49Ile (IGFBP-3 equivalent) and Leu73Trp
mutants show the comparable binding constants ($K_D$ of ca. 360 nM) while $K_D$ value for
Arg53Tyr is insignificantly higher (470 nM). Interestingly, Val49 corresponds to bovine
IGFBP-2 Val59, mutagenesis of which was without effect on IGF binding (Baxter, 2000).

Results obtained by means of surface plasmon resonance method only demonstrate that
every tested mutant binds to IGF-I with a lower affinity than mini-IGFBP-5 wild type. As
mentioned before, although all curves denoted the kinetic conditions, there was no fitting to a
Langmuir binding model, which resulted in lack of determination of $K_D$ values. However,
performed comparative studies allowed indicate the following order of activity of the mini-
IGFBP-5 mutants: Val49Ile > Leu73Trp > Arg49Tyr > Tyr50Lys, which practically agrees
with results procured with ITC experiments. The differences appeared in case of two other
mutants, which were examined by another investigator (Siwanowicz). ITC experiment has shown that Leu74Tyr mutant binds IGF-I more tightly than Val49Ile, and Leu73Trp binds better than Leu74Met, whereas the order demonstrated by BIAcore was inverted. These findings can be explained by the fact that under conditions of restricted volume and ligand concentration, a ligand (IGF-I) that is immobilized on the chip surface (BIAcore), might exhibit different binding kinetics than the corresponding ligand in solution (reservoir in ITC experiment).

The kinetic studies performed in this work have definitively confirmed the suggestion made by Kalus et al. (1998) and Zeslawski et al. (2001), that a specific group of hydrophobic amino acids within the N-terminal part of IGFBP-5 is essential for IGF-I binding. Some introduced substitutions, particularly Tyr50 to Lys, disturbed the hydrophobic character of the patch predicted from the NMR and X-ray model. In consequence this patch would have even been destroyed. The possibility that other changes in tertiary structure, such as alteration of the disulfide bond pattern or important folding disruptions, occurred as a result of these mutations can be excluded. Performed 1D-proton NMR experiment confirmed the correct folding of the miniBP5 mutants, therefore the idea that mutations disrupted the hydrophobic patch and that this is the principle reason that binding affinity is reduced seems to be reliable. Several observations can support this conclusion. First, it is based on a rational protein folding model proposed by both investigators (Kalus and Zeslawski). Second, no direct cysteine substitutions were performed, and thus there is no reason a priori to believe that the disulfide bond pattern would be altered. Finally, the substitutions refer to residues, which are in close physical proximity, making it less likely that some other important structural determinant at a distant site in the molecule was altered.
1.6 Results – RTS

To examine the influence of fusion tags on the production level of IGFBP-4 in the rapid translation system, four plasmids were constructed that produced IGFBP-4 fused with His-tag, Strep-tag, Avi-tag and GST-tag, respectively. All tested IGFBP-4 constructs have been successfully produced \textit{in vitro} in soluble form by means of RTS. The first experiments have shown nearly no influence of fusion tags on the synthesized protein, however the best expression was observed in case of the His-tagged IGFBP-4. It is known that shorter tags such as the 6xHis motif typically allow for larger amounts of protein production in comparison with larger tags (Luke, 2004). Unfortunately, beside the target protein, some fragments of IGFBP-4 with molecular weights lower than expected, were obtained. Moreover, under certain conditions, the yield of these fragments was much higher than the yield of the target protein itself (Figure 1.23, 1.25 A). To improve the quality of the desirable protein, the most optimal parameters for \textit{in vitro} reaction have been quested.

1.6.1 Optimisation of RTS conditions

Three different RTS kits (RTS100 HY, RTS500, and RTS500 HY) were used in order to optimize the conditions of production and solubility of IGFBP-4 constructs. RTS is an “open” system. It means that the protein production can be monitored during reaction by taking small samples from the reaction chamber without the risk of aggravation of protein expression level. Moreover, thanks to this feature the reaction conditions can be adapted in a protein-specific manner by adding chemicals or protein factors to the reaction mixture.

In this case the reaction time and temperature as well as the effect of addition of inhibitor proteases and chaperones to the reaction mixture were tested. The final results are summarized in Table 1.4, 1.5 and 1.6.

1.6.1.1 Time

To estimate the effect of incubation time on the production level of IGFBP-4, protein samples were taken from the reaction chamber at various time points and verified by Western blot as shown in Figure 1.23 and 1.24.
Figure 1.23. Western blot analysis of His-IGFBP-4 expression in RTS100 HY carried out at 30°C. **BP4**, IGFBP-4 standard; **1**, sample after 0.5h incubation; **2**, sample after 1h incubation; **3**, sample after 1.5h incubation; **4**, sample after 2h incubation; **5**, sample after 2.5h incubation; **6**, sample after 3h incubation; **7**, sample after 3.5h incubation; **8**, sample after 4h incubation.

Figure 1.24. Western blot analysis of His-IGFBP-4 expression in RTS500 carried out at 30°C. **BP4**, IGFBP-4 standard; **C**, control; **1**, sample after 2h incubation; **2**, sample after 3h incubation; **3**, sample after 4h incubation; **4**, sample after 5h incubation; **5**, sample after 6h incubation; **6**, sample after 7h incubation; **7**, sample after 22h incubation.

The immunoblot analysis has shown that incubation time distinctly affects the production level of IGFBP-4. It should be noticed that this effect varied not only for different RTS kits but also for the temperatures. Generally, the amount of expressed proteins (all tested constructs) increased in time in the reactions carried out at 30°C by means of RTS500. In case of GST-tagged IGFBP-4 (data not shown) and His-tagged IGFBP-4 (Figure 1.24) a significant rise of degradation products level was simultaneously observed. When the RTS500 HY was
employed, the amount of produced protein did not increase in time at 20°C and 30°C. Furthermore, the amount of degradation fragments rose emphatically and after 4 h incubation was much higher than the level of expressed full length IGFBP-4 itself (Figure 1.25 A and B). It should be pointed out that the maximal expression yield of the target protein was achieved after 1h incubation at both temperatures (20°C and 30°C) (Figure 1.25 A and B). Western blot analysis of His-IGFBP-4 expression in RTS100 HY carried out at 30°C allowed observing the occurrence of complete protein degradation process. The level of the target protein decreased, whereas the level of smaller fragments of His-IGFBP-4 increased in time. Moreover, after 3.5 h incubation only the presence of these fragments was noticed. Incubation time of 30 minutes seems to be the most suitable for the IGFBP-4 production because, at this point, the target protein achieves the maximum of its expression. During that time the amount of degradation products is the lowest (single band at ~14kDa).

1.6.1.2 Temperature

To estimate the effect of temperature on the production level of IGFBP-4, the RTS reactions were carried out at three different temperatures: 30°C, 20°C (RT) and 4°C. Protein samples were taken from the reaction chamber at various time points and verified by Western blot.
Figure 1.25. Western blot analysis of His-IGFBP-4 expression in RTS500 HY carried out at 30°C (A), RT (B) and 4°C (C). BP4, IGFBP-4 standard; C, control (positive in case of A and B, negative in case of C); P, pellet; S, supernatant; 1, sample after 1h incubation; 2, sample after 2h incubation; 3, sample after 3h incubation; 4, sample after 4h incubation; 5, sample after 5h incubation; 6, sample after 6h incubation.

Figure 1.26. Western blot analysis of GST-IGFBP-4 expression in RTS500 HY carried out at 4°C. BP4, IGFBP-4 standard; C, negative control; 1, sample after 1h incubation; 2, sample after 2h incubation; 3, sample after 3h incubation; 4, sample after 4h incubation; 5, sample after 5h incubation; 6, sample after 6h incubation.
Figure 1.27. Western blot analysis of Avi-IGFBP-4 expression in RTS500 HY carried out at 4°C. P, pellet after dialysis; S, supernatant after dialysis; 1, sample after 1h incubation; 2, sample after 2h incubation; 3, sample after 3h incubation; 4, sample after 4h incubation; 5, sample after 5h incubation; 6, sample after 6h incubation.

Figure 1.28. Western blot: comparison of in vitro expressed Avi-IGFBP-4 levels obtained at 30°C and at 4°C. BP4, IGFBP-4 standard; 1, sample after 6h incubation at 4°C; 2-3, samples after dialysis, RTS reaction carried out at 4°C; 4-5, samples after dialysis, RTS reaction carried out at 30°C.

The immunoblot analysis (Figure 1.25 A) has shown that the highest production level of IGFBP-4 in the RTS500 HY was obtained when the reaction was carried out at 30°C. Unfortunately, such a high temperature, although recommended by manufacturer, promoted the appearance of smaller fragments of the target protein. Mass spectrometry analysis confirmed the molecular weights of these fragments as 11.3 kDa, 14.4 kDa and 15.5 kDa. The maximal expression level of full length IGFBP-4 was achieved after 1h incubation in two tested temperatures (both 20°C and 30°C). At 4°C the expression level of His-tagged IGFBP-
4 maintained constant in time, whereas the expression level of GST-tagged IGFBP-4 and Avi-tagged IGFBP-4 slightly increased (Figure 1.26 and 1.27, respectively). It is interesting to point out that dialysis performed directly after RTS reaction did not affect the final level of the target protein. It means that the protein remained stable during this process. The comparison of the Avi-IGFBP-4 expressed in the reaction carried out at 30°C and 4°C is presented in Figure 1.28. This immunoblot emphasizes the differences in quality of the obtained target protein in these two temperatures. Lane 2 and 3 present the protein samples, which were dialyzed after RTS reaction performed at 4°C, whereas lane 4 and 5 show the protein samples, which were dialyzed after RTS reaction performed at 30°C. In both cases, the amount of the target protein was similar, but only in case of RTS at 30°C the presence of additional fragments of Avi-tagged IGFBP-4 was noticed. The molecular weights of these fragments approximate 11kDa and 15kDa, which is in good agreement with the mass spectrometry analysis performed before.

### 1.6.1.3 Proteases inhibitors and chaperones

In order to increase the amount of correctly folded full length IGFBP-4 and decrease the level of putative degradation IGFBP-4 products, protease inhibitors as well as chaperones (GroE) were added into RTS reaction. The GroE-system (GroEL/GroES) is a prokaryotic chaperones system, which is involved in *de novo* folding of 10-30% of all cytoplasmatic proteins (Ewalt *et al.*, 1997). According to manufacturer (Roche) the presence of GroE during cell-free protein synthesis can help in correct folding of proteins in the range between 20 and 60 kDa, therefore it is suitable in case of IGFBP-4 (29kDa). Smaller proteins are not recognized and larger do not fit into the cavity of chaperones.
Figure 1.29. Western blot analysis of His-IGFBP-4 expression in RTS100 HY carried out at 30°C. BP4, IGFBP-4 standard; 1, sample after 1h incubation; 2, sample with GroE after 1h incubation; 3, sample with inhibitor proteases after 1h incubation; 4, sample with GroE and inhibitor proteases after 1h incubation.

As shown in Figure 1.29 (lane 3), the addition of inhibitor proteases into RTS100 HY did not prevent protein degradation. The amount of full length His-tagged IGFBP-4 expressed under four different conditions seems to be similar. Surprisingly, the smaller fragments of target protein appear in force, particularly in case of the reaction carried out with addition of cocktail inhibitors and GroE-system (lane 4). In sum, protein of interest is not produced sufficiently in RTS 100 HY with addition of inhibitor proteases at 30°C.

1.6.2 Optimisation of RTS conditions – Summary

Tables presented below show the final results of optimisation trials of RTS conditions (RTS100 HY, RTS500 and RTS500 HY, respectively).

