Technische Universität München
Institut für Organische Chemie und Biochemie

Max-Planck-Institut für Biochemie
Abteilung Strukturforschung (NMR-Arbeitsgruppe)

Structural studies of cytoskeleton proteins, proteases and IGF-binding proteins

Grzegorz Maria Popowicz

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

Die Dissertation wurde am 01.07.2005 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie am 09.01.2006 angenommen.

For we know in part and we prophesy in part;
but when the perfect comes, the partial will be done away.

New Testament 1 Cor 13:9-10

Acknowledgements

I would like to thank all of those who have contributed to this work.

In particular I am most grateful to Professor Robert Huber for giving me the opportunity to work in his department and for being my Doktorvater.

Creation of this thesis was only possible because of the support of Doctor Tad A. Holak, my supervisor, to whom I am indebted for his scientific contribution, great support and care.

To the NMR group team: Dorota Książek, Igor Siwanowicz, Joma Joy, Loy D'Silva, Madhu Ghosh, Magda Wiśniewska, Mahavir Singh, Marcin Krajewski, Aleksandra Mikołajka, Przemyslaw Ozdowy, Sudipta Majumdar, Till Rehm Tomasz Sitar and Ania Czarny, for forming a great scientific team full of motivation and support.

My special thanks to Igor Siwanowicz, Tomasz Sitar, Magda Wisniewska, Przemyslaw Ozdowy for interesting talks as unrelated as possible with a subject of our work.

Last, but not least, I would like to thank my future wife, Alena Wantulokova, who paid a price of three years of separation to allow me to work on this thesis. For her, her love and care, my debt is infinite.

Publications

Parts of this thesis have been published or will be published in due course

Popowicz, G.M., Muller, R., Noegel, A.A., Schleicher, M., Huber, R., and Holak, T.A. (2004) Molecular structure of the rod domain of dictyostelium filamin. J. Mol. Biol. **342,** 1637–1646.

Popowicz, G.M., Dubin, G., Stec-Niemczyk, J., Czarny, A., Dubin, A., Potempa, J. and Holak, T.A. Functional and structural characterization of spl proteases from *Staphylococcus aureus.* (manuscript in preparation)

Popowicz, G.M., Schleicher, M., Holak, T.A. and Noegel, A.A. Structural bases of filamin function and organization (manuscript in preparation).

Dubin, G., Krajewski, M., Popowicz, G., Stec-Niemczyk, J., Bochtler, M., Potempa, J., Dubin, A., Holak, T.A. (2003) A novel class of cysteine protease inhibitors: solution structure of staphostatin A from *Staphylococcus aureus*. Biochemistry **42,**13449-56.

Dubin, G., Popowicz, G., Krajewski, M., Potempa, J., Dubin, A., Holak, T.A. (2004) ^{1}H, ^{15}N and ^{13}C NMR resonance assignments of staphostatin A, a specific *Staphylococcus aureus* cysteine proteinase inhibitor. J. Biomol. NMR **28,** 295-6.

Mavoungou, C., Israel, L., Rehm, T., Ksiazek, D., Krajewski, M., Popowicz, G., Noegel, A.A., Schleicher, M., Holak, T.A. (2004) NMR structural characterization of the N-terminal domain of the adenylyl cyclase-associated protein (CAP) from Dictyostelium discoideum. J. Biomol. NMR. **29,** 73-84.

Siwanowicz, I., Popowicz, G.M., Wisniewska, M., Huber, R., Kuenkele, K-P., Lang, K., Engh, R.A., and Holak, T.A. (2005) Structural basis for the regulation of insulin-like growth factors by IGF binding proteins. Structure ***13*** (in press).

Siwanowicz, I., Popowicz, G.M., Ghosh, M., Moroder, L., Dsilva, L., Joy, J., Majumdar, S., Wisniewska, M., Firth, S.M., Baxter, R.C., Huber, R., and Holak T.A. (2005) Molecular architecture of the insulin-like growth factor binding proteins (IGFBPs). J. Biol. Chem. (in press).

Benzinger, A., Popowicz, G.M., Joy, J.K., Majumdar, S., Holak, T.A., Hermeking, H. (2005) The crystal structure of the non-liganded 14-3-3sigma protein: insights into determinants of isoform specific ligand binding and dimerization. Cell Res. **15,** 219-27.

Arolas, J.L., Popowicz, G.M., Lorenzo, J., Sommerhoff, C.P., Huber, R., Aviles, F.X. and Holak, T.A. (2005) The three-dimensional structures of tick carboxypeptidase inhibitor in complex with a/b carboxypeptidases reveal a novel double-headed binding mode. J. Mol. Biol. (in press)

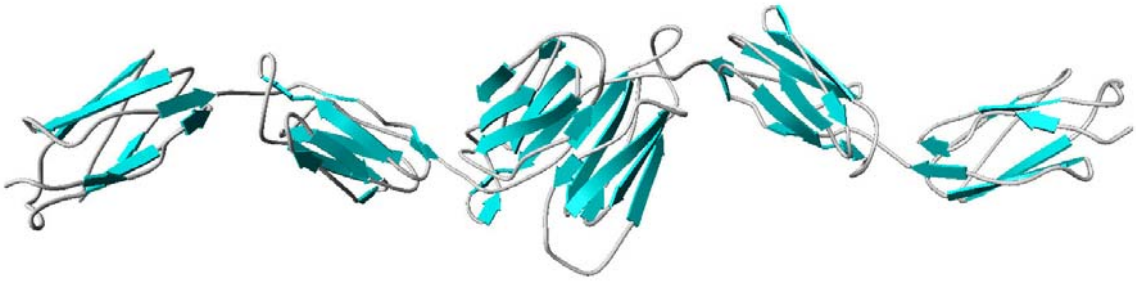
Arolas, J.L., Popowicz, G.M., Bronsoms, S., Aviles, F.X., Ventura, S., Huber, R. and Holak, T.A. (2005) Study of a major intermediate in the oxidative folding of leech carboxypeptidase inhibitor: contribution of the fourth disulfide bond. (manuscript in preparation)

Arolas, J.L., D'Silva, L., Popowicz G.M., Aviles, F.X., Holak, T.A. and Ventura, S. (2005) NMR structural characterization and computational prediction of the major intermediate in the oxidative folding of leech carboxypeptidase inhibitor. (manuscript in preparation)

Contents

Chapter 1

Actin cytoskeleton-related proteins



1.1 Introduction

1.1.1 Actin cytoskeleton

The cytoskeleton provides the foundation for spatial organization of living cells and their movement. The most important component of the cytoskeleton is the actin filament. In spite of our inclination to consider the cell "skeleton" as a rigid base, the real cytoskeleton is dynamic, undergoing permanent reorganization and modification. Actin filaments are elongated or cleaved by specific proteins. Their ends are protected against further elongation by capping proteins or are anchored to membranes. To strengthen cytoskeletal structures, proteins that cross-link actin filaments are also necessary. Spectrin, fimbrin, α-actinin, and filamin (FLN) belong to this group of actin cross-linkers. Most of these proteins are dimers with actin-binding and dimerization domains present in each monomer. While fimbrin and α-actinin are believed to form parallel actin bundles, filamin cross-links actin filaments at different angles.

Recent studies show that filamins are not only mechanical linkers for actin filaments but also serve as interaction partners for a number of proteins of a great functional diversity ranging from signal transduction to nuclear transcription factors (Feng and Walsh, 2004). Also recent genetic studies revealed significance of gene mutations in filamins to a number of diseases ranging from brain (Feng and Walsh, 2004, Fox et al., 1998, Sheen et al., 2001), to bone and cardiovascular systems (Stefanova et al., 2005; Robertson et al., 2003).

Although filamins are found in many organisms, best studied are those from D*ictyostelium discoideum* and mammals. These two prototypical filamins comprise an actin-binding domain (two tandem calponin homology domains) and an elongated rod domain built by six (in *Dictyostelium*) or 24 (in human) repeats of an immonoglobulin-like fold (Fucini et al., 1997). The last repeat of the rod domain is responsible for dimerization. Human filamin additionally has two unique long hinges between repeats 15-16 and 24-24, 27 and 35 residue long, respectively, which are postulated to be flexible (Stossel et al., 2001). The human filamin family has three members: filamins A, B and C, which share 70% homology of the sequence, except for the hinges, which have much less homology. Structures of filamin fragments known so far are presented on Figure 1.1.1.

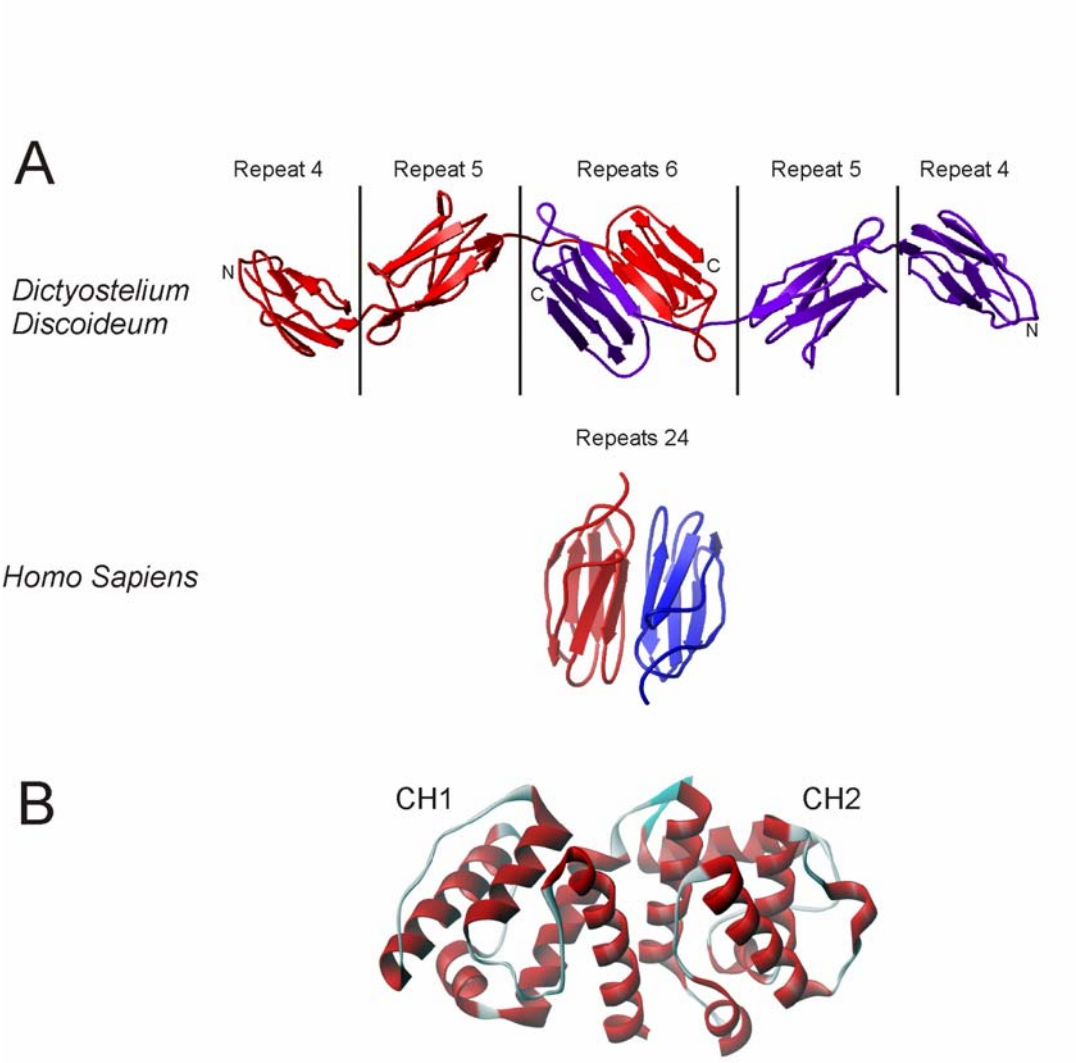


Figure 1.1.1. Summary of structural knowledge of filamins. (A) Structures of fragments of rod domains of *Dictyostelium* filamin comprising repeats 4, 5 and 6, and the *Homo sapiens* dimerizing repeat 24 (B) Ribbon plot of an actin binding domain from α-actinin (Franzot et a., 2005, PDB ID 1TJT). Based on sequential similarity, filamin is expected to have a similar domain at its N-terminal end.

1.1.2 Mechanism of filamin dimerization

The last C-terminal repeat of the rod domain is usually different from other repeats. Three structures of fragments of rod domains that included dimerization regions have been published until now; two of them from *Dictyostelium* and one human (McCoy et al., 1999, Popowicz et al., 2004). The amoeboidal filamin shows identical behavior of dimerization in both structures. Repeat 6 differs sequentially from its preceding repeat by lacking 12 residues at the N-terminus and shows no sequence homology up to the middle of the second strand; there is also one additional strand present at the very C-terminus of the repeat. The repeats form an antiparallel dimer

by β-sheet extension. The most extensive interactions are between the first and one before last strands in the repeat that binds to corresponding strands in the second molecule of the dimer. The dimer interface is large, comprising about 4500 A^2 (of which 69% is repeat's six primary dimerization area contribution and 22% is a buried linker contact surface). The two repeats of the dimer extend their β-strands to form a large β-sandwich of six strands at one side and eight at the other. The N-terminal parts of both repeats 6 embrace neighboring repeats and are also deeply buried between their strand-interconnecting loops. Such organization of the linker makes the fifth repeat to be closer to the sixth of the second molecule that to its own. This arrangement provides high degree of rigidity between repeats 5 and 6.

The fold of dimerization repeats of *Dictyostelium* and human filamins is similar; rms deviation for the core parts is only 1.27 Å. Significant differences affecting dimerization are located at C- and N-termini of the repeats (Figure 1.1.2).

The structure of dimerization repeats of human filamin shows a similar way of dimerization (β-strands extension) to that of *Dictyostelium* filamin, however important differences are seen. Both in *Dictyostelium* and human filamin they are arranged with an exact two-fold symmetry. Monomer binding takes place also by extending a β-sheet of each monomer to form an elongated β-sandwich, however, the strands responsible for direct interaction are different and internal organization of strands also differ significantly to that of *Dictyostelium*. Dimerizing strands of the repeat responsible for dimerization appear to be third and fourth. Their lengths are shorter than those of *Dictyostelium*. Additionally, there is no buried linker at the N-terminal part of the repeat. A 35 residue long hinge region precedes repeat 24 in vertebrates. The structure of this part is unknown. It also shows greater sequential divergence than Ig-like repeats. The function of this area in contrast to that of *Dictyostelium* is probably mostly mechanical, allowing high degree of flexibility and structural adaptability (Pudas et al., 2005), There are reports indicating a possible regulatory function of this region (Himmel et al., 2003). The interface of dimerization is much smaller than in *Dictyostelium*, the buried surface area is 1109 A^2, which takes 19% of the molecule surface. The crystallographic model was supported by mutagenesis studies, which excluded possibility of an artifact and proved that dimerization mechanism of the native vertebrate filamin is different and presumably weaker that that of *Dictyostelium*.

Overall comparison shows that dimerization interfaces in filamins were changed during evolution, although the same mechanism of a double-sided β-sheet extension of Ig-like repeats remained. However, rigid and very strong interaction of primitive organisms was replaced by more sophisticated, weaker and more flexible dimerization interfaces in higher organisms. Future studies on longer constructs of filamins are expected to give information regarding the spatial organization of its repeats and their rigidity.

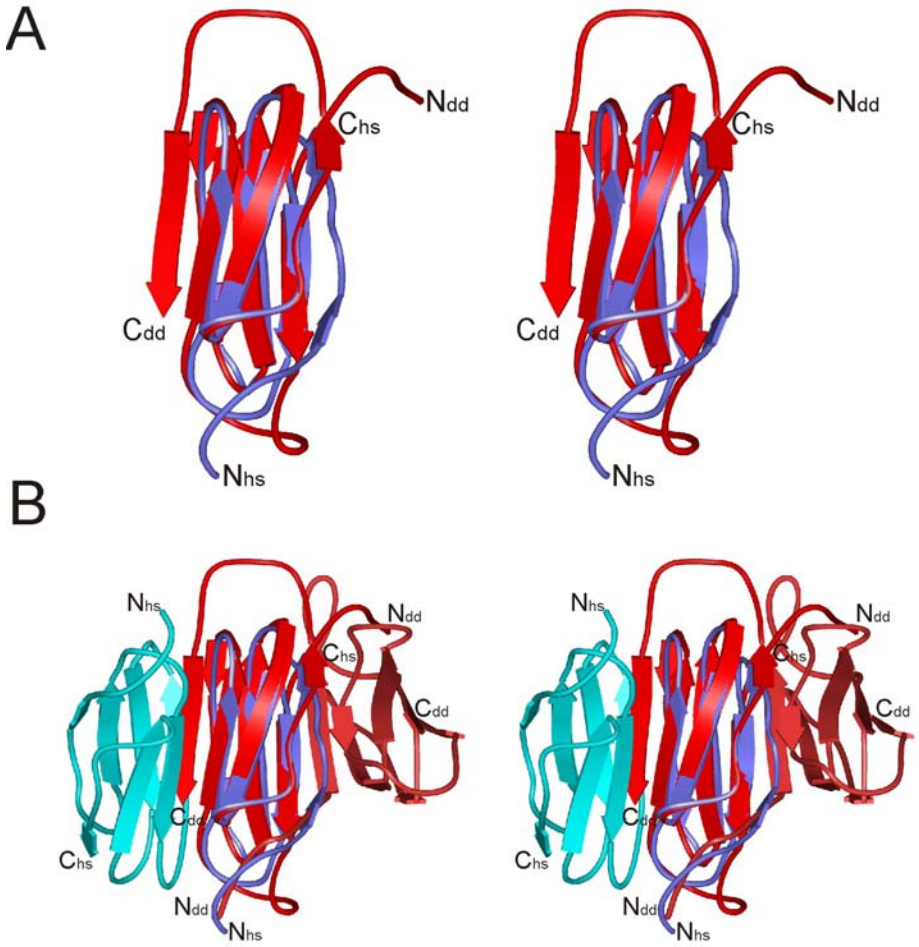


Figure 1.1.2. Structural comparison of dimerization repeats of *Dictyostelium* (red) and human (blue) filamins. (A) Monomer superimposition shows basically identical folds of the cores with differences seen at the C- and N terminal ends of repeats. The *Dictyostelium* has an additional long strand at C-terminus while the human protein has two additional short strands at the N-terminal part at the other side of the molecule. (B) Structures of homodimers superimposed on monomers as in (A) – despite of their structural similarity the dimerization area is different. The interacting surfaces are localized at opposite sides of repeats. There is no interaction between the N-terminal linker of repeat and the second molecule in the dimer in *hs*-filamin.

1.1.3 Inter-repeat organization

Repeats of the rod domain of *Dictyostelium* filamin are arranged in an extended way. Mechanical properties of the entire rod domain are essential for actin cross-linking capabilities and for properties of actin networks they create. Except for two hinge regions in human FLN, linkers are short among repeats. They comprise of only few residues rich in prolines. Additionally salt bridges were identified between neighboring repeats. The two structures of the D*ictyostelium* multi repeat constructs show that the rod domain remains in the same spatial position regardless of crystal packing and crystallization conditions. The structures maintain an exact 2-fold symmetry. rms deviation between the structures is only 1.27 Å. The rod domain appears to be a long extended spring-like structure with an inter-repeat tilt angle of about 115 degrees. Such conformations support a proposition that the whole *Dictyostelium* rod is an extended structure of limited flexibility that is capable of interconnecting distant actin filaments. Of course, certain flexibility must be involved to allow for binding and adaptation of the created actin network. This particular feature might be realized by a long linker between actin binding (ABD) and rod domains. In general it seems that the primary function of filamin in D*ictyostelium* is to reduce the number of degree of freedom in bound actin filaments.

There is no structural data regarding inter-repeat organization of vertebrate filamins. However, an almost five times longer rod domain of human filamin, relative to that of *Dictyostelium,* is likely to be organized in the same way. There are significant sequence similarities to support such a model. However, the two long hinge regions in human fialmin have larger amino acid diversity and their length would allow for the whole rod domain to be more flexible and to cross-link actin filaments in many orientations. Hinges may also present a potential proteolytic cleavage sites as the C-terminal (repeats 16-24) part of human filamin is reported to localize to the nucleus after cleavage (Loy et al., 2003. Furthermore, a region before repeat 24, considered to be flexible (Gorlin et al., 1990), is reported to regulate dimerization (Himmel et al., 2003). Besides its mechanical function, the rod domain of vertebrates has been shown to be an important scaffold for binding a great number of proteins important in a broad range of cellular processes (Stossel et al., 2001, Feng and Walsh, 2004).

1.1.4 The actin binding domain (ABD)

Members of the family of actin cross-linking proteins share the same mechanism of their interaction with actin. Filamins, α-actinin, spectrin, plectin, dystrophin and fimbrin utilize similar mechanism of two calponin homology domains for ABDs. A typical actin-binding domain has 250 residues and shares 20-60% of sequence identity with other ABDs in the family. Structures of the ABDs of α-actinin show that both subdomains form fully α-helical globular domain (Franzot et al., 2005). The subdomains appear to have either closed or open conformations, depending on contact area between them. These structures present a variety of possible subdomain organizations (Lehman et al., 2004). It is not known which one of those configurations is present in ABDs of filamins. The structure of α-actinin actin binding domain is shown in Figure 1.1.1B.

The interface between F-actin and filamin was studied extensively using peptides and antibodies of known specificities. The primary binding site of an ABD to actin is localized between residues 121 and 147. The interaction is mostly hydrophobic, but the susceptibility of the binding constant on ionic strength suggests also the presence of hydrophilic interactions. The actin residues involved in binding are between 112-125 and 360-372 (Lebart et al., 1993).

1.1.5 Unfolding of a rod domain of *Dictyostelium* filamin (ddFLN)

Detailed studies with atomic force spectroscopy have recently been carried out to obtain more information on mechanical properties of repeats of the *Dictyostelium* filamin. Atomic force spectroscopy indicates that individual repeats unfold under a force before the dimer is broken. To break the dimer about 200 pN is necessary. The fourth repeat shows an unfolding pattern different from all other repeats. It unfolds easiest and it appears that only this repeat has a stable folding intermediate. In the first stage of unfolding about 40 residues are stripped from the molecule and the remaining 60 form a folded intermediate. The forty residues correspond to the first two β-strands (Schwaiger et al., 2004). The intermediate is a stable structure able to fold upon itself. The two-stage folding of the fourth repeat is also the fastest process of folding observed for the rod domain (Schwaiger et al., 2005). The biological significance of this feature might be interesting. Easy unfolding and fast refolding of this repeat would allow for the whole rod to nearly double its length and then come

back to its native state. It is also possible that the folding speed of an elongated molecule is increased by presence of the intermediate. The free energy barrier between the unfolded state and the intermediate, and in the next step, the intermediate and a fully folded protein, is easier to break than for direct transition to the folded state. Another interesting feature of this unfolding mechanism is linked to proteins that can bind to the rod domain in the repeat 4 area. Mechanical stress causes unfolding of the repeat and dissociation of the binding partner or association of a new one to the intermediate. Bound or released interacting molecules may then be part of a signaling pathway. Such a mechanism would serve as a new cellular "sense" of mechanical forces present in the cytoskeleton.

1.1.6 The geometry of actin cross-linking

Filamin is a potent actin gelation factor; in fact only one molecule of FLN per actin filament is sufficient to induce gelation. Vertebrate filamins were studied extensively by electron microscopy. Human filamin appears to have a V shaped organization and is fundamentally more flexible, probably because of the presence of two hinges (Tyler et al., 1980) than *Dictyostelium* filamin. It appears that actin networks build have mostly perpendicular organization of actin filaments and the filamin acts as a bracket, holding actin filaments perpendicular to each other. It was shown that T, X or L shaped junctions created by human filamin has myosin arrows pointing towards the branch points of actin filaments (Tyler et al., 1980, Hartwig and Desisto, 1991).

1.1.7 Binding partners

The mechanical function of filamin undoubtedly remains important, however, in the light of recent publications, filamins reveal many new faces, as important as their primary function. Vertebrate filamins were found to interact with more than twenty of proteins; many of these interactions were unexpected. The recently published interaction partners are presented in Table 1.1.1. Filamins play of course, an important role in the cytoskeleton remodeling and signaling (Nagano et al., 2004, Ochta et al. 1999). It turns out that the postulated scaffolding function of filamins has also a broad range of applications in: the anchoring membrane receptors to the actin cytoskeleton, thus enabling precise localization and transport of the receptors (Arron et al., 2002, Awata et al., 2001, Cranmer et al., 2005, Gravante et al., 2004, Lin et al., 2001, Enz 2002, Seck et al., 2003, Sampson et al., 2003, Travis et al., 2004) and

other proteins (Liu et al., 1997), serving as a colocalization factor for signaling pathways (Tigges et al., 2003), and as a mechanical element in caveolae and membrane ruffle formation (Stahlhut and van Deurs 2000, Vadlamudi et al. 2002). An important role of filamins is also seen in cell-extracellular matrix interactions; here filamins, together with ECM receptors, serve as important interface between cyto- and exo-skeletons (Wu, 2005 Tu et al., 2003, Travis et al., 2004). Additionally, a number of unusual functions were identified. It appears that filamin may hold transcription factors in the cytoplasm compartment, thus disabling their activity in the nucleus (Yoshida et al., 2005). Surprisingly, filamin has recently been reported as a nuclear protein, binding to a number of transcription factors. It is postulated that an elongated filamin molecule may serve as a kind of a nuclear “cytoskeleton” for colocalization of nuclear functional complexes (Berry et al. 2004, Meng et al. 2004, Yuan and Shen, 2001) and can participate in the cytoplasm-nucleus transport (Ozanne et al., 2000). Nuclear functions are also present in the cleaved C-terminal part of human filamin, representing repeats 16-24, and reported to be present in the nucleus and to downregulate the androgen receptor (Loy et al., 2003).

Remarkably, most of the filamin-interacting proteins are reported to bind to filamin between repeats 16-24. The mapped binding regions of different binding partners overlap but are seldom identical. This leads to the postulate that the primary function of repeats 16-24 is binding to large proteins, like for example, receptors, while repeats 1-15 are bound by a number of smaller proteins participating in signaling processes. We can thus postulate that the actin-binding domain serves as an anchor, repeats 1-15 is a chain and the region of repeats 16-24 is a capstan. In light of the recently identified functions of filamins it is clear that their primal function (of actin cross-linking) was complemented in the evolutionary processes by additional tasks. Possible functions of filamin are presented on Figure 1.1.3.

So far, only one binding partner other than actin was identified for the *Dictyostelium* filamin. The FIP protein is reported to be important for the development of the cell, multicellular aggregates and phototactic behavior (Knuth et al., 2004). Its binding region includes the rod domain repeat 4, the fact that may support the model that unfolding of repeat 4 is important for ligand binding.

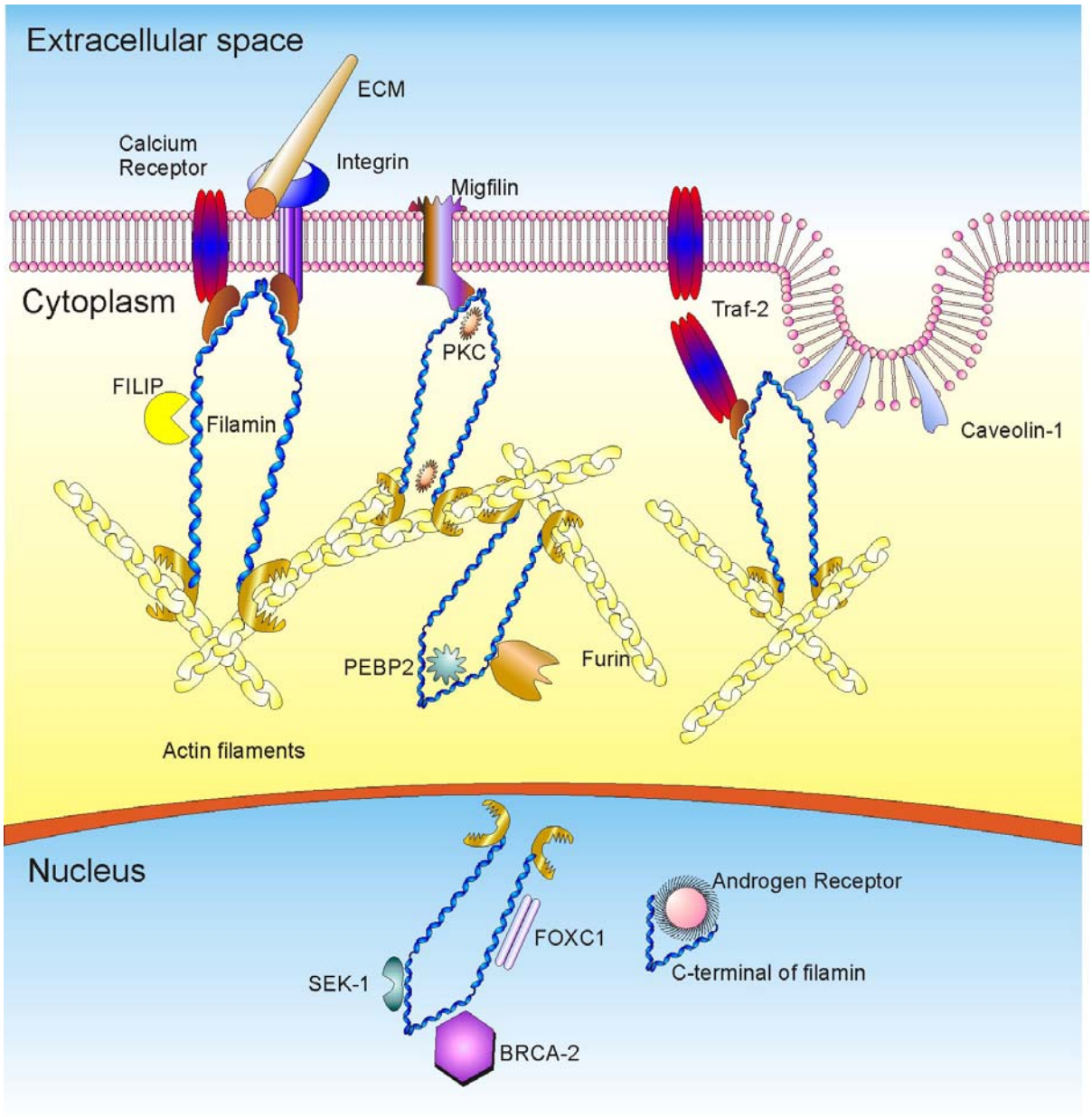


Figure 1.1.3. The versatile functions of filamins. The main use of filamins - to cross-link actin filaments - is supplemented with a variety of other actions. Filamins act as nuclear scaffolds for signal pathways, receptor anchoring and transportation. Filamins were also shown to be collocalized with transcription factors and nuclear receptors. In full or cleaved forms, they serve also as nuclear “skeletons” and regulators.

Table 1.1.1. Identified filamin interaction partners.

Binding partner	Function of binding partner	Function of filamin in complex with partner	Binding Site on filamin	Reference
Dictyostelium filamin				
FIP	Development	Membrane to actin link/signal transduction	Repeats 2-4	Knuth et al. 2004
Vertebrate filamins				
FILIP	Filamin degradation/regulation	Being downregulated by FILIP	?	Nagano et al. 2004
TRAF1 and 2	Inflammatory and immune signaling	Anchoring/receptor internalization and recycling	Repeats 15-19	Arron et al. 2002
CaR extracellular calcium receptor	G protein signaling	Receptor to actin anchoring	Repeats 14-16	Awata et al. 2001
Furin	Proteolytic maturation of proteins	Sorting, Compartmentalization and stabilization	?	Liu et al. 1997
FAP52	actin organization	?	Repeats 13-16	Nikki 2002
FOXC1	Transcription factor/cell differentiation	Nuclear scaffold (?)	aa 571-866 867-1154 1779-2284	Berry et al. 2004
glycoprotein Ibα	Platelet adhesion receptor	Transport of GPIbα from ER to cell surface	aa 567-571	Cranmer et al. 2005
SHIP-2	Cell adhesion/submembrane actin remodeling	Receptor to actin anchoring	?	Dyson 2003
HCN1	Pacemaker channels	Receptor to actin anchoring	Repeat 24	Gravante et al. 2004
D2/D3 dopamine receprors	Pre or post synaptic receptors	Receptor to actin anchoring	Repeat 19	Lin et al. 2001
Glutamate receptor type 7	Neurotransmitter receptor	Receptor to actin anchoring	Repeats 21-22	Enz 2002
calcitonin receptor	Calcium homeostasis	Anchoring/receptor internalization and recycling	Repeats 20-22	Seck et al. 2003
Androgen receptor	Nuclear transcription factor	Downregulates AR in nucleus	Repeats 16-24 after cleavage	Loy et al. 2003 Ozanne et al. 2000
SEK-1	Kinase	Tumor necrosis factor-α activation	Repeats 21-23	Marti 1997
BRCA-2	Tumor suppressor	Promotes recovery from G_2 arrest after DNA damage	Repeats 21-24 in nucleus	Yuan and Shen 2001 Meng et al. 2004
RalA	GTPase	cytoskeleton regulation/filopodia formation	Repeat 24	Ochta et al.1999
Kir2.1	Potassium channel	Receptor to actin anchoring	Repeats 23-24	Sampson et al. 2003
Smad	TGF-β signal protein	Anchoring/phosphorylation promotion	Repeats 20-23	Sasaki 2001

Caveolin-1	Membrane protein/caveolae formation	Anchoring caveolae to cytoskeleton	Repeats 22-24	Stahlhut and van Deurs 2000
Protein Kinase Cα	Signal transduction	Scaffold for signaling pathway	Repeats 1-3 and hinge2-repeat 24	Tigges et al. 2003
Integrin	ECM receptor	Receptor to actin anchoring	Repeats 19-24	Travis et al. 2004
Pak1	Cytoskeleton reorganization	Ruffle formation	Repeat 23	Vadlamudi et al. 2002
Migfilin	Actin remodeling/cell differentiation	Cell adhesion structure to cytoskeleton binding	Repeat 21	Wu 2005 Tu et al. 2003
PEBP2/CBF	transcription factor	Retains PEBP2 in cytoplasm inhibiting its nuclear activity	Repeats 23-24	Yoshida et al. 2005

1.1.8 Conclusions

Filamins present a wonderful example of multifunctionality. Its primary and primal functions of cross-linking actin filaments have been supplemented during evolution by various additional tasks. Medical relevance of filamins, especially in the field of developmental malfunctions is indisputable (Feng and Walsh, 2004). At the present stage of our structural knowledge on filamins it seems necessary to introduce a more holistic approach to this protein. Our knowledge of structural properties of filamins, and especially vertebrate filamins, is still too limited to link a variety of their biochemical functions to the structure. The most logical way of structural investigation appears to study filamin complexes with other proteins and search for a “filamin binding motif/domain” in its partners. Much can be done in this field; however also other properties of filamins, like for example: actin binding and inter-domain orientation, have to be studied.

1.2 Structure of ddFLN(4-6), implication for the molecular architecture of the major actin cross linking protein

1.2.1 Introduction

We studied a fragment of the rod domain of the *Dictyostelium* filamin comprising repeats 4, 5 and 6 (ddFLN(4-6). As the structure of the construct, containing repeats 5 and 6, was known (McCoy et al., 1999) we expected that molecular replacement would suffice to solve our structure. The presence of the fourth repeat in our protein was expected to say how repeats not involved in dimerization are organized. Direct comparison to known structures provides also important information about the mechanical stability of the whole rod domain.

1.2.2 Materials and methods

1.2.2.2 Protein preparation and characterization

The sequence corresponding to positions 1730 to 2670 of the cDNA was amplified using primers containing EcoRI and SalI add-on linkers. The EcoRI/SalI fragment was cloned into similarly cleaved pT7-7 (Tabor and Richardson, 1992) and the plasmid transformed into *E. coli* BL21. Soluble recombinant polypeptide was expressed at high levels after induction with IPTG (0.5 mM) and further growth at 37^{o}C for 4 hours. The protein was purified by anion exchange chromatography as followed by phenylsepharose chromatography. The protein was eluted using a linear gradient from 1.7 M to 0 M $(NH_4)_2SO_4$. It was further purified using gel filtration in the "crystallization" buffer (5 mM Tris, 50 mM NaCl, 1 mM BME, pH 8.0) and concentrated to 10 mg/ml. The identity of the protein was confirmed by MALDI mass spectrometry.

1.2.2.3 Crystallization and diffraction data collection

Crystallization of domains 4,5,6 of ddFLN protein was carried out with the hanging drop vapor diffusion method by mixing equal volumes of the protein and reservoir solution (27% 2-methyl-2,4-pentanediol, 0,1 M $CaCl_2$, 0,1 M sodium acetate trihydrate pH 4.8). Crystal needles appeared in the drop after 2-3 months. After two steps of macroseeding crystals grew to a final size of ~0.3× 0.1 × 0.1 mm. They belong to the space group $P2_1$ (with unit cell dimensions: a=56.320 Å, b=61.67 Å, c=119.03 Å, β=104.1˚). The crystal contains one dimer per asymmetric unit. A 2.8 Å

native data set was obtained from plunge-frozen crystals at synchrotron source. X-ray data were collected at the Swiss Light Source (Paul Scherrer Institute, Viligen) using the 156-mm MarCCD detector (Mar-USA, Evantson, IL) at 100K and were processed and scaled with XDS software package (Kabsch, 1993). The data statistics are summarized in Table 1.2.1

1.2.2.4 Structure determination and refinement

The structure was determined by molecular replacement. A known crystallographic structure of repeats 5 and 6 were used as a probe structure (McCoy et al., 1999). Rotation and translation searches were done using Molrep (CCP4, 1994). Rotation search in Patterson space yielded two peaks of height 10.01 σ and 9.87 σ over the highest noise peak of 4.10 σ. Translation search for the two rotation solutions gave peaks 18.53 σ and 15.57 σ, respectively, over the noise height of 5.4 σ. The two peaks correspond to two molecules in the asymmetric unit. They are related by a non-crystallographic dyad axis. The R-factor of the initial model without repeat 4 was 42%. The rigid body refinement was performed using CNS software (Brunger et al., 1998). An electron density map calculated at this stage allowed building of repeat 4 using its NMR model as a template and a rigid body refinement was then repeated. Solvent flattening and histogram matching were performed later with the program DM, (CCP4, 1994) assuming solvent content of 57.8%. The map quality was further improved by non-crystallographic symmetry averaging and restrained refinement with the program Refmac(CCP4, 1994). The model was completed and refined manually with help of the XtalView/Xfit software (McRee, 1999). Waters were added by Arp/warp (Lamzin and Wilson, 1993). The final model has R crystallographic factor of 25.7% and a free R factor of 26.3%. All residues have clear electron densities except the loop at the N-terminus of ddFLN(4-6) (residues 583-585). The residue 584 was therefore removed from the model. Data collection and refinement statistics are summarized in Table 1.2.1.