Table 1.4. RTS100 HY

<table>
<thead>
<tr>
<th>construct</th>
<th>His-IGFBP-4</th>
<th>Avi-IGFBP-4</th>
<th>Strep-IGFBP-4</th>
<th>GST-IGFBP-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>temperature</td>
<td>30°C</td>
<td>RT</td>
<td>4°C</td>
<td>RT</td>
</tr>
<tr>
<td>time</td>
<td>up to 6 h</td>
<td>n.t</td>
<td>4 h</td>
<td>n.t</td>
</tr>
<tr>
<td>stirrer speed</td>
<td>no</td>
<td>n.t</td>
<td>no</td>
<td>n.t</td>
</tr>
<tr>
<td>protein expression</td>
<td>yes</td>
<td>n.t</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 1.5. RTS500

<table>
<thead>
<tr>
<th>construct</th>
<th>His-IGFBP-4</th>
<th>Avi-IGFBP-4</th>
<th>Strep-IGFBP-4</th>
<th>GST-IGFBP-4</th>
</tr>
</thead>
<tbody>
<tr>
<td>temperature</td>
<td>30°C</td>
<td>RT</td>
<td>4°C</td>
<td>RT</td>
</tr>
<tr>
<td>time</td>
<td>ON</td>
<td>6 h</td>
<td>ON</td>
<td>n.t</td>
</tr>
<tr>
<td>stirrer speed</td>
<td>250 rpm</td>
<td>250 rpm</td>
<td>250 rpm</td>
<td>n.t</td>
</tr>
<tr>
<td>protein expression</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>n.t</td>
</tr>
</tbody>
</table>
### Table 1.6. RTS500 HY

<table>
<thead>
<tr>
<th>Construct</th>
<th>His-IGFBP-4</th>
<th>Avi-IGFBP-4</th>
<th>Strep-IGFBP-4</th>
<th>GST-IGFBP-4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temperature</strong></td>
<td>30°C</td>
<td>30°C</td>
<td>30°C</td>
<td>30°C</td>
</tr>
<tr>
<td><strong>Time</strong></td>
<td>4 h</td>
<td>4 h</td>
<td>6 h</td>
<td>n.t</td>
</tr>
<tr>
<td><strong>Stirrer Speed</strong></td>
<td>250 rpm</td>
<td>250 rpm</td>
<td>250 rpm</td>
<td>n.t</td>
</tr>
<tr>
<td><strong>Protein Expression</strong></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>n.t</td>
</tr>
</tbody>
</table>

n.t. – not tested

#### 1.6.3 Purification of IGFBP-4

His-IGFBP-4 and Strep-IGFBP-4 were purified by a one-step purification procedure as described before (section 1.3.9.7 and 1.3.9.8, respectively). Because of the small amount of protein, analysis was performed by means of Western blot. Purified His-IGFBP-4 was further subjected to verification by BIAcore biosensor instrument in context of the IGF-IGFBP-4 interactions.

![Western blot analysis of His-IGFBP-4 purified on Ni-NTA column. BP4, IGFBP-4 standard; P, pellet; S, supernatant; FT, flow through; 1, 2, 3, eluted fractions. His-IGFBP-4 was expressed in RTS500 HY at 4°C.](image)

**Figure 1.30.** Western blot analysis of His-IGFBP-4 purified on Ni-NTA column. BP4, IGFBP-4 standard; P, pellet; S, supernatant; FT, flow through; 1, 2, 3, eluted fractions. His-IGFBP-4 was expressed in RTS500 HY at 4°C.
Figure 1.31. Western blot analysis of Strep-IGFBP-4 purified on Strep-Tactin Sepharose column. BP4, IGFBP-4 standard; S, supernatant; FT, flow through; 1, 2, 3, 4, 5 eluted fractions. Strep-IGFBP-4 was expressed in RTS500 at 30°C.

1.6.3 BIAcore analysis

IGFBP-4 binding properties were analysed by means of BIAcore biosensor instrument. Figures presented below show the sensogram profiles of Avi-IGFBP-4 (Figure 1.32 A) and His-IGFBP-4 (Figure 1.32 B) binding to human IGF-I.
Figure 1.32. Representative sensograms of BIACore analysis of Avi-IGFBP-4 (A) and His-IGFBP-4 (B) binding to immobilized hIGF-I. The plots show the relative response in resonance units (RU) plotted against time. Avi-IGFBP-4 was expressed in RTS500 HY at 4°C and dialysed against Tris buffer, whereas His-IGFBP-4 was expressed in RTS500 HY at RT and purified on NiNTA column.
1.7 Discussion – RTS

Although bacterial \textit{in vivo} expression system is still the most popular method for the protein production, a significant increase of interest in the cell-free expression systems (based on both prokaryotic and eukaryotic cell lysates) is noticed (Madin \textit{et al.}, 2000; Tarui \textit{et al.}, 2000; Chekulayeva \textit{et al.}, 2001; Martemyanov \textit{et al.}, 2001; Tabuchi \textit{et al.}, 2002; Busso \textit{et al.}, 2003; Luke, 2004; Murthy \textit{et al.}, 2004). The standard \textit{in vivo} expression, because of many steps of the procedure, such as transformation, cultures growing and addition of the transcription inducing agents, is time consuming and often requires further laborious purification techniques (Murthy \textit{et al.}, 2004). On the contrary, the expression by using the cell-free systems is fast and easy. The lack of cell membranes and cell walls eliminates the lysis step in the purification procedure, which is also time consuming (usually, obtained lysate has to be centrifuged, filtrated and dialysed against appropriate buffer to further purification). Because of this, the genomic DNA contamination, insufficient lysis and the protein degradation due to lysis conditions are totally dispossessed in the cell-free system (Luke, 2004). In this thesis, the usefulness of the RTS, on the example of successful expression of the recombinant human IGFBP-4 (full length) was demonstrated. Our previous experiences with \textit{E. coli} system allow us to denominate this protein as an exceptionally difficult. As a member of IGFBPs family, IGFBP-4 contains the large number of disulfide bonds (due to 18 highly conserved cysteines and an additional two cysteines in the central region of the protein) that makes the high expression level and purification very difficult. When IGFBP-4 was expressed in \textit{E. coli} system as a His-tag fusion protein and purified under denaturing conditions, it had to be further refolded in order to get an active protein. Because the whole procedure was time consuming and the quality of obtained protein was not satisfactory (discussed below), we attempted to employ the brand new \textit{in vitro} technique offered by Roche Diagnostics. Rapid optimisation of IGFBP-4 expression was easily achieved thanks to the “open” nature of the RTS. The protein production could be monitored during reaction by taking samples from the reaction chamber without the risk of aggravation of protein expression level. Moreover, the reaction conditions could be adapted in a protein-specific manner by adding chemicals or protein factors to the reaction mixture (protease inhibitors and chaperones).

Our studies have demonstrated that the expression of the IGFBP-4 by means of the RTS kits was fast and mostly reproducible, however, because of the degradation problem, the final results were not fully acceptable.
Protein degradation

It should be emphasized that IGFBP-4 like the remained five IGFBPs is a very difficult case of protein. First of all, when expressed in a cell-based system, it is in ~30% unfolded that was confirmed by NMR studies. The NMR spectrum of the full length IGFBP-4 indicated that the central variable domain of the protein is in a random coil conformation, whereas the NMR spectrum of the C-terminal fragment (residues Gly151-Ala232, see Figure 1.3) showed some unstructured regions located also in this region (Figure 1.33). Nearly identical estimates were obtained from NMR spectra of IGFBP-5 (Siwanowicz et al., 2005b).

![Figure 1.33. Amide region of 1H NMR spectra of the full length IGFBP-4 and their isolated N- and C-domains.](image)
Inspection of such spectra yields semi-quantitative information on the extent of folding in partially structured proteins or their domains (Rehm et al., 2002). The spectrum of full-length IGFBP-4 exhibits a large peak at 8.3 ppm and some signals downfield (that is shifted to higher ppm values). The appearance of intensities at chemical shifts near ~8.3 ppm is an indicator for a disordered protein, as this is a region characteristic of backbone amides in random coil configuration (Wüthrich, 1986). On the other hand large signal dispersion beyond 8.5 ppm proves a protein to be folded. Due to the different chemical environment and thus the varying shielding effects the resonances of the single protons will be distributed over a wide range of frequencies. A typical intensity pattern of a folded protein can be observed in the spectra of the N-terminal domain of IGFBP-4 (residues 1-92).

It is known that such a rich in loops structure promotes the instability (protein is sensitive to proteolysis) and decrease the biological activity of protein. IGFBP-4 proteolytic degradation has been described in several in vitro models (in particular in human fibroblast-conditioned medium), as well as in vivo (human pregnant serum and human, porcine, bovine, ovine and equine follicular fluid of preovular follicles), and it has been shown that the pregnancy-associated plasma protein-A (PAPP-A) is responsible for the protein degradation (Conover et al., 1999; Lawrence et al., 1999; Overgaard et al., 2000; Byun et al., 2001; Mazerbourg et al., 2001). A number of studies have demonstrated that the linker region (L-domain between N- and C- domains, see Figure 1.3) is the site of proteolysis. For example, Chernausek et al. (1995) indicated the presence of three fragments (1-90, 1-120, and 132-C-terminal) of IGFBP-4 in the rat B104 neuronal cell line that confirmed the localization of cleavage site in the L-domain. In case of the IGFBP-4 produced in vitro with RTS, the proteolytic pattern seems to be similar. Molecular weights of the degradation products, determined by Western blot and mass spectrometry, have shown that the protein can be cleaved in the L-loop as presented in Figure 1.34.

![Figure 1.34. Scheme of human IGFBP-4. Regions of the amino acids homology among the IGFBPs are indicated as shaded boxes (violet and magenta for N-and C-terminus, respectively). Capital letters demonstrate central variable domain of protein (L-domain). The asterisk shows potential cleavage sites based on mass spectrometry analysis.](image-url)
Western blot analysis of the purified Strep-IGFBP-4 (Figure 1.31) and His-IGFBP-4 (data not shown) has indicated that the smaller fragments of the IGFBP-4 contain N-terminal His-tag and Strep-tag, respectively. Therefore, the idea that internal initiation site of translation is responsible for the appearance of these fragments can be excluded. This conclusion can be supported by the observations provided by the standard cell-based \textit{E. coli} system. There are no premises to say that such an occurrence happens in case of any IGFBPs.

**Disulfide bonds and protein activity**

The correct formation of disulfide bonds is essential for proper folding of the protein and consequently for its biological action. It should be noticed that the RTS is not able to produce proteins with the posttranslational modifications including not only disulfide bonds forming, but also glycosylation and phosphorylation (Betton, 2003; Luke, 2004). It means that during the RTS reaction the IGFBP-4 can be incorrectly folded. As it turned out, the lack of S-S bridges is crucial for the activity of protein that was confirmed by SPR studies. Activity assays using BIAcore instrument have shown that protein taken directly after RTS reaction did not bind to IGF-I. Only samples, which were purified or dialyzed before the measurement, were active (Figure 1.32 A and 1.32 B). It is clear that protein had to be oxidized to form disulfide bonds, and in this case S-S bridges were formed during purification (His-IGFBP-4) and dialysis (Avi-IGFBP-4) procedures.

**Protein expression at 4°C**

It is obvious that the high temperature promotes the protein degradation, whereas the low prevents it. However, the protein expression at 4°C is abnormal, even for an \textit{in vitro} system. Results presented here demonstrate that such a strange phenomenon is possible. Moreover, proteins obtained at 4°C seem to be stable and active that was confirmed by BIAcore experiments (Figure 1.32 A). Unfortunately, insufficient amounts of the protein produced in RTS did not allow verifying these disclosures by most detailed kinetic studies.