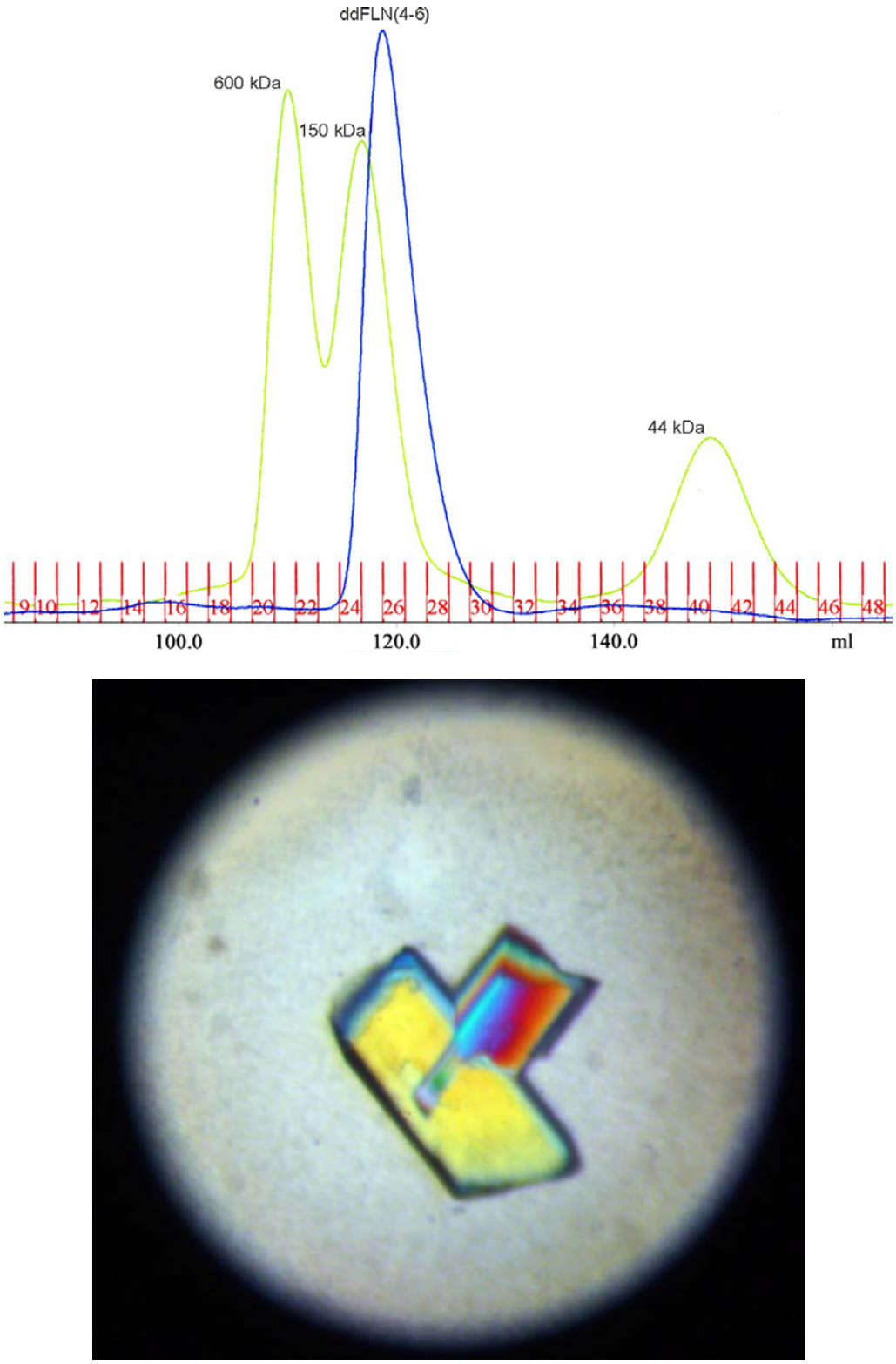


Figure 1.2.1. (Top) Chromatogram of the final purification step-gel filtration, indicating that ddFLN(4-6) is a dimer in solution. (Bottom) Crystals of the *D. discoideum* filamin rod domain repeats 4, 5 and 6. Crystals were grown in 27% 2-methyl-2,4-pentanediol, 0,1 M $CaCl_2$, 0,1 M sodium acetate trihydrate pH 4.8. Several steps of micro- and macroseeding were required to obtain crystals suitable for data collection.

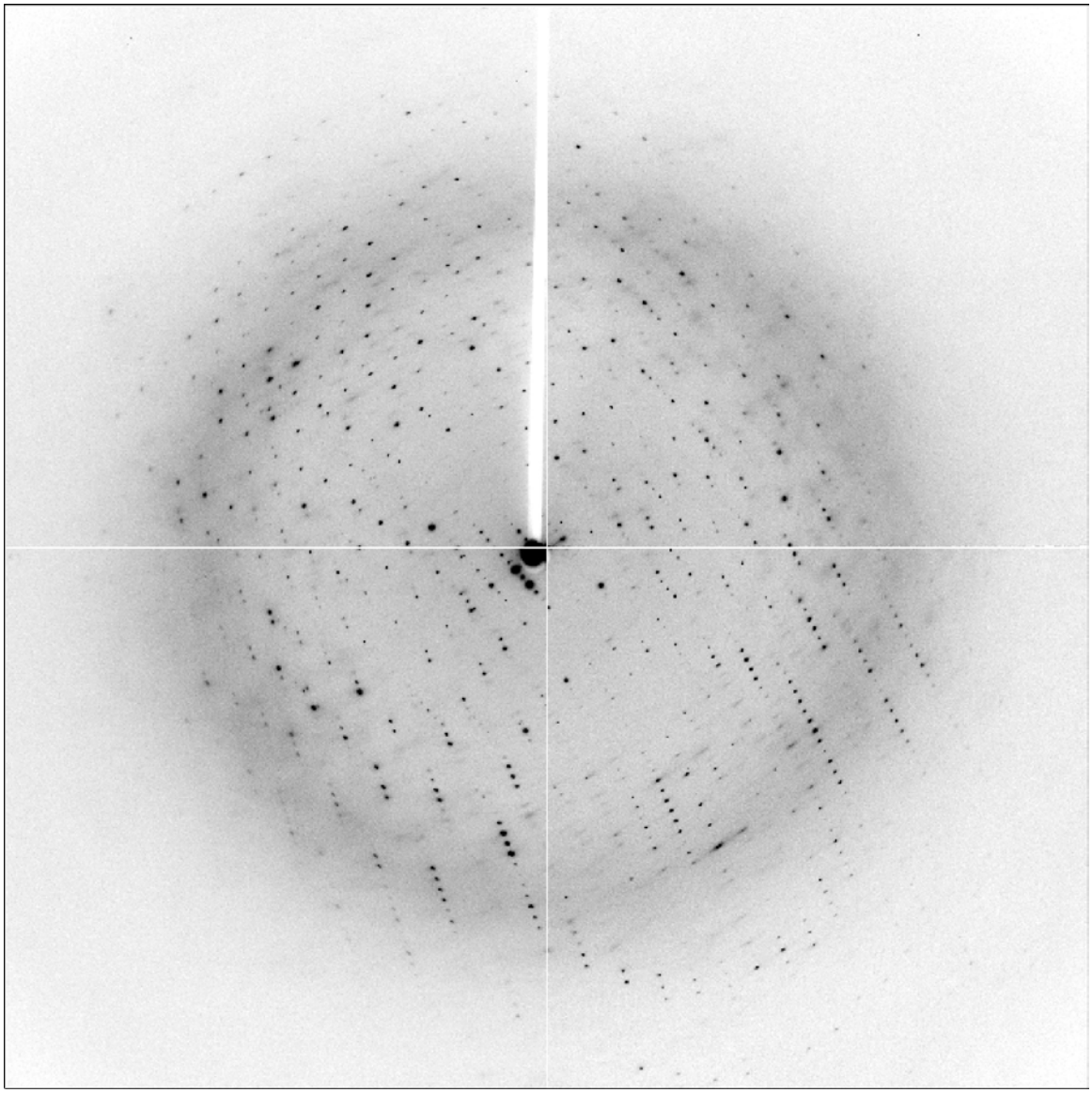


Figure 1.2.2. Diffraction image from the ADSC Q210 detector (ID29 beamline, ESRF, Grenoble, France) of the ddFLN(4-6) crystal. Edge of the detector is at about 2.7 Å. Space group is $P2_1$.

Table 1.2.1. Data collection and refinement statistics of ddFLN(4-6)

Data collection	
Space group	$P2_1$
Cell constants (Å)	a=56.32
	b=61.67 β=104.1
	c=119.03
Resolution range (Å)	30-2.8
Wavelength (Å)	0.97891
Observed reflections	68210
Unique reflections	20708
Whole resolution range:	
Completeness (%)	99.3
R_{merge}	2.6
I/σ(I)	18.1
Last resolution shell:	
Resolution range (Å)	2.85-2.8
Completeness (%)	90.1
R_{merge}	12.4
I/σ(I)	4.8
Refinement	
No. of reflections	20708
Resolution (Å)	30 – 2.8
R-factor (%)	25.7
R_{free} (%)	26.3
Average B (Å^2)	43.7
R.m.s bond lenght (Å)	0.015
R.m.s. angles (°)	2.6
Content of asymmetric unit	
RMSD of monomers (Å)	0.76
No. of protein molecules	2
No. of protein residues/atoms	614/4537
No. of solvent atoms	88

1.2.2.5 Coordinates

The coordinates of the structure have been deposited with the Protein Data Bank under the accession number 1WLH.

1.2.3 Results

1.2.3.1 Structure determination

The molecular mass of the ddFLN(4-6) construct determined by gel filtration was 65 kDa, proving that the protein had dimerized in solution. The space group was $P2_1$ and the cell dimensions were a=56.320 Å, b=61.67 Å, c=119.03 Å, β=104.1°. The model was built using molecular replacement techniques. The refinement statistics is shown in Table 1.2.1. The final structure consists of two chains starting at positions 547 and 549, ending at 857 of the wild type *Dictyostelium* ddFLN.

1.2.3.2 General structure description

A global view of the solved structure is shown in Figure 1.2.3. All repeats of the rod ddFLN(4-6) are full β-structured and share a similar immunoglobulin-like fold. The whole domain 4-6 forms a long extended structure that is about 200 Å long and 50 Å wide.

Repeats 5 and 6 in the current structure appear to be identical to those reported by McCoy et al 1999. for rod repeats 5 and 6; rms deviation for main chain atoms for these repeats is 1.26 Å. Repeat 6 of one monomer is antiparallel to repeat 6 of the other monomeric unit, and they interact with each other forming a symmetrical dimer (Figure 1.2.3 A and B). Repeats 5 and 6 in ddFLN(4-6) maintain also the same orientation relative to each other as found in the structure of repeats 5-6 only. The orientation of repeats was determined by measuring the "elbow" angle between the long axes of repeats. The long axes were defined by least-squares linear fit to the directions of all β-strands of each repeat. The "elbow" angle between the long axes of repeats 5 and 6 is approximately 93 degrees. This is much different from the angle between repeat 4 and repeat 5, which is 115 degrees. β-strand Lys641 – Lys646 ends repeat 4 and is connected to repeat 5 by a Pro rich sequence, which usually brings structural rigidity: the linking Pro647 is connected to a short strand Ala648-Pro649-Ser650 of repeat 5. There is only one residue not belonging to β-strands between the repeats. In addition, repeat 4 and 5 are connected by a salt bridge

formed by Glu565 and Arg734. The region of interaction between repeats 4 and 5 is shown in Figure 1.2.4.

Both monomers in the crystal structure maintain exactly the same organization in space. Thus the homodimer has a very precise non-crystallographic symmetry. This supports assumptions that the whole rod domain in a dimer is symmetrical along the plane and perpendicular to its long axis.

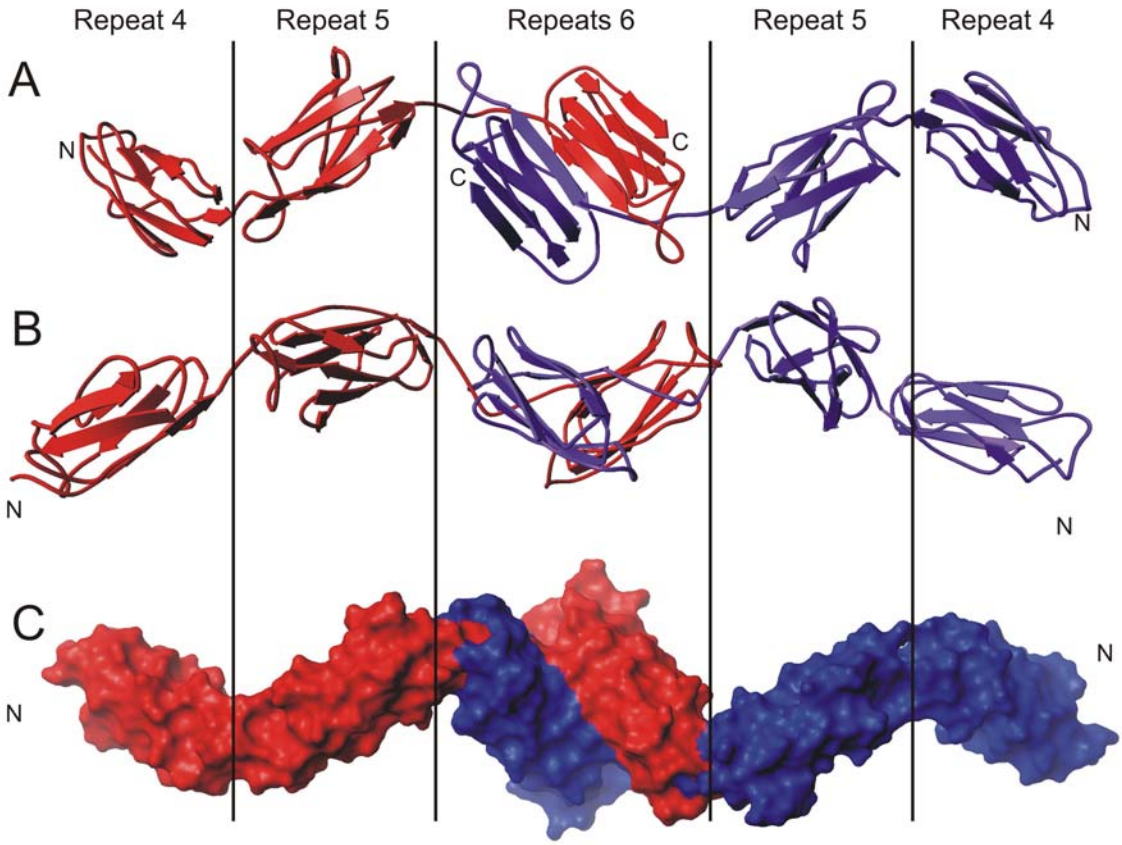


Figure 1.2.3 An overall view of ddfilamin(4-6). (A) and (B) show the molecule rotated by its long axis by about 90 degrees. Two chains forming a homodimer are shown in red and blue. All parts of the model maintain the same spatial organization symmetrical to the vertical axis of the Figure. (C) The molecular surface of the dimer orientation as in (A).

1.2.3.3 Structures of repeats 4 and 5; comparison with the NMR model

Our crystallographic structure of repeat 4 consists of 9 β-strands (β1-Tyr555-Glu557, β2-Gly564-Cys566, β3-Ser570-Ala576, β4-Gly590-Glu596, β5-Val604-Val606, β6-Gly611-Phe617, β7-Gly623-Asp632, β8-Asp634-Val636, β9-Lys641-Lys646). The fold of this repeat is a typical immunoglobulin-like β-sandwich resembling the C1 set of Igs (Fucini et al., 1997). Loops connecting strands appear to be stable and all are visible in the electron density map. Compared to the NMR model of this repeat, the X-ray structure shows only minor differences (Fucini et al., 1997). Overall rms deviation between mean NMR structure and crystallographic model is 1.7 Å. The fold pattern is exactly the same except of two strands Gly564-Cys566 and Asp634-Val636 not present in the NMR structure. It appears that interconnected strands Tyr555-Glu557 and Gly590-Glu596 are longer by one residue in the NMR structure whereas two strands are extended by two residues in the crystallographic model (strand Gly590-Glu596 at the N-terminal part and Gly623-Asp632 at the C-terminus) (Figure 1.2.5).

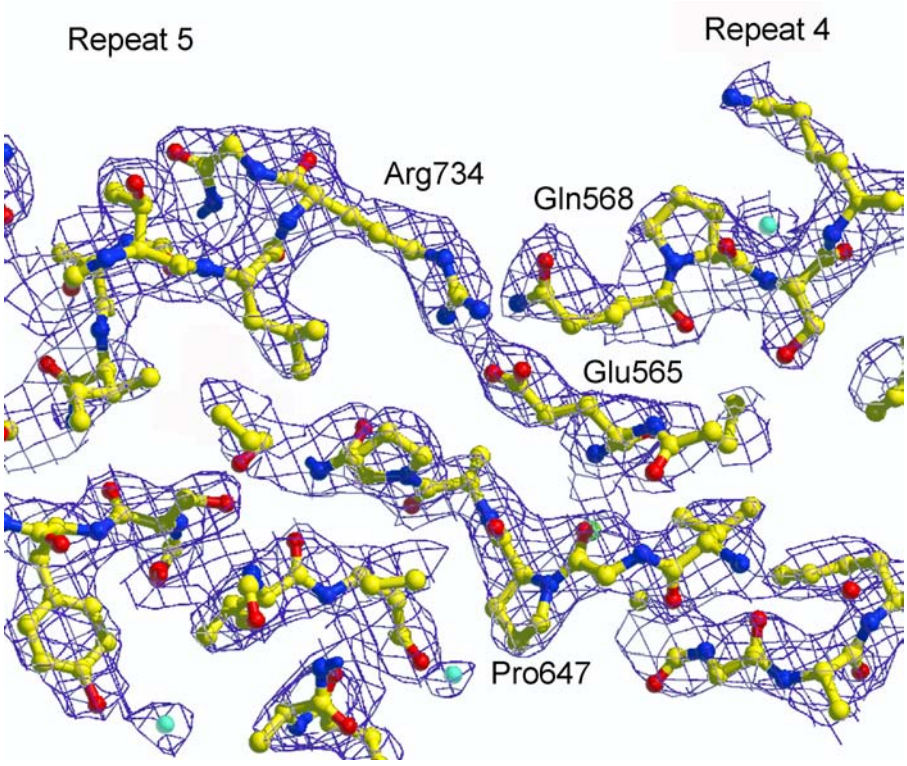


Figure 1.2.4 Interpreted electron density map of the border region between repeats 4 and 5. Pro647 placed between two β-strands connects repeats. Additionally side chains of Glu565 and Arg734 form a salt bridge.

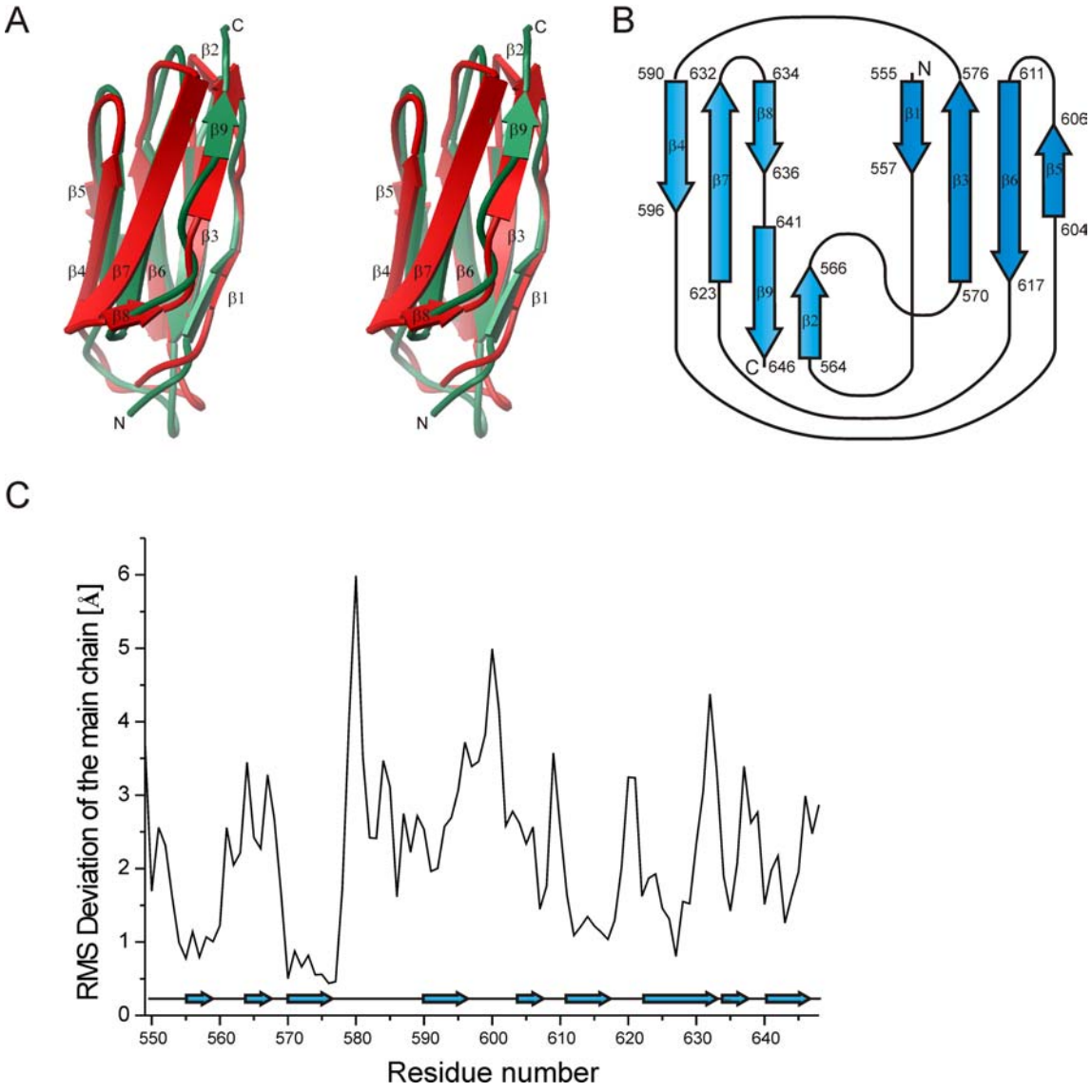


Figure 1.2.5 (A) Comparison of the NMR model (green) of repeat 4 with its crystallographic structure (red). (B) Diagram of the β-strands building repeat 4. Two groups of strands form a β-sandwich. (C) rms deviation plot of superimposed crystallographic and NMR models of repeat 4. Beta structured regions are marked by arrows at the bottom of plot.

As expected from their amino acid sequences (46% identity) the structures of repeat 4 and repeat 5 are similar. Rms deviation of main chain atoms between repeat 4 and 5 is 1.06 Å. Repeat 5 appears to have the β5 strand longer by one residue, the β2 strand shifted towards the C-terminal direction and the β3-β6 sheet is shorter. There is also an additional strand β8, which has no equivalent in repeat 5. A comparison of the two structures is shown in Figure 1.2.6. Repeat 6 differs from the two others by both sequence and spatial orientation. However the overall fold of the repeat is the same as for 4 and 5. Repeat 6 it has an additional long strand at the C-terminus and lacks strands β1 and β2. These are substituted by β9 from the monomer unit of the other repeat 6 in the dimer.

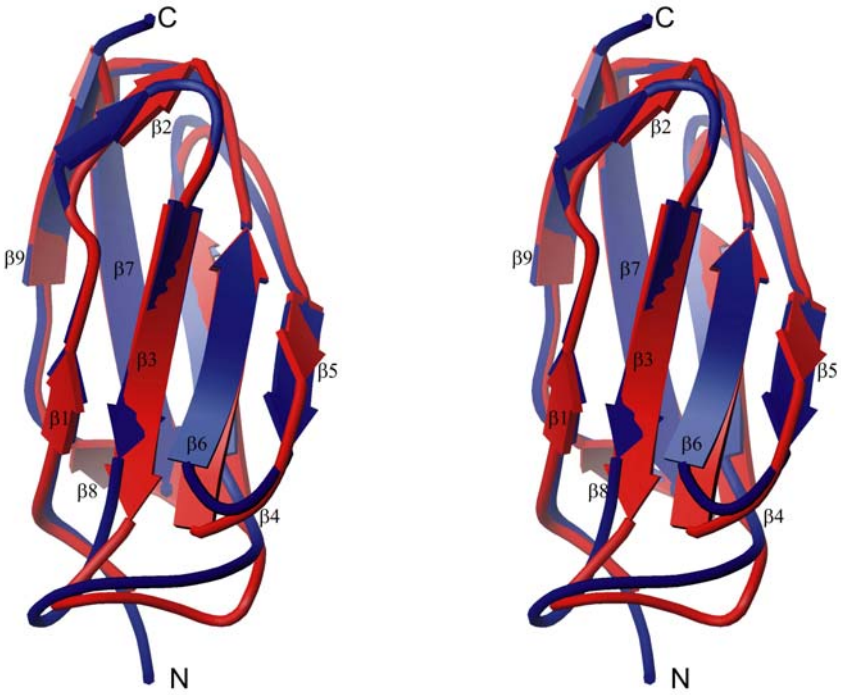


Figure 1.2.6 Superposition of repeats 4 and 5. Repeat 4 is shown in red and 5 in blue.

1.2.4 Discussion

1.2.4.1 Model for the dimerization of ddFLN

The present structure shows the way in which successive non-dimerizing repeats pack, the crucial information needed to produce a high-resolution model of the entire molecule. Solved structure of the rod repeats ddFLN(4-6) indicates that the arrangement of chains in the entire dimeric molecule is generally in agreement with the architecture proposed by McCoy et al. 1999 based on the structure of ddFLN(5&6). These two structures show that the two chains overlap only at repeat 6, which is very different from that of previously proposed models in which the rod domains of each chain overlapped for all or most of their length (Stossel et al., 2001, McCoy et al., 1999, Fucini et al., 1999).The present structure shows the way in which successive non-dimerizing repeats pack, the crucial information needed to produce a high-resolution model of the entire molecule. Our structure shows that the tilt angle between the non-dimerizing repeats of the rod is 115^{o}.

The complete filamin sequences (the actin binding and rod domains) are well known for their intrinsic flexibility. If one compares the F-actin crosslinking activities of filamin and α-actinin then filamin forms three-dimensional filament networks whereas α-actinin tends to bundle parallel filaments. In α-actinin the rod domain also consists of six repeats that are however based on a triple-stranded coiled-coil α-helix (Djinovic-Carugo et al., 2002). The rod domain in α-actinin is rigid and the linker region between repeating units is α-helical with no discontinuity in the secondary structure between the helical end of the linker of one repeat to the beginning of the next repeat. Still, there is a considerable elasticity observed in thin filaments crosslinked by α-actinin that has been attributed to the flexibility of the entire molecule near its actin-binding domains (Winkler et al., 1997). The linker sequence between the first repeat of the rod domain and the actin-binding domain is 25 amino acid long in α-actinin. In ddFLN the corresponding segment is also 25 residues long - the longest amino acid stretch among globular domains of the molecule (Figure 1.2.8). This fragment therefore can be a place of intra-molecular rotation.

Propensity of filamins to form three-dimensional filament networks is very pronounced and led to the description of ddFLN as an "actin-saving" protein, i.e. the viscosity of a rather dilute F-actin solution increases sharply of only tiny amounts of filamin are present (Winkler et al., 1997; Janssen et al., 1996). So far this was

explained by the putative flexibility of the rod domain in filamins, in contrast to the α-helical rod domain of α-actinin. Our crystallographic data do not indicate high flexibility of linker sequences that connect repeats, suggesting that filamin does not behave as a row of beads connected by freely-rotating linkers. The inter-repeat region between domains 4 and 5 is short and includes a stabilizing Pro-Ala-Pro sequence. Inter-repeat orientation is in case of repeat 5 and 6 exactly the same as in the previously published structure, suggesting that there is no inclination for reorientation. DdFLN(4-6) crystallizes in perfect non-crystallographic symmetry, which is not the case in highly flexible molecules. 1D proton NMR spectra also show no evidence of flexible segments in the protein (see Figure 2D in ref. 15). NMR spectroscopy is a powerful technique for detecting flexible structure and for studying folded-unfolded equilibrium in macromolecules (Rehm et al., 2002a; Weber et al., 2000). No pronounced intensities at chemical shifts near ~8.3 ppm, an indicator of amides in random coil configuration and no sharp peaks that could be attributed to the flexible peptide fragments are seen in the NMR spectrum (Rehm et al., 2002). All these observations however are not contradictory with a view of filamentous filamin having a plastic shape. For example, changes in interfilament spacing can easily be accommodated by reorientation on and/or bending of the crosslinking bridges without significant distortion in their length.

Figure 1.2.6 summarizes the current structural knowledge of the ddFLN dimer. The structure of the actin-binding domain can be modeled on the basis of its sequence similarity to the actin-binding domain of fimbrin (Goldsmith et al., 1997). The dimeric structure of rod domains 4, 5 and 6 described in this study completes the structural knowledge of the molecule, so that the conformation of all of the domains is either known directly or can be modeled with high confidence. Although the structures of rod domains 1–3 have not been experimentally determined to date, their sequences closely resemble those of rod domains 4 and 5. Electron microscopy of shadowed ddFLN molecules has indicated that they are between 350 and 490 Å long (Brink et al., 1990; Condeelis et al., 1984). The 12 Ig domains in ddFLN rod pack into roughly 410 Å in our present model. The actin-binding domain (ABD) of fimbrin has dimensions of ~60 Å × 45 Å × 40 Å. This implies that the ABD domain has to be positioned at an angle to the long axis of the molecule for the model to be approximately in agreement with electron microscopy measurements.

The tilting between the rod domains of ABP-120 is reminiscent of the packing of the Ig domains in the type III fibronectin (Leachy et al., 1996) and titin (Winkler et al., 1997). A four-domain fragment of titin showed a degree of bending and twisting of the domains with respect to each other, which resulted in an overall shortening of the length by 10–17% compared with that expected for a fully extended, linear configuration (Improta et al., 1998; Amodeo et al., 2001; Fraternali and Pastore, 1999). These authors observed also that in one and two-domain constructs the motions around the interdomain connecting regions were restricted.

The structure of repeats 4,5,6 makes a half of the ddFLN rod domain. It appears that this part forms an elongated element, which however is not in a fully extended, linear configuration. Instead the long axis of each Ig domain is arranged at an angle to the long axis of the rod. The tilt angle between non-dimerizing repeats 4 and 5 is 110^{o}. We expect that the whole rod domain maintains this shape and would be then approximately 410 Å long. This configuration is similar to that found in the antibody molecule. The variable V_L and constant C_L segments of Fab, for example, are oriented with respect to each other with an angle of ca. 100^{o} (Fuccini et al., 1997; Bork et al., 1994).

We can also assume that the human FLNa rod domain, much longer than *Dictyostelium* FLN, forms a straight and long structure able to crosslink F-actin chains at places as far away as about 1000 Å. Inter-repeat interactions should also be stabilized by the linker sequence Pro-Ala-Pro in a similar fashion as in ddFLN because this sequence is conserved in human FLNa (Figure 1.2.9).

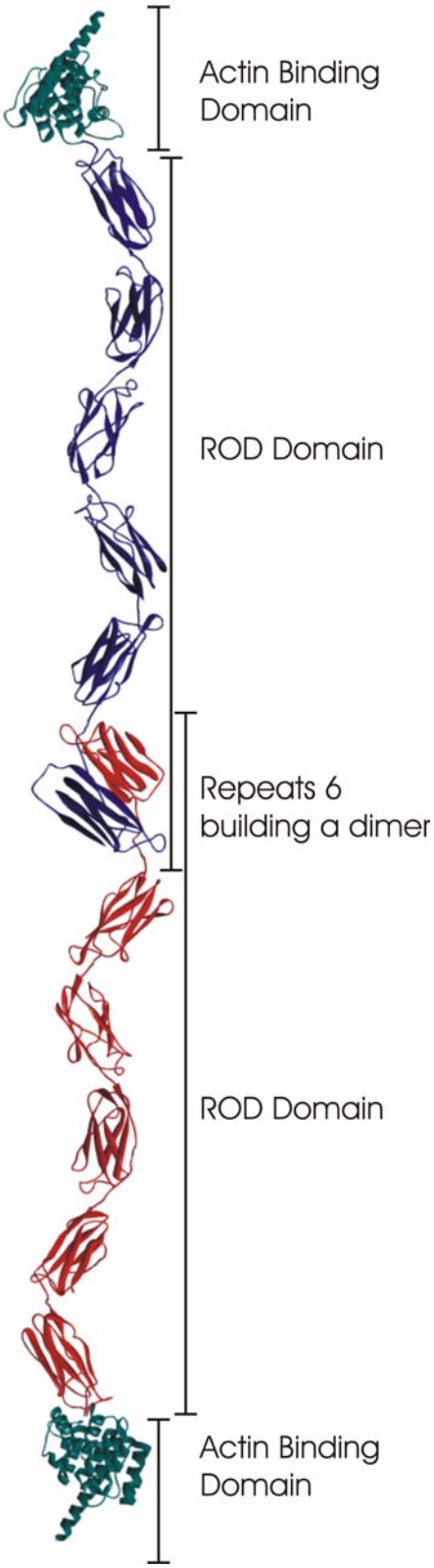
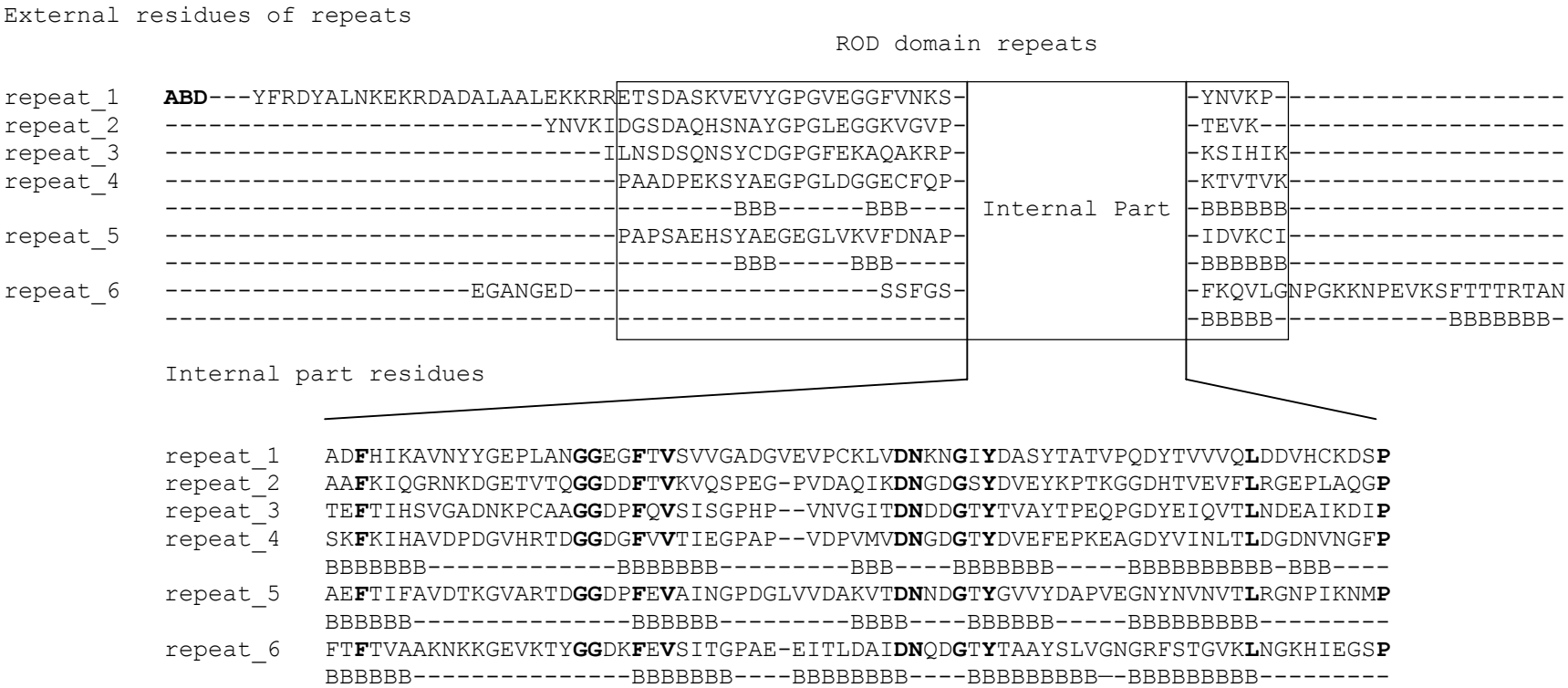


Figure 1.2.7 A homodimer model of the whole ddFLN protein. Inter-repeat angles are taken from repeats 4 and 5. Rod domain builds an extended chain interconnecting two actin-binding domains.

Figure 1.2.8

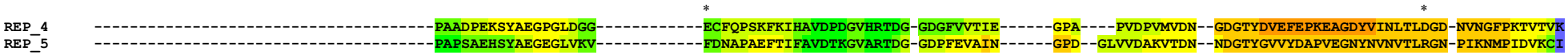
ddFLN:



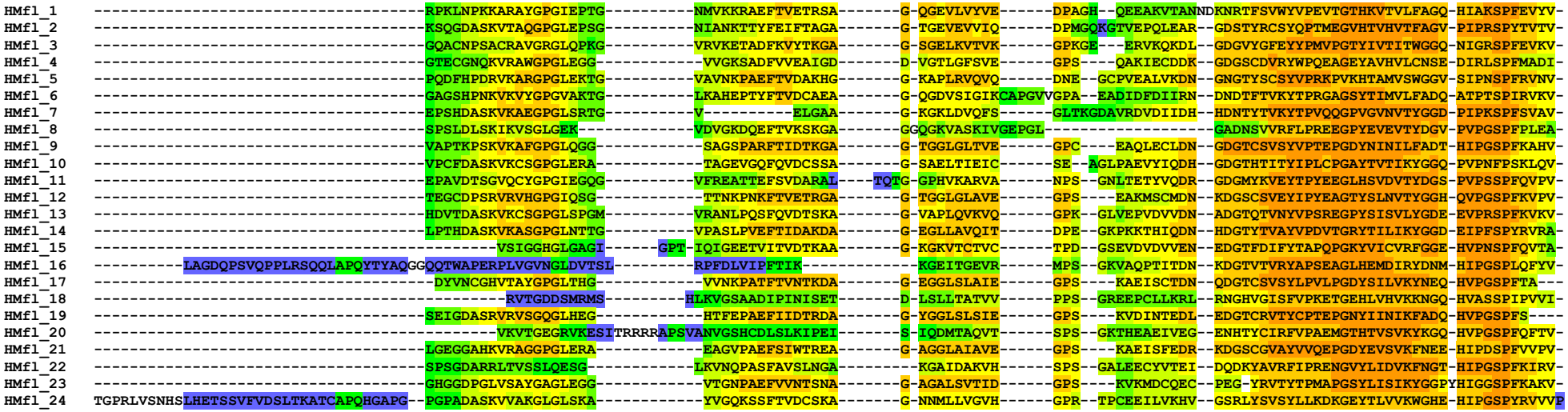
ABD: actin binding domain (250 amino acids)
BBBBBB: beta-structures

Figure 1.2.9

ddFLN:



hsFLNa:



1.3 Structure of ddFLN(1-6); building a complete model of filamin

1.3.1 Introduction

The structure of ddFLN(4-6) gave rise to new questions regarding the extent of rigidity of the entire rod domain. We have therefore undertaken structural characterization of the entire rod domain of *Dictyostelium* filamin.

1.3.2 Protein expression, purification and crystallization

A protein construct that contains a complete rod domain of ddFLN(1-6) was expressed and purified using the same procedure as that for ddFLN(4-6). After gel filtration, the protein was concentrated to around 1,6 mg/ml and used for initial screening for crystallization conditions. Screening was performed using a Cartesian robotic system to increase number of tested conditions and to minimize protein usage. The drop size was set up to 200 nl of protein solution plus 200 nl of buffer. Protein crystals appeared after six months in many conditions at 4°C; the protein was then checked by SDS-PAGE and N-terminal sequencing that revealed that the first repeat was not present. It was therefore assumed that the protein crystallized after the cleavage of repeat 1.

1.3.3 Data collection and structure determination

Among number of tested crystals only one was diffracting up to the resolution of 3.2Å on synchrotron radiation source, potentially allowing structure determination. The native dataset was collected at DORIS synchrotron beamline BW6 at 90K. The crystal appeared to belong to space group P21212 with unit cell dimensions a=62.72 b=108.46 c=140.00. Previously solved structure of ddFLN(4-6) was used as molecular replacement probe with program Molrep. Rotation solution gave two peaks of 4.39σ and 4.19σ over the highest noise peak of 3.81σ. Translation solutions were also unambiguous and yielded peaks of 45.61σ and 41.75σ over 10.19σ of noise. While the molecular replacement probe was refined, it appeared that due to a high model bias completion of a model was not possible. Therefore to resolve the structure of this construct new crystals that diffract to higher resolution are required. Crystallization trials are currently under way, however, because the precise cleavage site and its influence on crystallization has to be determined and additionally crystals need about half a year to grow to the sizes suitable for the data collection, the whole

structure solution process has been assumed to last more than a year. Thus, the structure of ddFLN(1-6) could not be included in this thesis and is a major goal of my postdoctoral work.

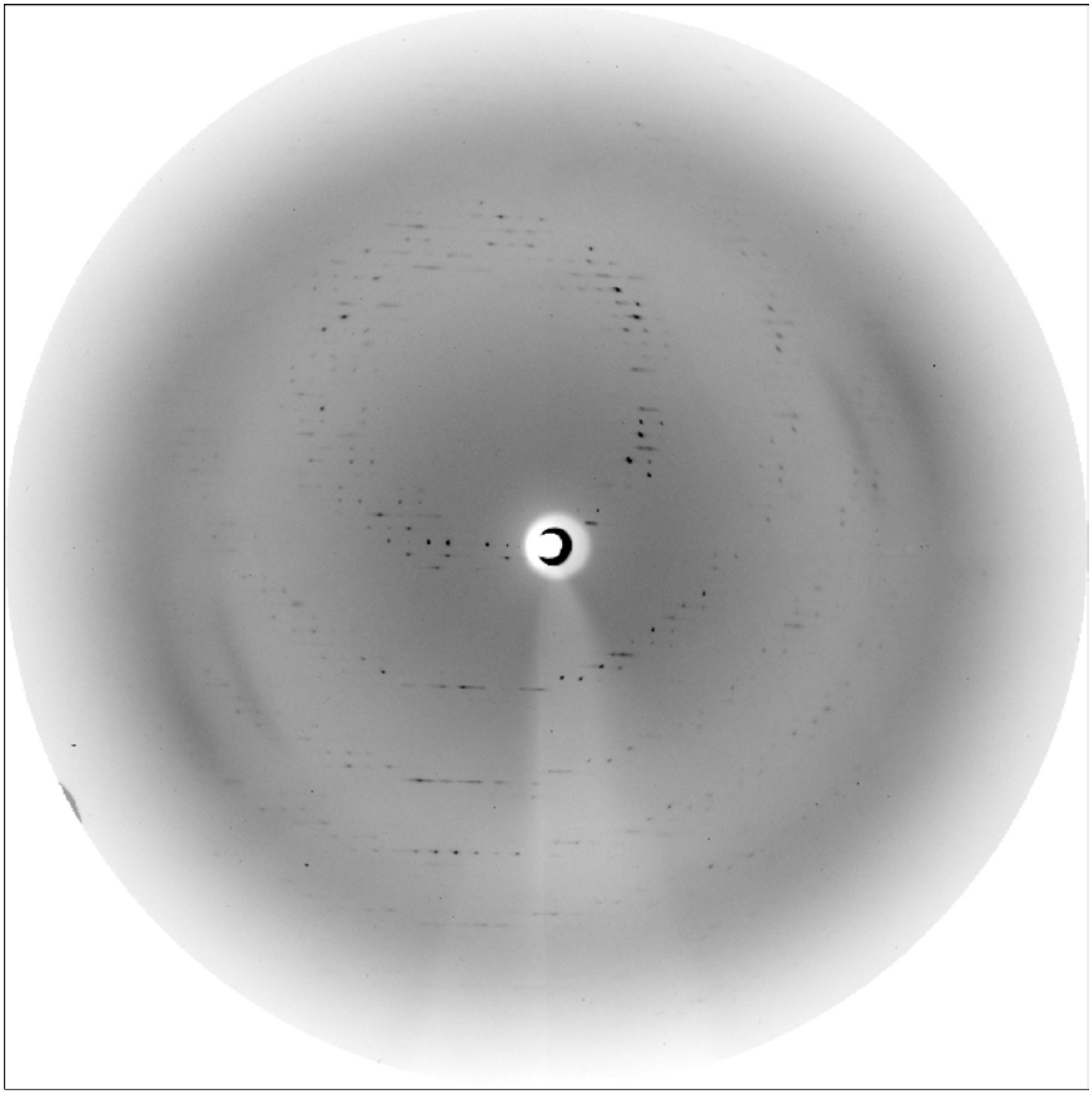


Figure 1.3.1. Diffraction pattern of the ddFLN(1-6) measured at BW6 MPG Beamline at DESY, Hamburg with MARCCD detector. Edge of the detector is at about 3 Å. The crystal belong to the space group P21212.

1.4 NMR and crystallographic structure of the N-terminal domain of the CAP protein

1.4.1 Introduction

Changes in the equilibrium between the monomeric (G) actin and polymerized (F) actin in response to extracellular signals affect the actin cytoskeleton, which controls cell shape, movement, polarity, and cytokinesis. Translation of extracellular and intracellular signals to changes in actin dynamics and organization is accomplished through an array of signaling pathways and proteins. Cyclase-associated proteins (CAPs) are developmentally important negative regulators of actin polymerization have signals from specific pathways been linked with elements of the cytoskeleton (Hubberstey and Mottillo, 2002, Field et al., 1990). Cyclase associated proteins are ubiquitous, highly conserved, proteins in eukaryotes and exhibit multifunctional activities due to the existence of domains involved in actin binding, adenylyl cyclase association, SH3 binding and oligomerization (Figure 1.4.1). Cyclase-associated proteins (CAPs) are known to modulate the actin-based cytoskeleton and play a role in Ras signaling (Hubberstey and Mottillo, 2002, Field et al., 1990).

The CAP protein of *Dictyostelium discoideum* is involved in the microfilament reorganization at anterior and posterior plasma membrane regions (Gottwald et al., 1996). The amino-terminal domain of *Dictyostelium* CAP encompasses residues 1–225, followed by the proline-rich region (residues 226– 255) (Figure 1.4.1). The amino-terminal region contains the adenylyl cyclase-binding site (Nishida et al., 1998), while the proline-rich middle domain interacts with proteins containing SH3 domains.

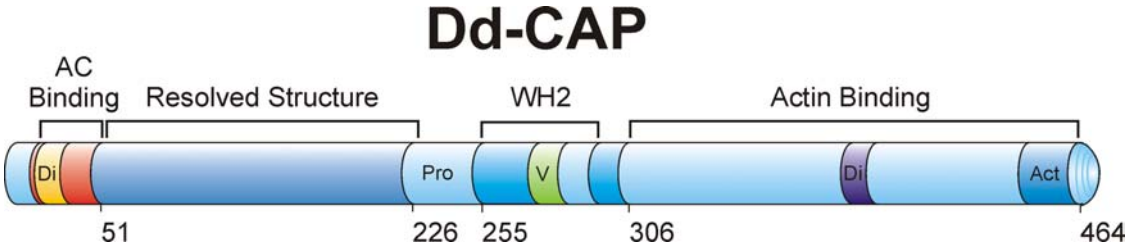


Fig 1.4.1. Domain structure of *D. discoideum* CAP (Gottwald et al., 1996; Hubberstey and Mottillo, 2002; Paunola et al., 2002). An adenylate cyclase binding domain (AC) and a dimerization domain (Di) are located at the amino terminus and are followed by the proline-rich region (Pro) and the WH2 domain (which includes a highly conserved verprolin homology region (V)). At the carboxy terminus is an actin binding domain (Act) and a second dimerization site (Di). The N-terminal domain consisting of residues 51-226 (CAP-N) used for our structure studies is highlighted.

1.4.2 Materials and methods

1.4.2.1 Sample preparation and NMR spectroscopy

The cDNAs encoding 176 residues from position 51–226 (CAPN151-678) were cloned into the *Nde*I and *Bam*HI restriction sites of the pT7-7 expression vector (Tabor, 1992). *E. coli* BL21 was used for an expression of protein. The following samples were available at concentrations ranging from 0.5 to 1.2 mM at pH 7.3 containing 10% D2O: Uniformly ^{15}N-, uniformly ^{15}N-^{13}C, selectively ^{15}N-Ala, ^{15}N-Phe, ^{15}N-Gly, ^{15}N-Ile, ^{15}N-Lys, ^{15}N-Leu, ^{15}N-Val, ^{15}N-Gly, ^{15}N- Ser-labeled and unlabeled protein samples. All NMR experiments were recorded at 300 K on Bruker DRX 600 and DMX 750 spectrometers. The sequence specific resonance assignment was accomplished as reported previously by Rehm et al. (2002b) using a pair of HNCA and CBCA(CO)NH triple-resonance spectra, with the help of ^{15}N-HSQC and ^{13}C-HSQC spectra of the uniformly labeled samples of CAP-N, and also with the ^{15}N-HSQC spectra of the amino acid type selectively labeled samples. HNCO, 3D ^{15}N-NOESYHSQC and ^{13}C-NOESY-HSQC spectra were also used for the assignment. The chemical shifts of CAP-N have been deposited in the BioMagResBank under the accession number 5393.

1.4.2.3 Input constraints and structure calculation

Interproton NOE distance constraints were generated from the integrated cross-peaks volumes from the 3D ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC spectra using the program Sparky (Goddard et al., 2000). 1500 NOE distance constraints and 358 torsion angles restraints were used for the structure calculation in the program CNS Solve 1.1 (Brünger, 1998). An initial ensemble of 30 structures was generated in the Cartesian coordinate space using standard protocols for simulated annealing (Nilges et al., 1988, Weber et al., 2000). A stepwise refinement protocol was performed using the Powell energy minimization algorithms to obtain the lowest possible energy for all the structures (Powell, 1977). A final selection of the CAP-N structures was based on the low total energy criteria and the converged structure ensemble was then analyzed with the software program Suppose for the rms deviation calculations (Table 1.4.1).

Table 1.4.1. Parameters characterizing the structures determination of CAP-N in solution at pH 7 and 300K

Parameters	<SA>
rms deviations from idealized geometry	
Bond lengths (Å)	0.0046 ± 0.000
Angles (deg)	0.6191 ± 0.031
Impropers (deg)	0.5332 ± 0.021
Energies (kcal.mol-1)	
E_{all}	340.4
E_{bond}	3.3
rms deviations of NOE violations	
Number of violations ≥ 0.2 Å	21.9 ± 4.150
Number of violations ≥ 0.5 Å	0.21 ± 0.043
rms deviations from experimental constraints (Å)	
NOE class all (1500)	0.062 ± 0.043

<SA> represents the ensemble of 18 structures. All parameters are calculated using the program CNS_solve 1.1.

1.4.3 NMR structure of CAP-N

Dictyostelium discoideum CAP protein is a fully helical structure (Figure 1.4.2). The three-dimensional structure of CAP-N indeed consist of six antiparallel helices, (Figures 1.4.2), each of them containing at least 10 to 20 amino acids. The helices are arranged into a six-helix bundle, which is connected in the complete protein to the C-terminal domain through a proline rich linker. In detail, the folded N-terminal domain consists of six helices in the regions extending from 52-73, 75-100, 107-128, 136-153, 158-180 and 185-208. In helix α1 one turn is distorted and φ and ψ angles for the two amino acids 61 and 62 do not fit the ideal helix conformation (φ, ψ : -91.2º, -70.1º; -118.5º, 1.2º, respectively), in the minimized averaged (SAm structure).

The pairwise backbone atomic rms difference of the structure ensemble is 1.4 ± 0.3 Å to the mean structure, excluding residues 100-105 and 213-226. The structure of the C-terminal domain of *S. cerevisae* CAP has been solved recently (Dodatko et al., 2004) (PDB ID: 1K4Z). In contrast to our N-terminal domain structure, the C-terminus of CAP is built solely by parallel β-strands that form a right-handed β-helix of six turns. The β-helix itself forms a homodimer with two β-structures arranged antiparallel to each other. It is interesting to note that the cyclase and actin binding sites are located in the whole protein on positions that are structurally independent from each other.

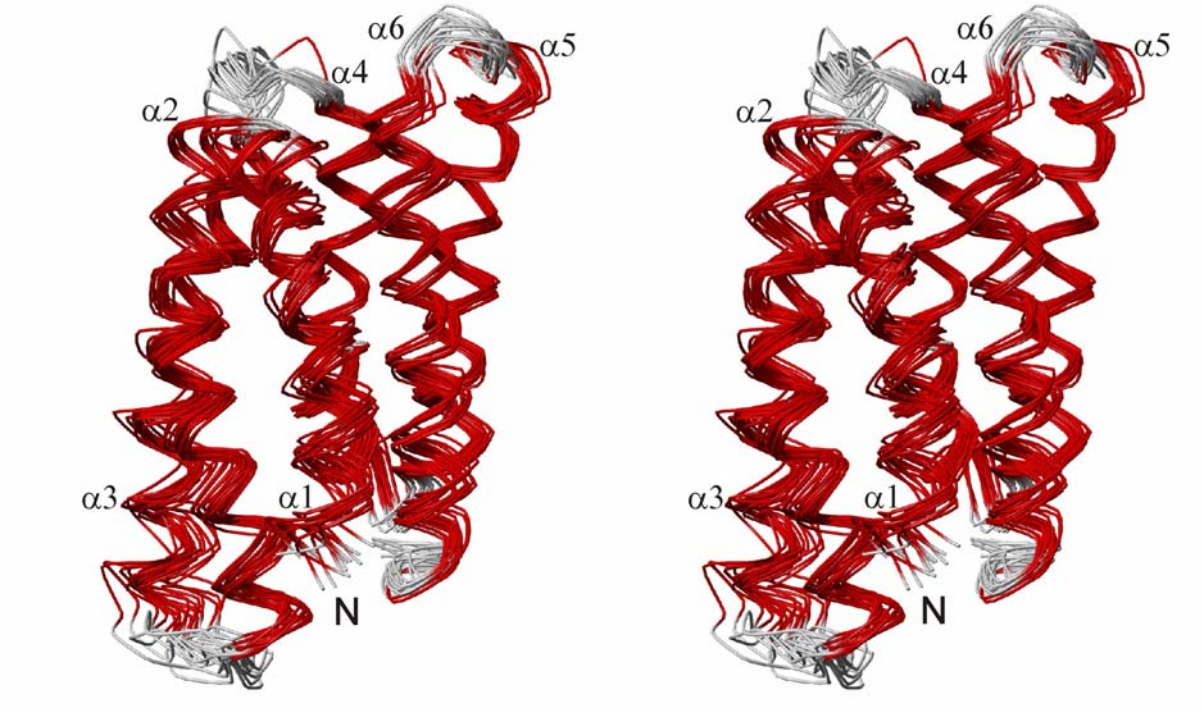


Figure 1.4.2. Stereoview of the backbone atoms (N, Cα, C and O) of all residues for the family of 18 structures of CAP best fit to N, Cα and C atoms of the regions with regular secondary structures (52-73, 75-100, 107-128, 136-153, 158-180 and 185-208).

1.4.4 Comparison to the X-ray structure

The overall folds of the structures solved by NMR and X-ray crystallography are very similar (Figure 1.4.4). The rms deviation between an averaged NMR structure and the X-ray model is 1.79 Å for all backbone heavy atoms (excluding the C-terminal residues 209- 226) and even lower (1.6 Å) for helices only. The structures differ mostly in loop regions that lack well defined secondary structure elements. The number of NOEs assigned to these regions is lower than that for the helical regions and larger rms differences are observed in the ensemble of the NMR structures in these parts. Therefore it is not possible to ascertain whether the difference is due to flexibility of these fragments or lack of the NMR constraining data (or both). On the other hand, crystal packing can restrict loop positions in the X-ray structure. For CAP-N this indeed seems to be the case as all loops and the C-terminus are in contacts with molecules from neighboring cell units. The α-helix between residues 53 to 72 (α1) is almost identical in both structures, so are helixes α3 and α6. The beginning of α2 (between residues 76-99) is located in a more external position in the NMR model, while its end is buried deeper into the center of the molecule compared with the X-ray structure. Helix α5 is bent in the NMR model, near residue 170, shifting amino acids 158-169 to a more external position, while the X-ray model shows a straight helix. The major difference in the structures pertains to helix α4. The NMR structure shows a continuous α-helix built by residues 136-153, while in the crystallographic model the helix ends at residue 143 and residues 144-158 form a long linker between the helices α4 and α5. CAP-N crystallized both as a dimer and a monomer from the same drop (30% PEG8000, 0.2 M $MgCl_2$, 10 mM β-mercaptoethanol and 0.1 M MES, pH 6.1; Ksiazek et al., 2003). The final X-ray structure of the dimer comprises CAP-N and a magnesium atom, and was solved to 1.4 Å resolution. The monomer crystals diffracted to 1.7 Å. Comparison of the monomer structure with that of the dimer showed that they were essentially identical. For the dimer, Arg 127, Asp 128 (end of α3) and Glu 144 from each monomer are at hydrogen bonding distances to water molecules, which surround the magnesium ion at an average distance of 2.1 Å. The interaction of Mg with Glu 144 could then explain why helix α4 is broken at this residue in the X-ray model of the dimer. No Mg has been seen in the X-ray model of the monomer structure. However, Mg still could be present at Glu 144 since the metal might not be recognized at the 1.7 Å resolution for the monomer crystals and lower occupancy of Mg would further weaken chances

of its detection. NMR titration of CAP-N with Mg^{2+} did not show any dimerization, which could be detected by changes in NMR line width or by induced chemical shifts. We propose that the NMR structure is closer to the native structure as crystals of both dimer and monomer were grown at a non-physiological concentration of $MgCl_2$ (0.2 M).

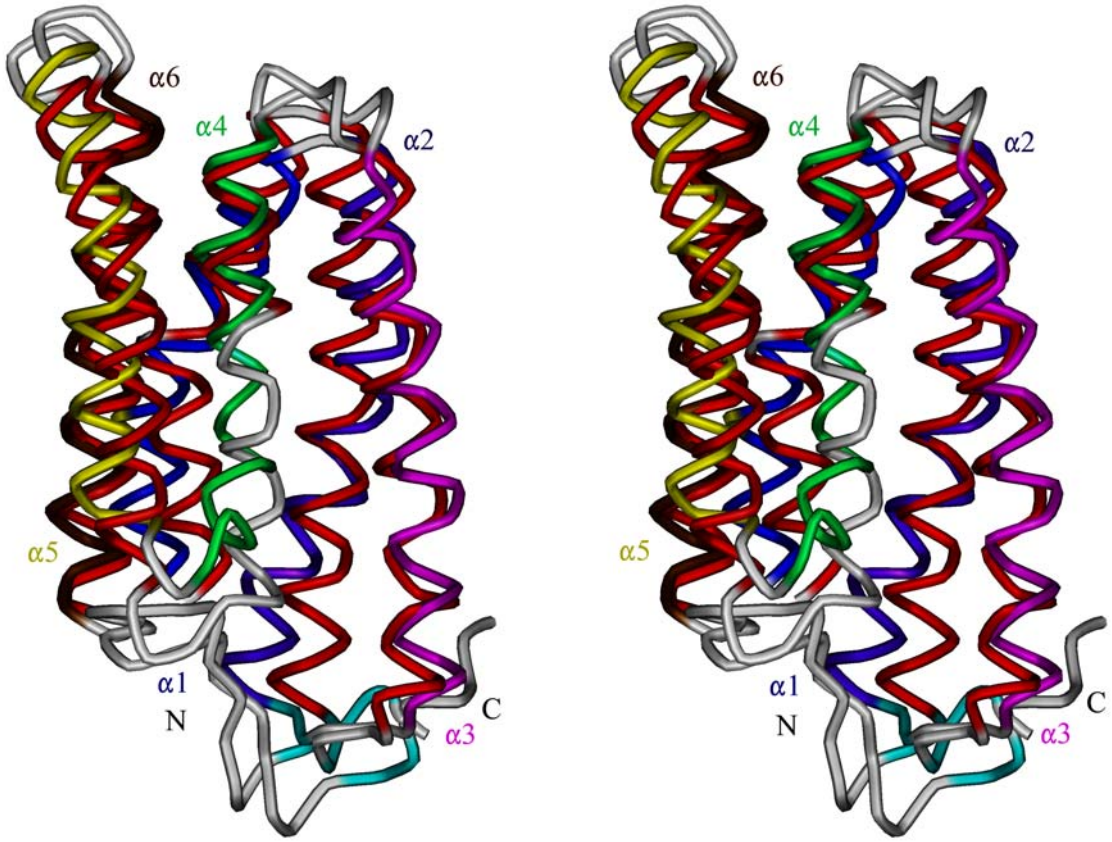
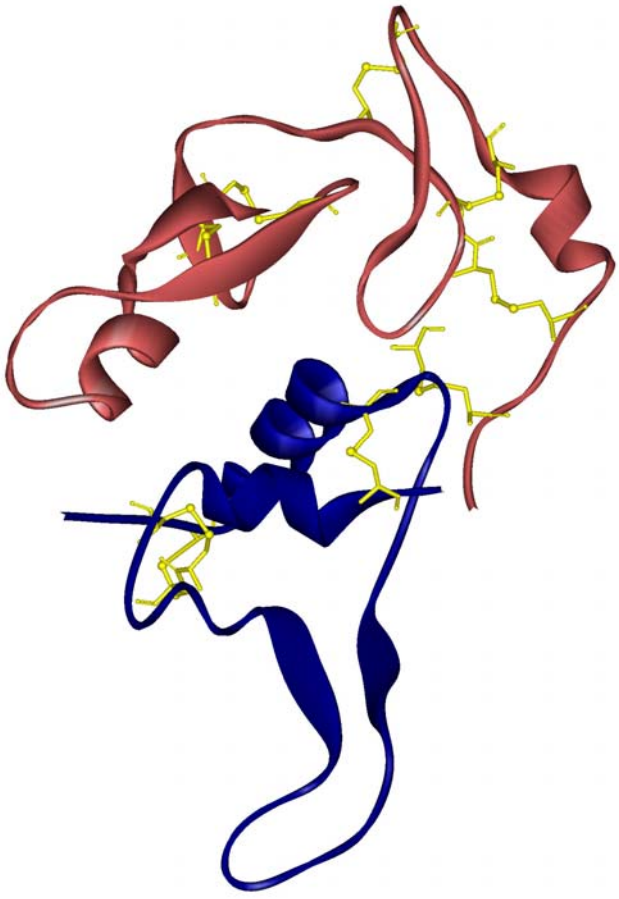


Figure 1.4.4. Stereoview of the Cα-backbone of the X-ray structure (helices in red) superimposed on the minimized averaged NMR structure (the six helices are shown each in different color).

Chapter 2

Insulin-like growth factor binding proteins (IGFBPs)



2.1 The IGF system

The type, function, and property of a growing and developing cell is predestined by external signals it receives from the surrounding environment. Contacts with neighboring cells and signals received from membrane receptors, which are activated by various specific growth factors, modulate gene expression pattern of the cell, leading to differentiation. The very same stimuli decide fates of the mature cell, i.e. the rate of protein and nucleic acid synthesis, cell cycle progression, and survival depend on the presence of cell-to-cell junctions and constant supply of the protein factors. In a cell deprived of this steady stimulation, anabolic pathways cease and suicidal pathways take over, leading to apoptosis. Deregulation of extracellular developmental signals may lead to an opposite extreme situation: quiescent cells commence excessive proliferation, the first step in cancerogenesis.

The IGF system is an ensemble of mutually interacting proteins that plays the key role in regulation of cell development, metabolism and survival. The IGF system involves two ligands, insulin-like growth factor-I and -II (IGF-I and IGF-II) that interact with specific type 1 or type 2 IGF receptors (IGF-IR and IGF-IIR). There are also six high affinity IGF binding proteins (IGFBPs) that modulate IGF availability and bioactivity.

2.1.1 IGFs

The IGFs were first identified by Salmon and Daughaday as “sulphation factor activity” (Daughaday et al., 1972). Later the term “somatomedin” was replaced by “insulin-like growth factor” because of its structural homology with insulin (Rinderknecht and Humbel, 1976). Rinderknecht and Humbel revealed the amino acid sequences of both IGF-I and IGF-II in 1978 (Rinderknecht and Humbel, 1978a; 1978b). Almost all cell types express IGF receptors, and therefore respond to IGF stimulation of growth and differentiation. Both IGF-I and –II are required for normal embryonic and fetal development. Postnatally IGF-I mediates the growth-promoting effect of growth hormone (Schlechter et al., 1986; Daughaday et al., 1999) and is the major regulator of animal skeletal growth, maintaining bone mass throughout life. IGF-I augments wound healing by acting at several sites during the repair process (Chan and Spencer, 1998; Suh et al., 1992). Also, nerve cell survival and regeneration is promoted by IGF-I (Leventhal et al., 1999; Jung et al., 1998). IGF-I is anti-apoptotic and autocrine production by tumors is common. In the kidney, IGF-I

increases the glomerular filtration rate, phosphate transport, and the activity of 25-hydroxyvitamin D-1-α-hydroxylase (Halloran and Spencer, 1988). The primary regulator of IGF-I expression is growth hormone. Developmental expression of IGF-I in various tissues precedes however that of GH, supporting an independent role of IGF-I in embryonic and fetal life.

Human IGF-I and IGF-II are single-chain polypeptides. IGF-I is a basic peptide and contains 70 amino acids, with IGF-II being slightly an acidic protein containing 67 amino acids. The amino acids of both IGF-I and -II are grouped into domains A and B (similar to insulin), domain C (analogous to the connecting peptide of proinsulin) and the C-terminal octapeptide (D region, 63–70) that has no counterpart in insulins and proinsulins (Baxter et al., 1992). Both IGF factors contain three disulphide bonds, and display approximately 62 % sequence homology with each other and 47 % with insulin.

2.1.2 IGFBPs

The insulin-like growth factor binding proteins (IGFBPs) are multifunctional proteins that modulate the biological effects of IGFs. Modulation is performed directly by sequestering IGFs, and indirectly by several mechanisms, including specific binding to a number of plasma, extracellular matrices and cell surface molecules by conditional proteolysis, rapid internalization into target cells, and translocation into the nucleus (reviewed in Firth and Baxter, 2002; Clemmons, 2001; Jones and Clemmons, 1995; Bach and Rechler, 1995). In addition to their IGF-modulating functions, IGF and IGF receptor-independent actions of IGFBPs are increasingly apparent (reviewed in Baxter, 2000; Firth and Baxter, 2002; Schedlich et al., 2004; Butt et al., 2000). In circulation and interstitial fluids, IGFBPs are the major carrier proteins for IGFs. IGFs are largely unavailable in these environments, as they are tightly associated with IGFBPs and are believed to be released after IGFBP proteolysis for binding to the cell surface IGF receptors (reviewed in Bunn and Fowlkes, 2003). Thus, by sequestering IGFs away from IGF receptors, IGFBPs may inhibit mitogenesis, differentiation, survival, and other IGF-stimulated events. Alternatively, the interaction of IGFBPs with cell or matrix components may concentrate IGFs near their receptor, enhancing IGF activity (reviewed in Clemmons, 1999, 2001). IGF- and IGF receptor-independent IGFBP actions are shown to have

major effects in regulation of cell adhesion and migration and in the cell cycle and apoptosis, which involves 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 carboxy-terminal basic motifs incorporating heparin binding domains 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 have not been identified as functional receptors. 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 LeRoith and Helman, 2004; Nemoto and Finkel, 2004; Pollak et al., 2004; Ali et al., 2003; Mazerbourg et al., 2004; Firth and Baxter, 2002; Cohen et al., 2000; Khandwala et al., 2000; Bach, 1999). The IGF-I receptor especially has come into focus in cancer research since it has elevated expression levels in most cancer cells (for recent reviews, see LeRoith and Helman, 2004; Pollak et al., 2004; Firth and Baxter, 2002; Khandwala et al., 2000). The IGFBP family comprises six proteins (IGFBP-1 to -6) that bind to IGFs with high affinity (for recent reviews, see Clemmons, 2001; Firth and Baxter, 2002). The precursor forms of all six IGFBPs have secretary signal peptides of between 20 and 39 amino acids, and the mature proteins have between 216 and 289 amino acids. All IGFBPs share a common domain organization and also a high degree of similarity in their primary protein structure (identities around 30%–40%) . The highest conservation is found in the N- (residues 1 to ca. 100) and C- (from residue 170) terminal cysteine-rich regions. Twelve conserved cysteines (ten in IGFBP-6) are found in the N-terminal domain, and six are found in the C-terminal domain. Both the N- and C-terminal domains participate in binding to IGFs (Payet et al., 2003; Shand et al., 2003; Firth and Baxter, 2002; Firth et al., 2001; Clemmons, 2001; Buckway et al., 2001; Baxter, 2000; Baxter et al., 1992), although the specific roles each of these domains play in IGF binding have not been decisively established. The central, weakly conserved part (L domain) contains most of the cleavage sites for specific proteases (Chernausek et al., 1995; Bunn and Fowlkes, 2003). Knowledge of the structure and mechanism of the IGFBPs binding to IGF-I and IGF-II is crucial for a

true understanding of their function. However, the complete structures of IGFBP and its complex with IGFs have not yet been determined. Several properties of IGFBPs and IGFs, such as the intrinsic disorder of the central one-third of the IGFBPs, the highly disulfide-bridged folded domains, and rapid aggregation of IGFs, have hindered structural studies with these proteins (reviewed in Clemmons, 2001; reviewed in Hwa et al., 1999; reviewed in Carrick et al., 2002; Cooke et al., 1991; Torres et al., 1995). Only in the last few years have reports on structural characterization of small domains of IGFBPs appeared (Kalus et al., 1998, Zeslawski et al., 2001; Headey et al., 2004, Yao et al., 2004). The miniIGFBP-5 structure comprising residue range Ala40-Ile92 of IGFBP-5 solved by Zeslawski et al. (2001) was used as a molecular replacement probe in our crystallographic calculations.

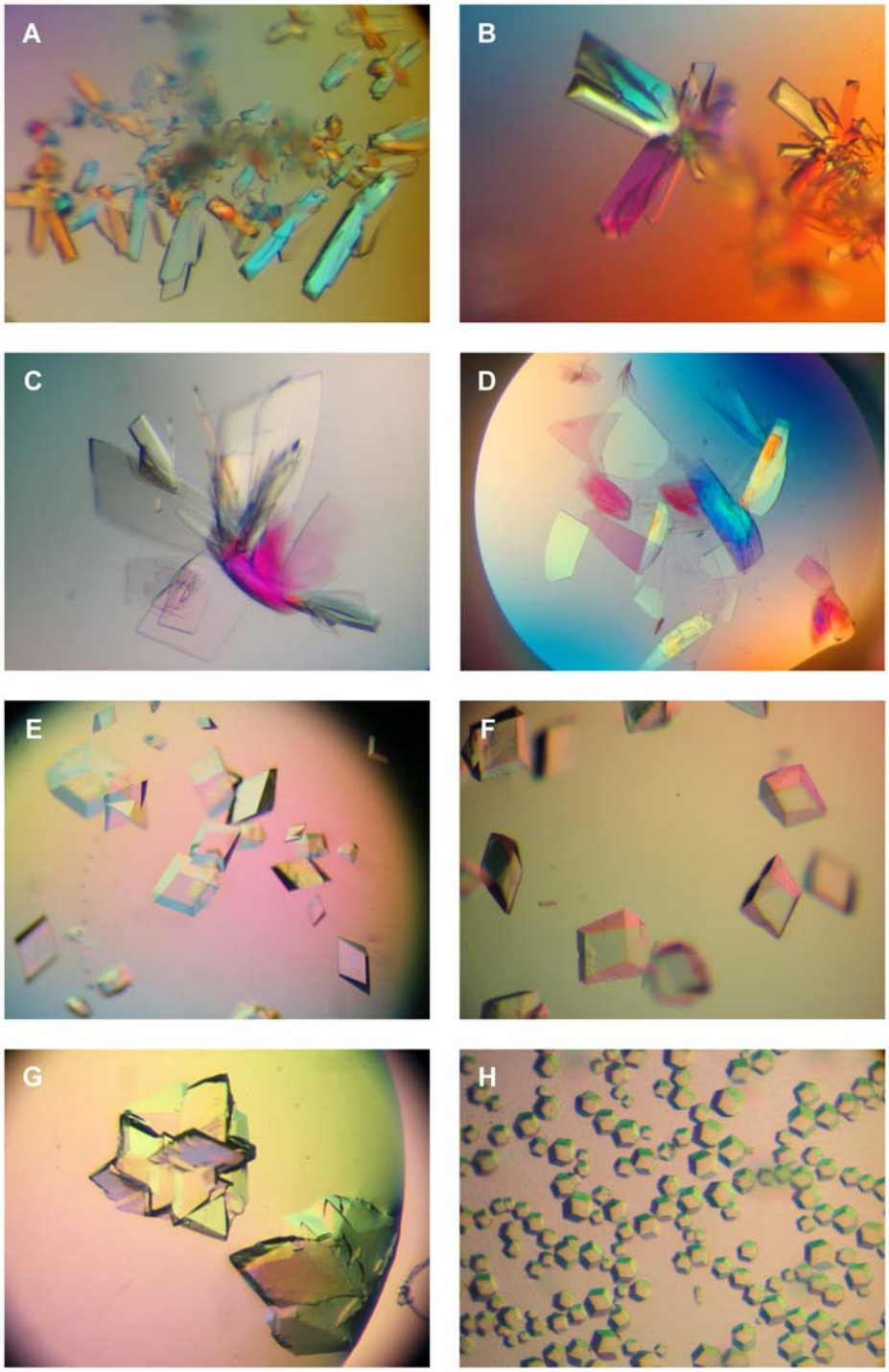


Figure 2.1. Examples of crystallization experiments. **A**; **B**: The X-ray quality orthorhombic crystals of the NBP-4(3-82)/IGF-I binary complex belonging to the $P2_12_12_1$ space group, grown from 23% PEG 1500, 50 mM Tris pH 8.2. The noticeable growth along a streak line arises from the micro-seeding technique used. **C**; **D**: the monoclinic plate crystals of the NBP-4(1-92)/IGF-I binary complex, obtained from 0.2 M lithium sulfate, 0.1 M Bis-Tris pH 5.5 and 25% w/v PEG 3350. The crystals belong to the $P2_1$ space group. **E**; **F**; **G**: rhombohedral crystals of the CBP-4/ NBP-4(3-82)/IGF-I ternary complex that appeared in 1.0 M lithium sulfate,1% w/v PEG 8000. Space group: R32. **H**: the ternary complex crystals from 1.0 M lithium sulfate, 15% w/v PEG 8000.

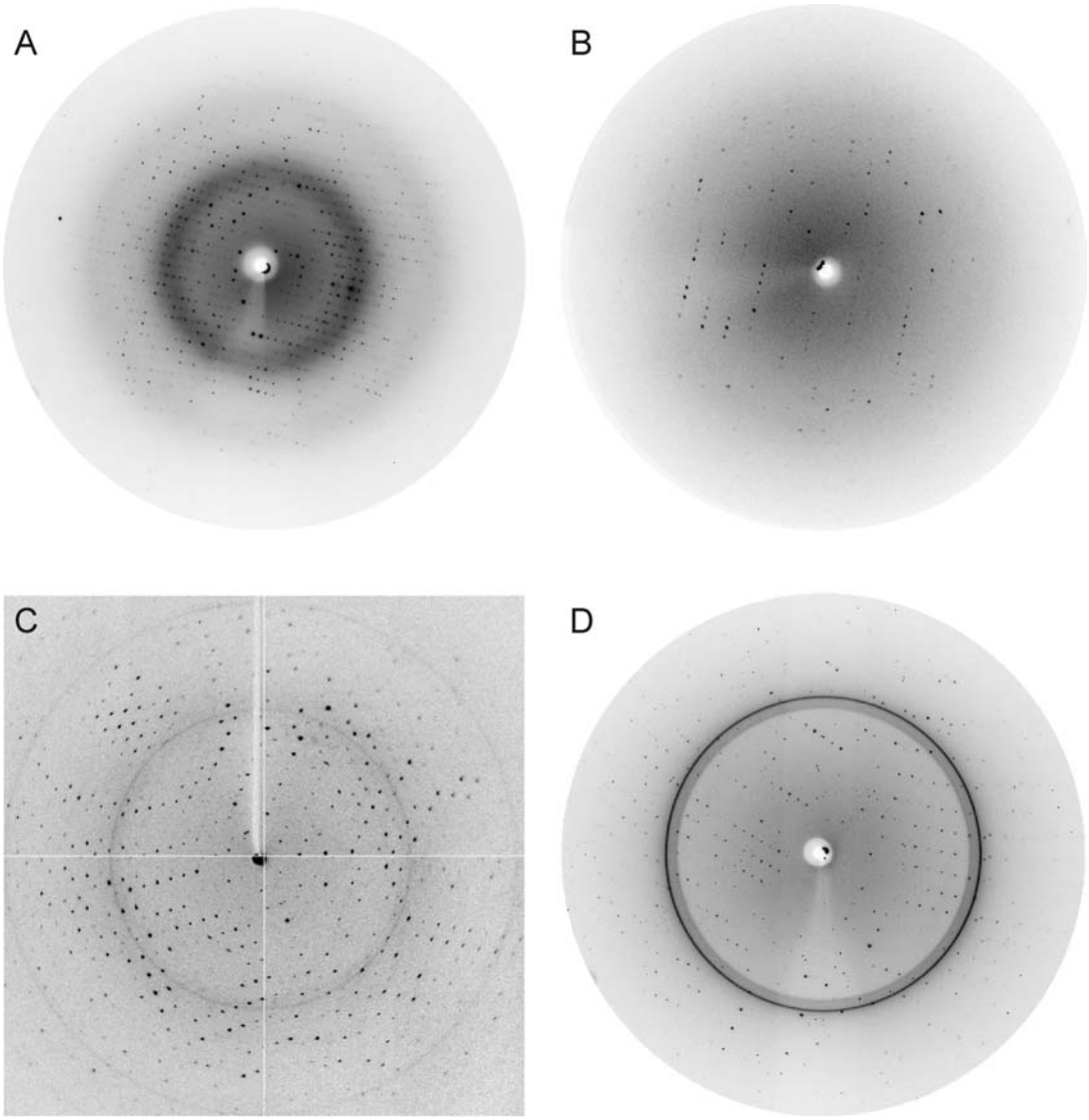


Figure 2.2. Diffraction patterns of crystals of the binary and ternary complexes. **A**: A frame from the MAR CCD165 (BW6, DESY, Hamburg). The crystal of NBP-4(3-82)/IGF-I (space group $P2_12_12_1$) was rotated 2°. The edge of the image is about 1 Å. **B**: A frame from the MAR 345 image plate, in-house rotating copper anode X-ray source. The $P2_1$ crystal of NBP-4(1-92)/IGF was rotated 3°. Resolution at the edge of the detector is ca. 1.5 Å. **C**: Diffraction image from the ADSC Q210 detector (ID29 beamline, ESRF, Grenoble, France). The C2 crystal of the ternary NBP-4(3-82)/CBP-4/IGF-I complex was rotated 2°. **D**: The diffraction pattern of the R32 crystal of the NBP-4(3-82)/CBP-4/IGF-I complex from the MAR CCD165 (BW6, DESY, Hamburg). Resolution at the edge of the detector is ca. 2.8 Å.