**General conclusions**

Rapid optimisation of IGFBP-4 expression is easily achieved thanks to the open nature of the RTS. We tested not only the expression of several IGFBP-4 constructs in different RTS kits,
but also the various reaction conditions (including temperature and time) to obtain a soluble and functional protein. Additionally, the enzymatic mono-biotinylation of the Avi-IGFBP-4 was successfully performed in the RTS500 and RTS500 HY. In sum, studies presented here demonstrate that the advantages resulting from the open character of the RTS system makes it a very attractive alternative for the standard cell-based expression systems.
Chapter 2                                                                 Structural studies on the CAS SH3 domain

2 Structural studies of SH3 domain of recombinant human Crk-associated tyrosine kinase substrate p130cas (CAS)

2.1 Biological background

2.1.1 p130cas (CAS)

The Crk-associated tyrosine kinase substrate p130cas (CAS) was first recognized as a phosphotyrosine (pTyr)-containing 130kDa protein that associated with two oncoproteins, pp60<sup>v-src</sup> (v-Src) and p47<sup>msg-crk</sup> (v-Crk) (Reynolds <i>et al.</i>, 1989; Matsuda <i>et al.</i>, 1990). Mutation in v-Src and v-Crk that abrogated binding of p130cas also abolished the transforming activity of these oncoproteins. This suggests that CAS plays an important role in the transformation process mediated by both of these oncoproteins (Kanner <i>et al.</i>, 1991). In normal cells, p130cas is localized to focal adhesions (FAs) (Harte <i>et al.</i>, 1996). FAs are specialized sites of adhesion developed by many cells in culture. They consist of aggregated extracellular matrix (ECM) receptors (integrins) that span the plasma membrane and interact outside with ECM components and inside with bundles of actin filaments (stress fibers). A variety of structural (actin, talin, vinculin, tensin, filamin) and signaling molecules (paxillin, focal adhesion kinase FAK, Src, Grb2, PI3K) have been identified in FAs, particularly at their cytoplasmic face (Burridge and Chrzanowska-Wodnicka, 1996).

Tyrosine phosphorylation of p130cas has been implicated as a key signaling step in integrin control of normal cellular behaviours, including motility, adhesion, proliferation and survival (Figure 2.1). There is also some considerable evidence linking CAS to the organization and regulation of the actin cytoskeleton.
**Figure 2.1.** Schematic representation of the integrin-mediated transduction signaling pathways showing the importance of the SH2 and SH3 domain-containing proteins (based on Turner, 2000; O’Neill et al., 2000)

CAS has the structural characteristics of an adapter protein, which can explain its role in the formation of multiprotein signaling complexes (Figure 2.2). CAS contains an SH3 domain followed by a short proline-rich segment (P), a large central substrate domain (SD) composed of fifteen repeats of a four amino acid sequence (YXXP), a serine-rich region (S)
and a carboxy-terminal domain which possess consensus binding sites for the SH2 and SH3 domains of Src (YDYV and RPLPSPP, respectively).

Figure 2.2. The structural organization of CAS protein and its binding partners. Key: SH3, Src-homology 3 domain; P, proline-rich region; SD, substrate domain; S, serine-rich region; SBD, Src binding domain. A partial list of references that address binding of these proteins to Cas includes: Fak (Burnham et al., 1996; Harte et al., 1996; Polte and Hanks, 1995), PTP1B (Liu et al., 1996), PTP-PEST (Garton et al., 1996), C3G (Kirsch et al., 1998), Crk (Burnham et al., 1996; Sakai et al., 1994a), Nck (Schlaepfer et al., 1997), PI3K (Li et al., 2000), SHIP2 (Prasad et al., 2001), 14-3-3 (Garcia-Guzman et al., 1999), Src family kinases (Burnham et al., 1996; Nakamoto et al., 1996; Sakai et al., 1994a), Grb2 (Wang et al., 2000).

The N-terminal SH3 domain mediates the interaction of CAS with FAK, the guanine nucleotide exchange factor C3G and the protein tyrosine phosphatases PTP1B and PTP-PEST, which promote the dephosphorylation of CAS. p130cas becomes phosphorylated in response to integrin engagement by a variety of ECM components such as fibronectin (FN), vitronectin, laminin and collagen (Vuori et al., 1995; Harte et al., 1996). In focal adhesions, phosphorylation of CAS is achieved by the interactions with Src family kinases, as well as FAK and its close relative PYK2 (Hamasaki et al., 1996). Two sites of tyrosine phosphorylation of CAS have been precisely characterized. Tyrosine residues from the substrate domain YXXP motifs when phosphorylated can serve as ligands for the SH2 or PTB domains of other cellular proteins, including Crk, Nck and SH2-containing inositol 5’-phosphatase 2 (SHIP2), which may act as downstream effectors in CAS signaling (Cantley et al., 1994; Schlaepfer et al., 1997; Prasad et al., 2001). The C-terminus of CAS represents the second site of tyrosine phosphorylation of this protein, and, as mentioned above, is
responsible for the interaction with Src family kinases as well as with some other proteins (Grb2, PI3K). It is known that tyrosine phosphorylation of CAS, besides the promotion of interaction with SH2-containing proteins serves also another function: it coincides with a relocalization of CAS from cytoplasmic to membranous, nuclear and insoluble cytoskeletal fractions (Sakai et al., 1994; Polte et al., 1997; O’Neill et al., 2000).

There are two other family members that share considerable structure and sequence homology with CAS: human enhancer of filamentation (HEF1) and embryonal Fyn-associated substrate/Src-interacting protein (Esf/Sin). In all three proteins, the greatest sequence similarity is observed in the SH3 domains. One of the most notable differences is found in the C-terminal Src binding sequences. The bipartite binding site (for both SH3 and SH2 domains of the Src) is present in CAS and Esf/Sin, whereas the SH3-binding sequence is absent in HEF1. Although the CAS family proteins share a conserved overall domain structure, they have potentially distinct functions that arise from differences in tissue distribution, subcellular localization, post-translational modification and primary sequence divergence (Bouton et al., 2001). CAS is expressed uniformly (Sakai et al., 1994), which may suggest a vital role for this protein in normal cell physiology. The tissue distribution in case of HEF1 (epithelial cells (Law et al., 1998), T and B cells (Minegishi et al., 1996)) and Esf (placenta, embryo, muscle and brain (Ishino et al., 1995) rather suggests some specialized functions of these molecules.

Deregulation of the signaling pathways that control cell adhesion, migration, proliferation and survival can result in oncogenic transformation and metastasis (Bouton et al., 2001; Ruest et al., 2001). Aberrant tyrosine phosphorylation of the CAS and other molecules participating in these pathways seem to be crucial for oncogenesis. In fact, numerous studies have shown that Src-CAS complex not only regulates the normal cell proliferation, but it can also promote unregulated growth and survival as a consequence of increased protein expression or prolonged signaling of this complex. Hence, it is not surprising that this adapter protein has been linked to the initiation and/or progression of human cancers including breast tumors, malignant melanoma (Eisenmann et al., 1999) and certain leukemias (Salgia et al., 1996; de Jong et al., 1997). The involvement of CAS and other SH3 and SH2 domain-containing proteins in oncogenesis, makes them potential valuable targets for the design of anti-proliferative agents against tumor cells (for details see 2.1.2)
2.1.2 SH3 domains as potential anti-tumor targets

The Src homology 3 domains (SH3) are small modular protein motifs about 55-70 amino acids long, which seem to be the best-characterized member of the growing family of protein interaction modules. These domains were found in signaling proteins (i.e. small adaptors such as Grb2), enzymes (kinases, lipases, phosphatases), and regulatory proteins (PI3K p85 subunit), as well as in cytoskeleton (α-spectrin) and oncoproteins (pp60c-Src involving in human breast cancer) (Vidal et al., 2001). By binding with moderate affinity and selectivity to proline-rich ligands, SH3 domains play critical roles in a wide variety of biological processes ranging from regulation of enzymes by intramolecular interactions, increasing the local concentration or altering the subcellular localization of components of signaling pathways, and mediating the assembly of large multiprotein complexes (Mayer, 2001). Numerous structural studies have shown that the surface of the SH3 domain is relatively flat and hydrophobic and the peptide ligand adopts an extended, left-handed helical conformation termed the polyproline type II (PPII) helix (Mayer and Eck, 1995; Mayer, 2001; Vidal et al., 2001). The ligand can interact with the SH3 domain in one of two orientations, depending on the position (N- versus C-terminal) of its proline-rich sequence. Core SH3 ligands are characterized by weak interactions (K<sub>d</sub> = 5–100µM) that show little binding selectivity within SH3 families (Scapin, 2002). Higher affinity, more selective ligands require additional flanking residues that bind to less conserved portions of the SH3 surface. Most of the efforts in designing potent and SH3-selective compounds have been based on the use of combinatorial libraries to target either the N-terminal end of the cognate peptide, or the internal Pro-Pro fragment (Combs et al., 1996; Feng et al., 1996; Nguyen et al., 1998). Although these efforts resulted in the synthesis of peptidomimetics, to date there are no reports of non-peptidyl small molecule SH3 inhibitors.
2.2 Aim of the project

This project was aimed at structural studies of the SH3 domain of recombinant human Crk-associated tyrosine kinase substrate p130cas (CAS). In this study, the CAS SH3 domain was successfully crystallized that allowed determining a high resolution X-ray structure of this protein. The further attempts were focused on the co-crystallization with the natural ligands for CAS SH3 domain (peptides from focal adhesion kinase containing proline-rich motifs), as well as with the putative inhibitors (chemical compounds selected by Pharma Research Penzberg).
2.3 X-ray crystallography as the most powerful method for structural studies*

2.3.1 General background

The word ‘crystal’ is derived from the Greek root ‘krystallos’ meaning ‘clear ice’. Like ice, crystals are chemically well defined, and many of them are of transparent and glittering appearance, like quartz, which was for a long time the archetype. Although crystallography, when compared with NMR, gives a more static description of the macromolecular structures, there are no limits in the size of the molecule to be analyzed. This makes X-ray crystallography the method of choice for studying large macromolecular complexes at the atomic level.

2.3.2 Crystal space groups and symmetries

Crystals are by definition regular, three-dimensional arrays of atoms, ions, molecules or molecular assemblies. The minimal structural part that repeats in all three dimensions to build up the crystal is defined as the unit cell. Within the unit cell, a crystal can contain further symmetry elements, dividing it into several asymmetric units, which form the most basic structural element, which is related to all other identical objects in the unit cell by symmetry elements. The geometry of the unit cell together with the possible symmetry operations defines the space group of the crystal. Proteins and nucleic acids do not crystallize in space groups with inversion symmetries because they are composed of enantiomers (L-amino acids and D-sugars, respectively). As a consequence, the number of possible space groups is reduced from 230 to 65, which are distributed between 7 crystal systems: triclinic, monoclinic, orthorhombic, tetragonal, trigonal, hexagonal and cubic. The combination of the 4 crystal lattices: primitive (P), body centred (I), face centred (F) and centred in the (010) planes (C) with the mentioned 7 crystal systems allows a total of 14 Bravais lattices. Identification of the correct space group is essential for indexing of diffraction patterns, which is the first step in understanding of crystal structure.

2.3.3 Crystal growth

Crystallization of macromolecules is a multiparametric process involving three main steps: nucleation, growth and cessation of growth. It is indispensable for crystallization to bring the
protein to a supersaturated state (Figure 2.3), which will force the macromolecules into the solid state - the crystal.

Figure 2.3. Solubility diagram. It represents the different steps occurring during crystal formation. The supersolubility curve separates the labile region, where nucleation occurs, from the metastable region, where crystals grow.