2.2 Structure of IGFBP-4

2.2.1 Preparation of the protein

For subcloning of NBP-4 (residues 3–82 of IGFBP-4) and CBP-4 (residues 151–232), their sequences were introduced into the BamHI and HindIII or BamHI and XhoI restrictions sites of the pET 28a-vector (Novagen, Canada) in-frame into a His-T7-tag, two sets of oligonucleotides were designed for amplification of cDNA by PCR: 5#NT-IGFBP-4: cgcggatccattgag ggtcgcGCCATCCACTGCCCGCCCTGC; 3#NT-IGFBP-4: cccaagcttt cattaCAGGTCCATGCACACGCCTTGCCCG; 5#CT-IGFBP-4: cgcg gatccctggtgccgcgcggctccTGCCAGAGCGAGCTGCACCGGG; 3#CTterminal IGFBP-4: ccgctcgagtcattaCAGCTCCATGCACACGCCTTGCCCG. Restrictions sites, stop codons, and the bases encoding an N-terminal factor Xa or thrombin cleavage sites were introduced by using overhangs shown in small letters. Vector BP4-2/pFDX500 (ROCHE Diagnostics, Penzberg) (0.5 µg) containing the complete sequence indicatof IGFBP-4 was used as a template. Standard procedures were used for construction and verification of vectors pET28a (NBP-4) and pET28a (CBP-4), which were transformed finally into the *E. coli* strain Bl21(DE3) (Novagen) for overexpression. The proteins were produced and purified essentially as described by Kalus et al. (1998). An additional step of chromatography on the 8 ml-MonoS paracation exchanger column (Pharmacia, Sweden) was introduced, in which the proteins were bound to the column material in 25 mM phosphate buffer (pH 7.1), 25 mM NaCl, and bound proteins were fractionated by a linear gradient of 25–700 mM NaCl over 20 column volumes.

2.2.2 Crystallization and structure solution

The binary complex of NBP-4 and IGF-I (GroPep, Australia) was prepared by mixing equimolar amounts of the components. The complex was separated from any excess of either protein by gel filtration on the Superdex S75 column. The buffer used contained 5 mM Tris (pH 8.0), 50 mM NaCl, and 0.01% NaN3. Crystallization of the complex was carried out with the sitting drop vapor diffusion method. Initial screening of crystallization conditions for the NBP-4(3-82)/IGF-I binary complex yielded 3 hits. Crystal Screen I solutions 42 (0.05 M KH_2PO_4, 20% PEG 8000) and 43 (30% PEG 1500) contained a quasi-crystalline material. The Index screen solution 37 (25% w/v PEG 1500) produced micro crystals. The latter condition was further optimized in terms of PEG and protein concentrations. The largest influence on crystal formation, however, had addition of the Tris buffer (even at concentrations

as low as 12.5 mM) promoting growth of large crystals. Crystallization occurred in various conditions of pH (7.25-8.45), PEG (20-23%), and Tris (12.5-100 mM) concentrations. Micro seeding had to be applied to induce crystal growth in most drops. The crystals were obtained from 23% PEG 1500, 25 mM Tris (pH 7.8). They appeared in several days and grew to a final size of ca. 0.4 × 0.2 × 0.2 mm. Prior to plunge freezing, the crystals were soaked for ca. 30 s in a drop of a reservoir solution containing 15% v/v ethylene glycol as cryoprotectant. The crystals belonged to the space group *P*212121 and contained one complex per an asymmetric unit. A low-resolution data set (up to 2.1 Å) was collected from a plunge frozen crystal at a rotating anode laboratory source. A high-resolution data set up to 1.6 Å was collected on the MPG/GBF beamline BW6 at DESY, Hamburg, Germany. Collected data were integrated, scaled, and merged by the XDS and XSCALE programs (Kabsch, 1993). The structure was determined by molecular replacement with the Molrep program from the CCP4 suite (CCP4, 1994). The structure of the complex of IGF-1 and a fragment of the N-terminal domain of IGFBP-5 (miniNBP-5, PDB entry 1H59) was used as a probe structure (Zeslawski et al., 2001). Rotation search in the Patterson space yielded one peak of height 7.56 σ over the highest noise peak of 3.75 σ. Translation search gave a 16.28 σ peak over the noise height of 9.49 σ. The initial R factor of the model was 0.49. At this stage, the free atom model improvement was performed by using Arp/wArp (Perrakis et al., 2001). As the next step, an iterative model building was performed using Arp/wArp (Perrakis et al., 1999). The initial model consisted of 137 residues and was completed and revised manually by using Xfit software (McRee, 1999). Arp/wArp was used to add solvent atoms (Lamzin and Wilson, 1993). The structure was finally refined by the Refmac5 program (CCP4, 1994). Final electron density maps were of high quality; there were, however, no interpretable densities for side chains of residues Glu11, Arg16, Leu42, Glu66, and Glu81 of NPB-4 and Lys27, Arg37, and Arg50 of IGF-I. These side chains were removed from the model. The final R crystallographic factor was 0.186, and Rfree was 0.2582. Data collection and refinement statistics are summarized in Tables 2.1 and 2.2.

Table 2.1. Data collection statistics for NBP-4(3-82)/IGF-I complex crystal. Dataset measured at MPG/GBF beamline BW6 at DESY using MARCCD detector.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 2.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	Rmrgd-F	S_norm/ S_ano
20.00	10	5	15	33.3%	4.0%	3.9%	9	18.31	5.6%	5.1%	0.00
15.00	84	19	19	100.0%	5.2%	4.7%	84	27.86	6.0%	4.2%	0.94
10.00	388	69	69	100.0%	4.8%	5.2%	387	27.99	5.3%	2.3%	1.14
6.00	1991	310	320	96.9%	5.7%	5.9%	1989	27.09	6.2%	2.7%	1.07
5.00	1880	278	287	96.9%	6.1%	6.0%	1880	26.68	6.6%	3.1%	1.05
4.00	4120	606	630	96.2%	5.9%	5.8%	4119	27.60	6.4%	3.1%	1.02
3.00	17046	1708	1723	99.1%	6.9%	7.7%	17042	28.88	7.2%	2.9%	1.00
2.50	21645	2115	2128	99.4%	9.8%	10.2%	21645	26.56	10.4%	3.8%	1.00
2.30	4861	1401	1412	99.2%	4.2%	4.5%	4765	22.86	4.9%	3.2%	1.02
2.10	6703	1980	2020	98.0%	4.7%	5.1%	6543	20.20	5.6%	3.7%	1.02
2.00	4144	1288	1315	97.9%	5.5%	6.2%	4023	15.73	6.5%	5.0%	0.99
1.90	4825	1534	1607	95.5%	7.0%	8.0%	4671	12.05	8.3%	6.7%	0.99
1.80	5716	1788	1978	90.4%	9.1%	10.6%	5555	9.39	10.8%	8.8%	1.03
1.70	6866	2110	2473	85.3%	12.9%	14.0%	6703	7.43	15.4%	12.4%	1.05
1.60	7847	2394	3114	76.9%	19.8%	20.3%	7665	5.45	23.6%	17.8%	1.03
total	88126	17605	19110	92.1%	7.0%	7.4%	87080	16.86	7.5%	4.4%	1.01

$$R_{observed} = \frac{\sum |I(h,i) - I(h)|}{\sum I(h,i)} \qquad R_{merge} = \frac{\sum_h \sum_i^{n_h} |\hat{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}} \qquad R_{meas} = \frac{\sum_h \sqrt{\frac{n_h}{n_h - 1}} \sum_i^{n_h} |\hat{I}_h - I_{h,i}|}{\sum_h \sum_i^{n_h} I_{h,i}}$$

(Diederichs and Karplus, 1997)

$R_{expected}$ - expected R-FACTOR derived from Sigma(I)
R_{meas} - redundancy independent R-factor (intensities)
S_{norm} - mean value of Sigma(I) for acentric reflections in each resolution shell assuming friedel's law is true.
S_{ano} - mean value of Sigma(I) for acentric reflections in each resolution shell assuming friedel's law is false.
Thus, anomalous scattering contributions to the intensities are indicated by S_norm/S_ano>1

Table 2.2. Data collection and refinement statistics for the NBP-4(3-82)/IGF-I complex

Data Collection	
Space group	$P2_12_12_1$
Cell constants (Å)	a=34.47 b=54.28 c=74.55
Resolution range (Å)	37-1.6
Wavelength (Å)	1.542, 1.05
Observed reflections	88126
Unique reflections	17605
Whole range:	
Completeness (%)	92.1
R_{merge}	4.4
I/σ(I)	16.86
Last shell:	
Resolution range (Å)	1.6-1.7
Completeness (%)	76.9
R_{merge}	17.8
I/σ(I)	5.45
Refinement	
No. of reflections	17388
Resolution (Å)	30-1.6
R-factor (%)	18.65
R_{free} (%)	25.82
Average B ($Å^2$)	24.51
R.m.s bond length (Å)	0.017
R.m.s. angles (°)	1.83
Content of Asymmetric Unit	
No. of protein complexes	1
No. of protein residues/atoms	144/1061
No. of solvent atoms	157

2.2.3 Structure of the NBP-4(3-82)/IGF-I binary complex

Figure 2.3 presents the structure of the NBP-4 (3-82)/IGF-I complex. NBP-4 has an L-like shape and covers both the N- and C-terminal parts of IGF-I. The global folds of NBP-4 (residues Ala39–Leu82) and miniNBP-5 (residues Ala40–Glu83) are almost identical in both complex structures, with a main chain rms Deviation of 0.77 Å; the rms Deviation of IGF/minNBP-5 versus IGF/ minBP-4 is 0.50 Å (Figure 2.4A).

The core of the NBP-4(3-38) subdomain presents a novel fold stabilized by a short two-stranded β-sheet and 4 disulfide bridges forming a disulphide bond ladder-like structure (Figure 2.4B). This structure is connected to the miniNBP fragment only by a short stretch of amino acids, which include Ala39, Leu40 and possibly Gly41. The X-ray structure shows extensive interactions between NBP-4(3-38) and miniNBP-4 around this region. There are several hydrogen bonds between Arg50 and Ser53 of the NBP-4(3-38) subdomain and Glu23, Glu27, and Glu29 of the miniNBP-4 fragment. Additionally the linker residues make backbone hydrogen bonds to Cys77, Met42, Gly52 and Arg50 of NBP-4(3-38). This network of H-bonding assures high degree of rigidity. The two subdomains are perpendicular to each other, creating the “L” shape for the whole N-domain. The miniNBP is globular, whereas in NBP-4(3-38) the β-sheet and disulfide bridges are all in one plane making the structure appear flat from one side. The core of the NBP-4(3-38) subdomain can be therefore visualized as a "palm" of a hand. The palm is extended with a "thumb" segment of 4-12 residues in various IGFBPs. The thumb segment consists of the very N-terminal residues up to the equivalent of Cys6 in IGFBP-4 and contains a consensus XhhyC motif, where h is a hydrophobic amino acid and y is positively charged. The solved structure of IGF-I is of the highest resolution to date (1.6 Å). Only very recently have X-ray structures of IGFs been published, with resolutions ranging from 1.8-2.5 Å (Vajdos et al., 2001; Brzozowski et al., 2002). In our structure a fragment of IGF-I consisting of amino acids Gly32-Gly42 is clearly seen in the electron density map. This segment was not previously defined but had apparently retained a natural flexibility in the crystal form in agreement with NMR data for free IGF-II (Torres et al., 1995). The fragment assumes a β-hairpin-like structure and appears to be stabilized by crystal packing, as several hydrogen bonds are present with neighboring NBP-4 molecules in the unit cell.

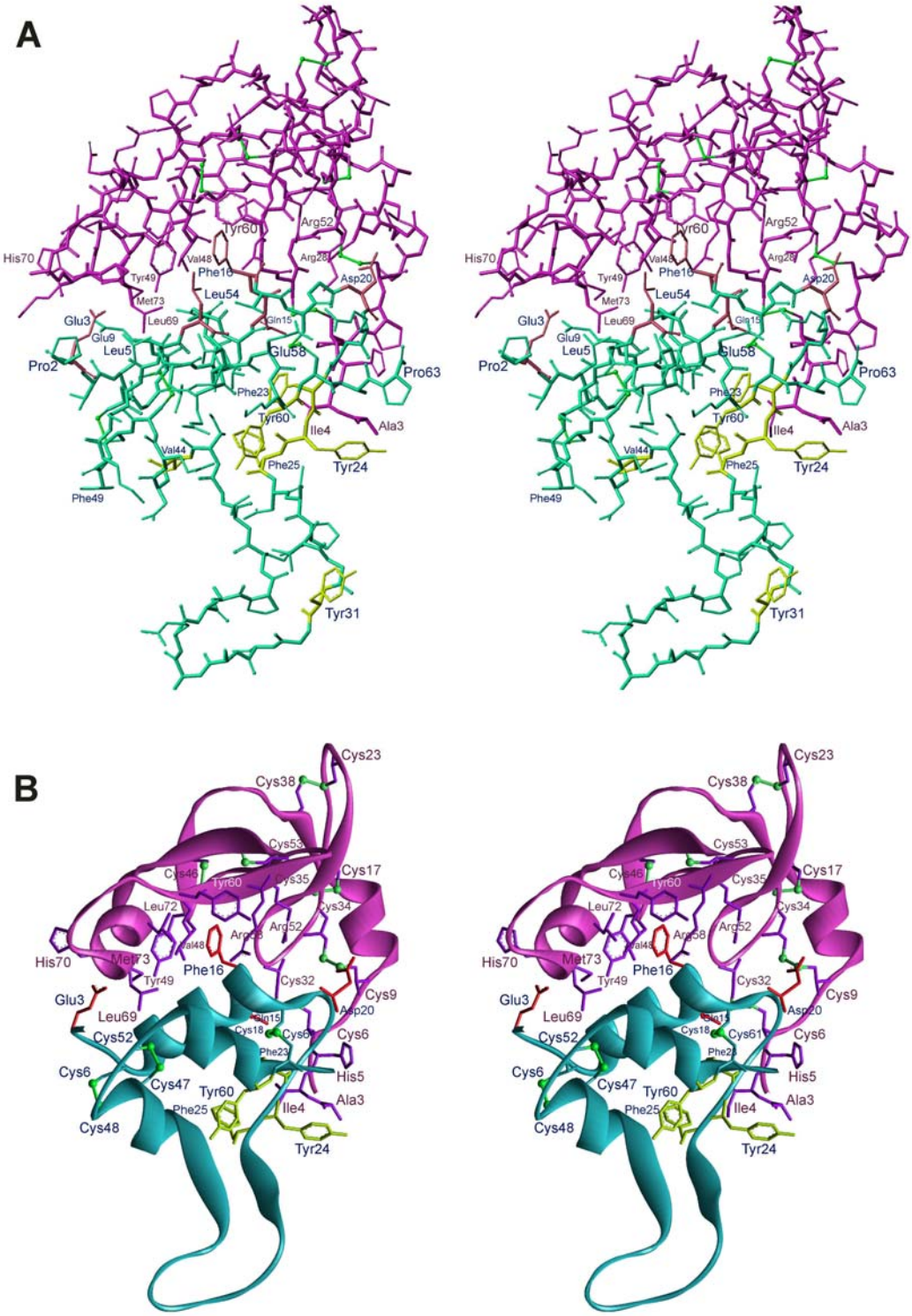


Figure 2.3. Structure of the NBP-4(3-82)/IGF-I complex. Heavy atom (**A**) and ribbon (**B**) plots of the binary complex. NBP-4 is shown in violet, IGF-I in green. Residues shown in red constitute the binding site for interaction with NBP-4. Residues marked in light green are determinants for binding to IGF-IR.

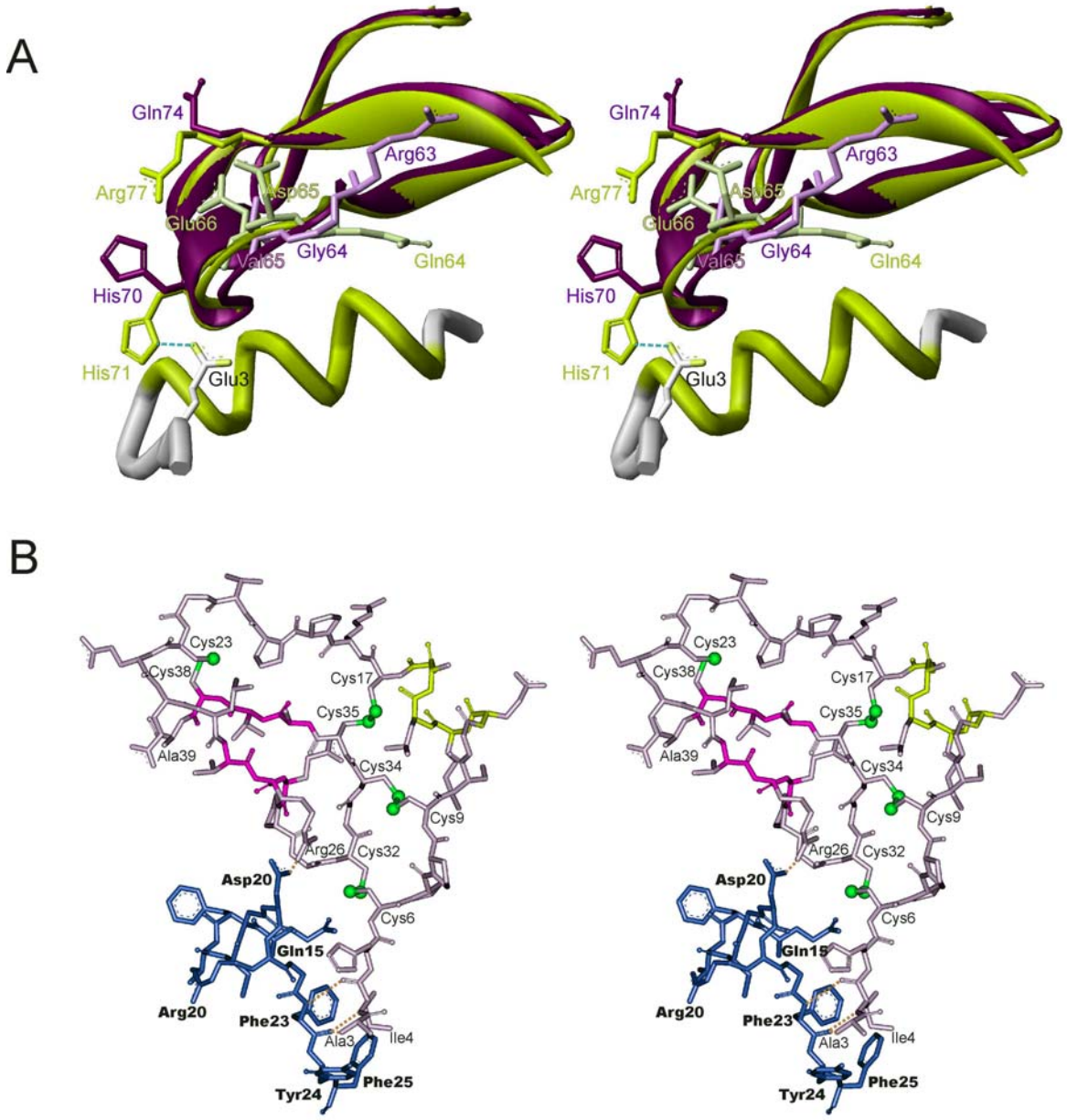


Figure 2.4. Structures of the NBP-4(3-82)/IGF-I complex. **A**: Comparison of miniNBP-5 and miniNBP-4 interactions with IGF-I. MiniNBP-5 (light green) and miniNBP-4 (violet) **B**: The structure of the N-terminal subdomain of NBP-4 and its interaction with IGF-I. Residues Ala3-Ala39 of IGFBP-4 are colored in gray, with secondary structure elements in pink (for β-sheet) and yellow (for α-helix). The IGF-I fragment (Gln15-Phe25) is in blue. The planar arrangement of disulphide bridges (green) is clearly seen.

2.3.1 The NBP4(1-92)/IGF-I

The first 92 residues of IGFBP-4 are 59% identical to the corresponding N-terminal residues of IGFBP-5, and the remaining residues are mostly functionally conserved. For miniNBP-5 (residues 40-92), the last 9 amino acids showed no electron density in its IGF complex structure (Zeslawski et al., 2001) and were unstructured as determined by NMR (Kalus et al., 1998). Equivalent residues therefore were not expressed in the construct NBP-4(3-82) to aid crystallization of the complex. However, residues Glu90 and Ser91 of IGFBP-4 were reported to be significant for high affinity binding with IGFs (Qin et al., 1998), and therefore we decided to include these residues in our extended N-terminal construct. The two first N-terminal residues were also added because the IGFBP-4(3-82)/IGF-I structure revealed the importance of the two N-terminal hydrophobic residues conserved among IGFBPs. Possibility also existed that eliminating the two first negatively charged residues, Asp1 and Glu2, at the N-terminus in the IGFBP-4 could have changed the properties of this amino terminal part. These residues were therefore added to the refined N-terminal construct, generating NBP-4(1-92).

2.3.2 Protein preparation and crystallization

DNA fragments corresponding to the NBP-4 residues 1-92 were generated by PCR amplification using human IGFBP-4 cDNA (vector BP4-2/pFDX500, ROCHE Diagnostics, Penzberg) as a template, The resulted PCR products were subcloned into the *BamH* I and *Hind* III restriction sites of the pET 28a vector (Novagen, Canada). The carboxyl-terminal domain of IGFBP-4 (residues 151-232), miniNBP-5 (40-92) and NBP-4(3-82) were cloned, produced and purified as for the NBP-4(3-82).

The complex of NBP-4 (1-92) and IGF-I (GroPep, Australia) was prepared by mixing equimolar amounts of the components. The complex was separated from any excess of either protein by gel filtration chromatography on the Superdex S75 prep grade 26/60XR column (Pharmacia, Sweden). The buffer used contained 5 mM Tris, pH 8.0, 50 mM NaCl and 0.01 % NaN_3. Crystallization of the complex was carried out with the sitting drop vapor diffusion method. The crystals were obtained from 23 % PEG 1500, 25 mM Tris pH 7 after 3 weeks in a form of plates measuring ca. 0.5 x 0.3 x 0.1 mm. Prior to plunge freezing, the crystals were soaked for ca. 30 s in a drop of a reservoir solution containing 20% v/v glycerol as cryoprotectant. The

crystals belong to the space group P21 and contained one complex per an asymmetric unit.

The data was collected from a plunge frozen crystal at a rotating anode laboratory source. The structure was determined by molecular replacement using the Molrep program from the CCP4 suite (CCP4, 1994). The structure of the complex of IGF-I and a fragment of the N-terminal domain of IGFBP-4 (residues 3-82) (entry 1WQJ) was used as a probe structure (Siwanowicz et al., 2005). Rotation search in Patterson space yielded one peak of height 12.11 σ over the highest noise peak of 4.21 σ. Translation search gave a 14.47 σ peak over the noise height of 4.49 σ. The initial R-factor of the model was 0.47. Model was completed and revised manually using Xfit software (McRee, 1999). Arp/wArp was used to add solvent atoms (Lamzin et al., 1993). The structure was finally refined by the Refmac5 program (CCP4, 1994). Final electron density maps were of good quality; there were however no interpretable densities for residue Pro63 and side chains of residues Glu11, Glu12, Lys13, Arg16, Trh37, Leu42, Glu66, His70, Gln76, Met80, Glu81 and Leu82 in NPB-4(1-92) model. The IGF-I model had no interpretable electron density for region Gly30-Pro39 and side chains of Arg50 and Glu58. These parts were removed from the model. The final R crystallographic factor was 0.24 and R_{free} 0.27. Data collection and refinement statistics are summarized in Tables 2.3 and 2.4.

Table 2.3. Data collection statistics for NBP-4(1-92)/IGF-I complex crystal. Dataset measured at Cu rotating anode laboratory source and MAR345 detector.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 2.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	Rmrgd-F	S_norm/ S_ano
20.00	53	9	9	100.0%	3.5%	3.4%	53	45.88	3.8%	2.2%	1.30
10.00	524	82	83	98.8%	3.2%	3.4%	524	45.58	3.5%	1.6%	1.10
6.00	2182	311	322	96.6%	4.2%	4.2%	2182	36.56	4.5%	2.2%	1.10
5.00	1975	275	284	96.8%	4.7%	4.6%	1975	36.89	5.0%	2.7%	1.05
4.00	4465	621	642	96.7%	4.5%	4.5%	4465	36.89	4.9%	2.3%	1.01
3.00	12159	1693	1778	95.2%	7.3%	7.3%	12159	23.71	7.8%	4.3%	1.00
2.80	4365	609	708	86.0%	17.5%	17.3%	4365	11.57	18.8%	8.9%	1.00
2.70	2652	368	422	87.2%	20.6%	20.6%	2652	9.90	22.2%	11.7%	1.01
2.60	3065	437	515	84.9%	25.6%	25.2%	3065	8.07	27.6%	12.8%	1.02
2.50	3564	510	585	87.2%	29.8%	29.7%	3564	6.76	32.2%	14.9%	0.99
total	35004	4915	5354	91.8%	7.4%	7.4%	35004	21.64	8.0%	5.2%	1.02

Table 2.4. Data collection and refinement statistics for the NBP-4(1-92)/IGF-I complex

Data Collection	
Space group	$P2_1$
Cell constants (Å)	a=32.33, b=38.99, c=61.33, β=99.89
Resolution range (Å)	20-2.5
Wavelength (Å)	1.542
Observed reflections	35004
Unique reflections	4915
Whole range:	
Completeness (%)	91.8
R_{merge}	5.2
I/σ(I)	21.64
Last shell:	
Resolution range (Å)	2.5-2.6
Completeness (%)	87.2
R_{merge}	14.9
I/σ(I)	6.76
Refinement	
No. of reflections	4915
Resolution (Å)	30 – 2.5
R-factor (%)	23.8
R_{free} (%)	27.0
Average B ($Å^2$)	36.5
R.m.s bond length (Å)	0.007
R.m.s. angles (°)	1.09
Content of Asymmetric Unit	
No. of protein complexes	1
No. of protein residues/atoms	179/1063
No. of solvent atoms	31

2.3.3 NBP4(3-82)/IGF-I vs. NBP4(1-92)/IGF-I

Examination of the structures of the two complexes reveals virtually no changes in positions of the backbone and most of the side chain atoms of both NBP-4 and IGF. rms deviation for all atomsis 0.75 Å. It can therefore be concluded that the presence of residues 1, 2 and 83-92 does not influence the fold of NBP-4 or IGF-I (Figure 2.5). Complexes of a longer and shorter NBP-4 with IGF-I crystallized in two different space groups; identity of the two structures therefore proves that crystal packing has no influence on their folds. Crystal packing, however, had influence on the loose parts of the IGF-I molecule: the flexible loop spanning residues 30-40, clearly defined in the NBP-4(3-82)/IGF-I, showed no electron density in NBP-4(1-92)/IGF-I.

The sequence Ala83-Leu92, of which the fragment Glu84-Glu90 forms a short helix, does not contact IGF directly. In the study of Qin et al. (1998), deletion of Glu90 and Ser91 led to the reduced IGF-I and –II binding activity, suggesting functional significance of these residues. Our crystallographic structure, however, shows no contribution of these two residues in formation of the IGF binding site. The presence of the 10-amino acid-long fragment may have, however, an indirect influence on IGF binding: side chains of Ile85, Ile88, and Gln89 shield Tyr60 side chain from the solvent and constrain its conformation of Tyr60 that otherwise would point away from the IGF surface, as can be seen in the NBP-4(3-82)/IGF-I complex structure. Tyr60 along with Pro61 form small hydrophobic cleft, in which Leu54 of IGF-I is inserted, thus extending the hydrophobic contact area of the two proteins.

The position of the His70 side chain in NBP-4(3-82) was rotated ca. 180° relative to the corresponding His71 of miniBP-5. In the structure of NBP-4(1-92), the imidazole ring of the histidine is however flipped back to the configuration observed in the miniNBP-5/IGF-I complex and forms a network of hydrogen bonds with side chains of Glu3 and Glu9 of IGF-I. Similarly, in the NBP-4(3-82)/IGF-I complex structure, IGFs' Phe16 was found rotated ca. 20° clockwise about the α-helix axis, when compared to its position in the miniBP-5/IGF-I complex. In the structure of IGF-I bound to NBP-4(1-92), Phe16 returns to the position found in the miniNBP-5/IGF-I.

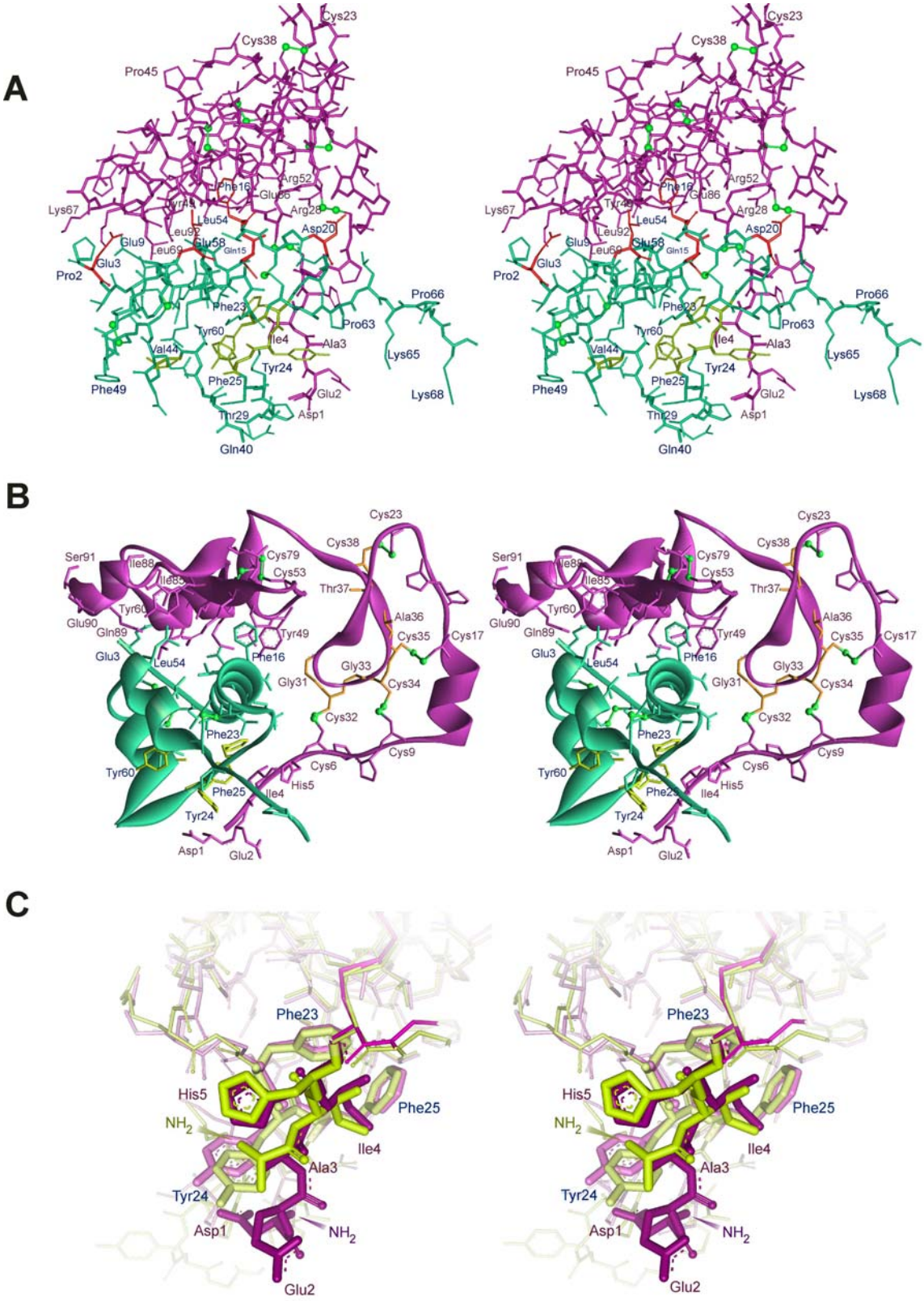


Figure 2.5. Structure of the NBP-4(1-92)/IGF-I complex. Heavy atom (**A**) and ribbon (**B**) plots of the complex. NBP-4 is shown in violet, IGF-I in green. Residues shown in red constitute the binding site for interaction with NBP-4. Residues marked in ligh green are determinants for binding to IGF-IR. The GCGCCXXC consensus motif is shown in yellow (**B**). Comparison of the IGF-I residues interacting with the “thumb” region in complex with NBP-4(3-82) (yellow) and NBP-4(1-92) (violet) (**C**).

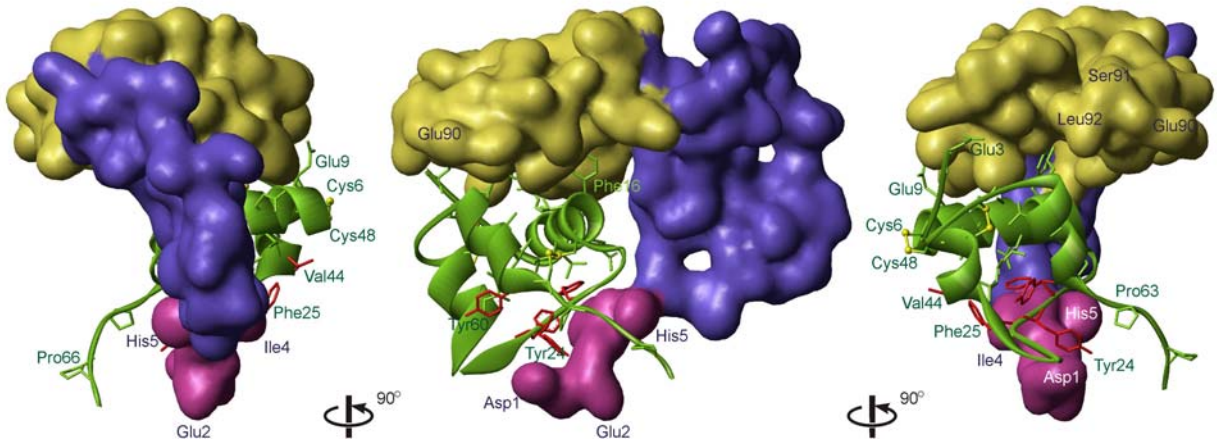


Figure 2.6. Bipolar binding of IGF-I by NBP-4(1-92). Three views of the binary complex; NBP-4(1-92) is presented as a surface: the "base" region is colored in yellow, "palm" in blue and "thumb" in magenta. IGF-I is shown as a green ribbon.

2.4.1 Ternary complex NBP-4(3-82)/IGF-I/CBP-4(151-232)

We have also obtained crystals of a ternary complex of NBP-4(3-82) and the C-terminal domain, CBP-4(151-232), with IGF-I. Proteins used for crystallization were prepared same way as for binary complexes.

Crystals of the ternary complex were obtained in two conditions. One form diffracted up to 2.9 Å and belonged to the space group R32 and the second crystal with the space group C2 diffracted up to 2.1 Å. Molecular replacement trials using the NBP4/IGF-I model were unsuccessful. Experimental phasing was then performed using R32 crystals because it was not possible to derivatize C2 crystals. Soaking crystals with 2 mM potassium hexabromoplatinate (IV) or 2 mM uranylacetate for 24 hr created two isomorphous derivatives. Native and derivative datasets were collected at the BW6 beamline in DESY, Hamburg. Datasets were indexed, integrated, and scaled using a XDS software package. Heavy atom search and phase improvement were performed with the SHARP program (de La Fortelle and Bricogne, 1997.) using both anomalous and isomorphous signals. The phases obtained were of good quality up to the resolution of 3.5 Å. The model of the NBP4/IGF-I complex was fitted to this map. All parts of the complex maintain the spatial organization found in the crystals of the NBP4/IGF-I binary complex. After placing the NBP-4/IGF part, polyalanine fragments were fitted into broken electron density associated with the C-terminal domain. This model of the complex was later used as a molecular replacement probe for the C2 crystal data. The replacement

solution was clear, however further model building and completion of the C-terminal part was not possible. The best R-factor of the ternary complex model obtained was 34%. Statistics of data collection, phasing and refinement are summarized in Table 2.3.

Table 2.3. Data collection, phasing and refinement of the the NBP-4(3-82)/CBP-4 /IGF-I complex

	CRYSTAL FORM I			CRYSTAL FORM II
Space group	R32			C2
Unit cell constants [Å]	a=b=127.9 b=75.77			a=74.4;b=50.25 c=64.3 β=115.3
Dataset	Native	Platinum	Uranium	Native
Resolution [Å]	30-2.9	30-3.0	30-2.9	30-2.1
Whole resolution range				
Completeness [%]	87	91	97	98
Unique reflections	9112	6351	8997	11290
R-merge [%]	4.5	2.7	3.9	3.6
I/σ(I)	26.15	34.56	21.24	152
Highest resolution shell				
Resolution [Å]	2.9-3.0	3.0-3.2	2.9-3.0	2.1-2.2
Completeness [%]	69	62	62	87
R-merge [%]	24.7	23.8	21.9	9
I/σ(I)	4.51	4.6	4.54	17
Phasing statistics				
Number of sites	-	1	1	-
Phasing power isomorphous	-	1.17	1.27	-
Phasing power anomalous	-	0.47	0.69	-
Mean FOM	-	0.58		-
Refinement				
R-factor [%]	30.2			34.4
R_{free} [%]	38.6			40.1

The structure of the binary segment NBP-4(3-82)/IGF-I in this complex could be readily solved by molecular replacement. Unfortunately, the C-domain produced electron densities in disconnected patches that could not be interpreted in detail. Only fragments of the domain in contact with or in proximity to NBP-4/IGF-I showed electron density sufficiently defined to allow incorporation of fragments of a polyalanine model. Identities of individual residues could not however be reliably established. The data allows identification of the C-domain interface surface of IGF-I and NBP-4 (Figures 2.7). CBP-4 appears to exhibit a flat structure that has far-reaching interactions with the N terminal part of NBP-4 from Ala3 up to Tyr48 and also extensive contacts with IGF-I

2.4.2 The IGF-I/NBP-4 interaction – implications for IGF-I binding to its receptor

The principal interface between IGF-I and NBP-4 is located in the miniNBP subdomain and comprises a hydrophobic sandwich that consists of interlaced protruding side chains of IGF-I and solvent-exposed hydrophobic side chains of the NBP-4. The side chains of IGF-I Phe16 and also Leu54 are inserted deep into a cleft present in NBP-4. This cleft is formed by side chains of Arg52 and Arg58 on one side of the molecule and by Val48, Leu69 and Leu72 on the opposite inner side, with a base formed by residues Gly47, Cys59 and Tyr60. Phe16 of IGF makes direct contacts with the backbone and side chain of Val48, and with Cys59 of NBP-4. The IGF Phe16 NBP-4 interaction is closed on the solvent side by side chains of Glu3 and Glu9 of IGF-I, and by Tyr49, Thr50 and Pro30 of NBP-4.

The N-terminal palm subdomain (between residues 6-38) and the thumb segment (residues 3-5) make several additional contacts to IGF-I. The "thumb" Ala3 and Ile4 of NBP-4 make major interactions by filling the hydrophobic patch created by IGF Phe23, Tyr24 and Phe25. The interaction is further strengthened by hydrogen bonding (Ile4 C=O to Phe23 NH, Ile4 NH to Phe23 C=O). The palm contacts IGF only from one edge mostly through BP-4 Arg28 (Arg28 NH1 to the IGF Asp 20 OD1).

The C-domain appears to make extensive contacts to IGF. Contact residues on IGF are Cys6-Ala8, the sidechain of Glu9, Val11 and Asp12, Tyr24, Phe25, Asn26, Val44, Cys48 and Phe49 (Figures 2.7 and 2.8).

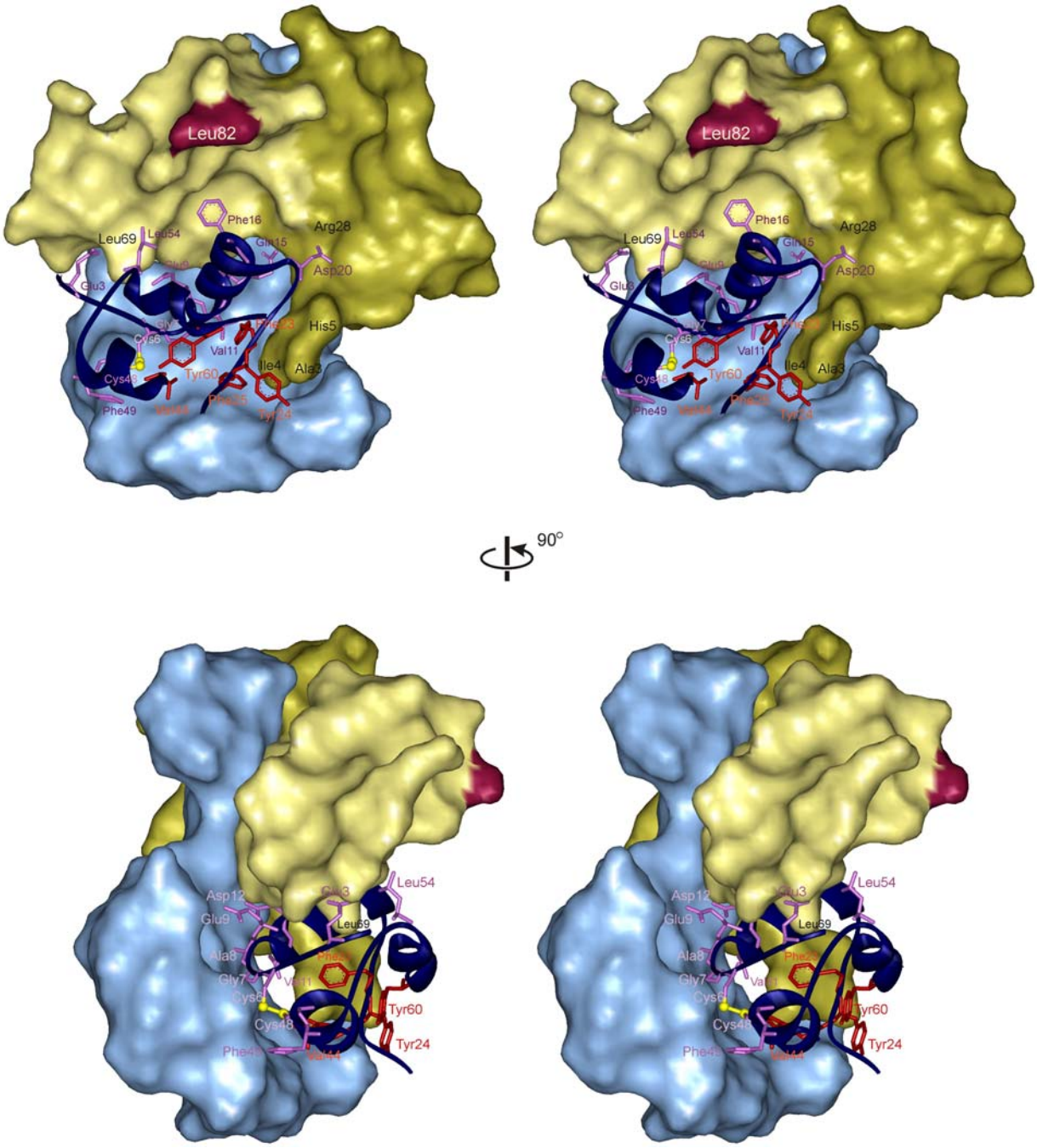


Figure 2.7. Surface plot of the structure of the IGF-I/CBP-4/NBP-4 ternary complex. Side-chains of IGF residues important for IGF-IR binding are shown in red. Primary determinants of the IGF/IGFBP-4 interaction are shown in violet. The “base” region of NBP-4 is colored in lighter, “palm”-“thumb” in darker shades of yellow, IGF-I (dark blue ribbon); key IGF residues are numbered, IGF residues responsible for IGF-IR binding are shown in red; CBP-4 is presented as a light blue surface.

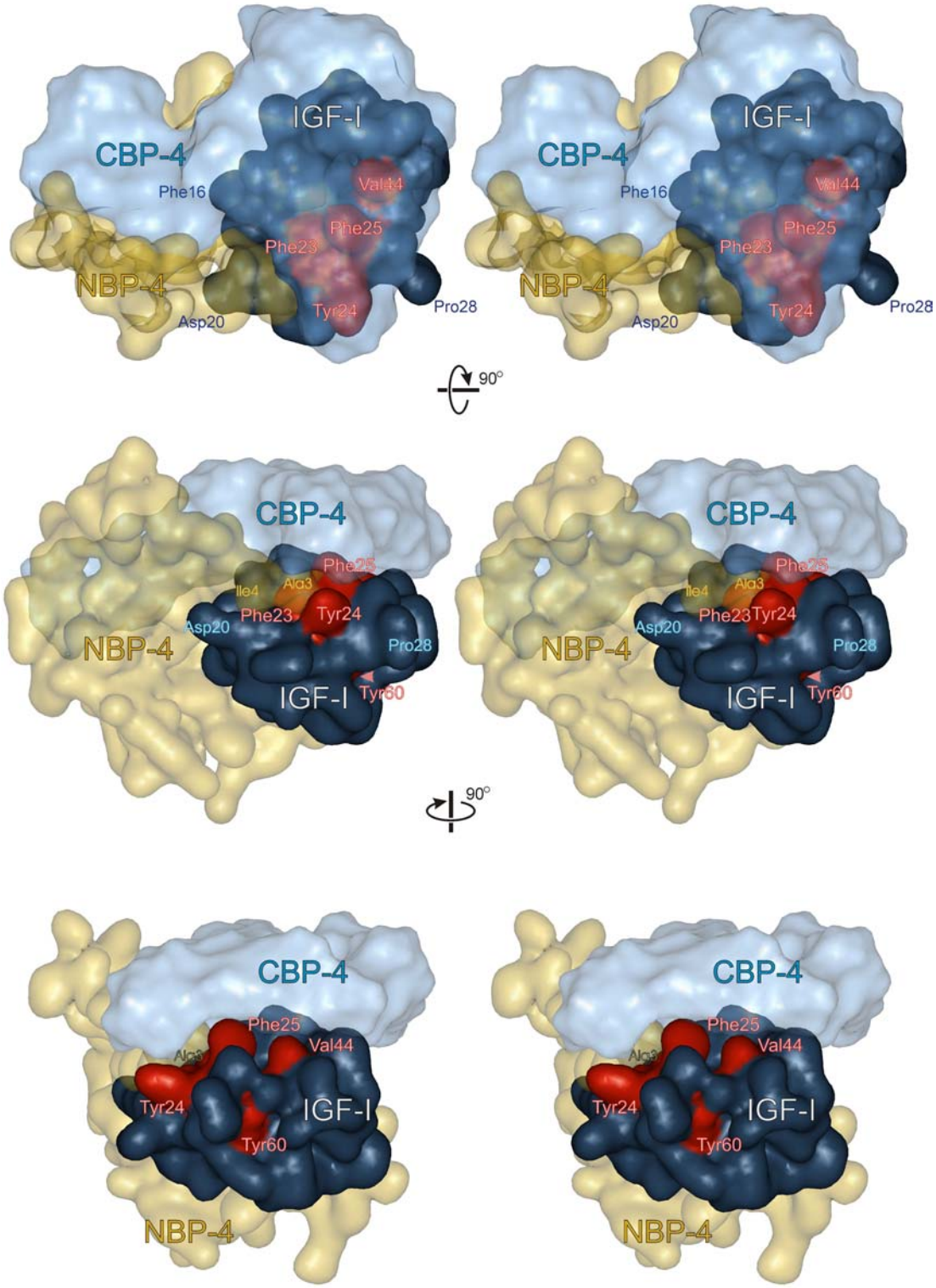


Figure 2.8. A surface plot of the structure of the IGF-I/CBP-4/NBP-4 ternary complex. IGF-I (solid dark blue); key IGF residues are numbered, IGF residues responsible for IGF-IR binding are shown in red; the NBP-4´s transparent surface is shown in yellow; CBP-4 is presented as a transparent light blue surface.

2.5 Conclusion

Mapping of binding sites of IGFs for IGF-IR and the insulin receptor (IR) showed that the major determinants of binding in IGF-I are Tyr60, Tyr24, Phe23, Val44 (Cascieri et al., 1988; Bayne et al., 1990; Hodgson et al., 1996; Sakano et al., 1991; Perdue et al., 1994), and possibly Phe25 and Arg21 (Hodgson et al., 1996; Torres et al., 1995; Cooke et al., 1991) (Figure 2.9). The most notable feature evident from the earlier structural studies on the miniNBP-5/IGF-I complex was that Phe23-Tyr24-Phe25, Val 44, and Tyr60 form a fully solvent-exposed hydrophobic binding site for IGF-IR that is located on the opposite side of IGF to that used for the binding to mini-IGFBP-5 (Zeslawski et al., 2001). The additional N-terminal subdomain 1-38 of NBP-4 on the other hand reaches the IGF-I Tyr24 and its surroundings primary through the "thumb" segment residues Ala3 and Ile4 of NBP-4 that fill a hydrophobic cleft of IGFs' Phe23, Tyr24 and Phe25. We would therefore expect to see some inhibition of activity of IGF-IR for the entire NBP-4 (e.g. which include the "thumb" residues) compared to the miniNBP-4 domain only.

Both IGF-I and IGF-II bind to the IGF-I receptor with high affinity (K_D = 1.5 and 3.0 nM, respectively; Loddick et al., 1998). In our studies, a full-length IGFBP-4 inhibited IGF-IR autophosphorylation with IC_{50} of about 2.4 nM. The N-domains showed some inhibition, with the longer NBP-4 fragment NBP-4(1-92) (and NBP-5(1-94)) exhibiting superior inhibition relative to that of NBP-4(3-82). This trend parallels stronger IGF binding of these fragments relative to NBP-4(3-82). MiniNBP subdomains showed no inhibition, indicating that the N-terminal "thumb" part of BPs is important for inhibition of IGF action.

As discussed above, the N-terminal domain of IGFBPs is characterized by high degree of rigidity. This may be of significance when the competition with IGF-IR for IGF binding is concerned. To displace the hydrophobic thumb that covers the primary IGF-IR binding site of IGFs (IGF-I, Phe23-Phe25), the receptor has to lift also the rest of the N-terminal domain, which is bound on the opposite side of the IGF-I molecule, and does not prevent receptor binding on its own (Kalus et al. 1998). Thus, the thumb does not have to significantly contribute to the overall binding affinity of IGFBPs for IGFs. This mechanism is expected to be shared by all IGFBPs given the conserved arrangement of the N-terminal cysteine residues and the consensus presence of two hydrophobic residues at positions -2 and –3 with respect to the first N-terminal Cys residue.

The structure of the ternary complex suggests that addition of CBP-4(151-232) to the NBP-4(3-82)/IGF interaction should enhance inhibition of IGF-IR activity, and indeed a 6-fold increase is seen. Our data are in agreement with that of Payet et al. (2003) who have recently demonstrated that isolated amino-terminal and carboxy-terminal domains of IGFBP-3, residues 1-88 and 185-264, respectively, cooperate in the presence of IGFs to form high-affinity complexes that retain the ability to block IGF activity. We observe no inhibition for miniNBP-4(39-82) and CBP-4, indicating again a crucial role of the N-terminal "thumb" residues of BPs in blocking IGF activity.

Proteolytic cleavage has gained wide acceptance as the predominant mechanism for IGF release from IGFBPs (Bunn and Fowlkes, 2003). The current paradigm for IGF release proposes that proteolysis takes place at the central L-domain producing low affinity N- and C-terminal fragments that cannot compete with IGF receptors for IGFs. The data presented by Payet et al. (2003), together with our results, clearly indicate that IGFBP fragments still retain properties to inhibit IGF activity, and therefore this paradigm has to be revised. The structures of the IGF/BP-4 complexes presented here suggest a following model for IGF binding proteins actions (Figure 2.10): The data indicates that IGFs, which are approximately globular structures, are encapsulated by IGFBPs. The N-terminal domain of IGFBPs plays an important role in binding IGFs. The globular C-terminal part of this domain, miniNBP, serves as a primary docking site for IGFs; the N-terminal subdomain, which consists of a "palm" and a "thumb", reaches a hydrophobic IGF receptor binding site of IGFs with the conserved thumb hydrophobic residues. Next, the C-terminal domain, which covers roughly an equal surface of IGF to that of the N-terminal domain, shields the IGF receptor binding site of IGFs, close to the N-terminal thumb fragment residues, thus cooperatively enhancing inhibition of IGF receptor activity. The central domain, which is unstructured and flexible, acts as a "mechanical flap" that covers most of IGF not yet covered by N-terminal and C-terminal domains.

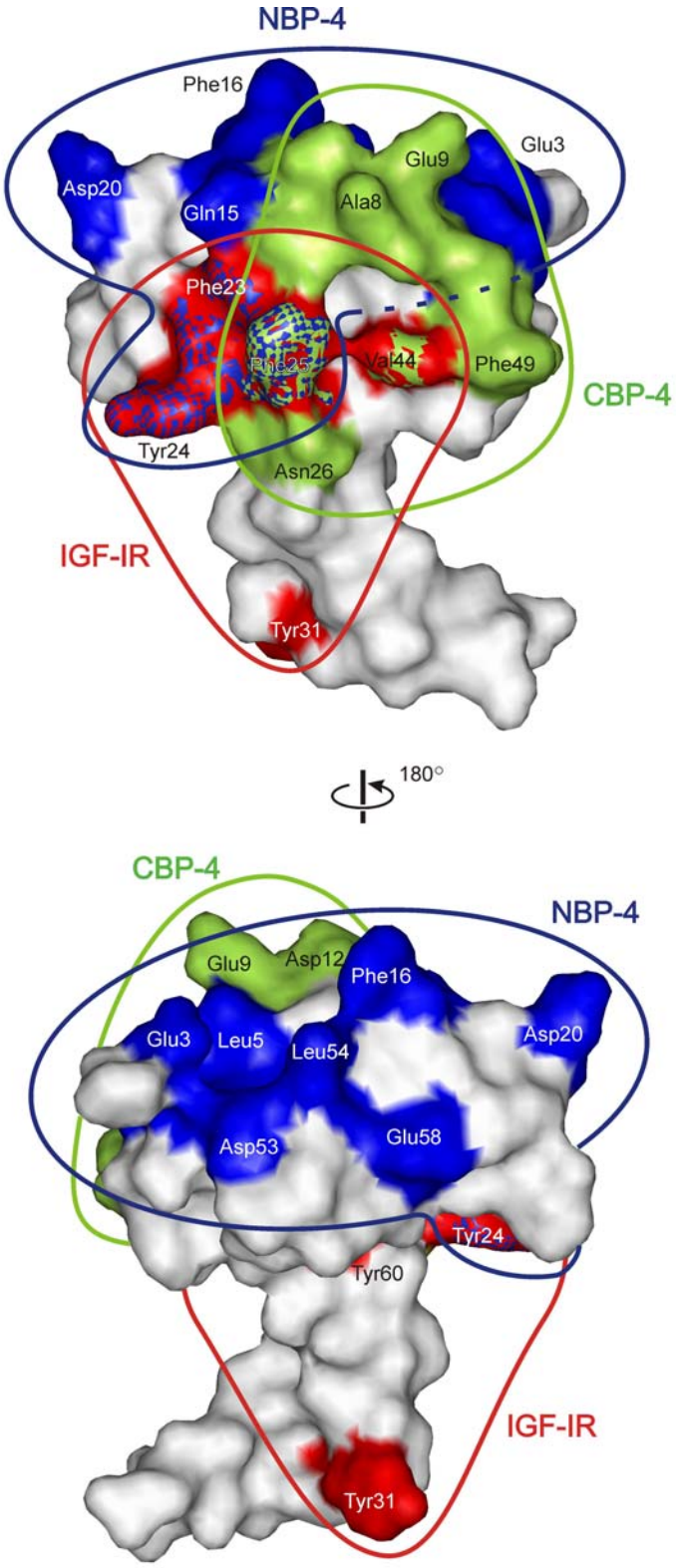


Figure 2.9. Interaction of IGF-I with IGFBP-4. Schematic of the surface of IGF-I showing the binding site for the IGF-IR (red), N-terminal domain of IGFBP-4 (blue) and CBP-4 (green).

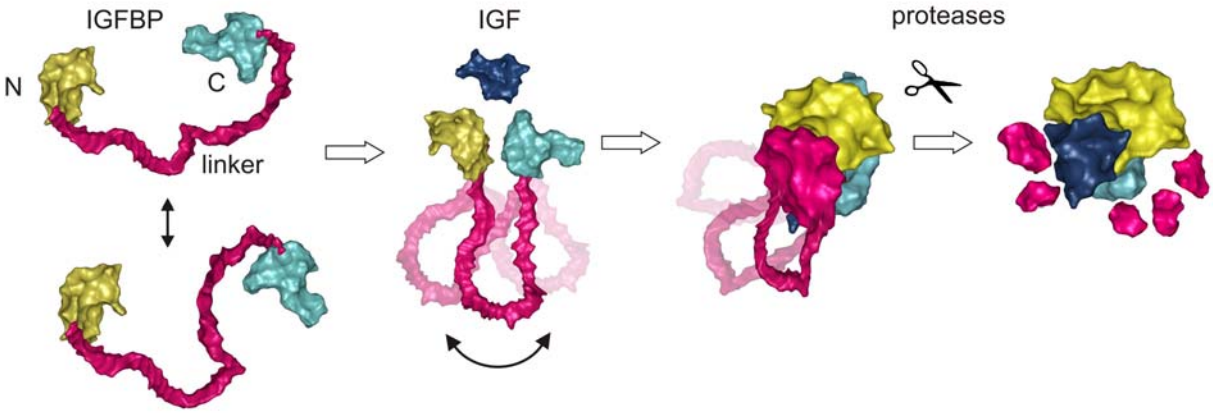


Figure 2.10. A model of IGFBP/IGF complex and its regulation by proteolysis. IGF is in dark blue, NBP in yellow, and CBP in light blue. The linker domain (in purple) is hypothetical and its flexibility is emphasized. The last step of the reaction depicts proteolysis of IGFBP.

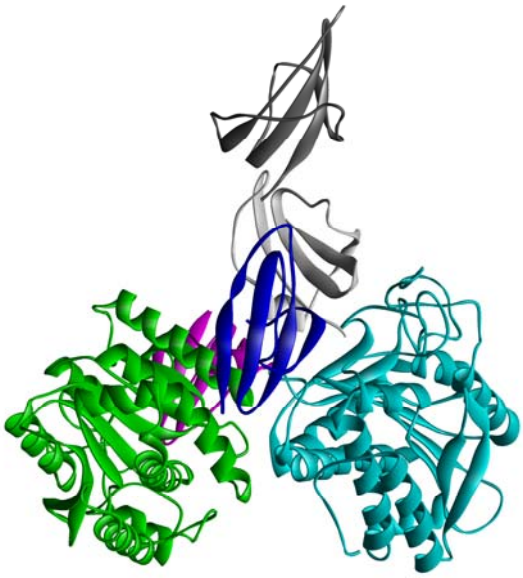
Proteolysis of the IGF/IGFBP capsule would first remove the central domain residues by degradation. This partial removal of the capsule exposes IGF but still maintains IGF inhibition as long as the thumb and/or CBP fragments of IGFBPs are not removed.

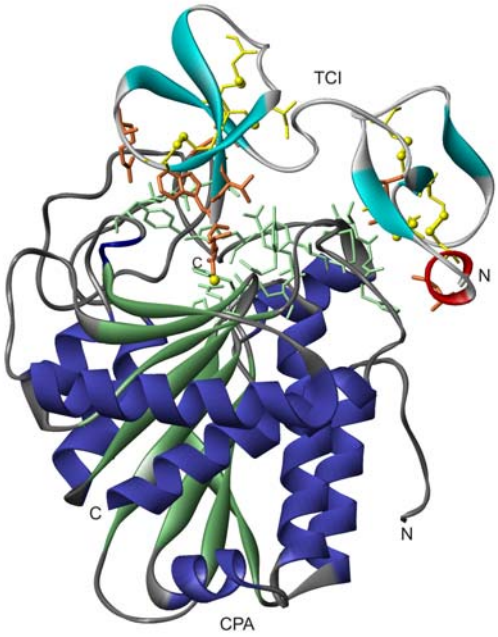
Data on structural determinants of IGFBP/IGF binding can be used for design of targets that could regulate the actions of IGFs. This can be carried out either by site-directed mutagenesis of IGFBPs and thus producing IGFBPs of modified binding affinity, or by design of novel, low molecular weight ligands. For example, IGF-I and IGF-II exhibit neuroprotective effects in several forms of brain injury and neurodegenerative disease. This implies that targeted release of IGF from their binding proteins might have therapeutic value for stroke and other IGF responsive diseases.

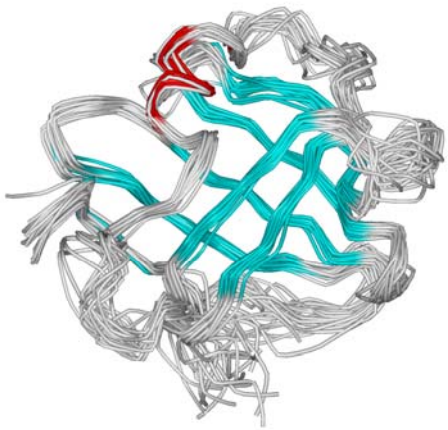
Inhibition of IGF-stimulated cancer cell growth via receptor intervention is a major strategy in anticancer drug discovery as shown by the efforts of pharmaceutical companies to develop both protein and small molecule antagonists toward the receptor (LeRoith and Helman, 2004; Ali et al., 2003; Mazerbourg et al., 2004; Firth and Baxter, 2002; Cohen et al., 2000; Khandwala et al., 2000; Bach, 1999). Mutagenesis of the thumb fragment to increase the IGF/BP fragments binding or to enhance their inhibition of IGF-IR action can therefore offer a therapeutic strategy against cancers that are the type I IGF receptor dependent.

Chapter 3

Structural analysis of serine proteases and carboxypeptidase inhibitors







3.1 Serine protease inhibitors from *Staphylococcus aureus*

Unlike other bacteria *Staphylococcus aureus* is able to colonize and infect virtually every tissue of the human body and is therefore responsible for a wide variety of human diseases. In addition, increasing antibiotic resistance makes this versatile pathogen an emerging plaque of 21st century calling for new treatment strategies (McGahee and Lowy, 2000). In turn, the development of such strategies requires understanding of this bacteria physiology, which, although still very basic, already points to some interesting targets. Highly regulated expression of a multitude of extracellular and cell surface associated proteins has been demonstrated to be necessary for virulence (Abdelnour et al., 1993). Among the *Staphylococcal*-secreted proteins, the proteases of three catalytic classes, including serine, cysteine, and metalloenzymes, have been indirectly associated with many different functions in the infection process. Only recently information became available for a cysteine protease, staphopain B of *S. aureus.* This organism produces two distinct cysteine proteases, staphopains A and B. The latter protease is encoded in one operon (*ssp*)1 with *Staphylococcal* serine protease (SspA). The insertional inactivation of the *ssp* operon leads to decrease in virulence in several animal infection models (Coulter et al., 1998). In contrast, strains lacking only SspA remain virulent, at least in the mice skin abscess model (Rice et al., 2001). These results seem to point directly to an important role of staphopain B, but not SspA, in infection. Several lines of evidence, including protection of the rat skin from *Staphylococcal* infection by P-cystatin R, an effective inhibitor of staphopain A elastolytic activity (Takahashi et al., 1994), seem to implicate this protease as an important virulence factor. In addition, a staphopain A homologue of *Staphylococcus epidermidis* was identified as an etiologic agent in perifollicular macular atrophy (Varadi and Saqueton, 1968). Moreover, similar to other virulence factors, staphopains are tightly regulated. First, the protein production is controlled on the expression level by several regulatory loci, including *agr* and *sar*, the two major global virulence gene expression regulators (Novick, 2000). Second, the enzymes are secreted as inactive zymogens undergoing activation in a complicated network of interaction with other proteases. Recently specific, intracellular, protein inhibitors of these enzymes were discovered. The *Staphylococcal* *s*erine *p*rotease operon (*ssp*) encodes, besides SspA and staphopain B, a third protein designated SspC, a potent inhibitor of staphopain B. Similarly, in the *Staphylococcal* cysteine protease operon (*scp*), an ORF downstream

of the staphopain A gene (*scpA*) encodes staphostatin A, a specific inhibitor of this enzyme (Massimi et al., 2002; Rzychon et al., 2003). Although these cysteine proteases share significant primary and high tertiary structure homologies (Filipek et al., 2003), their specific inhibitors show only limited similarities at the amino acid sequence level. Thus the inhibitors were classified into one class of cysteine protease inhibitors; however, staphostatin A does not inhibit staphopain B, and SspC is not active against ScpA (Rzychon et al., 2003). Moreover, the inhibitors do not inhibit any other tested cysteine proteases, although staphopains are not structurally distinct to papain (Filipek et al., 2003; Hofmann et al., 1993). Most protein inhibitors of cysteine peptidases described up to now are structurally related and therefore were grouped into one superfamily designated cystatins. These homologous proteins are further divided into three families on the basis of their molecular structures. The inhibitors belonging to the first two families are about 12 kDa and differ in the content of disulfide bridges. The third family encompasses proteins of much larger molecular sizes, containing several copies of a cystatin domain, as well as unrelated domains. Cystatins have been described from several higher animal species, but examples are also found in many lower organisms (Abrahamson, 1994; Barrett et al., 1986). Amino acid sequences of staphostatin, however, are not similar to the sequences of cystatins or other known protease inhibitors. The recently determined structure of staphostatin B shows a fold unrelated to that of cystatins but similar to the fold of lipocalins and in particular triabin, a serine protease inhibitor, and von Ebner protein, a cysteine protease inhibitor (Rzychon et al., 2003). Nevertheless, the inhibition mode of staphopain B by staphostatin B is unique (Filipek et al., 2003). The staphostatin, therefore, becomes the first member of a novel class of cysteine protease inhibitors, staphostatin A being a very probable follower. With the growing body of evidence indicating *Staphylococcal* cysteine proteases as virulence determinants it becomes apparent that this class of proteins would make a good target for the development of a novel antistaphylococcal therapy.

3.1.1 Protein expression and purification

The wild-type inhibitor was purified according to the procedure described by Rzychon et al. 2003, with some modifications. The plasmids were transformed into BL21 Star (DE3)pLysS cells (Invitrogen). The bacteria were grown at 37°C, 160 rpm, according to the needs either in LB, M9, or in the selective labeling media to the

OD_{600} equal to 1.0. The temperature was lowered to 30 °C, and the expression was induced with 1 mM IPTG when the OD600nm reached 1.2. Cells were harvested by centrifugation after 3 h in the case of LB and M9 media or after 2 h when grown in the selective labeling medium. All steps were carried out at 4 °C, unless indicated otherwise. Cells were resuspended in PBS and lysed by sonication. The debris was centrifuged out at 75000*g*, and the supernatant was recirculated overnight over glutathione-sepharose 4B (Pharmacia). After extensive washing with PBS, followed by a short wash with 50 mM Tris-HCl, pH 8.0, the protein was eluted with 10 mM reduced glutathione in the same Tris buffer. To cleave the fusion GST tag, thrombin (Sigma, T-9681) was added (250 units/15 mg of fusion protein), and the mixture was dialyzed at room temperature for 72 h to a 20-fold excess of 50 mM Tris- HCl, pH 8.0, containing 2 mM reduced and 2 mM oxidized glutathione and 0.5 mM EDTA. The latter buffer favors cysteine oxidation (Ruoppolo et al., 1996), and the time scale was chosen to allow for complete cleavage and oxidation. When the fusion protein was eluted with reduced glutathione and cleaved in the elution buffer, without dialysis, a mixture of oxidized and reduced forms resulted. To obtain the inhibitor with fully reduced cysteine residues, elution and the following steps were omitted, and the cleavage was carried out on the column by recirculating 50 mM Tris-HCl, pH 8.0, containing thrombin (100 units/L of starting culture) over glutathione-Sepharose with the bound fusion protein. In either case gel filtration was used to recover the free inhibitor and to exchange the buffer to 140 mM NaCl, 2.7 mM KCl,10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, and 0.05% NaN_3, pH 7.5 (staphostatin A slowly aggregates in low-salt buffers). The samples were subsequently concentrated to approximately 8 mg/mL (0.7 mM), and D_2O was added to 10% to allow for NMR measurements. The protein could be easily concentrated to more than 20 mg/mL and was stable for several hours; however, a heavy precipitate formed after a few days, the concentration in the supernatant remaining at about 8 mg/mL. The yield ranges from 7 mg/l from the LB medium to 11 mg/l from the media used for selective labeling.

3.1.2 NMR spectroscopy

All NMR experiments were recorded at 27°C on Bruker AMX500, DRX600, or DMX750 spectrometers equipped with triple resonance probe heads and pulsed-field

gradient units. The sequence-specific assignment of ^{1}H, ^{15}N, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$, resonances was accomplished with the use of HNCA (Grzesiek and Bax 1992; Kay et al., 1990) and CBCA-(CO)NH (Grzesiek and Bax 1992; Zweckstetter and Holak, 1999) spectra of the uniformly $^{15}N/^{13}C$/70% D-labeled sample and ^{1}H-^{15}N HSQC (Mori et al., 1995) with spectra of ^{15}N uniformly and ^{15}N Tyr, Lys, Val, Ile, Gly/Ser, Leu, and Phe selectively labeled samples. In addition, ^{15}N-edited 3D NOESY (Piotto et al., 1992, Sklenar et al. 1993), 2D TOCSY (Dhalluin et al., 1996), and 2D NOESY (Piotto et al., 1992) in 10% D_2O/90% H2O and 100% D_2O were used. HSQC, HNCA, and CBCA(CO)NH experiments were recorded with respectively 128, 64 x 64, and 64 x 64 increments in indirect dimensions, and linear prediction was used to double these numbers. NOESY planes of 720 x 4096 points were recorded. For all experiments zero filling was used prior to the Fourier transform. The HNHA experiment (Geerten and Bax, 1993) was acquired to obtain the $^{3}J_{HN\alpha}$ coupling constants for the determination of torsion angle constraints. Furthermore, modified versions of ^{1}H-^{15}N hetNOE (Farrow et al., 1994, Muehlhahn et al., 1996; Renner et al., 2002) and T_1 relaxation time experiments (Skelton et al., 1993) were measured to investigate the polypeptide backbone flexibility. For ^{1}H-^{15}N hetNOE experiments, amide protons were presaturated with 120 deg pulses for 2.5 s prior to the experiment. To determine T_1, six ^{1}H-^{15}N planes (256 x 2048 points) were recorded in an interleaved manner, with relaxation delays of 12.4, 384.4, 756.4, 1128.4, 1500.4, and 12.4 ms, and peak heights were taken for further analysis. All spectra were processed with the XWinNMR software of Bruker.

3.1.3 Assignment and structure calculation

The Sparky software (Goddard and Kneller) was used for data handling and assignment of the ^{1}H-^{15}N HSQC spectrum of staphostatin A, which showed well separated signals with only few overlaps. The backbone resonance assignments were obtained for 95 residues of a total of 107 residues of the engineered staphostatin A (105 residues of a native protein). The resonances of the three N-terminal residues (of which the two most N-terminal were added in the cloning process), N17 to E19, and six residues distributed throughout the protein sequence did not appear in NMR spectra. The data of the ^{1}H, ^{15}N and ^{13}C chemical shifts and

$^3J_{HN\alpha}$ coupling constants of staphostatin A have been deposited in the BioMagResBank under accession number 5810 (http://www.bmrb.wisc.edu).

For structure calculations the simulated annealing function of the CNS package (Brunger et al. 1998) and energy minimization functions of SPDBV were utilized. Molmol2k, SPDBV, and WebLab Viewer Pro were used for visualization of calculated structures and for statistics generation. NOE distance constraints for structure calculations were derived from 2D NOESY spectra in H_2O and D_2O and from the ^{15}N-edited NOESY spectrum. For the quantification of peak intensities peak heights were used, and three classes of NOEs were distinguished: 4 Å for weak, 3 Å for medium, and 2 Å for strong signals. The uncertainties of the distances were set to ±2, ±1 and ±1 Å, respectively, in preliminary calculations, and some were later tightened near the ideal values (±0.3 Å) in well-defined secondary structures. Pseudoatom corrections were added to the upper bounds where eligible (Wüthrich, 1986). Several rounds of calculations were performed, repeatedly violated NOEs were rechecked and removed if proven wrong, and new constraints assigned on the basis of preliminary structures were included in subsequent iterations. For the final round of calculations distance constraints were supplemented with torsion angle constraints derived from the HNHA experiment (Vuister and Bax, 1993). Structure calculations were performed using standard protocols for simulated annealing (Holak et al., 1989) without the preceding distance geometry step. During the last step 200 conformers were generated, and the 20 best representatives were chosen on the basis of total energy value. The final family of structures was deposited in the Protein Data Bank under accession number 1OH1.

3.1.4 Three-dimensional structure of staphostatin A

All structures of the selected set fully satisfy the experimentally determined constraints, confirming that the final model corresponds well to the real structure of the protein in solution. The staphostatin A polypeptide chain folds into a slightly deformed, eight-stranded, β-barrel (Figure 3.1.1). Strands β4 through β8 form a standard antiparallel sheet while the N-terminus is best described as a ψ-loop motif (Hutchinson and Thornton, 1990). The connection between the first and second, β-strands is extended and includes also a short helix. The C-terminal-most strand is short, owing to an extended β7-β8 loop. The barrel-closing contacts between the N-

and C-terminal parts consist of hydrogen bonds and a disulfide bridge. Thus, the overall fold of staphostatin A is identical to that of staphostatin B, providing unquestionable evidence that both proteins belong to the same, new class of cysteine protease inhibitors.

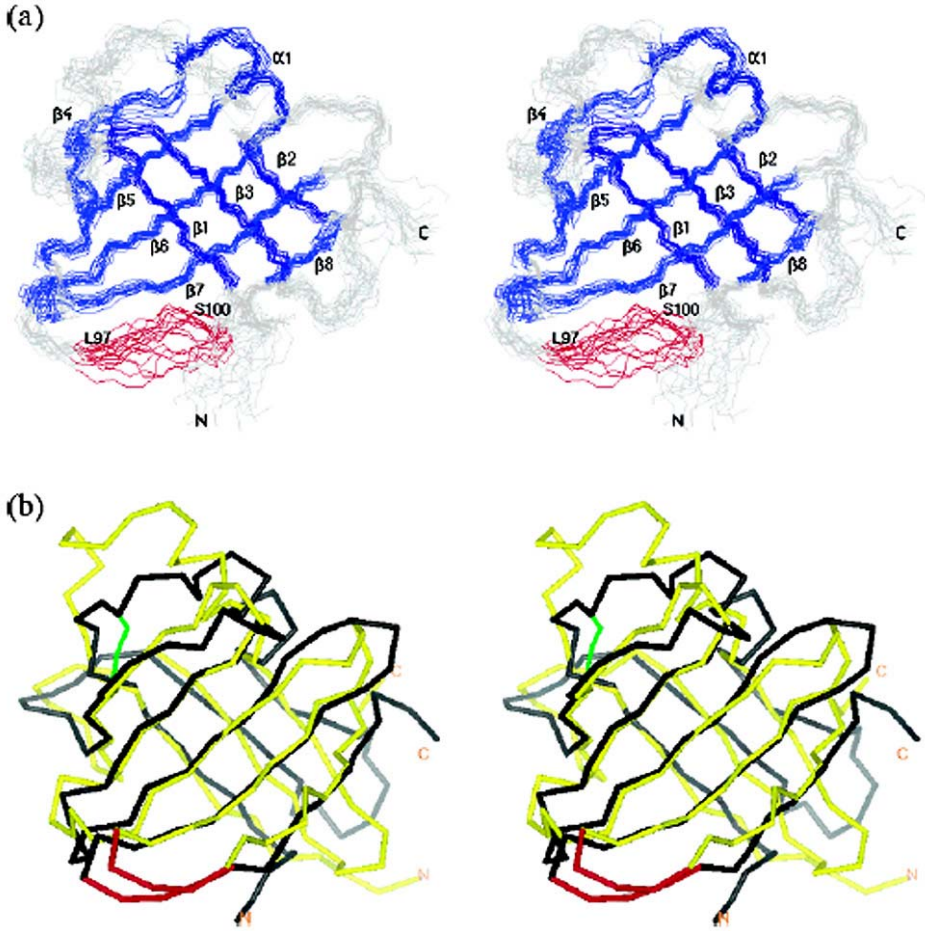


Figure 3.1.1.: (a) Stereopair of overlays of the backbone (N, Cα, C) atoms of the 20 final models of the solution structure of staphostatin A. Well-defined secondary structures are in blue, and poorly defined regions are in gray. The protease active site contacting residues are in red. (b) Stereoview of the best backbone fit of a mean structure of staphostatin A (black; calculated for a family of 20 NMR structures, 1OH1) and of staphostatin B (yellow, 1NYC). Cα traces are shown. The disulfide bridge present in staphostatin A is colored green. The N and C termini are indicated.

3.2 Structure of the *Staphylococcus aureus* splC serine protease

3.2.1 Introduction

A substantial effort has been made into understanding *Staphylococcus aureus* pathobiology for the selection of novel vaccine candidates and drug targets. The struggle has been recently facilitated by genomic approach. The genomes of seven different *S. aureus* strains (Lindsay and Holden, 2004) and two *S. epidermidis* ones (Gill *et al*., 2005) are currently publicly available providing a plethora of information on the bacteria biology and genetics. Only a brief comparison of the syntenic (collinear) core of those chromosomes allows distinguishing a handful of *S. aureus* genes not present in *S*. *epidermidis* that had been previously attributed to the pathogenic potential of *S. aureus* (Lindsay and Holden, 2004). The primary source of intra- and interspecies variation in pathogenicity and resistance, however, originates from large nonsyntenic regions – the genome islands. The islands cluster well-documented virulence factors such as a toxic shock syndrome toxin, leukotoxin and leukocidin or exo- and enterotoxins to mention only some (Gill et al., 2005). Many of those genes appear in tandem repeats of slightly different forms suggesting some islands may act as virulence 'gene nurseries' (Lindsay and Holden, 2004). The genomic approach allowed moreover identification of several poorly characterized genes that are hosted at the islands among the thoroughly characterized deleterious factors. It is possible therefore to hypothesize that those proteins may contribute largely to *Staphylococcal* virulence. Among the identified genes a cluster of six spl proteins seems to be the most interesting in terms of its potential role in virulence.