The Debye-Hückel theory describes how the solubility of the protein depends on the present ions. At low ionic strength the protein solubility increases and this phenomenon is called salting-in. On the contrary, in the salting-out phenomenon, the protein at high ionic strength behaves as a neutral dipole and solubility is mainly governed by hydrophobic effects. Thus, salting-out corresponds to a decrease of protein solubility at high salt concentration. The crystallographer can shift the equilibrium from solution to supersaturation by increasing or reducing the ionic strength of the protein solution. The main parameters that influence protein solubility are temperature, pH and presence of precipitating agents (polymers and organic solvents) such as PEG, iso-Propanol, or MPD.

As shown in Figure 2.3, crystal growth can be divided into two steps. First, a spontaneous nucleus formation occurs in the supersaturation area followed by formation of small aggregates. After the critical amount of aggregated molecules (10-200) is surpassed, the crystal growth becomes an energetically privileged process. Crystal growth always needs a lower degree of supersaturation than nucleus formation. Crystals should grow slowly enough to achieve the maximum possible internal order.

The protein solution used in crystallization experiment must be pure and homogenous. It means that protein samples should be cleaned from all potential aggregates and should contain a population with the same protein conformation. In order to obtain satisfied structural data, only best quality crystals should be chosen.
2.3.4 X-ray diffraction by crystals

Bragg’s law (Bragg and Bragg, 1913) interprets X-ray diffraction by a crystal lattice as a conjunct of reflections from different planes of atoms in the crystal (Figure 2.4). For a constructive interference, this can be written as follows:

\[ 2d \sin \theta = n \lambda. \]

where \( d \) is the distance between successive planes of atoms, \( \theta \) is the angle of incidence of the X-rays that equals the angle of reflection, \( n \) is an integer, and \( \lambda \) is the wavelength of the X-rays (usually corresponding to the Cu K\( \alpha \) radiation = 1.5418 Å. Thus, as Bragg pointed out, X-ray diffraction can be regarded as the reflection of the beam of X-rays from the planes of points in the crystal lattice.

**Figure 2.4.** Scheme explaining Bragg’s law. X-rays (\( X_1, X_2 \)) that are reflected by lattice planes (A, B) with distance \( d \) have a difference in path length that is equal to \( 2d \sin \theta \). A prerequisite for constructive interference is, that this difference in path is an integer multiple \( n \) of the used wavelength \( \lambda \).

2.3.5 How to solve the phase problem?

When waves are diffracted from a crystal, they give rise to diffraction spots. Each diffraction spot corresponds to a point in the reciprocal lattice and represents a wave with the amplitude and a relative phase. Knowledge of both, amplitudes and phases, allows the reconstitution of the electron density of the crystal. The amplitudes can be deduced from the intensities of the diffracted X-rays but the phases cannot be directly measured. This is known as the “phase problem”.

75
To overcome this problem several approaches are applicable:

- Molecular Replacement (MR)
- Single Isomorphous Replacement (SIR)
- Multiple Isomorphous Replacement (MIR)
- Multiple-Wavelength Anomalous Dispersion (MAD)
- Direct methods

The method of **Molecular Replacement (MR)** depends on the availability of a sufficiently homologous model structure. That is, if the structure of a protein (search molecule) is known that is homologous to the crystallized protein (target molecule) with unknown structure, the former can be used as a model to calculate a starting set of phases that can be interactively refined (Hoppe, 1957; Huber, 1965).

If successful, a preliminary model of the target structure will be obtained by correctly orienting and positioning the search molecule in the target cell. This solution can be optimised by rigid body refinement. Finally, the target structure can be put through cycles of map calculation, model fitting and refinement, which help to reduce the bias introduced by the starting model. Even in the absence of a suitable search molecule, the self-rotation function can be used to determine the direction and nature of non-crystallographic symmetry elements. Generally there are two steps in molecular replacement and these are known as the rotation and translation functions. In the rotation step, the spatial orientation of the known and unknown molecule with respect to each other is determined, while in the next step the translation, which is needed to superimpose the now correctly oriented molecule onto the other is calculated. The basic principle of the MR method can be understood with help of the Patterson function of a protein crystal structure. The Patterson map is a vector map: vectors between atoms in the real structure show up as vectors from the origin to maxima in the Patterson map. If the pairs of atoms belong to the same molecule, then the corresponding vectors are relatively short and their end-points are found not too far from the origin in the Patterson map; they are called self-Patterson vectors. If there were no intermolecular vectors (cross-Patterson vectors), this inner region of the Patterson map would be equal for the same molecule in different crystal structures, apart from a rotation difference. For homologous molecules it is not exactly equal but very similar. Therefore, the self-Patterson vectors can supply us with the rotation relationship between the known and the unknown molecular structures. The solutions of the rotation and translation functions are not always found in a
straightforward way. In some cases, it can be necessary to modify the model, for instance, by ignoring the side chains and deletions/insertions in the model, and/or by systematically varying the resolution range of the X-ray data used in the search. With the rapid increase in the number of successful protein structure determinations, molecular replacement has become an extremely used technique for protein phase angle determination.

Without any previous knowledge of the structure **Multiple Isomorphous Replacement (MIR)** is still the most commonly used method. In this method, the idea is to make a change to the crystal that will perturb the structure factors and, by the way that they are perturbed, to make some deductions about possible phase values. It is necessary to be able to explain the change to the crystal with only a few parameters, which means that the heavy atoms must be used. The way to introduce such an atom is either soaking the crystal or co-crystallization along with the protein. Soaking with the help of the large solvent channels present in protein crystals is usually the method of choice.

If there are two crystals, one containing just the protein (native crystal) and one containing in addition bound heavy atoms (derivative crystal), diffraction data from both can be measured. The differences in scattered intensities will largely reflect the scattering contribution of the heavy atoms, and these differences can be used to compute a Patterson map. Because there are only a few heavy atoms, such a Patterson map will be relatively simple to deconvolute.

If the heavy atom doesn't change the rest of the structure (this is what the term "isomorphous" refers to), then the structure factor for the derivative crystal ($F_{PH}$) is equal to the sum of the protein structure factor ($F_P$) and the heavy atom structure factor ($F_H$)

$$F_{PH} = F_P + F_H$$

**Figure 2.5.** A vector diagram illustrating the native protein structure factor ($F_P$) and heavy atom structure factor ($F_H$) contributions to the structure factor for the derivative crystal ($F_{PH}$).
Because the structure factors can be thought of as vectors, then equation defines a triangle. There are two ways to construct such a triangle, which means that there are two possible phases ($\alpha_P$) for $F_P$, which cannot be distinguished with a single isomorphous derivative. The phase ambiguity for acentric reflection can be illustrated geometrically in a Harker diagram (Harker, 1956) as shown in Figure 2.6.

**Figure 2.6.** The Harker diagram for protein phase determination. 

A. Case where only one heavy atom is present. A circle of radius $|F_P|$ representing the native protein is drawn (black) and from its centre the vector $-F_H$. A second circle (in red) with radius $|F_{PH}|$ is drawn at the endpoint of $F_H$. The intersections of the two circles correspond to two equally probable protein phase angles. Both triangle fulfil the condition $F_{PH} = F_P + F_H$

B. Case where two different heavy atoms are present. In this situation it is easier to elucidate which is the correct solution because the drawn circles ideally intersect at the same position.

The ambiguity of the phase is theoretically solved when a second, unrelated heavy atom derivative is obtained. In practice the experimental data are affected by a number of errors arising from the crystal growth procedure, the soaking (or co-crystallisation) process, non-isomorphism, and errors derived from the collection of X-ray diffraction data. To obtain an interpretable electron density map usually more than two isomorphous heavy derivatives are necessary. Additionally, the phasing ambiguity can be solved by incorporation of
anomalous scattering data from different wavelength. By density modification methods (e.g. solvent flattening) or by averaging the calculated electron density when non-crystallographic symmetry is present, phases can be improved.

The anomalous scattering effect depends on the frequency of oscillation being similar to the natural frequency for the atom. So clearly the strength of the anomalous scattering effect depends on the wavelength of the X-rays, which will change both the normal scattering and the out-of-phase scattering of the anomalous scatterers. By collecting data at several wavelengths near the absorption edge of an element in the crystal, the phase information analogous to that from MIR is obtained. This technique is called MAD, for multiple-wavelength anomalous dispersion. One popular way to use MAD is to introduce selenomethionine in place of methionine residues in a protein. The selenium atoms (which replace the sulfur atoms) have a strong anomalous signal at wavelengths that can be obtained from synchrotron X-ray sources.

*This introduction was prepared based on Protein Crystallography Course presented to staff and students of CIMR, University of Cambridge and the MRC-LMB. For more details about crystallization procedures please refer to McPherson (1999) and for the basis of X-ray crystallography to Drenth (1994) or Giacovazzo (1992).
2.4 Expression and purification of CAS SH3 domain

The CAS SH3 domain was expressed in *E.coli* strain BL21 (DE3) as a fusion protein with glutathione S-transferase. Bacteria were grown in 2x YTA medium with 100 µg/ml ampicillin and protein expression was induced with 1 mM isopropyl-β-thiogalactopyranoside (IPTG). Bacterial pellets were suspended in 1x PBS pH 7.4, 1 mM EGTA, 1 mg/ml aprotinin, 100 mM PMSF and lysed by means of French Pressure Cell technique (SLM-Aminco). GST-SH3 fusion protein was purified using GSH-Sepharose (Amersham Biosciences). Bound GST-SH3 was eluted with 10 mM glutathione, 50 mM Tris pH 8.0. In order to remove the GST tag, thrombin cleavage was performed. Subsequently, SH3 domain was purified by size exclusion chromatography (Sephadex-75, Amersham Biosciences) and concentrated for crystallization experiments. The purity and monodispersity of the protein were confirmed by SDS PAGE electrophoresis as well as by ESI-MS.

2.5 Crystallization of the CAS SH3 domain

2.5.1 Co-crystallization trials

The best crystallization conditions previously found for native CAS SH3 domain were used in the initial co-crystallization experiments.

<table>
<thead>
<tr>
<th>Condition A</th>
<th>Condition B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1 M MES, pH 6.5</td>
<td>0.1 M acetate, pH 4.5</td>
</tr>
<tr>
<td>1.4 M sodium acetate trihydrate</td>
<td>2 M sodium formate</td>
</tr>
</tbody>
</table>

2.5.1.1 Compounds-putative inhibitors for CAS SH3

Chemical compounds, in concentration of 10mM, were kindly provided by Roche. In the preliminary crystallization trials compounds were diluted 5x in crystallization buffer (condition A or condition B) to obtain the final concentration of ~2mM. Crystallization was carried out by means of sitting drop and hanging drop vapour diffusion methods by mixing equal volumes of the protein and inhibitor solution (at a 1:2 molar ratio). Because the crystals appeared only in the ‘native’ drops, another co-crystallization conditions were tested.
Inhibitor solution (10 mM stock in DMSO) was mixed with protein solution in such a way that molar ratio 1:2, 1:2.5 and 1:3 was obtained for protein and inhibitor, respectively. Subsequently, the screening procedure was performed using 16 different crystallization buffers. Crystals of different morphologies (Figure 2.7) were obtained from several conditions at room temperature with the sitting drop vapour diffusion method by mixing equal volumes of the protein and reservoir solution (1µl each). The best conditions (0.1 M acetate, pH 4.5, 0.2 M MgCl₂, 30% PEG 8000) were further optimized in order to achieve the crystals of good quality.