The first report on spl proteins was contributed by Rieneck and collaborators (1997), who have screened a *Staphylococcal* expression library with serum of an endocarditis patient to find a high level of antibodies against the SplC protein, thus indirectly demonstrating its expression in the course of disease. The authors have also pointed to the homology of the identified protein with glutamylendopeptidase (V8 protease) and epidermolytic toxins. The latter proteins are specific serine proteases with documented role in *Staphylococcal* virulence. V8 protease, cleaving only after dicarboxylic aminoacids, was ascribed to a number of rather unspecific deleterious processes during infection associated both with limited proteolysis of host proteins (though due to narrow substrate specificity it was never considered solely as a supplier of nutrients) and the modification of bacteria protein repositories (Arvidosn S.O., 1983; Dubin G., 2002). Apart from its apparent role in pathogenesis

the protease found a widespread use for protein fragmentation in MS and amino acid sequencing. The epidermolytic toxins are even more specific proteases, responsible for *Staphylococcal* Scalded Skin Syndrome - a disease affecting only young infants, characterized by splitting of the cellular layers of skin. Desmoglein 1, a desmosomal cadherin mediating the cell-cell adhesion was proven to be the target protein of epidermolytic toxins. Except for the above case and a seemingly physiologically irrelevant cleavage of melanocyte stimulating hormones no other substrates were identified to date (Amagai *et al*., 2000).

Genetic characterization of the entire spl operon and the credit for the name of the group (Spl – Serine protease like) goes to Reed and colleagues (2001). The authors confirmed secretory production of Spl proteins and demonstrated a positive regulation by the accessory gene regulator (*agr*) similar to that observed for a multitude of documented virulence factors. Moreover, SplB and C were shown to truly posses the speculated proteolytic activity. Furthermore, the authors generated an Spl operon knockout strain and assessed its virulence in a crude murine intraperitoneal injection model to find no significant difference with the wild type. However, in the view of above presented genetic and comparative data and as the authors themselves acknowledge, to unambiguously demonstrate the role of spl proteins in *Staphylococcal* virulence, studies in models more closely mimicking the normal infection are needed.

The above briefly reviewed contributions are the only reports known to us that were published up to date on spl proteases. The unanswered questions, together with the described revelations uncovered by genomic studies and a general need to understand the *Staphylococcal* biology and pathogenesis, prompted us to closer examine the spl proteases. We have developed an efficient expression system for recombinant production of SplB and SplC proteases, which may be easily extended to other members of the group to facilitate future studies on substrate specificity and assessment of potential role of those enzymes. Further, we describe a unique activation mechanism to protect the cell from a premature activity of those enzymes. We also solved a crystal structure of the SplC protease demonstrating that like V8 protease and epidermolytic toxins the Spls possess a highly restricted substrate specificity achieved by a unique mechanism of activation of the active site of the enzyme. The SplC protease catalytic triad is blocked by an unusual conformation of

an adjacent flexible loop. Folding of this loop upon binding of a substrate allows the enzyme to become fully active.

3.2.2 Materials and methods

3.2.2.1 Purification of His-tagged proteins

The His-tagged proteins were purified form *E. coli* BL21, carrying appropriate plasmids. The bacteria were grown overnight, harvested by centrifugation and lysed by sonication in buffer E (50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl and 10 mM imidazole). The recombinant protein was separated on chelating sepharose charged with nickel ions and eluted with 250 mM imidazole in buffer E. If cleavage was necessary, the preparation was dialysed into appropriate buffer and incubated with the protease until the process was completed. This was followed by dialysis to buffer F (50 mM sodium acetate, pH 5.0). Otherwise the protein was dialysed directly to buffer F. The preparation was then applied to a Source S column and bound proteins were eluted with a linear gradient of buffer G (F+1M NaCl). Fractions containing the desired protein peak were pooled and gel filtration in either PBS or crystallization buffer (5 mM Tris pH 8.0, 50 mM NaCl) was carried out on superdex s75pg.

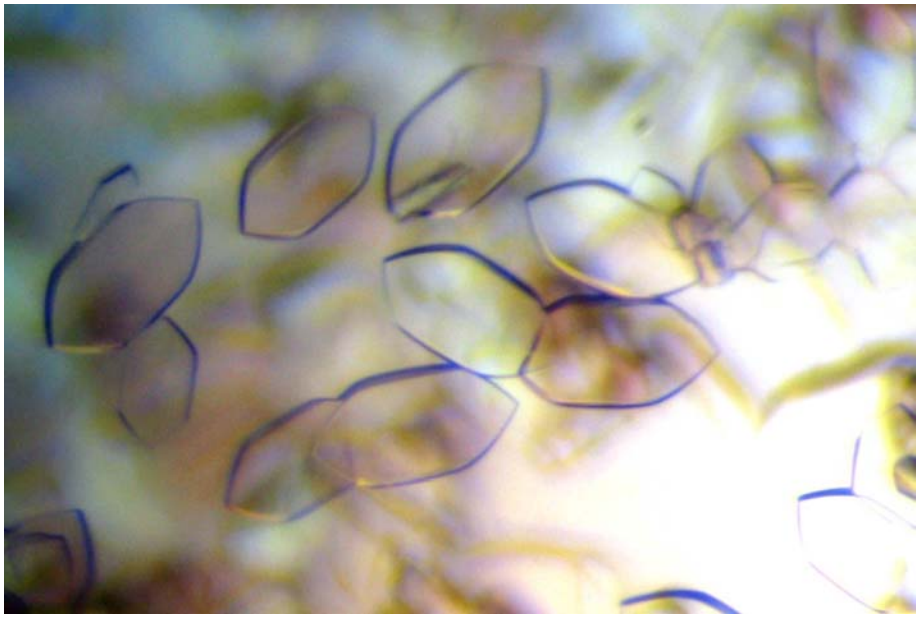


Figure 3.2.1. Crystals of the SplC protein.

3.2.2.2 Protein crystallization and structure solution

The purified protein was concentrated to 10 mg/ml in the crystallization buffer and used for initial screening. Crystals appeared in several different conditions after a few weeks. The best buffer was Hampton Research Crystal Screen solution number 45 (0.2 M zinc acetate, 18% PEG 8000, 0.1 M sodium cacodylate pH 6.5). These conditions were further optimized (the final solution 6% PEG 8000, 0.2 M zinc acetate, 0.1 M sodium cacodylate pH 7.5) to obtain single monocrystals of about 0.2 mm (Figure 3.2.1). Crystals were plunge-frozen after 30 s in the cryoprotectant solution containing 20% of glycerol in the mother liquor. The diffraction data was measured on the MPG/GBF beamline BW6 at DESY. Data were indexed, integrated and scaled with XDS package (Kabsch, 1993). The native dataset, collected at 1.05 Å had a significant anomalous signal from zinc and therefore, was treated as a remote-SAD experiment. Six heavy atom sites were found using the SHELXD (Schneider and Sheldrick, 2002) software. Initial atom positions were refined using autoSHARP software package (La Fortelle and Bricogne, 1997). Resulting phases were improved by the DM program (CCP4, 1994) and used for an automated model building with Arp/Warp software (Perrakis et al., 1999). The resulting model of about 80% completeness was inspected and finished manually with Xfit program (McRee, 1999). Restrained refinement by the Refmac5 software (CCP4, 1994) was followed by addition of water molecules by Arp/Warp (Lamzin and Wilson, 1993). Data collection, phasing and refinement statistics are given in Tables 3.2.1 and 3.2.2. Most of the model has a clear and well interpretable electron density with the exception of the Gly176-Glu183 loop in B chain indicating high flexibility of this region. Additionally, some solvent exposed sidechains had no interpretable electron density. The above parts were omitted in the final model. The R-factor of the presented structure is 21.6% and R-free 24.9%.

Table 3.2.1 Data collection statistics for SplC crystal. Dataset measured at MPG/GBF beamline BW6 at DESY using MARCCD detector.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 2.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	Rmrgd-F	S_norm/ S_ano
10.00	1861	454	482	94.2%	3.2%	4.2%	1856	29.83	3.8%	2.2%	1.24
6.00	7497	1730	1757	98.5%	3.9%	4.5%	7496	27.71	4.5%	2.8%	1.17
5.00	7074	1632	1661	98.3%	4.2%	4.7%	7074	26.58	4.8%	2.9%	1.10
4.00	15855	3667	3704	99.0%	4.1%	4.6%	15855	27.43	4.7%	2.8%	1.06
3.50	15763	3676	3738	98.3%	4.5%	5.0%	15763	25.11	5.2%	3.2%	1.06
3.00	27851	6475	6665	97.1%	5.1%	5.5%	27851	22.38	5.8%	3.8%	1.07
2.60	38807	9071	9676	93.7%	7.5%	7.6%	38807	16.94	8.5%	5.6%	1.06
2.50	13359	3135	3432	91.3%	9.3%	9.3%	13359	14.39	10.7%	7.0%	1.05
2.20	53637	12758	14526	87.8%	11.5%	11.5%	53637	11.92	13.2%	8.7%	1.04
2.10	24015	5766	6849	84.2%	14.0%	14.0%	24015	10.02	16.1%	10.3%	1.03
2.00	26881	6565	8312	79.0%	18.0%	17.5%	26881	8.16	20.8%	13.2%	1.02
1.90	29051	7116	10069	70.7%	22.6%	22.0%	29049	6.78	26.0%	16.3%	1.02
1.80	29099	7173	12513	57.3%	28.7%	27.5%	29098	5.45	33.2%	20.3%	1.02
1.70	26776	6763	15580	43.4%	37.8%	36.0%	26750	4.15	43.7%	26.3%	1.01
total	317526	75981	98964	76.8%	7.3%	7.6%	317491	13.43	8.4%	7.8%	1.06

Table 3.2.2. Data collection, phasing and refinement statistics for SplC protein crystals

Data collection	
Space group	$P2_12_12_1$
Cell constants (Å)	a=53.12
	b=81.4
	c=107.39
Resolution range (Å)	50-1.7
Wavelength (Å)	1.05
Observed reflections	317526
Unique reflections	98964
Whole resolution range:	
Completeness (%)	76.8
R_{merge}	7.8
I/σ(I)	13.4
Last resolution shell:	
Resolution range (Å)	1.7-1.8
Completeness (%)	43.4
R_{merge}	26.3
I/σ(I)	4.15
Phasing	
Number of sites found/present	6/7
Phasing power	0.98
FOM	0.38
Solvent content [%]	48.7
Refinement	
No. of reflections	49311
Resolution (Å)	30 – 1.7
R-factor (%)	21.59
R_{free} (%)	24.98
Average B (Å^2)	19.54
R.m.s bond lenght (Å)	0.008
R.m.s. angles (°)	1.151
Content of asymmetric unit	
RMSD of monomers (Å)	0.56
No. of protein molecules	2
No. of protein residues/atoms	402/3156
No. of solvent atoms	224
No. of Zn atoms	7

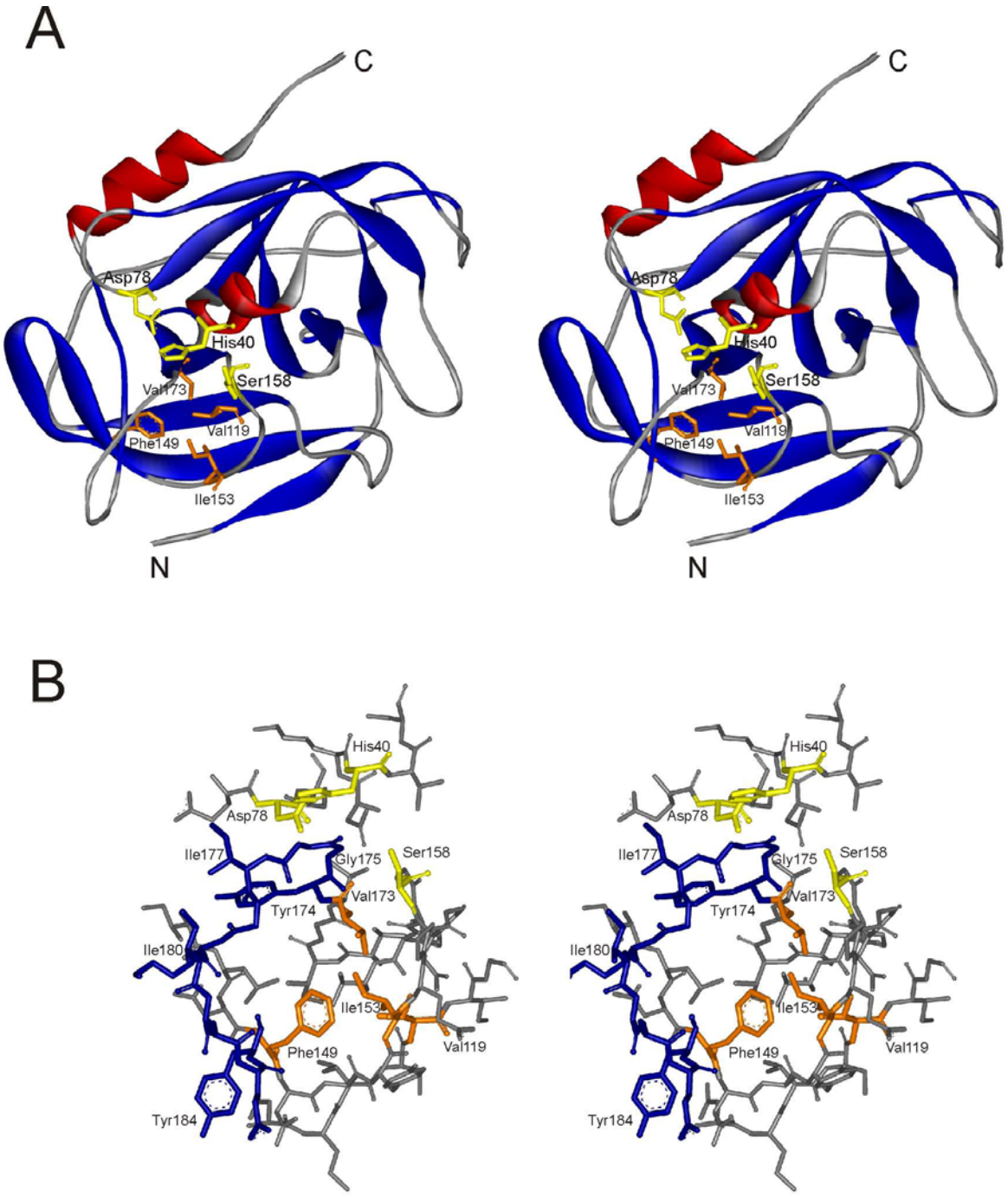


Figure 3.2.2.(A) Overall ribbon plot of the SplC molecule in stereo. The catalytic triad is presented in yellow. The primary specificity region in orange. (B) A close-up on the active site of the protease. Steric conflict of His40 with Gly176-Gly176 is evident. A part of the structure crucial for proteolytic activity is well formed. Coloring as in (A). The Tyr174-Tyr184 region, crucial for substrate recognition and enabling activity, is presented in dark blue.

3.2.3 Activation mechanism

Since 1D proton NMR spectra of SplC and SplB showed that the proteins were properly folded, the extra amino acids from tags remained a last conceivable reason for the observed lack of activity of spls. Analysis of the structures of homologous serine proteases - trypsin and epidermolytic toxins - made this interference with the activity of the enzyme rather not probable. However, when the structure of V8 protease appeared (Prasad *et al*., 2004), documenting the role of the first N-terminal residue in substrate specificity, we decided to remove the extra residues from pSplB. Two derivatives were constructed by site directed mutagenesis – pSplB(X) and pSplB(E). The first carried a factor Xa cleavage site, while the latter enterokinase. Both enzymes are superior to thrombin. They are unspecific for the primed sites and accept most residues thus being suitable for generating the N-terminus of a recombinant protein. Both SplB(X) and SplB(E) were active after purification and cleaving off the tag in zymographic analysis. Moreover, our results demonstrate that a precise removal of the secretion sequence by signal peptidase activates the enzyme only in the extracellular compartment thus protecting the cytoplasm from its premature activity.

We have also engineered a factor X cleavage site into pSplC, between the signal sequence and the mature protease in the plasmid. However, after purification and the cleavage the protein was still inactive in zymographic analysis. The C-terminal His-tag was also removed by cloning splC gene into pGEX-5T with factor Xa cleavage site directly preceding the mature protein sequence as previously for SplB. Nevertheless, though properly folded and entirely similar to a native protein in terms of the aminoacid sequence, the protein after purification and cleavage was inactive in zymographic analysis. These results suggest either a total lack of activity or, more probably, a restricted substrate specificity of the SplC protein.

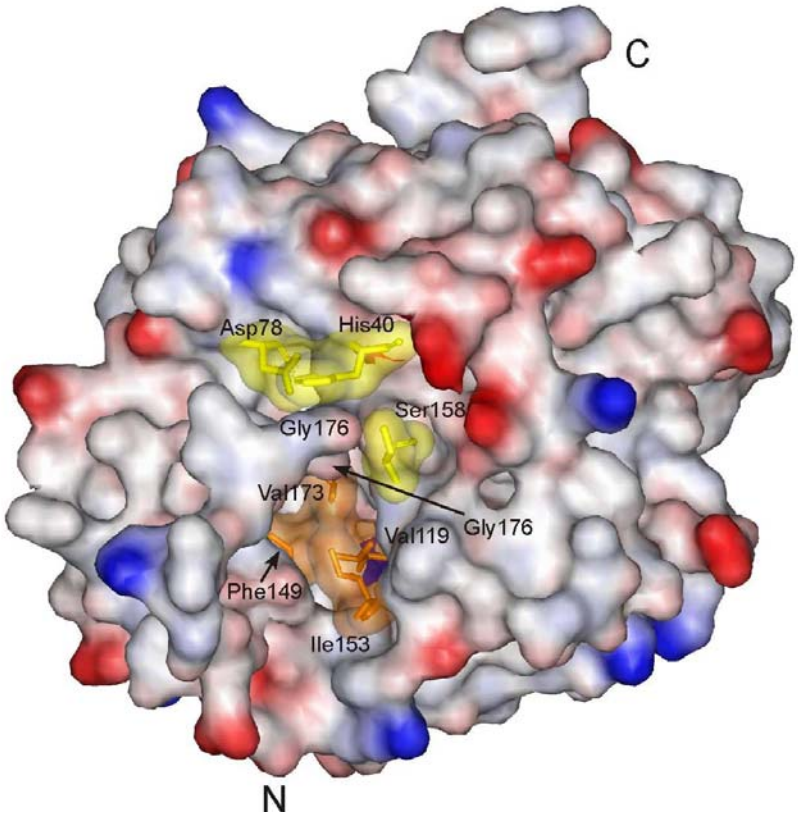


Figure 3.2.3. Surface representation of the SplC molecule. The catalytic triad (yellow) is blocked by Gly175-Gly176 and the whole Tyr174 -Tyr184 loop region interferes with enzymatic activity. The specificity region enables binding of a medium-sized hydrophobic residue.

3.2.4 The crystal structure of the SplC protease

To determine the structural basis for the lack of detectable activity of SplC, we have determined the crystal structure of the splC. The overall structure of SplC reveals a trypsin-like fold. The structure consists of two perpendicular β-barrels, with the helix Lys39-Asp44 (stabilizing the catalytic His40) and a second one at the C-terminal end (Pro192-Lys200). The active site is located at the interface of the barrels. The standard serine protease catalytic triad is well preformed by His40, Asp78, Ser158 residues, while the oxyanion hole is shaped by Gly156. It appears that this protease is not in an active state due to a side chain of the catalytic histidine being rotated outside the catalytic cleft (Figure 3.2.2 and 3.2.3), because of a steric conflict between the active site His 40 and Gly175-Gly176 residues. Moreover, the Asn38 sidechain is rotated to the space freed by the imidazole ring of His40. The residue preceding glycines (Tyr174) has its main chain angles of φ=-50º and ψ=150º, while

in the V8 protease φ=-125º and ψ=-35º. This causes Gly175-Gly176 to push away the imidazole ring of His40. In addition, the glycine residues begin a much longer loop (Tyr174 -Tyr184) of a significantly different conformation than the one seen in other trypsin-like serine proteases, having an elongated β-strand instead of a flexible loop at this position. This region has a clear electron density map only in one of the two molecules present in the asymmetric unit, thus, indicating the flexibility of this area. However, the position of Gly175 is defined and identical in both molecules regardless of crystal packing. The loop has two hydrophobic isoleucine residues (Ile177 and Ile180) unusually exposed to solvent. Two valine residues in positions 172-173 most probably cause such an atypical configuration. The valines are unique to SplC while other Spls, V8, ETA and ETB have isoleucine-histidine in corresponding positions. The Tyr174 side chain is placed similarly to the tryptophan imidazole ring of V8 protease, however the sidechain conformation of this residue is completely different. Thus, a significant change in this entire area has to take place to move glycines 175 and 176 from the inhibiting positions and allow the His40 sidechain to shift into a catalytic orientation thus enabling the proteolytic activity.

Substrate specificity is determined by S1 residues in standard trypsin-like proteases. The S1 pocket of SplC is formed by Val173, Ile153, Val119 and Phe149. It is completely hydrophobic; suitable to accept a medium sized nonpolar side chain of the substrate.

The N-terminal Met1 residue is a cloning artifact and remains outside the molecule. This residue appears to have a flexible side chain, as it is not seen in the electron density map. At the C-terminus three out of five histidines of the his-tag are visibly bound to zinc atoms of the buffer.

3.2.5 Comparison of SplC and other trypsin-like proteases

The overall folds of SplC and the recently published V8 protease (Prasad et al., 2004) are almost identical. The superimposition of the SplC structure with that of protease V8 gives backbone rms deviation of 1.1 Å and with that of ETA 1.26 Å, while with porcine trypsin the value is 1.31 Å. When compared to the V8 protease, most significant differences appear in the regions between Glu85-Asn101 and Gly176-Tyr184 (Figure 3.2.4). The Glu85-Asn101 loop is located on the opposite side of the molecule to the active site; therefore it is unlikely that it affects the protein function. Superimposition of the backbone trace and important residues is presented in Figure 3.2.4. The positively charged S1 specificity pocket of protease V8 has similar size to that of the SplC protein, however no N-terminal penetration is seen in the SplC molecule. The N-terminus in the V8 protease is reportedly responsible for the substrate specificity determination, which is obviously not the case in SplC and is not a result of an artifficial Met1 residue. The Tyr174 -Tyr184 region of SplC is longer than the corresponding Trp185-Phe192 region of V8 protease. In V8, it forms a β-strand, while in the SplC molecule it is flexible. Also in the ETA structure, this area is β-folded (Cavarelli et al., 1997). In V8, the His184 residue is important for stabilization of a polar S1 pocket. In SplC it is replaced by Val173 and the entire S1 pocket is hydrophobic. Unlike the ETA structure, where the oxyanion hole is blocked by Pro192, SplC shows proper, undisturbed fold in this area.

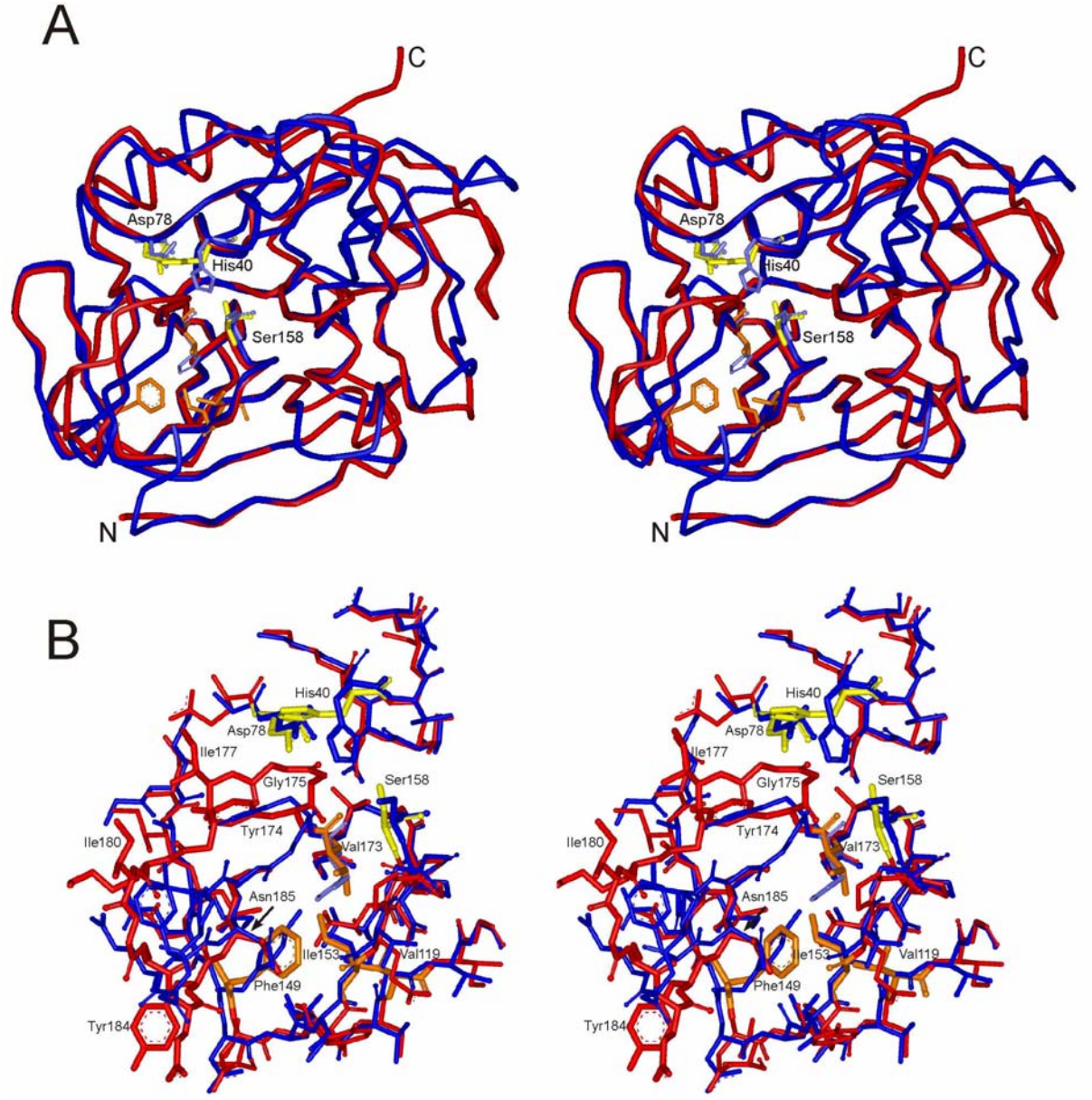


Figure 3.2.4. Structure of SplC (red) superimposed on the V8 protease (blue). (A) An overall view of the whole model showing high degree of structural similarity. (B) The detailed view of the active site and specificity region. The structures appear to be almost identical except for the Tyr174-Tyr184 loop, which is β-structured in V8 while flexible in SplC. The "locked" state of the catalytic His40 is evident.

3.2.6 Conclusions

3.2.6.1 Protection of cytoplasm against misdirected Spls – activation mechanism

The two constructs (pSplB(X) and pSplB(E)) (engineered to express the SplB protein exactly collinear with the native one) produced an active Spl protease. We have therefore demonstrated that a precise processing of the N-terminals is necessary for the Spl activity, similarly to the S1A family of proteases, but the function of SplB is most probably regulated in a different manner than that of SplC.

All previously characterized *Staphylococcal* proteases are produced as inactive zymogens to ensure the protection of the cytoplasm in an event the enzyme becomes misdirected. The activation takes place only extracellularly in a proteolytic cascade. The indirect demonstration of this mechanism comes form heterologous expression studies where active enzymes are deleterious to *E. coli* when expressed intracellularly. The *Staphylococcal* cysteine proteases (staphopains A and B), in addition to propeptides, possess specific cytoplasmatic inhibitors. Mutants deficient in the latter show markedly decreased growth rates while heterologous coexpression of an active protease and its inhibitor becomes possible. Epidermolytic toxins are the only exception; this is however justified, considering extremely limited substrate specificity (Amagai et al., 2000). The Spl proteases, on the other hand, present detectable activity but do not carry propeptides of any kind nor intracellular serine protease inhibitors were described in *S. aureus*. Our data explains this apparent discrepancy. The signal sequences of these enzymes serve a double role not only directing the protease to the extracellular compartment, but also assuring their inactive states while inside the cell. Only after an exact N-terminal processing by a signal protease, the enzyme activity is liberated while already in the extracellular compartment.

We have attempted to produce the SplC protease using the same protocol as for SplB. Unfortunately, we were unable to show the enzymatic activity of two constructs: the His-tagged and collinear with the native one. Our NMR and crystallographic studies showed that the lack of activity was not due to improper folding. Based on the structural data, we propose that SplC is most probably a protease with a very limited substrate specificity and thus its activity could not have been demonstrated so far in the performed assays. This is in discordance with the work of Reed and collaborators (2001) where they demonstrate the SplC activity on

a casein zymogram analysis. We could explain this findings by contaminations of their preparation with other proteases.

3.2.6.2 The crystal structure of the SplC protease

The structure of the SplC protein remains in an agreement with known biochemical findings; however, the activation mechanism of splC appears to be significantly different from splB. The protein appears to be a serine protease with a well-formed active site. In the free form, the site is “locked” by an unusual Gly175-Tyr184 loop, pushing the side chain of the catalytic triad His40 away from the active conformation. This kind of internal inhibition is similar to the one previously described for the human pro-granzyme K. Pronounced changes in the conformation of the loop enable the proteolytic function of this inhibited protease (Hink-Shauer et al., 2002). Similarly in SplC, to allow the His40 imidazole ring to enter the active site cavity and form hydrogen bonds with Asp78 and Ser158 and push Asn38 away, the Gly175-Ser182 loop must acquire different conformation. This may be achieved by rotating the Tyr174 main chain bonds. It is reasonable to suggest that upon such a change the proteolytic activity may be developed, as other functionally essential parts of the proteolytic machinery are well preformed. We propose that the binding a specific substrate to the protease can most likely cause the described changes in the Gly175-Tyr184 region. This assumption is further supported by the presence of two hydrophobic residues, Ile177 and Ile180 exposed to solvent. These residues would interact with a ligand and allow for activation of the proteolytic functionality. The S1 specificity pocket is suitable to accept only a medium sized hydrophobic side chain. All the above features allow us to conclude that SplC is a highly specific protease, targeted to cleave after a hydrophobic region of the substrate.

3.3 The three-dimensional structures of tick carboxypeptidase inhibitor (TCI) in complex with bovine carboxypeptidase A and human carboxypeptidase B

3.3.1 Introduction

Metallo-carboxypeptidases (CPs) comprise a family of zinc-containing exopeptidases that catalyze the cleavage of C-terminal peptide bonds in proteins and peptides. These enzymes can be subdivided structurally into the pancreatic/digestive or CPA/B subfamily and the regulatory or CPN/E subfamily (Vendrell et al., 2000). The members of the first subfamily are synthesized as inactive zymogens called procarboxypeptidases (PCPs) and are subsequently activated by limited proteolysis (Aviles et al., 1993). Traditionally, these carboxypeptidases have been classified into A and B forms based on their substrate specificity: CPA has a preference for aliphatic and aromatic C-terminal residues, and CPB for basic C-terminal residues (Guasch et al., 1992). The members of the CPN/E subfamily, commonly referred to as “regulatory” carboxypeptidases, do not appear to be produced as inactive precursors, and are located in subcellular compartments to prevent inappropriate cleavages that would otherwise damage the cell (Gomis-Rüth et al., 1999; Reznik and Fricker, 2001).

Normally, the biological actions of proteases are controlled by specific interactions with proteinaceous inhibitors (Bode and Huber, 2000). So far, however, only a few protein inhibitors have been identified for metallo-carboxypeptidases (Vendrell et al., 2000). Among them, the potato and leech carboxypeptidase inhibitors (PCI and LCI) have been structurally studied in detail by NMR (Clore et al., 1987; Reverter et al., 2000; Gonzalez et al 2003) and by X-ray crystallography in their complexes with bovine CPA and human CPA2 (Rees and Lipscomb, 1982; Reverter et al., 2000). Both proteins inhibit carboxypeptidases via a substrate-like interaction of their C-terminal tail with the active site groove of the enzyme (Bode and Huber, 1992; Vendrell et al., 2004).

Numerous protease inhibitors have been described in hematophagous organisms, most notably those directed toward their host’s blood coagulation factors such as thrombin and factor Xa (Salzet, 2001; Mans and Neitz, 2004). The structures of several of these inhibitors in complex with different target proteases have been solved by X-ray crystallography. These include hirudin, ornithodorin, triabin, and tick anticoagulant peptide, among others (Rydel et al., 1990; van de Locht et al., 1996; Fuentes-Prior et al., 1997; Wei et al., 1998). A common feature of these non-

homologous inhibitors is their small size and their stabilization by several disulfide bridges.

The tick carboxypeptidase inhibitor (TCI) is a novel metallo-carboxypeptidase inhibitor recently identified in the ixodid tick *Rhipicephalus bursa* (Arolas et al., 2005). TCI is a 75-amino acid protein stabilized by 6 disulfide bridges that inhibits various members of the A/B carboxypeptidase subfamily with nanomolar affinity, including plasma CPB (also called thrombin-activatable fibrinolysis inhibitor, TAFI). Its amino acid sequence shows no homology with other known proteins except at its C-terminus, which resembles that of PCI and LCI, suggesting a similar mechanism of inhibition. Most likely, ticks secrete TCI during feeding and digestion to maintain the liquid state of blood because TCI strongly inhibits TAFI, a well-known attenuator of fibrinolysis (Bajzar et al., 1995; Bouma and Meijers, 2003). Indeed, we have recently shown that recombinant TCI stimulates fibrinolysis *in vitro* and thus may have potential for applications to prevent or treat thrombotic disorders (Arolas et al., 2005). In addition, TCI could also modulate host defense mechanisms during the parasitic infection by affecting mast-cell CPA and/or TAFI, both of which are thought to be involved in inflammatory processes (Springman et al., 1995; Myles et al., 2003).

3.3.2 Protein expression and purification

Expression and purification of recombinant TCI were carried out as described by Arolas et al. (2005). TCI was produced in the *Escherichia coli* strain BL21(DE3) using the pBAT-4-OmpA plasmid for periplasmic expression, and was purified from the culture medium using a Sep-Pak C_{18} cartridge followed by cation-exchange chromatography and reversed-phase high performance liquid chromatography on a Vydac C_4 column.

Recombinant human pancreatic PCPB and the derived active CPB were produced as published in Reverter et al. (1998) and in Ventura et al. (1999). Bovine pancreatic CPA was purchased from Sigma and used without further purification. The identity and purity of the proteins was verified by mass spectrometry, SDS-PAGE, and N-terminal sequence analysis.

3.3.3 Complex formation

The TCI-bCPA and TCI-hCPB complexes were prepared by incubating equimolar quantities of inhibitor and enzyme in 10 mM Tris-HCl (pH 7.5), 50 mM NaCl for 2 hours at 20ºC. The complete inhibition of the carboxypeptidases was monitored by continuous photometric assays with specific chromogenic substrates (Arolas et al., 2005). The complexes were purified by gel-filtration chromatography (Superdex 75 HiLoad 26/60), and finally concentrated to about 10-12 mg/ml.

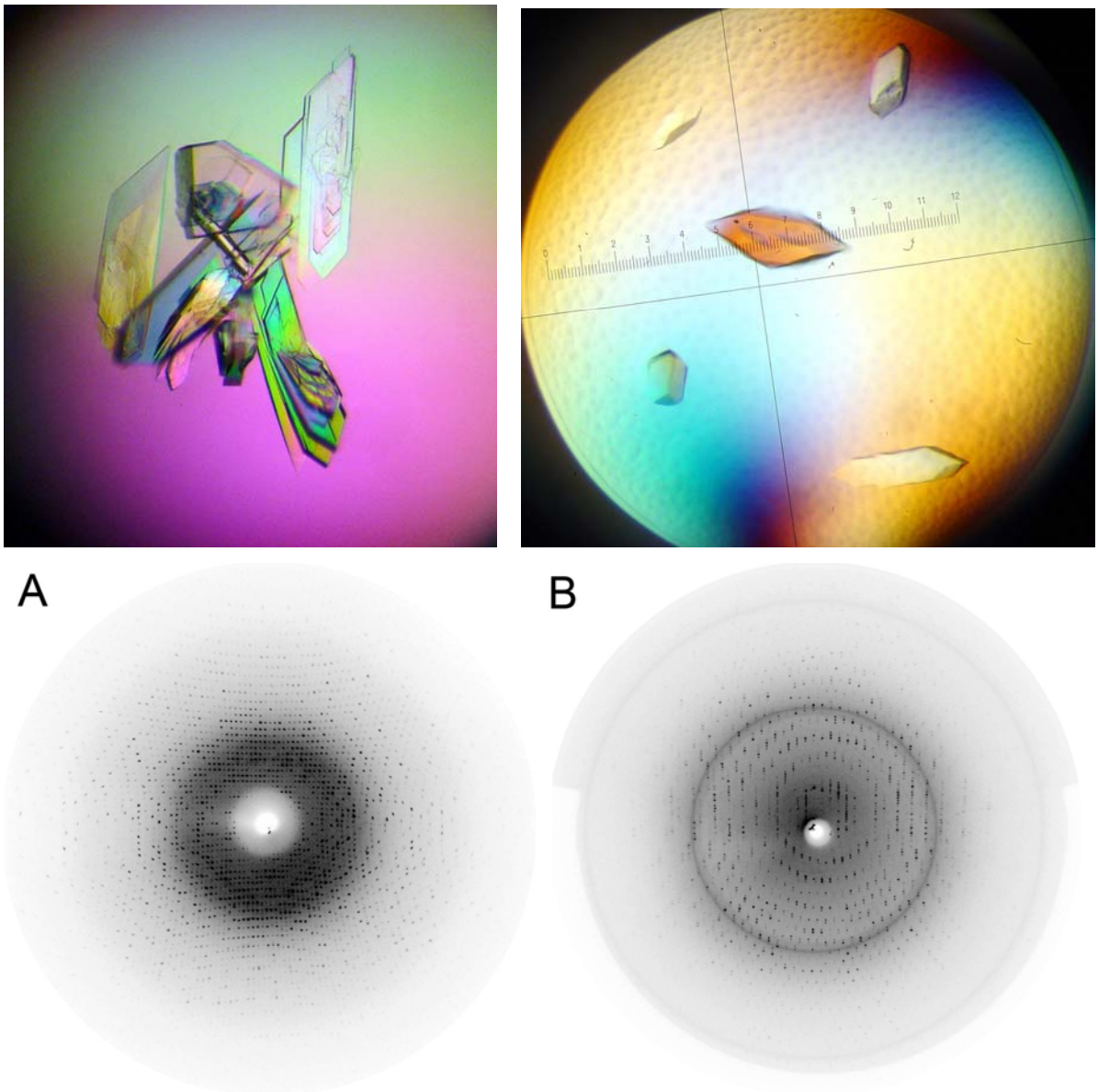


Figure 3.3.1. (Top) Crystals of the CPA-TCI(left) and CPB-TCI(right) complexes. (Bottom) Diffraction patterns of CPA-TCI (A) and CPB-TCI (B) crystals. For CPA-TCI space group is $P2_1$ and maximal resolution 1.7 Å. CPB-TCI crystallized in $P4_32_12$ spacegroup and diffracted up to 2.0 Å.

3.3.4 Crystallization and diffraction data collection

For the TCI-bCPA complex, crystals suitable for data collection were obtained after a few days at 20ºC using the hanging-drop vapor-diffusion method. Drops were prepared by mixing equal volumes of protein solution and reservoir buffer containing 0.2 M sodium cacodylate (pH 6.5), 0.2 M zinc acetate dihydrate, 7% (w/v) PEG 8000, and 10% (w/v) dried dioxan. For the TCI-hCPB complex, single crystals of suitable size were grown within a few weeks at 20ºC using the sitting-drop vapor-diffusion method. Drops were made by mixing equal volumes of one of the assayed reservoir solutions (0.1 M bis-tris (pH 5.5), 0.2 M lithium sulfate monohydrate, 25% (w/v) PEG 3350) and the protein solution.(Figure 3.3.1)

Both TCI-bCPA and TCI-hCPB crystals contained one complex per asymmetric unit. 1.7 Å (TCI-bCPA) and 2.0 Å (TCI-hCPB) native data sets were obtained from plunge-frozen crystals at 100 K. Crystals were measured using an in-house rotating anode X-ray source. Rotation images were collected on Mar345 image plate. Data were indexed, integrated, and scaled using XDS software (Kabsch, 1993). The data statistics and crystals' space groups are summarized in Tables 3.3.1 and 3.3.2.

3.3.5 Structure determination and refinement

Both structures were determined by molecular replacement, however experimental phasing on Zinc atoms bound to the molecules was also possible. The known crystallographic structures of bovine CPA (pdb code 1M4L) and human PCPB (pdb code 1KWM) were used as probes (Kilshtain-Vardi et al., 2003; Barbosa Pereira et al., 2002). The N-terminal pro-segment of hPCPB (95 residues long) was removed from the search model. Rotation and translation searches were performed using Molrep (CCP4, 1994). For the TCI-bCPA complex, rotation search in the Patterson space yielded a peak of height 21.2 σ over the highest noise peak of 4.8 σ. Translation search gave peaks of 20.9 σ over the noise height of 4.9 σ. Corresponding values for CPB were 12.33 σover 6.36 σ, and 51,75 σ over 28,04 σ. The R-factors of the initial models without the inhibitor were 43% for bCPA and 46% for hCPB. At this stage, a free atom model improvement was introduced using Arp/wArp (Perrakis et al., 2001) for both structures. As the next step, iterative model building was performed for CPA using Arp/wArp (Perrakis et al., 1999). The model thus obtained (360 residues) was completed and revised manually using Xfit

software (McRee, 1999). Similarly, the structure of CPB was built using Arp/wArp, but the initial phases calculated from the improved model were corrected by density modification in the DM program. The resulting model of 356 residues was also completed using Xfit. For both structures, Arp/wArp was used to add solvent atoms (Lamzin et al., 1993). Structures were finally refined with Refmac5 (CCP4, 1994). Refinement statistics are shown in Table I. Final electron density maps were of high quality. The TCI-bCPA model was built from residues Ser3 to Leu305, and has no interpretable density for residues Ser134-Ser135 and side chains of residues Lys177, Lys190, Lys216, Gln221, Lys239, Arg276, Gln13i, and Gln58i (residues of TCI are suffixed with an i). The TCI-hCPB model has all residues present from Thr5 to Tyr309 but the following amino acids have no interpretable electron density in the side chain area: Arg92, Glu93, Lys122, Lys237, Arg 276, Lys298, Lys6i, Gln13i, Gln18i, and Glu19i. These side chains were removed from the models. The Ramachandran plot calculated for CPA model shows only one residue (Ser199) in a "disallowed" region of the ϕ, ψ plane. It agrees with the previously reported structures of native CPA. The TCI-bCPA contains four additional surface-bound zinc atoms probably as an artifact caused by the use of zinc in the crystallization buffer. Those atoms were the source of strong anomalous signal while the crystals were later measured at synchrotron source. The coordinates of the structures reported in this work have been deposited with the Protein Data Bank (accession code 1ZLH for the TCI-bCPA complex and 1ZLI for the TCI-hCPB complex).

Table 3.3.1. Data collection statistics for CPA-TCI and CPB-TCI complex crystals. Dataset measured at Cu rotating anode laboratory source and MAR345 detector.

CPA-TCI Crystal:
SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 2.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR COMPARED observed	I/SIGMA	R-meas	Rmrgd-F	S_norm/ S_ano
30.00	2	1	8	12.5%	0.3%	2	16.08	0.4%	0.3%	0.00
20.00	67	20	20	100.0%	4.3%	66	21.70	5.2%	3.1%	1.14
10.00	842	174	183	95.1%	8.8%	841	21.00	9.8%	2.7%	0.97
6.00	3726	706	718	98.3%	9.5%	3725	31.15	10.4%	3.3%	0.97
5.00	3566	665	678	98.1%	8.7%	3559	30.27	9.6%	2.9%	0.96
4.00	11948	1476	1486	99.3%	9.5%	11940	26.10	10.1%	2.7%	0.96
3.00	38608	4092	4162	98.3%	13.0%	38583	20.30	13.7%	3.5%	0.96
2.70	24720	2571	2645	97.2%	13.5%	24708	19.23	14.3%	4.0%	0.97
2.50	16667	2440	2555	95.5%	14.3%	16641	12.94	15.4%	6.1%	0.96
2.30	18450	3300	3504	94.2%	14.2%	18405	12.84	15.6%	6.7%	0.96
2.10	24790	4577	4995	91.6%	15.1%	24717	11.40	16.6%	7.4%	0.96
1.90	32683	6352	7254	87.6%	16.4%	32515	9.68	18.2%	8.8%	0.96
1.80	19588	4016	4903	81.9%	18.9%	19428	8.44	21.1%	10.8%	0.97
1.70	11971	3786	6154	61.5%	17.6%	11511	6.48	20.9%	13.4%	0.99

CPB-TCI Crystal:
SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 2.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR COMPARED observed	<I/SIGMA>	R-meas	Rmrgd-F	S_norm/ S_ano
20.00	136	30	48	62.5%	3.0%	135	43.61	3.5%	2.2%	1.41
10.00	2674	275	276	99.6%	3.0%	2674	61.26	3.2%	1.3%	1.06
6.00	11109	1006	1014	99.2%	3.7%	11109	55.01	3.9%	1.5%	1.06
5.00	10191	890	898	99.1%	3.8%	10191	55.09	4.0%	1.4%	1.04
4.00	23126	1987	2012	98.8%	3.7%	23125	57.05	3.9%	1.4%	1.02
3.00	63215	5395	5487	98.3%	4.5%	63213	43.97	4.7%	2.0%	1.01
2.60	56792	4823	5022	96.0%	7.6%	56792	29.73	7.9%	3.5%	1.01
2.50	19567	1666	1774	93.9%	10.3%	19567	24.19	10.8%	4.5%	1.01
2.40	22860	1946	2077	93.7%	12.0%	22860	21.38	12.5%	5.3%	1.01
2.30	26466	2265	2484	91.2%	13.5%	26466	19.01	14.1%	6.0%	1.01
2.20	30884	2649	2923	90.6%	17.5%	30883	15.91	18.3%	7.2%	1.01
2.10	35262	3038	3495	86.9%	21.5%	35262	12.89	22.5%	9.1%	1.00
2.00	18227	1680	4215	39.9%	25.9%	18213	10.88	27.1%	10.7%	1.00

Table 3.3.2. Data collection and refinement statistics

	TCI-CPA	TCI-CPB
A. *Data collection*		
Space group	$P2_1$	$P4_32_12$
Cell constants (Å)	a=47.07 c=57.89 b=68.60 β=104.72	a=b=74.20 b=163.55
Resolution range (Å)	20-1.7	20-2.0
Wavelength (Å)	1.542	1.542
Observed reflections	215232	339432
Unique reflections	36248	29538
Whole range		
Completeness (%)	98.9	99.9
R_{merge}	6.7	4.6
$I/\sigma(I)$	12.9	28.5
Last shell		
Resolution range (Å)	1.7-1.8	2.0-2.1
Completeness (%)	61.5	39.9
R_{merge}	13.4	10.7
$I/\sigma(I)$	6.48	10.88
B. *Refinement*		
No. of reflections	36248	27853
Resolution (Å)	20-1.7	20-2.0
R-factor (%)	15.4	16.1
R_{free} (%)	18.5	21.5
Average B ($Å^2$)	15.2	19.38
r.m.s.d. bond length (Å)	0.007	0.010
r.m.s.d. angles (°)	1.112	1.273
C. *Content of asymmetric unit*		
No. of protein complexes	1	1
No. of protein residues/atoms	375/2908	380/2963
No. of solvent atoms	379	505
Other atoms or molecules (Zn)	5	1

3.3.6 Crystal structure of TCI

The structures of TCI are very similar in its complexes with bCPA and hCPB, with a rms deviation of 0.51 Å for the backbone chain atoms and 0.87 Å for both backbone and side chains. The structure consists of two clearly separated regions named here as the "N-terminal" (residues Asn1-Lys36) and "C-terminal" (residues Gly39-Leu74) domain that are linked by residues Leu37 and Thr38 (Figure 3.3.2). Although both domains show only a low degree of sequence identity they are structurally very similar, displaying an r.m.s. deviation of 1.26 Å for main chain atoms (Figure 3.3.2), and thus could have arisen by gene duplication. Similarly, several thrombin inhibitors from blood-sucking organisms such as *rhodniin, ornithodorin* and *savignin* are double-headed inhibitors (van de Locht et al., 1995; van de Locht et al., 1996; Mans et al., 2002). The N-terminal domain of TCI consists of a short α-helix located between residues Glu2 and Ser5 and an antiparallel triple-stranded β–sheet involving residues Gly9-Leu11 (β1), Ala20-Leu22 (β2) and Val30-Asp33 (β3) with α-β1-β2-β3 topology (Figure 3.3.2). In the C-terminal domain, residues Glu46-Leu50 (β1'), Lys55-Leu57 (β2') and Gly 67-Val72 (β3') also form a triple-stranded antiparallel β–sheet, which is connected to the N-terminal domain by a loop that is structurally similar to a small α-helix (residues Gly39-Gly44).

The disulfide pairing of TCI was determined during the three-dimensional structure calculations. Disulfide bridges are formed between Cys3-Cys31, Cys10-Cys27, Cys16-Cys32, Cys40-Cys70, Cys47-Cys64, and Cys54-Cys71 (Figure 3.3.2 lower panel). The large number of disulfides probably contributes to the high stability of the protein against temperature and denaturing agents observed in previous biochemical studies (Arolas et al., 2005). Also, it is probably responsible for the high percentage of residues in loops and for the absence of a defined hydrophobic core. However, approximately 38% of the TCI residues belong to regular secondary structure elements, a high content when compared to other proteins of similar size. A computer search employing DALI server (Holm and Sander, 1993) showed that the structures of both domains and their disulfide pattern are similar to those of proteins of the β-defensin-fold family (Torres and Kuchel, 2004). The fold of these proteins generally consists of a short helix or turn followed by a small twisted antiparallel β-sheet stabilized by six cysteine residues that are paired in a 1-5, 2-4, 3-6 fashion and

maintain a compact structure. It is remarkable that the superposition of each TCI domain with some members of this family (e.g. the venom toxin crotamine and human β-defensin-2) displays an rms deviation of only 1.3-1.5 Å for backbone atoms.

For both the potato (PCI) and leech (LCI) carboxypeptidase inhibitor, the overall structure in solution is very similar to that bound to carboxypeptidases (Reverter et al., 2000; Gonzalez et al., 2003). However, NMR studies indicated that the N- and C-terminal tails of both inhibitors are highly flexible but become more rigid upon binding to a carboxypeptidase. Unlike PCI and LCI, the N- and C-terminal ends of TCI appear more rigid due to their involvement and proximity to secondary structure elements and disulfides. The N-terminus is located within an α-helix, while the C-terminal tail protrudes from the β3-strand of the second domain that is greatly stabilized by both secondary structure interactions and disulfide bridges (Figure 3.3.2). It is worth mentioning that some flexibility is expected between the two domains of TCI because their relative position is slightly different when bound to bCPA and hCPB and no inter-domain interactions are observed in the TCI structure. Thus, the domains probably can adopt different orientations when binding to different carboxypeptidases. In addition, unlike in the cases of PCI and LCI that are compact one-domain proteins, conformational changes could also take place upon binding to the enzyme. A better understanding of these changes would require the determination of the structure of TCI in its free state and this is a next goal of our studies

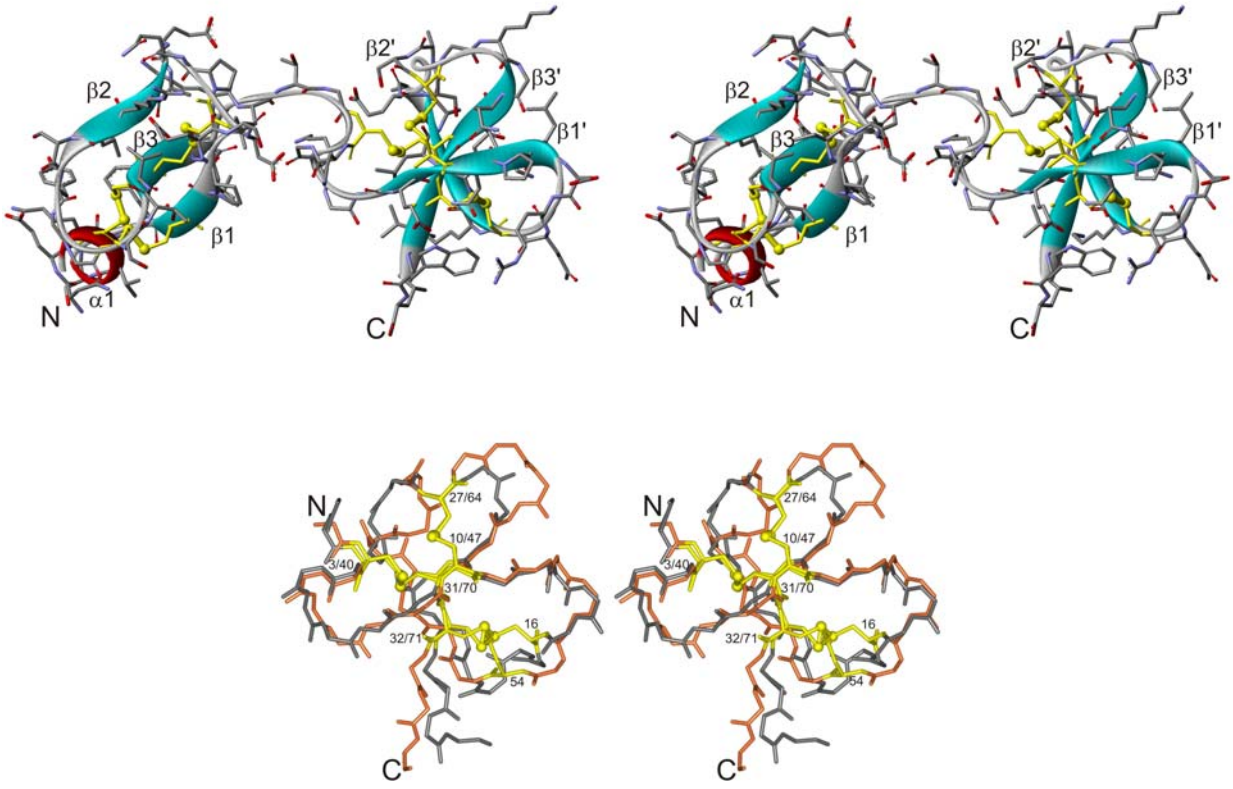


Figure 3.3.2. The three-dimensional structure of TCI. *Upper panel.* Stereo view of TCI shown in ribbon representation. The helix (α1) and β-strands (β1-β3 and β1'-β3') are colored red and light blue, respectively. The six disulfide bridges are shown in yellow. N and C indicate the N- and C-termini of TCI. *Lower panel.* Superposition of the backbone chains of the N-terminal (gray) and the C-terminal (orange) domains of TCI shown in a stereo view. The disulfide pairing of both domains is shown in yellow.

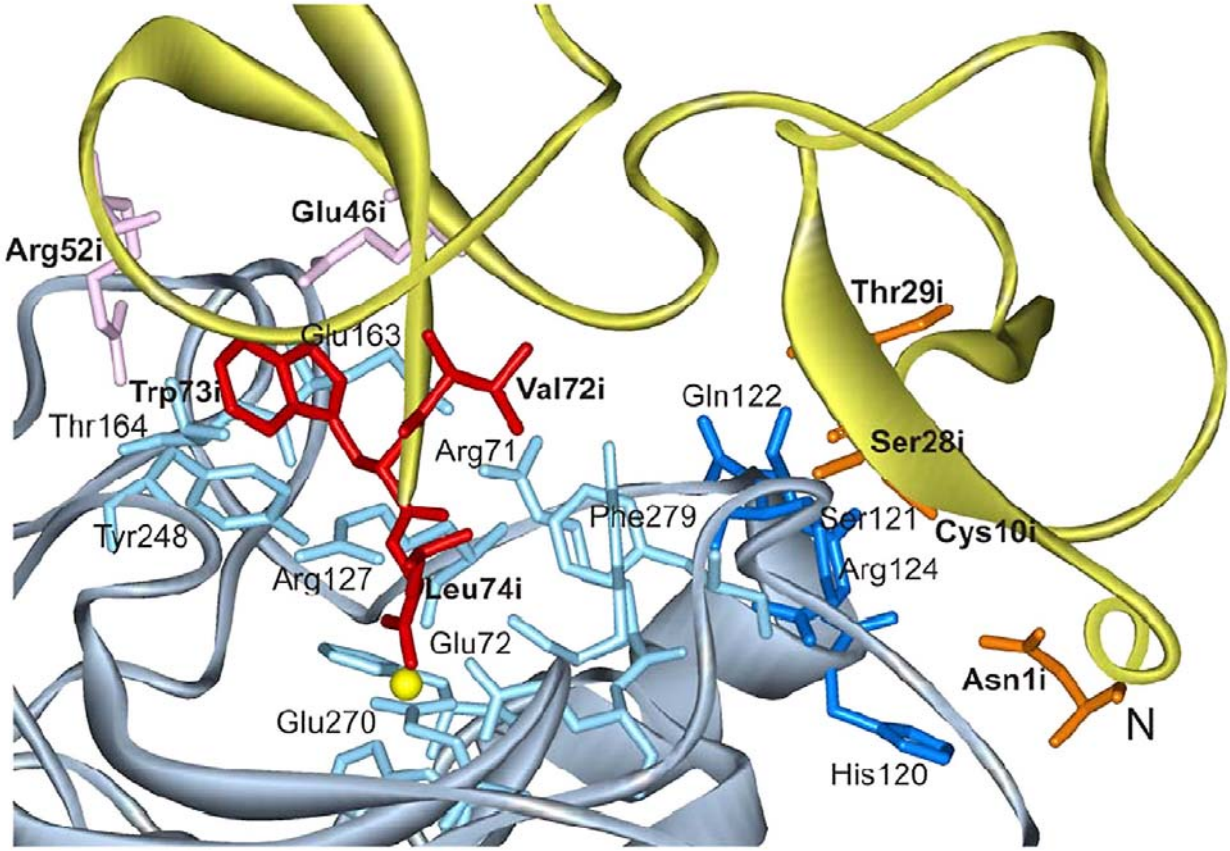


Figure 3.3.3. A close-up view of the double-headed binding of TCI to CPA. TCI (yellow) and CPA (gray-blue) are shown in ribbon representation. The N-terminus of TCI is labeled and the zinc atom of the CPA active site is represented by a yellow sphere. The residues involved in binding are labeled (TCI residues with a suffixed i), and their side chains are shown as stick models. Residues constituting the "primary binding site", i.e. those of the TCI C-tail and the CPA active site, are colored red and light blue, respectively. Residues forming "secondary contacts" are colored in light purple (TCI C-terminal domain) and light blue (CPA), and those that define the "exosite" on the CPA surface in orange (N-terminal domain of TCI) and dark blue (CPA).

3.3.7 Crystal structures of bovine CPA and human CPB

The structure of bovine CPA in its complex with TCI is almost identical to that of the unliganded enzyme described previously (Kilshtain-Vardi et al., 2003; r.m.s. deviation 0.43 Å for the backbone atoms). Bovine CPA, one of the most thoroughly studied carboxypeptidases, consists of a central mixed eight-stranded twisted β-sheet that is surrounded by eight α-helices forming a globular α/β motif (Figure 3.3.4). In contrast to the unbounded form, in the complex the catalytic water of the active site Zn^{2+} is substituted by the C-terminus of TCI. The largest differences between the free and TCI-bound structures of CPA are seen in residues 246 to 249 (shifted ~ 1.1 Å), a region directly involved in binding TCI. Among these residues, the movement of the Tyr248 side chain from the native "up" conformation (in isolated CPA) to the "down" position (in the complex) is noteworthy. This conformational change has also been reported in the complexes of other carboxypeptidase with protein inhibitors (Reverter et al., 2000). Additional differences are found in the side chains of residues Arg127, Arg145, Thr164 and Glu270, residues important for substrate binding and catalysis (Vendrell et al., 2000).The structure of human CPB in its complex with TCI is very similar to that of the pro-enzyme form (Barbosa Pereira et al., 2002), with a rms deviation between active enzymes of 0.48 Å for main chain atoms. The structure of the catalytic domain shows the characteristic α/β hydrolase topology of carboxypeptidases described above for bCPA. However, hCPB contains two additional disulfide bridges (Cys66-Cys79 and Cys152-Cys166). As seen in the TCI-bCPA complex, the catalytic water of the active site is substituted by the C-terminus of TCI. The largest differences between the native proform and the TCI-bound active hCPB are again found in the region involved in binding of TCI, i.e. residues 245 to 250 that are shifted by ~ 1.45 Å. In PCPB the Tyr248 presents a side chain flip and is in the "down" conformation when compared to the porcine counterpart or bCPA. Upon binding to TCI, the Tyr248 side chain is in a "down" conformation, with localization even deeper inside the carboxypeptidase. The side chains of residues Arg127, Arg145, Cys152 and Tyr165 are also hidden due to the binding of TCI. The side chain of Glu163 is moved to allow the "new" location of Arg127, while the Ile247 and Glu270 side chains are in a different conformer when compared to that in the free enzyme.

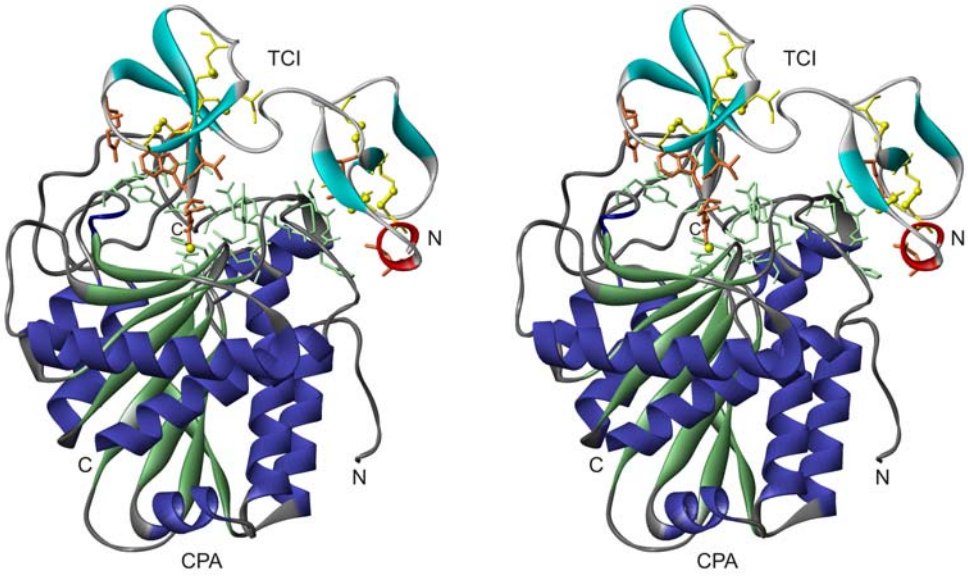


Figure 3.3.4. Structure of the TCI-CPA complex shown in a stereo ribbon representation. The helix and β-strands of TCI are shown in red and light blue, respectively, and the disulfide bridges are indicated in yellow. The helices and β-strands of CPA are shown in dark blue and dark green, respectively, and the catalytic zinc atom is represented by a yellow sphere. The side chains of the residues involved in the interaction are colored in orange (TCI residues) and light green (CPA residues). The N- and C-termini of TCI and CPA are labeled.

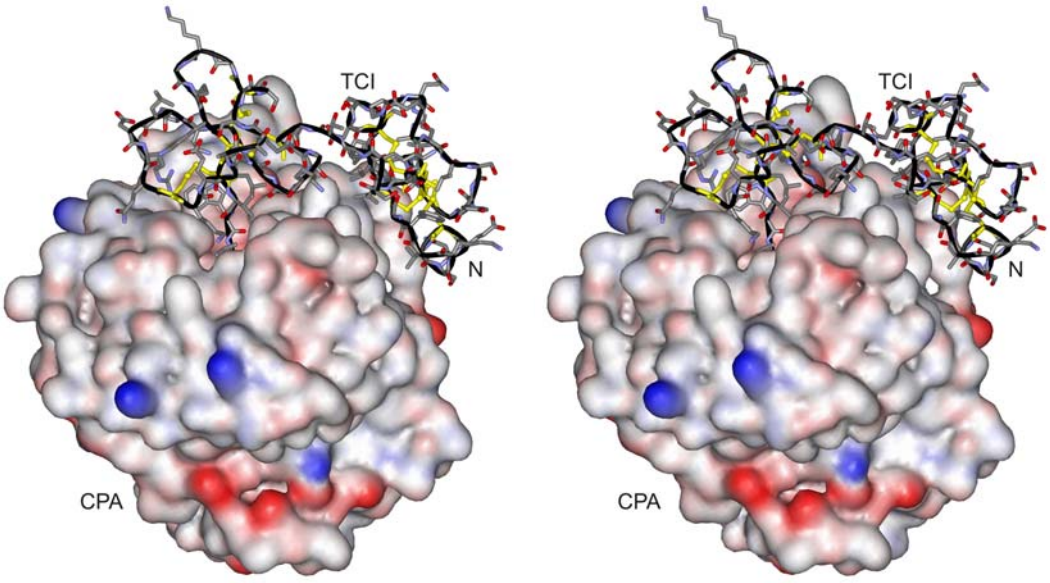


Figure 3.3.5. Interaction of TCI with CPA. CPA is shown in a solid surface representation; positive surface charges are colored in blue and negative charges in red. TCI is shown in a gray stick model and the disulfide bridges are highlighted in yellow. The N-terminus of TCI is labeled.

3.3.8 Binding interactions between TCI and the carboxypeptidases

TCI interacts extensively with both carboxypeptidases, with contact areas of 1229 and 1476 $Å^2$ in its complex with bCPA and hCPB, respectively. The interaction of the C-terminus of TCI with the active site groove of the carboxypeptidase constitutes the "primary binding region" (Figure 3.3.3). The binding subsites S1, S2 and S3, which are conserved between bCPA and hCPB, can be located based on their interaction with the C-terminal residues Leu74, Trp73, and Val72 of the inhibitor, respectively. After leaving the active site groove, the polypeptide chain of TCI first turns away from CPA/CPB but later approaches the enzymes' surface again to form a separate contact region (Figures 3.3.3, 3.3.4 and 3.3.5).

Unlike those of PCI and LCI, the C-terminal residue of TCI (His75i; in this section residues of TCI are suffixed with an i), which is cleaved off by the enzyme, is not trapped in the S1' subsite of CPA/CPB (formed by Asn144, Arg145 and Tyr248). The new C-terminal Leu74i residue coordinates with the active site Zn through one carboxylate C-terminal oxygen atom (2.1 Å) in the TCI-CPA complex and two carboxylate C-terminal oxygen atoms (2.51 Å for O, and 2.23 Å for OT) in TCI-CPB. Thus, in these complexes the coordination shell of Zn consists of five and six protein atoms, respectively, i.e. four from CPA or CPB (Oε1 and Oε2 of Glu72, and the Nδ1 atoms of His69 and His196), and one or two from TCI (O and OT of Leu74i). In comparison, in the LCI-CPA2 and PCI-CPA complexes two and one oxygen atoms from the inhibitor are coordinated to zinc, respectively. In the TCI-CPA/CPB complexes, the catalytic water, which in the free enzyme attacks the C-terminal peptide bond during the enzymatic reaction, is displaced by TCI. Water molecules are bound to the guanidinium group of Arg145, and to Asn144, Tyr248 and Glu270.

The S1 subsite of CPA/CPB (Arg127 and Glu270) is shielded by Leu74i, which interacts with the enzyme through several hydrogen bonds, i.e. two with the side chain of Glu270 (2.83/2.86 Å and 2.96/2.98 Å for TCI-CPA/CPB) and one with the guanidinium group of Arg127 (2.83/3.0 Å) (Figure 3.3.3). These important interactions are also present in the LCI-CPA2 complex. Leu74i forms an additional hydrogen bond with the side chain of Glu72 (3.20/3.19 Å), which formally does not belong to the S1 subsite but is conserved in both carboxypeptidases. The main chain of Leu74i is also hydrogen bonded with the side chain of Tyr248 (2.92/2.79 Å).

The major interaction in the S2 subsite (formed by Arg71, Ser197, Tyr198 and Ser 199) is the hydrogen bond between the carbonyl oxygen of Trp73i and the guanidinium group of Arg71 (3.04/3.07 Å). The side chain of Trp73i is also in contact with Thr164 (conserved in both carboxypeptidases) and forms an internal hydrogen bond with the side chain of Glu46i. Equivalent internal hydrogen bonds are found in the PCI/LCI-CPA complexes (Reverter et al., 2000). Another remarkable intramolecular hydrogen bond is established between Val72i and Lys55i. There are no hydrogen bonds between Val72i and the S3 subsite (Phe279), although the Val72i residue is in contact with the phenyl ring of Tyr198 and Phe279. In both complexes several polar residues of TCI (Lys41i, Glu46i, Arg52i, Glu53i and Lys55i) surround the TCI C-terminus, thereby covering the remainder of the CPA/CPB active site.

The only C-terminal residue of TCI that is likely to be charged is the carboxylate group of Leu74i (after cleavage). The active site of the carboxypeptidase contains several charged groups that are buried in both complexes, i.e. the guanidinium groups of Arg71, Arg124, Arg127 and Arg145 and the carboxylate group of Glu270. These charged residues could be important for the proper approach of substrates to the active site, for their proper positioning in it, and for their cleavage (Aviles et al., 1993).

Similar to PCI and LCI, residues from TCI establish “secondary contacts” with regions close to the carboxypeptidase active site groove (Figure 3.3.6). Thus, e.g. side chains of residues Glu46i and Arg52i of the C-terminal domain form hydrogen bonds with Glu163 and Thr164 of CPA, respectively. However, unlike other inhibitors, TCI binds to carboxypeptidases in a double-headed fashion: The N-terminal domain of TCI establishes interactions with the enzyme surface away from those made by its C-terminal domain or those made by PCI and LCI (Figure 3.3.4 and 3.3.5). These additional “secondary contacts” of the N-terminal domain define an “exosite” on the enzyme surface not identified previously, that is located around the S4 subsite of the carboxypeptidase (formed by Gln122/Lys122, Arg124 and Lys128). The interaction of TCI with this exosite comprises several hydrogen bonds between Asn1i and His120/Trp120 (3.26/3.06 Å for TCI-CPA/CPB), Cys10i and Arg124 (2.8/2.81 Å), Ser28i and Ser121/Thr121 (3.06/2.95 Å), and Thr29i and Gln122/Lys122 (2.71/2.58 Å).

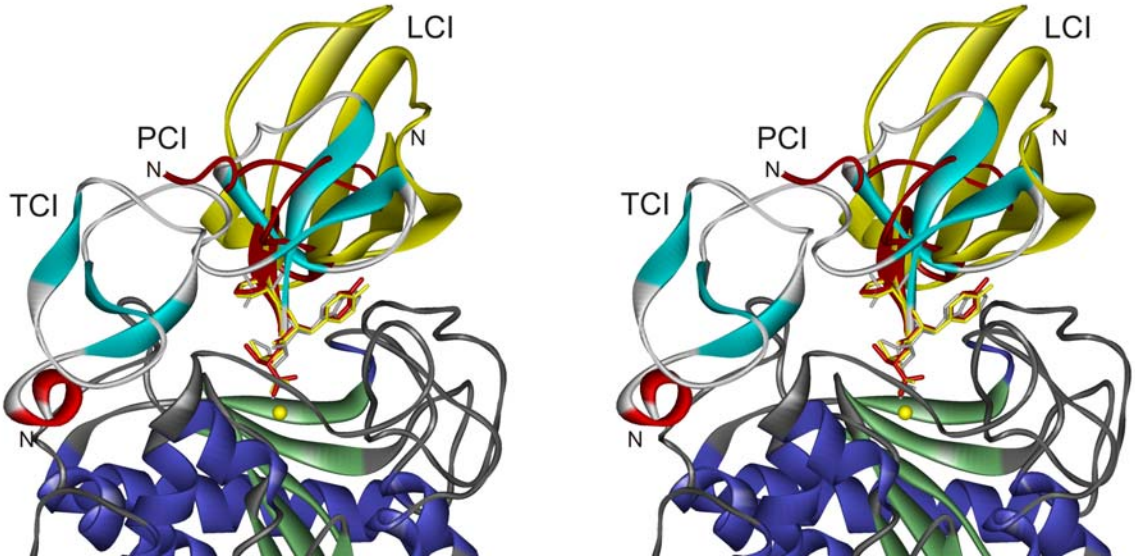


Figure 3.3.6. Comparison of the binding of TCI, PCI and LCI to CPA. The CPA parts of crystal structures were superimposed; only the CPA moiety of the TCI-CPA complex is shown with helices and β-strands colored in dark blue and dark green, respectively. The catalytic zinc atom is represented by a yellow sphere. The inhibitors TCI (gray, light blue and red), PCI (dark red) and LCI (yellow) are represented by ribbons; their binding is shown in two different orientations (rotated by ~180º). For all inhibitors the side chain of the C-terminal residue is shown as stick model and the N-terminus is labeled.

3.3.9 Mechanism of inhibition of carboxypeptidases by TCI

Many of the features of the inhibition of carboxypeptidases by TCI characterized previously in enzymatic studies of TCI can be explained based on the three-dimensional structures reported here and comparisons with related inhibitors. The product generated after complex formation, TCI-desHis75, has the same inhibitory activity as native TCI (Arolas et al., 2005), indicating that both forms bind similarly to the enzyme. After the initial trimming, TCI-desHis75 remains intact. This is not surprising given the excellent fit of the surfaces of TCI and the carboxypeptidase that hinders any further penetration (Figure 3.3.5). A highly efficient inhibition is probably due to the abundant interactions established between the inhibitor and the enzyme, which affect or shield many residues of the carboxypeptidase that are essential for substrate binding and catalysis. The coordination of the enzyme's catalytic Zn with the carboxylate group of Leu74 also contributes to the stabilization of the complex. Taken together, numerous interactions account for the formation of a tight and stable complex.

The interaction interfaces in both complexes are nearly indistinguishable, demonstrating that TCI binds and inhibits forms A and B of carboxypeptidases by the same mechanisms. However, the position of the two TCI domains is slightly different in the two complexes suggesting that the inhibitor might adjust its structure, i.e. the position of its two domains relative to each other, for proper interaction with different carboxypeptidases. Curiously, such a behavior is reminiscent of a strategy recently introduced in drug design, called “tethering” (Erlanson et al., 2004), in which synthetic inhibitors are optimized by combining a small fragment that binds to the enzyme’s active site with an adaptable exosite ligand.

Overall, the number of enzyme-inhibitor contacts seen in the complexes solved in this work is larger than those seen in the PCI-bCPA and LCI-hCPA2 complexes (Figure 3.3.6). However, many similarities in binding mode are observed among all of them. The C-terminal tail also constitutes the primary binding site of PCI and LCI, docking into the active site of the enzyme and leading to a stopper-like inhibition (Aviles et al., 1993; Vendrell et al., 2004). The C-terminal residue of PCI and LCI (Gly39 and Glu66) also is cleaved off by the carboxypeptidase, but unlike in TCI-CP complexes, this last residue remains bound in the S1’ subsite of the enzyme’s active site (Rees and Lipscomb, 1982; Reverter et al., 2000). The number of interactions between this cleaved-off residue and the active site of the carboxypeptidase might determine its presence (PCI, LCI) or absence (TCI) after the crystallization process. In addition, in TCI-CPA/CPB Trp73 may also prevent its presence in the active site due to the larger size of the Trp side chain compared to the Tyr37 in PCI and Tyr65 in LCI.

Significantly, the overall folds of TCI, PCI and LCI in the complexes are completely different except at their C-tail, where a similar sequence accounts for an almost identical backbone fold (Figure 3.3.6). This is the only motif conserved among the carboxypeptidase inhibitors isolated from evolutionarily highly distant organisms, i.e. blood-sucking animals and *Solanacea* plants, representing a good example of convergent evolution dictated by the three-dimensional structure of the target enzymes.

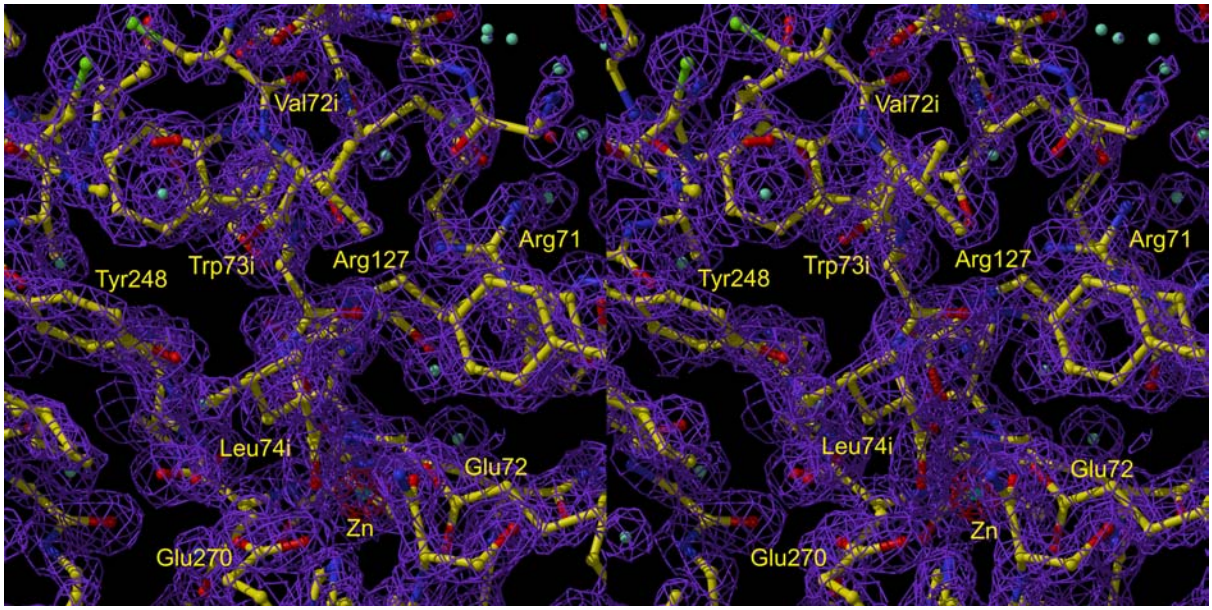


Figure 3.3.7. Stereo plot of the final ($2F_{obs}$-F_{calc}) electron density map (dark violet) of the C-terminal tail of TCI bound within the active site groove of CPA. Residues of TCI are suffixed with an i. The zinc atom is represented by a red sphere.