Figure 2.7. Crystals of the CAS SH3 domain. They grew from: 0.1 M acetate, pH 4.5, 0.2 M MgCl₂, 30% PEG 8000 (C), (D), (E), (F); 2 M (NH₄)₂SO₄ (A); 0.1 M acetate, pH 4.5, 2M sodium formate (B). B, D and E show the crystals of protein produced in the RTS.
2.5.1.1.1 Soaking with compounds

In the initial soaking experiments, native crystals of CAS SH3 domain were soaked by direct addition of inhibitor solution to the crystallization wells. Unfortunately, because the working inhibitor solutions were prepared in DMSO, the crystals cracked and dissolved.

2.5.1.1.2 Cross-linking trials

Cross-linking of protein by glutaraldehyde was performed in the following manner. 1-2µl of 0.25-0.5% glutaraldehyde was added to the wells containing the native protein crystals. After ~1h incubation, 1µl of the inhibitor (10mM in 100% DMSO) was added to the drops with cross-linked crystals which remained solid. In comparison to the previous soaking trials, crystals did not crack and/or dissolve (they remained stable for several weeks even in 100% DMSO and diffracted.).

2.5.1.2 Peptide inhibitors

Two peptides from focal adhesion kinase (FAK), in concentration of 1mM each, were used in co-crystallization experiments. Protein was mixed with peptide ligand EAPPKPSRP (designed as P1) or EAPPKPSRPGYPSP (designed as P2) at a 1:1 or 1:2 molar ratio and equilibrated at 4°C for several hours before setting up crystallizations. Crystallization was achieved at room temperature using the sitting drop technique by mixing equal volumes of the protein and reservoir solution consisting of 0.1 M acetate, pH 4.5, 0.2 M MgCl₂, 30% PEG 8000 (condition 1) and 2 M (NH₄)₂SO₄ (condition 2). Monoclinic crystals appeared using the first conditions within several hours, whereas orthorhombic crystals appeared with the second conditions after 2-3 weeks.

2.5.2 Crystallization of the CAS SH3 domain expressed in the RTS

CAS SH3 domain produced in the RTS was successfully crystallized using the sitting drop vapour diffusion methods by mixing equal volumes of the protein and reservoir solution consisting of 0.1 M acetate, pH 4.5, 2M sodium formate.
2.6 Results and discussion

2.6.1 Data collection and structure determination

The best X-ray data set up to 1.1 Å resolution was collected from a monoclinic crystal at the MPG/GBF beamline BW6 at DESY, Hamburg. For protection, the measured crystal was earlier soaked for 1-2 min in a solution containing 80 % (v/v) reservoir solution and 20 % (v/v) ethyleneglycol. Other data sets (resolution in range 1.3-2.1 Å) were collected from the monoclinic and orthorhombic crystals at the Swiss Light Source (SLS), Villigen, Switzerland and at the European Synchrotron Radiation Facility (ESRF), Grenoble, France, as well as at a rotating anode laboratory source. In the last case, the crystal was measured using capillary system. The monoclinic and orthorhombic crystals provide high resolution structures of the CAS SH3 domain in dimeric (dCAS-SH3) and monomeric (mCAS-SH3) asymmetric unit arrangements, at 1.1 and 2.1 Å resolutions, respectively (Table 2.1). The structure of the $C222_1$ crystal as well as $P2_1$ was determined by molecular replacement using MolRep from the CCP4 program suit. As search model the $\alpha$-spectrin (PDB ID: 1SHG) was chosen. Both CAS SH3 models were refined to a crystallographic R-factor of 20.0 % and 17.6 %, respectively with good agreement to standard stereochemical parameters. The Ramachandran plots shows that all residues have $\phi/\psi$ angles within allowed regions (Figure 2.8 and Table 2.1). The final dCAS-SH3 model consists of 742 protein atoms, one magnesium ion, one ethylene glycol molecule and 165 water molecules whereas the mCAS-SH3 comprises 394 protein atoms and 35 water molecules. As it turned out, neither compound nor peptide bound to the SH3 domain of CAS. It should be emphasized that the electron densities of all but two residues within the solved structures are extremely well defined; the C-terminal residues Lys68B and Pro69B of the dimer are flexible, with weaker electron density and high thermal parameters. Comparisons of the monomer and the two dimer structures show only few differences (discussed below).

Figure 2.9. The electron density map (1.1 Å resolution, contoured at 1σ above the mean) around the tryptophan 41 residue. The atom colours are as follows: carbon, yellow; nitrogen, blue; oxygen, red. The figure was produced with Raster3D (Merrit and Bacon, 1997).
Figure 2.8. Ramachandran plots for the refined mCAS-SH3 and dCAS-SH3 structures.

Table 2.1. Data collection and refinement statistics

<table>
<thead>
<tr>
<th>Data collection</th>
<th>monomer</th>
<th>dimer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C2221</td>
<td>P21</td>
</tr>
<tr>
<td>Cell (a, b, c) (Å)</td>
<td>38.7, 70.7, 57.0</td>
<td>38.0, 56.2, 38.0</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>35.4 – 2.1</td>
<td>34.1 – 1.1</td>
</tr>
<tr>
<td>Completeness (%) [last shell]</td>
<td>99 [99]</td>
<td>93 [90]</td>
</tr>
<tr>
<td>I/σ(I) [last shell]</td>
<td>5.8 [1.6]</td>
<td>4.2 [1.8]</td>
</tr>
<tr>
<td>R_{sym} [last shell]</td>
<td>0.093 [0.417]</td>
<td>0.073 [0.40]</td>
</tr>
</tbody>
</table>

Refinement

| Reflections used                | 4620          | 46200         |
| R-factor (%)                   | 20.0          | 17.6          |
| Free R-factor (%)              | 22.7          | 19.1          |
| Free R-value test size (%)     | 4.7           | 5             |
| Number of protein molecules    | 1             | 2             |
| Number of protein residues/atoms | 69/394       | 139/742       |
| Number of solvent atoms        | 105           | 495           |
| Other atoms or molecules       | Mg, EGL       |               |
Standard deviation from ideal values

<table>
<thead>
<tr>
<th>Bond length (Å)</th>
<th>0.017</th>
<th>0.010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond angles (°)</td>
<td>1.64</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Ramachandran Plot, the percentage of residues in the regions of plot (%)

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Most favoured</td>
<td>94.7</td>
<td>93.9</td>
</tr>
<tr>
<td>Additionally allowed</td>
<td>5.3</td>
<td>6.1</td>
</tr>
</tbody>
</table>

2.6.1.1 Data collection- crystals of the protein produced in the RTS

X-ray data sets collected at the Swiss Light Source (SLS) indicated that the crystals produced by means of RTS were of very high quality. They diffracted to a resolution limit of 1.7 Å (Figure 2.10) that is comparable to the resolution obtained with crystals of CAS SH3 domain expressed in *E.coli* system.

![Figure 2.10.](image)

Figure 2.10. Diffraction pattern of the CAS SH3 domain crystals. The resolution at the edge of the images is 1.7 Å.
2.6.2 Molecular structure

The SH3 domain of CAS is formed from five anti-parallel β-strands arranged as two orthogonal β-sheets (Figure 2.11). Each sheet consists of three strands; one long strand contributes different segments to each sheet. Naming the strands sequentially as β1, β2, β3, β4 and β5, one of the sheets is formed by β2, β1 and β5 and the other by β2, β3 and β4. We observe additionally that the residues C-terminal to the standard definition of the SH3 domain region form a strand β6.

![Figure 2.11. Stereo ribbon representation of the CAS SH3 domain structure. The core of the domain consists of two perpendicular, anti-parallel, three-stranded β-sheets. Naming the strands sequentially as β1, β2, β3, β4 and β5, one of the sheets is formed by β2, β1 and β5 and the other by β2, β3 and β4. Positions of the loops are also indicated.](image)

The first two strands (β1 and β2) are linked by the RT loop (named for the RT residues found here in src-SH3, residues 10 to 25), which forms a hairpin-like structure. This loop is stabilized by the extensive network of intra-loop hydrogen bonds as well as by the loop-protein core hydrogen bonds (Table 2.2). These stabilize the structure of the RT loop (see discussion of flexibility, below) and is important for the function of the SH3; the most highly conserved SH3 domain residues are located on it. Two of these (Tyr10, Glu19) are
likely involved in protein-peptide ligand interactions. In the asymmetric dimer crystal packing arrangement, two monomers interact via their two RT loops with six hydrogen bonds involving side-chain atoms of Asp11A, Asp11B, Glu15A, Arg23A, Arg23B with the main-chain atoms of Val13A, Val13B and Ala14A and amide of Arg23A (Table 2.2). Additionally, a hydrated magnesium ion interacts in this region, which was present at a 200 mM concentration (as MgCl$_2$) in the crystallization buffer. The magnesium ion is bound asymmetrically between two chains with most interactions to chain B. The residues involved in the metal binding site (Asp11 from monomer A and Asn12, Val13, Glu15, Glu19 from monomer B) are at hydrogen bond distances to water molecules, which surround the ion at the average distance of 2.1 Å (Figure 2.12 and 2.13).

![Figure 2.12.](image)

**Figure 2.12.** Magnesium binding site. Residues involved in the metal binding site (Asp11A and Asn12B, Val13B, Glu15B, and Glu19B are shown as sticks. Chain A, cyan; chain B, blue; magnesium ion, yellow; water molecules, red. Hydrogen bonds are indicated as dashes lines.)
Figure 2.13. Stereo view of a representative part of electron density map around the magnesium binding site. The electron density map (1.1 Å resolution, contoured at 1σ above the mean) corresponds to residues Asp11A and Glu15B and six coordinating water molecules. The atom colours are as follows: carbon, yellow; nitrogen, blue; oxygen, red; magnesium, cyan. The figure was produced with Raster3D (Merrit and Bacon, 1997).

The second strand (β2), comprising residues 26-34, is connected with β3 strand by a short n-src loop, which is stabilized in both chains by several hydrogen bonds. For example Thr35 forms H-bonds with Leu38, Gly37 and Trp41. The next loop, linking strands β3 and β4, is known as a distal loop and creates Iβ turn. Other two β strands, as well as loops which connect them, are shorter and their architecture is stabilized by the network of hydrogen bonds.

Table 2.2. Hydrogen bonds involving residues on the RT loop

<table>
<thead>
<tr>
<th>Chain A RT loop- chain B RT loop</th>
<th>hydrogen bonds</th>
<th>Distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-V13B</td>
<td>OD1-D11A</td>
<td>2.85</td>
</tr>
<tr>
<td>NH2-R23A</td>
<td>O-A14B</td>
<td>2.87</td>
</tr>
<tr>
<td>N-V13A</td>
<td>OD1-D11B</td>
<td>3.05</td>
</tr>
<tr>
<td>NH1-R23B</td>
<td>OE1-E15A</td>
<td>3.3</td>
</tr>
<tr>
<td>NH1-R23B</td>
<td>OE2-E15A</td>
<td>3.1</td>
</tr>
<tr>
<td>NE-R23A</td>
<td>O-V13B</td>
<td>3.1</td>
</tr>
</tbody>
</table>
## Intra-loop hydrogen bonds

<table>
<thead>
<tr>
<th>Donor</th>
<th>Acceptor</th>
<th>Distance (Å)</th>
<th>Chain A</th>
<th>Chain B</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-D18</td>
<td>O-S16</td>
<td>3.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-E19</td>
<td>O-S16</td>
<td>3.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ND2-N12</td>
<td>O-L20</td>
<td>2.93</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-F22</td>
<td>O-N12</td>
<td>3.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-N12</td>
<td>O-F22</td>
<td>2.83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-K24</td>
<td>OD2-D11</td>
<td>2.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OG-S16</td>
<td>OE1-E19</td>
<td>2.77</td>
<td></td>
<td>-----</td>
</tr>
<tr>
<td>N-E15</td>
<td>OE2-E19</td>
<td>-----</td>
<td></td>
<td>2.92</td>
</tr>
<tr>
<td>ND2-N12</td>
<td>OE1-E19</td>
<td>-----</td>
<td></td>
<td>2.86</td>
</tr>
</tbody>
</table>

## Loop-protein core hydrogen bonds

<table>
<thead>
<tr>
<th>Donor</th>
<th>Acceptor</th>
<th>Distance (Å)</th>
<th>Chain A</th>
<th>Chain B</th>
</tr>
</thead>
<tbody>
<tr>
<td>N-G25</td>
<td>O-A8</td>
<td>2.9</td>
<td></td>
<td>2.85</td>
</tr>
<tr>
<td>NH1-R57</td>
<td>OD1-N12</td>
<td>3.13</td>
<td></td>
<td>3.17</td>
</tr>
<tr>
<td>NH1-R57</td>
<td>O-D11</td>
<td>-----</td>
<td></td>
<td>2.57</td>
</tr>
<tr>
<td>N-I52</td>
<td>O-D18</td>
<td>2.81</td>
<td></td>
<td>2.82</td>
</tr>
</tbody>
</table>

### 2.6.3 Water molecules

The crystal structure of the dimeric form of CAS SH3 domain has been refined at 1.1 Angstroms including 165 water molecules. Some appear in the interior of the protein molecule and link the fold. For example, waters are involved in connecting the following pairs of residues: Gln50 and Pro17, Glu15 and Val13 as well as Asn3 and Leu31. The presence of the six water molecules in the ion binding area was indicated (Figure 2.12 and 2.13). The crystal structure of monomer SH3 domain has been refined with 35 water molecules. As with the dimeric CAS crystal structure, three water molecules are interior to the fold and link disparate fold elements. However, in the monomer crystal structure, these are
found at different positions and bridge different protein atoms or different residues. In the mCAS-SH3 waters link Met64 with Leu61, Glu19 with Asn12 and Glu19 with Glu15. The first bridge can be important for stabilizing the loop near C-terminus (between strands 5 and 6) of the molecule, whereas the two remaining bridges can stabilize the side chains of both glutamic acids. In case of the Glu19 it is exceptionally important because this residue seems to be involved in protein-ligand interactions.

An interesting feature of the dimer structure is the presence of an ethylene glycol molecule, which was used as a cryoprotectant. Two oxygen atoms of ethylene glycol form two hydrogen bonds with residue Gln50 of chain A, with distances 2.8 and 2.7 Å, respectively.

2.6.4 Crystal packing effects

The role of the SH3 domain of CAS as a docking molecule involved in numerous protein-protein interactions is well established (Polte and Hanks, 1995; Burnham et al., 1996; Garton et al., 1997; Kirsch et al., 1998). Such interactions generally have the potential to modulate to the protein fold. Therefore, the observation of different crystal packing arrangements offers an opportunity to gauge the plasticity or deformability of the fold. As was mentioned above, in the dCAS-SH3 crystal form the main interaction interface forms the asymmetric dimer and involves the RT loops. All other crystal contacts are due to symmetry related molecules. In the dSH3 we can observe approximately 420 atom-atom contacts (defined as distances less than 4 Å, excluding water molecules), while in the mSH3 the number of these contacts is distinctly lower (140). Some of them are hydrogen bonds (48 and 20 H-bonds in dSH3 and mSH3, respectively). In the monomer, the C-terminal residues including Val62, Gly63, Tyr65, Asp66 and Lys67, as well as amino acids from the other regions of the molecule (Glu15 and Asp18 from RT loop, Gln33 from strand 3, Gln36 from n-src loop, and Gln50 and Ile60 from strands 4 and 5, respectively) are involved in hydrogen bonds.

2.6.5 Comparing CAS SH3 structures from the three different packing environments

The general folds of the two CAS SH3 domains in the dimeric arrangement (dCAS-SH3) and the monomeric domain (mCAS-SH3) are essentially identical, including the side-chain orientations, with several notable exceptions (Figure 2.15).
Figure 2.14. The CA traces of the superposed SH3 domains of CAS (green, cyan and blue indicate the monomer, chain A and B, respectively), PI3K (orange) and α-spectrin (red).

The two chains of dSH3 show an overall RMS deviation for the main-chain atoms of 0.69 Å. The biggest differences are observed at C-terminus (K68 and P69 with 4.4 and 5.6 Å, respectively), the n-src loop (G37 with 1.7 Å) and the first portion of RT loop (E15 with 1.9 Å) (Figure 2.15). (The deviations for K68 and P69 should be understood in light of the apparent intrinsic flexibility of this segment as shown by the poorly defined electron density map here. If these residues are excluded, the total RMS deviation is lowered to 0.56 Å.) The variations in the position of Glu15 may be a consequence of the hydrogen bond between Glu15A and R23B in dCAS-SH3. Two other residues, Arg49 and Gln50 from β4 strand, also display the relatively high values of RMS deviations (0.94 Å and 0.91 Å, respectively) (Figure 2.15 and 2.16). The comparison of the monomer with each chain of the dimer demonstrates a quite high structural alignment with an overall RMS deviation for main chain atoms of 0.94 Å (chain A) and 0.6 Å (chain B). In both cases, major differences are noticed at n-src loop (residues Gly37, Leu38, Asp39) and in position of Gly48, which is localized at the distal loop. Additionally, the comparison of the mSH3 with chain A of the dSH3 shows the distinct deviations in positions of Glu15 and Ser16, whereas the comparison of the mSH3 with chain B shows the highest deviations at C-terminus (~5 Å) and first portion of β4 strand. The different pattern of hydrogen bonds in chain B in this region is reflected by the large value of the RMSD. First of all, the H-bond between Arg49 and Pro17 formed in both chain A and the monomer is absent in chain B. Further, Gln50 of chain A makes hydrogen bonds...
with ethylene glycol molecule, whereas Gln50 of the monomer forms two H-bonds with the residue Gln33 of the symmetry related molecule. The average temperature factors (B-factors) of main chain atoms in dCAS-SH3 are essentially the same (12.4 Å² for chain A, and 12.7 Å² for chain B), whereas it is distinctly higher (39 Å²) in mCAS-SH3. In each case, the highest B-factors of main chain atoms are observed in the loops regions, but only in chain A is it higher at the distal loop than at the n-src loop. Functional interactions may require such structures, that are generally stabilized by H-bonds but retain the flexibility necessary for adaptation to binding partners. A dramatic example of such functional flexibility was observed in the case of the Eps8 SH3 domain, which formed a strand-exchanged intertwined dimer (Kishan et al., 1997, 2001).

![Figure 2.15. RMS deviations for main chain atoms between the chain A and chain B, chain B and monomer, and chain A and monomer, respectively. The RMS deviations were calculated for each residue of SH3 domain by the program Superpose Molecules of the CCP4 crystallography package.](image-url)
Figure 2.16. Superposition of the residues which show the highest RMSD values. There are Glu15, Ser16, Gln36, Gly37, Leu38, Asp39, Gly48, Arg49 and Gln50 from the monomer (green), chain A (magenta) and chain B (blue) of the dimer SH3 domain, respectively.

Figure 2.17. Superposition of the conserved residues showing the low RMSD values. There are Tyr10, Glu19, Trp41 and Pro54 from the monomer (green), chain A (magenta) and chain B (blue) of the dimer SH3 domain, respectively.

Additionally, the CAS SH3 domain was compared with two other crystal structures of SH3 domains: α-spectrin and phosphatidylinositol 3-kinase p85 subunit (PI3K) (Figure 2.14). Due to its sequence similarity to CAS SH3 domain, α-spectrin (Musacchio et al., 1992) was chosen and successfully used in the molecular replacement solution of the structure. Comparisons with mCAS-SH3 and with each chain of the dCAS-SH3 highlights its structural
similarity, with an overall RMS deviation for main chain atoms of ~1.1 Å (220 atoms were considered). The lowest values of RMSD are at the regions of the highest sequence homology (β1 strand, first portion of the β2, β4 and β5 strands, as well as the second portion of the RT loop). The biggest differences are observed at the n-src loop region, with its lack of conservation and occurrence of an insertion in CAS. The most highly conserved residues (Lys9, Tyr10 and Gln19, numbering for CAS) show a relatively low RMSD for main chain atoms, however, for conserved Trp41 and Pro54 they are anomalously high.

The SH3 domain of PI3K p85 subunit was chosen for the comparison studies because it binds to the same proline-rich sequences of the focal adhesion kinase (FAK) as the CAS SH3 domain (Zachary, 1997). As in case of the α−spectrin, the greatest structural deviations occur at the n-src loop. In the PI3K SH3 domain this loop is extremely long (15 residues) and contains additionally one regular α−helix as well as two 3_10 helices (Liang et al., 1996). According to Yu (1994), two classes of proline-rich peptides can interact with the PI3K SH3 domain. FAK, the natural binding partner for both SH3 domains, possess the consensus sequence PXFXPF (PKPKPSR) termed as a class II site. According to its definition, class II binds to the SH3 domains in an amino to carboxyl (N to C) orientation. We suggest that the CAS SH3 domain can interact with its binding partner in the same way as discussed below.

### 2.6.6 Implications for protein-ligand interactions

Several X-ray and NMR studies of SH3 domains (i.e. α-spectrin, Crk, Ab1, Fyn, Grb2, PI3K) have characterized the general SH3 folding and ligand binding pattern. Despite relatively low sequence homologies between different SH3 domains (Figure 2.18), all share an overall topology of two small β sheets packed against each other at approximately right angles and three variable loops (a long hairpin RT loop, n-src loop and distal loop).
Figure 2.18. Structural and sequence alignment of several SH3 domains. There are: human p130 Cas, N-terminal SH3 domain (residues 1-63), chicken Src-tyrosine kinase transforming protein (residues 79-146), human Crk-like protein, N-terminal SH3 domain (residues 125-181), human Lck tyrosine kinase (residues 51-119), human Fyn tyrosine kinase (residues 84-142), chicken spectrin (residues 2-62), Ab1 tyrosine kinase (residues 84-141), human Grb2 protein, N-terminal SH3 domain (residues 1-59), p85 subunit of human phosphatidylinositol 3-OH kinase (residues 1-80). The most highly conserved residues are shown with a green background. The blue colour indicates generously conserved amino acids. Key: o, alcohol; l, aliphatic; a, aromatic; c, charged; h, hydrophobic; -, negative; p, polar; s, small; u, tiny; t, turnlike; ., any.

RT and n-src loops are situated on the ligand-binding face of the domain and are crucial for modulation of protein-ligand interactions. The distal loop lies on the opposite site and can interact with other regions of the SH3 containing or PRM containing molecule but its precise role is still unknown (Mayer, 1995). The notably conserved residues, in particular Lys9, Tyr10, Gln19, Trp41 and Pro54 (numbering for CAS), form the hydrophobic peptide ligand binding cleft. A number of previous structural studies have shown that these clefts surrounded by charged residues are very similar in different SH3 domains, underscoring the fact the SH3 domains bind a common peptide motif. Indeed, as mentioned above, the SH3 domains interact with proline-rich ligands (PXXP, or ΦPpXP, whereby Φ represents any hydrophobic residue and p represents moderate conservation as proline) that adopt the polyproline type II helix (PPII) conformation. This structural motif provides a common mechanism for the molecular recognition via SH3 domain binding. Peptides that have this conformation form an extended molecular scaffold that is well suited for the grooves of the SH3 ligand-binding site (Yu, 1994). Two different classes of the proline-rich ligands are commonly distinguished. Class I has the general consensus sequence +xΦPxΦΦ (where + represents a positively charge residue) and binds to the SH3 domains in a carboxyl to amino (C to N) orientation, whereas class II has the general consensus sequence ΦPxΦPx+ and binds in the opposite orientation (N to C). Because ligand sequences can be highly symmetrical, it is often possible that a SH3 domain binds different ligands in opposite orientations. This was observed in case of the interactions between SH3 domain of Src and the peptides: RPLP_ and the “reversed” APPLPR, as well as between SH3 domain of PI3K and the peptides RXLP and XPPLXP (conserved prolines are underlined, X denotes any amino acid) (Feng, 1995; Liang, 1995). Generally, the orientation of ligand binding is determined largely by the presence of a positively charged residue, which can be either N- or C-terminal to the core of the proline rich motif (PRM). We propose here that SH3 domain of CAS interacts
with its natural ligand in an N to C orientation because a positively charged arginine of the ligand is localized in the C-terminus. In our model (Figure 2.19, 2.20 and 2.21) we used the peptide from FAK protein (sequence EAPPKPSRP), which binds CAS SH3 domains with a ca. 6 micromolar affinity as shown by fluorescence correlation spectroscopy (FCS). These affinity assays also confirmed the crucial role of the conserved tryptophan in the protein-peptide interactions, as mutation of the tryptophan to lysine in position 41 weakened the binding between the ligand and CAS SH3 domain by an order of magnitude. A longer FAK peptide (sequence EAPPKPSRPGYPSP) also binds with greater affinity (2 micromolar) than its shorter counterpart, implying that the residues surrounding the proline-rich core are also important for the protein-peptide interactions. These additional flanking amino acids can bind to less conserved portions of the SH3 surface, increasing the binding affinity and sequence selectivity.

In our model of the CAS SH3-peptide interactions we hypothesize that the two pairs of amino acids Ala FP2-Pro FP3 and Lys FP5-Pro FP6 (FP denotes FAK peptide) bind to the protein clefts termed pocket 1 (or site 1, S1) and pocket 2 (or site 2, S2), respectively. In all known structures of SH3 domain, these pockets have the hydrophobic and aromatic character of the conserved constituting residues of tyrosine and tryptophan. For example, in the SH3 domain of PI3K S1 is composed of the two tyrosine (Tyr12, Tyr73) and one leucine (Leu11) residues, whereas S2 is formed by the two tyrosine (Tyr14, Tyr73) and one tryptophan (Trp55) residues. In addition, proline (Pro70), threonine (Thr72) and glycine (Gly54) residues line the pocket (Liang, 1996). Concerning the CAS SH3, the architecture of these pockets seems to be similar, while their chemical character is uniquely polar due to the exchanges of the conserved tyrosine (sometimes also phenylalanine) residues to asparagine at position 12 and to arginine at position 57. Although arginine 57 would seem to place a repulsive charge near the Lys FP5, and additional charged residue, glutamic acid 15, is also in this region and might provide the counter charge to Lys FP5. Position 15 is highly variable among SH3 domains, with a negative charge occurring infrequently, but notably also in PI3K (Figure 2.18), which like CAS binds a PXKP consensus sequence. In addition to the hydrophobic contacts involving mainly the prolines, we predict the following interactions to occur between the FAK peptide and the SH3 domain. At the N-terminal side of the FAK binding peptide, an intermolecular hydrogen bond should exist between the side chain of Tyr10 and the backbone carbonyl of the Glu FP1. SH3 domains bind to their ligands also via the pocket 3 (or site 3, S3), which seems to play a crucial role in binding specificity because of its structural diversity. S3 is formed by the highly conserved tryptophan Trp41 or Trp55 (numbering for
CAS and PI3K, respectively), negatively charged Glu19 (CAS) or Asp21 (PI3K), as well as Ser16 (CAS) or Arg18 (PI3K). In our model, Arg8 of the FAK peptide can make a salt bridge with the conserved glutamic acid residue (Glu19), which is localized in this pocket. Such a specific interaction was previously observed in other SH3-PRM complexes (i.e. C-Src SH3-APP12 complex) (Feng, 1995). Moreover, this orienting salt bridge was found to be an important specificity determinant for cellular proteins binding to SH3 domains (Weng et al., 1995). In addition, the indole NH of Trp41 is at the hydrogen bond distance to the carbonyl oxygen of Ser FP7 and the nitrogen of Pro FP9. The interactions between the CAS SH3 protein and the peptide may also be mediated by hydrogen bonds involving the side chains of Ser16 and Arg FP8, Asp39 and Pro FP9, as well as by the salt bridge between Lys24 and the N-terminal glutamic acid GluFP1.

**Figure 2.19.** The CA traces of the superposed CAS SH3-FAK peptide complex (magenta) and C-Src SH3-App12 complex (cyan). Model of CAS SH3-FP complex was prepared based on the structure of the SH3-App12 complex (Brookhaven Protein Data Bank entry code 1QWE, Feng et al, 1995). The positions of the peptide residues were modeled using XtalView/xfit (McRee, 1997) as well as WebLab ViewerPro.
Figure 2.20. Stereo view of the model of the FAK peptide bound to the CAS SH3 domain. Peptide residues are shown in green and numbered from 1 to 9. Residues that are involved in ligand binding are shown in ball and stick representation (the highly conserved amino acids are in cyan).
Figure 2.21. Interface of the CAS SH3-peptide complex interactions. SH3 domain is shown as a surface plot (residues in red, negatively charged; blue, positively; white, neutral), FAK peptide is shown in green.

2.6.7 General Conclusions

The crystal structures of the p130cas SH3 domains described here provide the basis for a structural understanding of the interactions between the p130cas protein and several of its many docking partners. The protein kinase FAK (focal adhesion kinase) phosphorylates p130cas via binding to the SH3 domain, the protein tyrosine phosphatases PTP1B and PTP-PEST also dock to the p130cas SH3 domain and dephosphorylate p130cas, while CAS may localize the guanine nucleotide exchange factor C3G to specific cell compartments, similarly via p130cas SH3 domain docking. These docking interactions occur via the SH3 domain ligand sequence of XXPXKPX. The lysine, unusual among SH3 ligand sequences, is shown here to likely bind via polar interactions in a network formed among others by Asn12, Glu15, and Arg57, residues either highly conserved as aromatics (Asn12 and Arg57) or non-negative residues (Glu15) in other SH3 domains. This provides an explanation of the selectivity of the CAS SH3 domain.

Strategies in SH3-directed drug discovery are generally based on peptidic ligands that contain a proline-rich sequence. Because neither the affinities nor selectivities of these peptides are typically sufficient for likely success as a therapeutic, modified ligands with
additional interactions are required. Several studies have demonstrated the efficacy of incorporation of non-peptide elements into SH3 domain ligands using combinatorial chemistry, but the resulting ligand affinities remain in the micromolar range (Vidal, 2001). To date there are no reports of non-peptidyl small molecule SH3 domain inhibitors. The atomic resolution crystal structures of this protein as reported here should assist in the discovery of effective SH3 domain inhibitors.
3 Summary

The thesis presented here is a result of the studies carried out in the Department of Structural Research at the Max Planck Institute for Biochemistry from June 2001 to December 2004. The work has been divided into two practically independent parts. The first project concerned the biochemical studies on IGF and IGF-binding proteins interactions in which kinetic, as well as structural investigations were undertaken. Additionally, the use of in vitro transcription/translation protein expression system was tested. The second project focused on X-ray structure determination of the SH3 domain of recombinant human Crk-associated tyrosine kinase substrate p130cas (CAS).

Insulin-like growth factor binding proteins (IGFBPs) control the extracellular distribution, function and activity of IGFs, which are known to be the strong mitogens for a wide variety of cancer cell lines such as sarcoma, leukaemia, and cancers of the prostate and breast. Therefore, IGFBPs regulate the processes of growth, differentiation, and survival mediated by IGF-I and –II. The knowledge of the structure and the mechanism of IGFs-IGFBPs interactions is essential for a true understanding of their function. The information obtained from NMR and crystallographic studies gave the opportunity to further investigation of these interactions. Based on the knowledge of the mini-IGFBP-5 structure, several mutants have been designed in order to modulate and preferably enhance the IGF binding to the mini-IGFBP-5. Because the previous NMR and crystallographic studies had shown that hydrophobic residues beyond the 9th cysteine (Cys47) in IGFBP-5 formed the core of the IGF binding site, four residues in that region were mutated.

Kinetic analysis, using two methods: surface plasmon resonance (SPR) and isothermal titration calorimetry (ITC), have shown that mini-IGFBP-5 mutants have lower binding affinities for IGF-I than mini-IGFBP-5 wild type, and ITC measurements allowed determination of the of binding constant values. Performed studies have definitively confirmed the suggestion made previously by Kalus et al. (1998) and Zeslawski et al. (2001), that a specific group of hydrophobic amino acids within the N-terminal part of IGFBP-5 is essential for IGF-I binding.

The rapid translation system is a high-yield protein expression system that uses an enhanced E. coli lysate for an in vitro transcription/translation reaction. In presented thesis,
the usefulness of RTS, by successful expression of soluble and active protein (IGFBP-4 full length), has been demonstrated.

In summary, the IGF system plays a central role in many aspects of tumorigenesis. A better understanding of this complex system will facilitate the development of novel approaches to diagnose and treat the various human cancers.

The Crk-associated tyrosine kinase substrate p130cas (CAS) is a docking protein containing an SH3 domain near its N-terminus, which mediates the interaction of the whole protein with several proteins involved in signaling pathways, such as focal adhesion kinase (FAK), tyrosine phosphatases PTP1B and PTP-PEST, and the guanine nucleotide exchange factor C3G. As a homolog of the corresponding Src docking domain, the CAS SH3 domain binds to proline-rich sequences (PXXP) of its interacting partners that can adopt a polyproline type II helix. A high resolution X-ray structure of the recombinant human CAS SH3 domain has been determined. The domain, residues 1-69, crystallized in two different space groups, $\text{P}2_1$ and $\text{C}22_1$, that provided diffraction data to 1.1 Å and 2.1 Å, respectively. The crystal structure shows the SH3 domain architecture of two roughly perpendicular, anti-parallel, three-stranded $\beta$-sheets. The conserved amino acids are located in close proximity to form the protein ligand binding surface which is atypically charged. The structure enables visualization of the docking interactions to its ligands, for example from focal adhesion kinase, and the role of the CAS specific binding residues

Because neither the affinity nor selectivity of known peptide inhibitors is satisfactory, some other ligands are required. It is interesting to point out that to date there are no reports of non-peptidyl small molecules inhibiting the activity of SH3 domains. Atomic resolution crystal structures of this protein as reported here should assist in the discovery of effective SH3 domain inhibitors.
4 Zusammenfassung


Das \textit{rapid translation system} (RTS) ist ein einen hohen Ertrag abwerfendes Proteinexpressionssystem, das ein erweitertes \textit{E. coli} Lysat für in vitro Transkription/Translation verwendet. Die vorliegende Doktorarbeit zeigt den Nutzen des RTS mit der Darstellung einer erfolgreichen Expression von löslichem und aktivem Protein (IGFBP-4, volle Länge).

Crk-assoziiertes Tyrosinkinase Substrat p130cas (CAS) ist ein Dockingproteing, das mit mehreren Proteinen interagiert, die an Signalwegen beteiligt sind, wie beispielsweise \textit{focal adhesion kinase} (FAK), Tyrosinphosphatasen PTP1B und PTP-PEST und der Guanin-Nukleotid Austauschfaktor C3G. Die CAS SH3 Domäne erkennt Prolin-reiche Sequenzen (PXXP) seiner Bindungspartner, die Polyprolin Typ II Helices bilden können. Eine hoch
aufgelöste Röntgenstruktur der rekombinant humanen CAS SH3 Domäne wurde in zwei Kristallformen, die bis zu 1.1 beziehungsweise 2.1 Å streuten, bestimmt. Diese mehrfachen hoch aufgelösten Strukturen bilden ein Rahmenwerk für Untersuchungen der CAS SH3 Domäne Protein-Protein Wechselwirkungen einschließlich Anwendungen für Drug Design.
5 References


Chapter 5


ITC manual


RTS instruction manual (2001)


6 Appendix

6.1 Amino acid sequences of the constructs used in a work

**MiniBP5 WT**
GSALAEQSCGVYTERCAQGLLCLPRQDEEKPHALLHGRGVCNLNEKSYREQVKI

**MiniBP5V49I**
GSALAEQSCGIGYTERCAQGLLCLPRQDEEKPHALLHGRGVCNLNEKSYREQVKI

**MiniBP5Y50K**
GSALAEQSCGVKTERCAQGLLCLPRQDEEKPHALLHGRGVCNLNEKSYREQVKI

**MiniBP5Y50A**
GSALAEQSCGVATERCAQGLLCLPRQDEEKPHALLHGRGVCNLNEKSYREQVKI

**MiniBP5R53Y**
GSALAEQSCGVYTEYCAQGLLCLPRQDEEKPHALLHGRGVCNLNEKSYREQVKI

**MiniBP5L73W**
GSALAEQSCGVYTERCAQGLLCLPRQDEEKPHAWLHGRGVCNLNEKSYREQVKI

**IGFBP-4**
Tag_factorXa_DEAIHCPPSEKLRACRPPVGVCEELVREPGGCATCALGLGMPCGVYTPRCCGSGLRCYPPRGEVKPHLTMHGQGVCMELAEIAIEAESLQPSDKDEGHPNNSFSPCSAHDRRCLQKHFAKIRDRTSGKMKVNGAPREDARPVGQSCQSELHRALELAASQSRTHEDLYIIIPCMCDRNGFHPQCHPALDGGRGKCVCVRKTGVKLPGLGEPKGELDCQLADSRE

Legend:
Tag_factorXa_MSMSGSHHHHHHTHSHSSGIEGR
Tag_factorXa_MSSWSHPQFEKIEGR
Tag_factorXa_MGLNDIFEAQKIEWHEVDIIEGR
Tag_factorXa_MSGSHHHHHHSSHSHMHSPILGYKIKGLVOPTRLLLEYLEEKEYEEHLYERDEGDKWRNKFFELGFPNLPYYIDGDKLTQSMAIRY1ADKHNMLGGCKPERAEISMLEGAVLDIRGYVSRAYSKEFETLKVDLSKLPELMKMFEDRLCHKTYLNGDHTHPDFMLYDADVLYMDPMCLDAFPKLVCFKRIEAIPIOIDKYLKSSKYIAWPIQGQATFGGDHPKSDMHSSNNNNNNNNNNLGIEGR

Colours as follows: red: His-tag, green: Strep-tag, violet: Avi-tag, blue: GST-tag, magenta: factorXa cleavage site, linkers are underlined
SH3 domain of p130cas

MSPILGYWKIKLVQPTRLLEELYEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPPYYIDGVKLTQSMAIRYIADKHMLMGCPKERAEMEGAVLDIRYGVSIAYSKDFETLKVDFLSKLPMLKMFEDRLCHKTYLNGDHTHPDFMLYDAVLYMDPMLDAPKLVCFKKRIEAPIQDKYLKSSKYIAWPLQGWQATFGGDHPKSLVLVRHLNVLAKALYDNVASEPDELSFRKGDMTVLEQDTQGDLGWLCMDKPRGQGVNRLKILYVMDKPG

Colourcode: black: SH3 domain, blue: GST-tag, orange: thrombin cleavage site, grey: additional sequence of full length protein

6.2 List of chemicals

- acetic acid
- acrylamide
- L-arginine
- aminohexanoic acid
- ammonium chloride, NH₄Cl
- ammonium persulfate, APS
- biotin
- calcium chloride, CaCl₂
- citric acid
- coomassie Brillant Blue R-250
- copper (II) chloride, CuCl₂
- desthiobiotin
- dimethylsulfoxide, DMSO
- dipotassium hydrogenphosphate, K₂HPO₄
- disodium hydrogenphosphate, Na₂HPO₄
- dithiothreitol, DTT
- ethanol
- ethylenediaminetetraacetic acid, disodium salt, EDTA
- ethyleneglycotetraacetic acid, EGTA
- N-ethylsulfonic acid morpholine, MES
- formaldehyde
- D-glucose
- glutardialdehyde
- L-glutathione, oxidized, GSSG
• L-glutathione, reduced, GSH
• glycerine
• glycine
• guanidinium chloride
• hydrochloric acid, HCl
• N-(2-Hydroxyethyl)piperazine-N’-(2-ethanesulfonic acid), HEPES
• imidazole
• isopropanol
• isopropyl-β-D-thiogalactopyranoside, IPTG
• magnesium chloride, MgCl₂
• magnesium sulfate, MgSO₄
• β-mercaptoethanol, β-ME
• methanol
• N,N’-methylenebisacrylamide
• phenylmethylsulfonylflouride, PMSF
• polyethylene glycol 400, 6000, 8000
• potassium chloride, KCl
• potassium dihydrogenphosphate, KH₂PO₄
• silver nitrate, AgNO₃
• sodium acetate trihydrate, CH₃COONa.3H₂O
• sodium azide, NaN₃
• sodium cacodylate, C₂H₆AsNaO₃.3H₂O
• sodium chloride, NaCl
• sodium dihydrogenphosphate, NaH₂PO₄
• sodium dodecylsulphate, SDS
• sodium formate, NaCOOH
• sodium hydroxide, NaOH
• surfactant P20
• N,N,N’,N’-tetramethylenethylenediamine, TEMED
• thiamin
• tricine
• tris-(hydroxymethyl)-aminomethane, TRIS
• triton X-100
• tryptone
• yeast extract
• zinc acetate trihydrate

6.3 Laboratory Equipment

Consumables:
Centripreps YM3: Amicon (Germany)
Centrifugal filter devices 5 kDa: Millipore (Germany)
Cryschem 24 well plates: Hampton Research (USA)
Crystal screens: Hampton Research (USA)
Dialysis tubing Spectra/Por MW 3500: Roth (Germany)
Falcon tubes, 15 ml, 50 ml: Becton Dickinson (Germany)
Gene pulser cuvettes: Biorad (USA)
Nitrocellulose membrane: Amersham
NMR-tubes, 5 mm: Wilmad (USA)
Parafilm: American National (Canada)
Pipette tips 10 µl, 200 µl, 1000 µl: Gilson (France)
Plastic disposable pipettes 1 ml, 5 ml, 10 ml, 25 ml: Falcon (Germany)
Reaction cups 0.4 ml, 1.5 ml, 2 ml: Eppendorf (Germany)
Sterile filters Millex 0.22 µm, 0.45 µm: Millipore (Germany)
Syringes 1 ml, 2 ml, 10 ml, 20 ml, 60 ml: Braun (Germany)
Ultrafiltration membranes YM3: Amicon (Germany)

Chromatography equipment, columns
ÄKTA explorer 10: Amersham Pharmacia (Sweden)
Peristaltic pump P-1: Amersham Pharmacia (Sweden)
Fraction collector RediFrac: Amersham Pharmacia (Sweden)
Recorder REC-1: Amersham Pharmacia (Sweden)
UV flow through detector UV-1: Amersham Pharmacia (Sweden)
BioloLogic LP System: Biorad (USA)
HiLoad 26/60 Superdex S75pg: Amersham Pharmacia (Sweden)
HiLoad 10/30 Superdex S75pg: Amersham Pharmacia (Sweden)
Mono S HR 5/5, 10/10: Amersham Pharmacia (Sweden)
<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>NiNTA-agarose</td>
<td>Qiagen (Germany)</td>
</tr>
<tr>
<td>Strep-Tactin Sepharose</td>
<td>IBA GmbH</td>
</tr>
<tr>
<td><strong>Miscellaneous:</strong></td>
<td></td>
</tr>
<tr>
<td>Autoclave</td>
<td>Bachofer (Germany)</td>
</tr>
<tr>
<td>Balances PE 1600, AE 163</td>
<td>Mettler (Germany)</td>
</tr>
<tr>
<td>BIAcore 300 instrument</td>
<td>BIAcore (Australia)</td>
</tr>
<tr>
<td>BioPhotometer</td>
<td>Eppendorf (Germany)</td>
</tr>
<tr>
<td>Centrifuge Avanti J-30I</td>
<td>Beckman (USA)</td>
</tr>
<tr>
<td>Centrifuge Microfuge R</td>
<td>Beckman (USA)</td>
</tr>
<tr>
<td>Centrifuge 3K15</td>
<td>Sigma (Germany)</td>
</tr>
<tr>
<td>Centrifuge 5415D</td>
<td>Eppendorf (Germany)</td>
</tr>
<tr>
<td>Centrifuge 5804R</td>
<td>Eppendorf (Germany)</td>
</tr>
<tr>
<td>Chambers for SDS PAGE</td>
<td>BioRad (USA)</td>
</tr>
<tr>
<td>Ice machine Scotsman AF 30</td>
<td>Frimont (Italy)</td>
</tr>
<tr>
<td>MARresearch image plates, mar345</td>
<td>MARresearch (Germany)</td>
</tr>
<tr>
<td>Magnetic stirrer Heidolph M2000</td>
<td>Bachofer (Germany)</td>
</tr>
<tr>
<td>Mastercycler</td>
<td>Eppendorf (Germany)</td>
</tr>
<tr>
<td>NMR-spectromete DRX600</td>
<td>Bruker (Germany)</td>
</tr>
<tr>
<td>pH-meter pHM83</td>
<td>Radiometer (Denmark)</td>
</tr>
<tr>
<td>Pipettes 2.5 µl, 10 µl, 20 µl, 200µl, 1000 µl</td>
<td>Eppendorf (Germany)</td>
</tr>
<tr>
<td>RTS 500 instrument</td>
<td>Roche (Germany)</td>
</tr>
<tr>
<td>Quarz cuvettes QS</td>
<td>Hellma (Germany)</td>
</tr>
<tr>
<td>Shaker incubator</td>
<td>Adolf-Kühner AG (Switzerland)</td>
</tr>
<tr>
<td>Sonicator, Sonifier 250</td>
<td>Branson (USA)</td>
</tr>
<tr>
<td>Spectrophotometer</td>
<td>Amersham Pharmacia (Sweden)</td>
</tr>
<tr>
<td>Thermomixer</td>
<td>Eppendorf (Germany)</td>
</tr>
<tr>
<td>Ultrafiltration cells, 10 ml, 50 ml, 200 ml</td>
<td>Amicon (Germany)</td>
</tr>
<tr>
<td>VP-ITC system</td>
<td>MicroCal (USA)</td>
</tr>
<tr>
<td>Vortex</td>
<td>Cenco (Germany)</td>
</tr>
<tr>
<td>X-ray generator RU2000, 45 kV, 120 mA</td>
<td>Rigaku (Japan)</td>
</tr>
</tbody>
</table>
Chapter 6

6.4 Computer programs

ARP/wARP (Perrakis et al., 1999)
MolRep
PROCHECK (Laskowski et al., 1993)
Refmac 5 (Murshudov et al., 1997)
all from CCP4 program suite (Collaborative Computational Project, Number 4, 1994)
Adobe Photoshop 8
Corel Draw 12
MOSFLM (Leslie, 1991)
Moloc (www.moloc.ch)
Raster3d (Merritt and Bacon, 1997)
Swiss-Pdb Viewer (www.expasy.org/spdbv)
XtalView/Xfit (McRee, 1999)
WebLab Viewer