3.3.10 Biomedical implications

Besides their classical role in alimentary protein degradation, metallo-carboxypeptidases also participate in more selective regulatory processes like blood coagulation/fibrinolysis, inflammation, local anaphylaxis, prohormone/neuropeptide processing, and insect/vegetal attack-defense strategies, among others (Aviles et al., 1993; Reznik and Fricker, 2001). Knowledge of the control mechanisms of such enzymes as well as of the structures of carboxypeptidases and their inhibitors are essential for the rational design of improved and/or minimized drugs for biotechnological and biomedical applications.

One of the enzymes studied, bovine CPA, potentially is the natural target of TCI because *R. bursa* ticks usually infect cows. The other carboxypeptidase, human CPB, was selected not only because the three-dimensional structure of its activated form has not yet been described, but also because it is the carboxypeptidase most closely related to plasma carboxypeptidase B or TAFI with 48% of sequence identity (Figures 3.3.8 and 3.3.9). TAFI has attracted considerable interest in recent years due to its biomedical implications: Activated TAFI (TAFIa) inhibits blood fibrinolysis by removing lysine and arginine residues from partially degraded fibrin, thereby lowering its binding affinity for plasminogen and inhibiting tissue plasminogen

activator-mediated fibrin degradation (Wang et al., 1998; Bouma and Meijers, 2003). Consequently, the inhibition of TAFIa by PCI, LCI and TCI results in an enhanced tissue plasminogen activator-induced clot lysis (Nagashima et al., 2000; Walker et al., 2003; Salamanca et al., manuscript in preparation; Arolas et al., 2005). Also, the plasma concentration of TAFI is a novel parameter of medical interest, as elevated TAFI concentrations increase the risk of thrombosis and coronary artery disease (van Tilburg et al., 2000; Franco et al., 2001; Eichinger et al., 2004) and the enzyme is also thought to be involved in inflammatory processes (Myles et al., 2003; So et al., 2003; Bajzar et al., 2004).

Within this context, the three-dimensional structure of TAFI would be of great interest, but attempts to crystallize human TAFI have been unsuccessful so far, probably due to the glycosylation and limited solubility of the protein as well as its intrinsic instability (Boffa et al., 1998, Boffa et al., 2000; Marx et al., 2000). Recently, human CPB has been used to generate TAFI-CPB chimeras with enhanced stability (Marx et al., 2004). The information derived from this work thus may help to better understand the biological action of TAFI and can be useful to control this enzyme by using more specific and potent drugs, e.g. bivalent inhibitors.

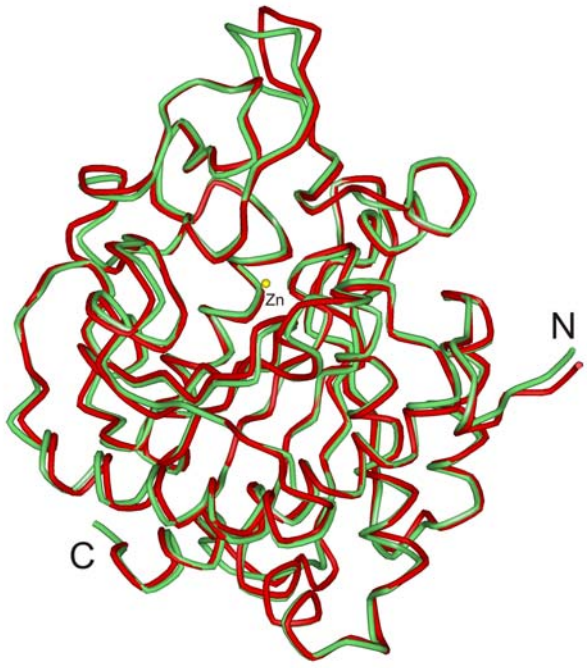


Figure 3.3.8. Structure superimposition of human pancreatic carboxypeptidase B (hCPB) and predicted human TAFIa (hTAFIa), great structural similarity is evident. hCPB is shown in light green and hTAFIa in red.

Finally, it is worth mentioning that the domains of TCI are structurally related to the β-defensin-fold family (Torres and Kuchel, 2004). This fold is found in a wide range of small disulfide-rich proteins from e.g. venoms of snakes (Nicastro et al., 2003), sea anemones and platypus. Proteins presenting this fold have numerous pharmacological activities including ion-channel inhibiting, analgesic and myonecrotic actions. Several anti-microbial defensins from vertebrates also belong to this family (Hoover et al., 2000). Because hard ticks feed for several days or even weeks with their mouthpart embedded in their vertebrate hosts, it appears possible that TCI, primarily a pro-fibrinolytic compound, could also participate in the modulation of inflammation and host defenses.

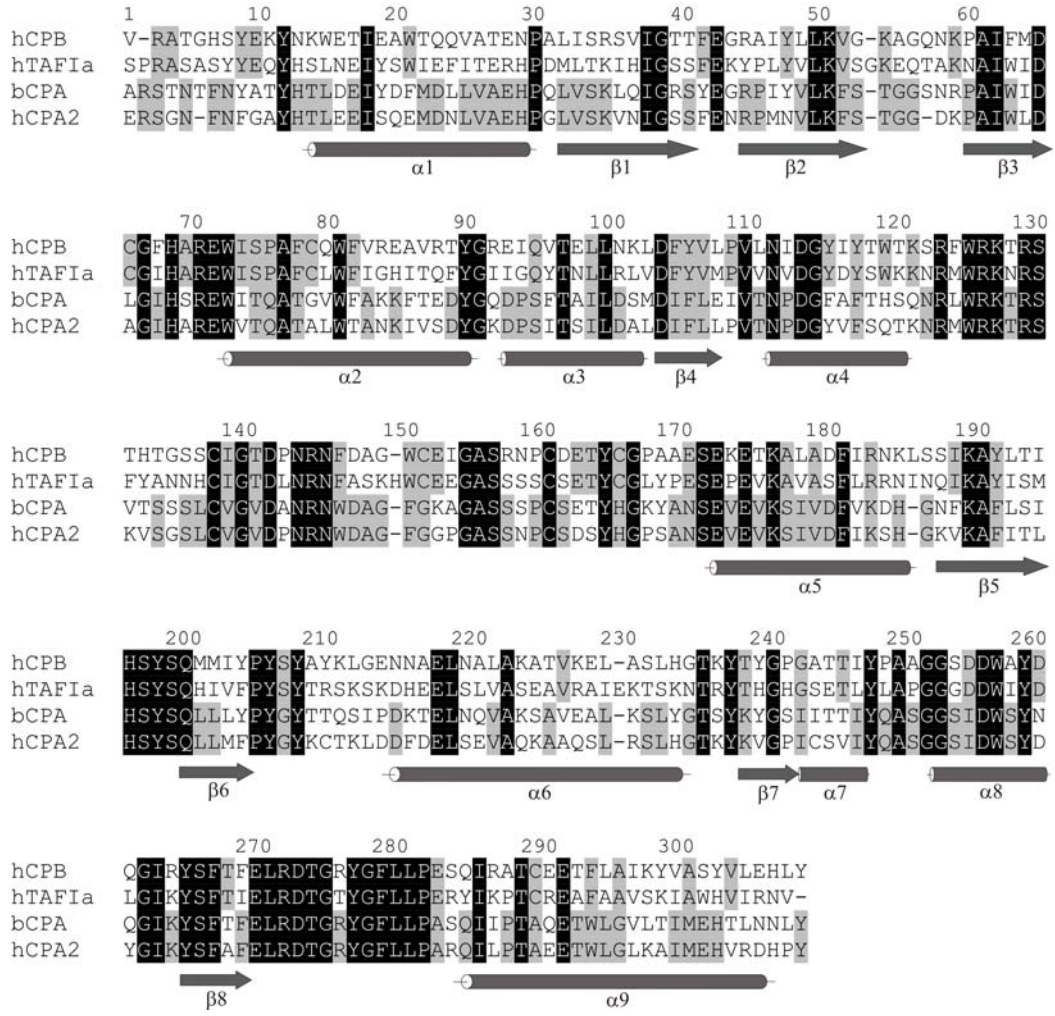


Figure 3.3.9. Amino acid sequence alignment of human pancreatic carboxypeptidase B (hCPB), human TAFIa (hTAFIa), bovine pancreatic carboxypeptidase A (bCPA), and human pancreatic carboxypeptidase A2 (hCPA2). Residues that are identical in all sequences are printed as white letters on black background. Residues identical in hCPB and hTAFIa or bCPA and hCPA2 are shaded. The helices and β-strands are indicated by cylinders and strands, respectively.

3.4 Structures of the analog of a major Intermediate in the oxidative folding of leech carboxypeptidase inhibitor (LCI)

3.4.1 Introduction

Leech carboxypeptidase inhibitor (LCI) is a cysteine-rich polypeptide of 67 residues that behaves as a tight binding inhibitor of different metallo-carboxypeptidases (Reverter et al., 1998). Assuming that leeches secrete LCI during feeding, this inhibitor seems to maintain blood in the fluid state by inhibiting plasma carboxypeptidase B, also known as thrombin-activatable fibrinolysis inhibitor (TAFI), which acts as a potent attenuator of fibrinolysis (Wang et al., 1998; Bouma and Meijers, 2003). Indeed, LCI significantly enhances the *in vitro* clot lysis induced by tissue plasminogen activator, suggesting a potential use in the prevention or treatment of thrombotic disorders (Silveira et al., 2000; Eichinger et al., 2004). The three-dimensional structure of LCI shows that it folds in a compact domain consisting of a five-stranded antiparallel β-sheet and a short α-helix (Reverter et al., 2000), with the occurrence of four disulfide bridges between cysteines 11-34, 19-43, 22-58, and 18-62, all of them located within regular secondary structure elements.

This protein folding proceeds through a series of intermediates that define the folding landscape from the unfolded polypeptide to the native structure (Dill and Chan, 1997; Honig, 1999). Understanding the fundamental relationship between the amino acid sequence and the three-dimensional structure of the native protein requires structural analysis of those folding intermediates. Although characterizing the intermediates is usually a difficult task due to their short half-life, studies of proteins stabilized by disulfide bonds have provided considerable insight into the field of protein folding (Creighton et al., 1996). In disulfide-rich proteins, the coupling of the covalent chemistry of disulfide bond formation to the non-covalent folding processes makes it feasible to trap, isolate and characterize their intermediates (Creighton et al., 1986). However, the role and significance of many structured intermediates that accumulate along folding processes are still largely unknown.

The III-B intermediate has been determined by X-ray crystallography, and III-A was studied in solution by NMR.

3.4.2 Structure of III-B intermediate

3.4.2.1 Protein expression and purification

The synthetic gene for LCI (Reverter et al., 1998) was cloned into the pBAT4 plasmid (Peranen et al., 1996), fused in frame to the OmpA signal sequence for extracellular expression. The C19A/C43A analog was constructed using a two steps PCR method for site-directed mutagenesis. All constructs were verified by DNA sequencing. Wild-type and C19A/C43A LCI were obtained by heterologous expression in *Escherichia coli* strain BL21(DE3) using M9CAS medium containing 0.5% glycerol. Proteins were purified from the culture medium using a Sep Pak C_{18} cartridge (Waters), followed by anion-exchange chromatography on a TSK-DEAE 5PW column (Tosohaas), and by RP-HPLC on a 4.6 mm Protein C4 column (Vydac). Protein identity and purity (>98%) were confirmed by MALDI-TOF MS on a Bruker Ultraflex spectrometer and automatic Edman degradation using a Precise 492 Protein Sequencer (Applied Biosystems), respectively.

3.4.2.2 Crystallization and structure determination

C19A/C43A LCI was crystallized in complex with bovine CPA. The complex was obtained by mixing both proteins in Tris-HCl buffer (50 mM, pH 7.5) containing NaCl (100 mM). The analog was added until complete inhibition of the enzyme was reached. The complex was maintained for 2 hours at 20ºC and purified by gel-filtration chromatography on a Superdex 75 HiLoad 26/60 column (Amersham Biosciences) using the same buffer, and further concentrated to about 10-12 mg/ml. Suitable crystals for data collection were obtained at 20ºC using the sitting-drop vapor-diffusion method in a few weeks growing period. Drops were obtained by mixing equal volumes of protein solution (C19A/C43A LCI-CPA) and reservoir buffer containing Lithium Sulfate monohydrate (1.5 M) and Tris (100 mM, pH 8.5).

A dataset up to 2.8 Å was collected on the MPG/GBF beamline BW6 at DESY, Hamburg (Germany). Diffraction data for the structure refinement was collected at 90K. Diffraction images were taken on MARCCD Detectors. The summary of the data collection is shown in Tables 3.4.1 and 3.4.2. The collected data were integrated, scaled and merged by XDS and XSCALE programs (Kabsch, 1993). The structure was determined by molecular replacement using the Molrep program from the CCP4 suite(CCP4, 1994). The structure of the LCI-CPA complex taken from the PDB entry 1DTD (Reverter et al., 2000) was used as a probe after removing the

inhibitor part. The initial R-factor of the model was 0.46. The model was then refined by Refmac5 .(CCP4, 1994) and rebuilt by XtalViev/Xfit (McRee, 1999) and by a subsequent Refmac5 refinement. Waters were added by Arp/warp (Lamzin and Wilson 1993). The final R crystallographic factor was 0.19 and Rfree 0.23. The asymmetric unit contains two complexes of C19A/C43A LCI- CPA and two additional C19A/C43A LCI molecules imprisoned in crystal lattice. The unbound inhibitor molecules display significant flexibility in the loop regions and therefore their models are not completely built because of the lack of interpretable electron density. Most of the other molecules had a clear and interpretable electron density. However, the loop region between Val132-Ser136 is missing on the map in both CPA molecules. There are also solvent-exposed side chains with missing density; these parts were omitted in the final model.

Table 3.4.1. Data collection statistics for CPA-LCI complex crystal. Dataset measured at MPG/GBF beamline BW6 at DESY using MARCCD detector.

SUBSET OF INTENSITY DATA WITH SIGNAL/NOISE >= 2.0 AS FUNCTION OF RESOLUTION

RESOLUTION LIMIT	NUMBER OF REFLECTIONS OBSERVED	UNIQUE	POSSIBLE	COMPLETENESS OF DATA	R-FACTOR observed	R-FACTOR expected	COMPARED	I/SIGMA	R-meas	Rmrgd-F	S_norm/ S_ano
20.00	637	130	156	83.3%	2.8%	3.1%	636	43.07	3.1%	1.9%	1.18
10.00	5770	1104	1116	98.9%	2.7%	3.2%	5768	43.79	3.0%	1.8%	1.06
6.00	23689	4517	4586	98.5%	3.6%	4.0%	23686	33.49	4.0%	2.8%	1.05
5.00	21762	4150	4248	97.7%	4.6%	4.7%	21761	29.17	5.1%	3.4%	1.03
4.00	49652	9497	9660	98.3%	3.9%	4.3%	49649	31.99	4.3%	2.9%	1.02
3.00	132606	25485	27074	94.1%	6.6%	7.0%	132601	20.50	7.4%	5.4%	1.01
2.70	74355	14647	17412	84.1%	15.2%	15.6%	74353	10.49	17.0%	11.3%	1.01
2.60	28158	5734	7675	74.7%	20.6%	21.3%	28157	7.96	23.2%	14.4%	1.01
2.50	32254	6561	9004	72.9%	24.1%	24.4%	32245	7.06	27.0%	16.6%	1.01
2.40	35349	7355	10518	69.9%	27.7%	28.4%	35335	6.15	31.3%	19.0%	1.00
2.30	37596	7950	12444	63.9%	32.1%	32.3%	37577	5.28	36.2%	21.3%	1.01
2.20	37780	8113	14861	54.6%	37.2%	36.6%	37733	4.61	42.1%	24.2%	1.00
2.10	13935	3206	17775	18.0%	43.4%	41.2%	13871	3.97	49.5%	28.8%	1.01
total	493543	98449	136529	72.1%	8.2%	8.5%	493372	15.60	9.1%	8.7%	1.01

Table 3.4.2 Data collection and refinement statistics

Data collection	
Space group	$P4_32_12$
Cell constants (Å)	a=124.93
	b=124.93
	c=154.90
Resolution range (Å)	36-2.2
Wavelength (Å)	1.05
Observed reflections	493543
Unique reflections	98449
Whole resolution range:	
Completeness (%)	97.7
R_{merge}	2.9
I/σ(I)	31.99
Last resolution shell:	
Resolution range (Å)	2.2-2.3
Completeness (%)	63.9
R_{merge}	21.2
I/σ(I)	5.28
Refinement	
No. of reflections	66228
Resolution (Å)	30-2.2
R-factor (%)	18.9
R_{free} (%)	23.3
Average B (Å^2)	28.2
R.m.s bond lenght (Å)	0.012
R.m.s. angles (°)	1.48
Content of asymmetric unit	
RMSD of complexes (Å)	0.43
No. of protein molecules	6
No. of protein residues/atoms	835/6966
No. of solvent/Zn atoms	538/2

3.4.2.3 Crystal structure of C19A/C43A LCI

Both free and bound analog forms were crystallized together; each asymmetric unit in the crystal contains two complexes of C19A/C43A LCI-CPA and two free C19A/C43A LCI molecules. The C19A/C43A LCI structure determined at 2.2 Å resolution shows that this analog is very similar to the wild type form (Figure 3.4.1A).(Reverter et al., 2000) Its structure consists of a five-stranded antiparallel b-sheet with a β3-β1-β2-β5- β4 topology, and a short a-helix that packs into the most compact part of the β-structure 11 (Figure 3.4.1B). This antiparallel b-sheet involves residues Glu7-Gln13 (β1), Gln16-Arg23 (β2), Glu33-His37 (β3), Val51-Tyr53 (β4) and Gly 56-Ile63 (β5), while the short a- helix is located between residues Pro41 and Gly46. The disulfide pairings of this mutant, as determined from its three-dimensional structure calculations, are Cys11- Cys34, Cys18-Cys62, and Cys22-Cys58, and completely agree with the results previously obtained by digestion with thermolysin. The structure of the analog upon interaction with CPA greatly resembles that of the wt form. The root mean square (r.m.s.) deviation between the bound analog and bound wt form (0.42 Å for backbone) is lower than the rms deviation between the bound and free forms of the analog (0.91 Å for backbone). Direct comparison of temperature factors (B-factors) clearly shows that the analog is stabilized upon complex formation, with average B-value of 29 $Å^2$ for the bound form and 40 $Å^2$ for the free molecule. It is worth to mentioning that the free form of the mutant shows a higher flexibility in the Pro38-Trp50 region and displays a shifted a-helix (1.8 Å) out of the molecule center when compared to the bound state. C19A/C43A LCI interacts extensively with the carboxypeptidase (Figure 3.4.2). The occlusion of the C-terminus of this analog in the active site groove of the carboxypeptidase constitutes the “primary binding region”. The location of binding subsites S1, S2 and S3 on the enzyme active site may be identified from the binding interactions of the C-terminal residues of the inhibitor (Val66, Tyr65, Pro64) with this region (Figure 3.4.2). The interactions formed between the residues of the analog C-tail and the CPA active site are nearly indistinguishable from those found in the wt LCI-CPA2 complex(Reverter et al., 2000). However, unlike in LCI-CPA2 complex, in C19A/C43A LCI-CPA the residue Glu67 is cleaved off and does not remain in the active site of the enzyme. The “secondary contact region”, established by

interactions between the carboxypeptidase surface and the neighboring residues of the mutant, is also highly similar to that of the wt LCI-CPA2 complex.

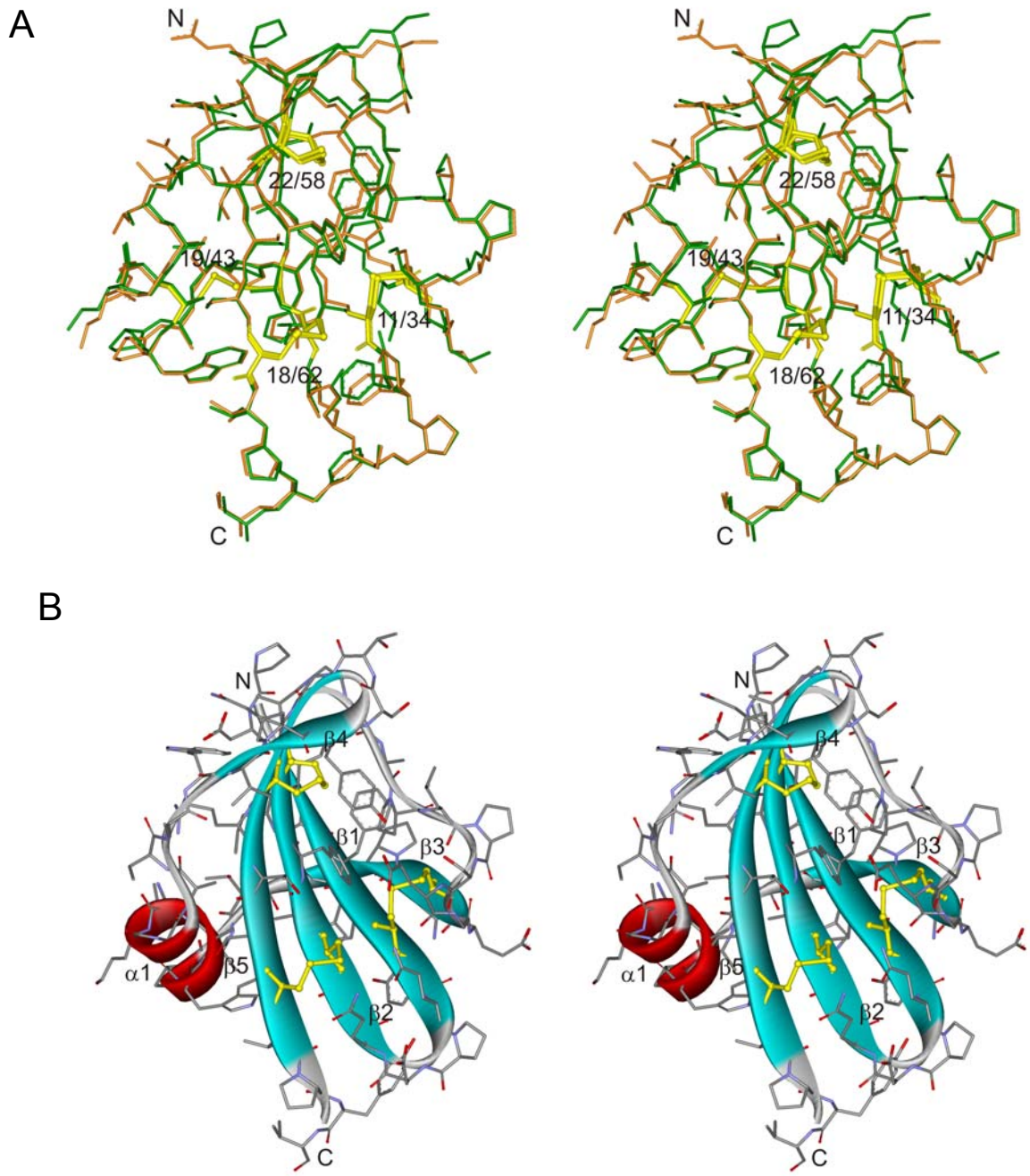


Figure 3.4.1. Three-dimensional structure of C19A/C43A LCI. **A**. Stereo view of the overlapping between the backbone atoms from C19A/C43A (green) and wt LCI (orange). The disulfide pairings of each protein are shown yellow in the structure. **B**. Stereo representation of the C19A/C43A LCI ribbon. The helix and β-strands (β1-β5) are represented in red and light blue, respectively. The three disulfide bridges of this mutant (Cys11-Cys34, Cys22-Cys58 and Cys18-Cys62) are represented in yellow. *N* and *C* indicate the location of N- and C-terminal tails of C19A/C43A LCI.

3.4.2.4 Conclusion

The structure of wt LCI consists of a five-stranded antiparallel b-sheet and one short a-helix.16 The intermediate III-B has two free cysteines, Cys19 and Cys43, which connect the a-helix and the b-sheet in the native structure of LCI, suggesting that its structure might display a disconnected a-helical section and a main core of a five antiparallel b-sheet stabilized by the remaining three native disulfide bonds. The fact that this intermediate elutes in the RP-HPLC very close to other species precludes its homogeneous purification and further precise structural characterization; therefore an analog was constructed by replacing the two free Cys with Ala (C19A/C43A LCI). The X-ray structure of C19A/C43A LCI alone and in complex with its target, carboxypeptidase A, shows that its overall chain fold is very similar to that of wt LCI. The comparison between wt LCI-CPA2 complex and C19A/C43A LCI-CPA complex reveals that despite the missing disulfide bond in the analog both, backbone and side chains adopt a conformation that closely resembles that in the wt form. This includes those residues close to the carboxypeptidase or near the mutation point, with an rms deviation between both molecules of 0.42 Å for the backbone. Despite the similarity of both crystal structures, close examination reveals that the B-factor of the atoms in the side chains of the analog near the place of mutation is clearly higher than those in the wt structure. The B-factor characterizes not only thermal motion but also the lack of static ordering in the crystal, and could be indicative of net gain in flexibility in the regions spatially adjacent to the mutation sites. This increase in flexibility around the mutation points becomes more obvious in the crystal structure of the uncomplexed analog form. Although the backbone of the bound wt LCI and unbound analog are very similar, the electronic densities of the latter are absent in the residues conforming the last turn of the a-helix (residues 45-47), clearly indicating a higher flexibility in this region. This result is sharply coincident with our observation that in the solution structure of III-A intermediate, the other major kinetic intermediate in LCI folding reaction, the absence of the fourth disulfide bond results in a native-like structure with an increase of backbone flexibility around the free cysteines (Cys22 and Cys58). According to our data, the docking of the analog to the carboxypeptidase turns into a reduction of conformational flexibility and thus in a net gain of structural specificity. Overall, the fourth disulfide provides LCI with conformational stability as well as reduced flexibility, that is, increased structural specificity. These qualities appear to be highly desirable for its function *in vivo* since

LCI is a protease inhibitor from leech saliva evolved to act in blood. Lower conformational stability or higher backbone fluctuation in a 3-disulfide version of this molecule would probably render a protein more susceptible to proteolytic attacks. The advances made in understanding the stability and folding behavior of LCI have provided a greater insight into the nature of this protein and constitute a basis for the development of variants of this molecule with enhanced activity and/or stability. This is of great interest given that carboxypeptidase inhibitors such as LCI or PCI markedly accelerate the tissue plasminogen activator induced lysis of human plasma clots by modulating TAFI activity, (Nagashima et al., 2000; Walker et al., 2003) and may be used as lead compounds for the optimization of thrombolytic therapies.

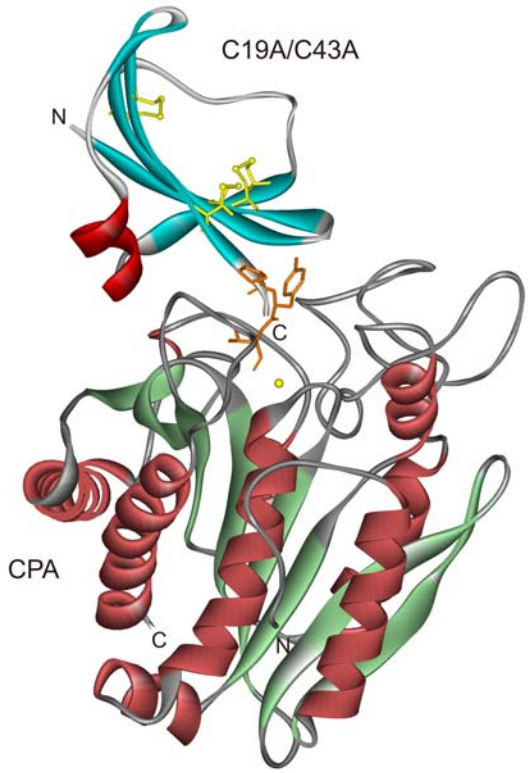


Figure 3.4.2. Ribbon plot of the complex formed between C19A/C43A LCI and CPA. The helix and β-strands of C19A/C43A LCI are shown in red and light blue, respectively, and the disulfide bridges are in yellow. The helices and β-strands of CPA are in dark red and dark green, respectively. The side chains of C19A/C43A LCI residues involved in the interaction with CPA are explicitly shown in orange. The zinc atom of CPA is represented by a yellow sphere. The N-terminal and C-terminal domains of the mutant and the enzyme are depicted.

3.4.3 NMR studies on the III-A folding intermediate of leech carboxypeptidase inhibitor

We also carried out NMR structural analysis of the III-A folding intermediate of LCI. This intermediate has been directly purified from the oxidative folding reaction using RP-HPLC, and its structure characterized by NMR and compared to that of native LCI.

3.4.3.1 Protein preparation

The synthetic gene for LCI (Reverter et al., 1998) was cloned into the pBAT4 plasmid (Peranen et al., 1996), fused in frame to the OmpA signal sequence. Recombinant ^{15}N-labeled LCI was obtained by heterologous expression in *Escherichia coli* strain TG1. Cells were grown in M9 media containing ^{15}NH4Cl as the only nitrogen source and 0.1 mg/mL carbenicillin. This second preculture was continued overnight, and the cells contained in 10 mL were used to inoculate 1 L of the same minimal media. Protein expression was induced in late phase (OD600 = 1.0) by adding IPTG (1 mM final concentration). LCI was purified from supernatant as described (Reverter et al., 1998). In summary, the protein was initially purified using a Sep-Pak C18 Cartridge (Waters), followed by anion-exchange chromatography on a TSK-DEAE 5PW column (Tosohaas), and by RP-HPLC on a 4.6 mm Protein C4 column (Vydac). The 15N LCI labeling was almost heterogeneous (>99%) as deduced by MALDI-TOF MS analysis on a Bruker Ultraflex spectrometer. Briefly, native ^{15}N-labeled LCI was reduced and denatured in 0.1 M Tris-HCl (pH 8.4) containing 8 M guanidine hydrochloride and 50 mM dithiothreitol, at 22ºC for 2 h. To initiate folding, the sample was passed through a PD-10 column (Sephadex-25, Amersham Biosciences), previously equilibrated with 0.1 M Tris-HCl (pH 8.4). Reduced and denatured LCI was recovered and immediately diluted to a final protein concentration of 0.5 mg/ml in the same Tris-HCl buffer. Folding intermediates of LCI were trapped after approximately 8 h of refolding by 22 mixing aliquots of the sample with 2% trifluoroacetic acid (TFA). The trapped III-A intermediate was purified by RP-HPLC using the following conditions: solvent A was water containing 0.1% TFA and solvent B acetonitrile containing 0.1% TFA. A linear 20-40% gradient of solvent B was applied over 50 min, with a flow rate of 0.75 ml/min. The column used was a 4.6 mm Protein C_4 (Vydac).

3.4.3.2 NMR experiments and structure calculation

Protein samples for NMR experiments were prepared by dissolving lyophilized ^{15}N LCI and ^{15}N III-A in either H2O/D2O (9:1 ratio by volume) or D2O, at a concentration of 1 mM and pH 3.5. All experiments were carried out at 600 MHz on a Bruker DRX-600 spectrometer, at 27ºC. The spectrometer was equipped with a triple resonance, triple gradient probe head. The TOCSY experiments (Rance, 1987) were performed with different mixing times between 20 and 40 ms, while the NOESY experiments (Kumar et al., 1980) were carried out with a mixing time of 120 ms. 4096 complex data points were recorded in the time domain t2 and 700 increments in the t1 domain. Water suppression was achieved using the WATERGATE pulse sequence (Piotto et al., 1992). The ^{1}H-^{15}N HSQC spectra (Bodenhausen and Ruben, 1980) were also recorded at the same temperature with 2048 complex data points in the t2 domain and 128 points in the t1 domain, with 256 scans. The 3D NOESY-HSQC spectra (Marion et al., 1989) were performed with a mixing time of 100 ms, and 4096 complex data points were recorded in the t3 domain. For the amide proton exchange experiments lyophilized samples of 15N native LCI and III-A were dissolved in D2O at pH 3.5, 27ºC. A series of consecutive 2D heteronuclear ^{1}H-^{15}N HSQC experiments were acquired with increased delays for up to 3 days. 23 The collected spectra were processed by using the standard XWinNMR software package of Bruker and analyzed with the SPARKY software (Goddard and Kneller, 2000). Chemical shifts were assigned applying a combination of TOCSY/NOESY techniques (Wüthrich, 1986). Peaks were classified according to their intensities as weak (3.8-5 Å), medium (2.8-3.8 Å), and strong (2.0-2.8 Å). A total of 20 structures were calculated by the simulated-annealing method with the program CNS (Brünger et al., 1998). Structure calculations were carried out essentially according to the basic protocol described previously (Holak et al., 1989). For the final refinement, NOE tables were supplemented with constraints for several hydrogen bonds identified from the determined secondary structure. Parameters of the resulted 20 energy minimized conformers for N LCI and III-A intermediate are summarized in Table 3.4.2. The atomic coordinates of the bundle of 20 conformers of native LCI and III-A determined at pH 3.5 have been deposited in the Protein Data Bank, accession codes 1ZFI and 1ZFL, respectively.

Table 3.4.2. Statistics for NMR structures calculation

	Native	**Intermediate**
Distance restraints		
Total NOE distance restraints	250	259
Short range	67	62
Medium range	168	184
Long range	15	13
Hydrogen bond restraints	30	18
Violations >0.5A	0	0
Ramachandran analysis [%]		
Residues in favored regions	60.5	48.5
Residues in allowed regions	33.2	40.2
Residues in generously allowed regions	5.4	9.5
Residues in disallowed regions	0.9	1.8
RMSD from ideal geometry		
Bond lengths [A]	0.0063	0.006
Bond angles [°]	0.61	0.59
Average pairwise rms deviations [A]	1.4	1.4

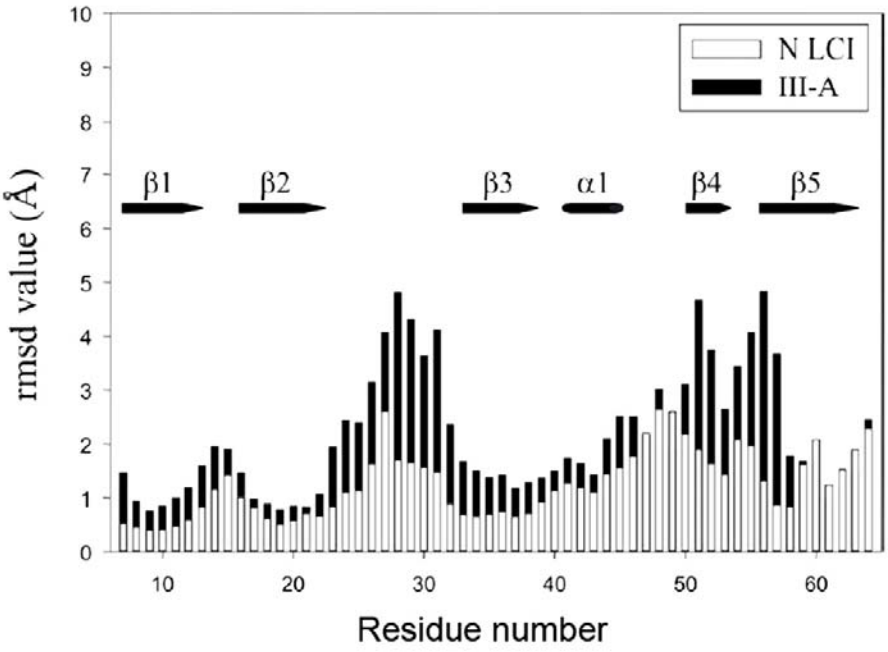


Figure 3.4.3. Comparison of local rms deviation values for backbone atoms of native LCI and III-A. The rms deviations of the backbone atoms from the 20 calculated structures are plotted vs. the residue number (residues 6-64). Secondary structure elements for native LCI are indicated inside the graphic.

3.4.3.3 Three-dimensional structure calculations

The structures of native LCI and III-A intermediate were calculated using the simulated annealing method with the program CNS (Brünger et al., 1998). For native LCI, with the exception of the five N-terminal and the three C-terminal residues, the ensemble of 20 calculated structures is well defined (Figure 3.4.4), with an average backbone rms deviation of 1.40 Å (residues 6-64). Figure 3.4.3 indicates the rms deviation of each residue in the bundle of 20 structures and clearly shows the most flexible regions in native LCI. The three-dimensional structure of this molecule at pH 3.5 is very similar to that calculated previously at pH 6.5 (Reverter et al., 2000): a five-stranded antiparallel β-sheet with a β3-β1-β2-β5-β4 topology, and a short α-helix that packs onto the most compact part of the β-structure interacting with the end and the beginning of the β1 and β2 strands, respectively (Figures 3.4.4 and 3.4.5). A high percentage of residues belong to regular secondary structure elements (nearly 45%), which are cross connected and stabilized by the presence of four disulfide bridges: Cys11-Cys34 (β1- β3), Cys18-Cys62 (β2-β5), Cys19-Cys43 (β2-α1), and Cys22-Cys58 (β2-β5). They provide high stability and compactness to the protein. The calculation of the structure of the III-A intermediate confirms that it also possesses a well-defined globular conformation that includes a four-stranded antiparallel β-sheet with a β3-β1-β2-β5' topology. However, some parts of this molecule are very flexible. This is shown by the higher average backbone rms deviation value for residues 6-64 of the 20 calculated structures: 2.47 Å. The rms deviation of each residue is shown in Figure 3.4.3 and provides evidence for the presence of highly flexible regions, mainly between residues Arg23-Gly32 and Arg44-Gln57. The disulfide pairings of III-A were unambiguously determined during three-dimensional structure calculations and were in complete agreement with a previous assignment carried out by digestion of the vinylpyridine-derivatized intermediate with thermolysin and analysis of the resulting disulfide-containing peptides by MALDI-TOF MS and automated Edman degradation (Arolas et al., 2004): Cys11-Cys34 (β1-β3), Cys18-Cys62 (β2-β5'), and 12 Cys19-Cys43 (β2-"α"). The missing disulfide bond established in the native form between Cys22 and Cys58 seems to account for the lower compactness of such intermediate as compared to that of the native protein. However, the structural similarities between the native protein and the III-A intermediate are striking.

3.4.3.4 The role of the III-A intermediate in the folding pathway of LCI

Based on X-ray study of III-B intermediate it was hypothesized that the way the oxidative folding pathway of LCI proceeds depends on the ability of its secondary structure elements to protect progressively native disulfide bonds from rearrangement in the interior of a totally or partially folded structure (Arolas et al., 2004). This would result in a final native structure in which disulfide bonds should be highly protected. Our view is strongly reinforced here by the results of amide proton exchange experiments on native LCI and III-A intermediate. Both forms were dissolved in D2O at pH 3.5 and several ^{1}H-^{15}N HSQC spectra were recorded over time. For native LCI, maximum exchange was achieved after approximately 10 h and the spectra did not significantly change after that time point. The following residues were found protected: Phe9-Gln13 (β1), Gln16- Arg23 (β2), Asn35 and His37 (β3), Cys43 (α1), Val51 (β4), Cys58-Arg59, Thr61-Ile63 and Tyr65 (β5). Thus, the protected residues are located within all secondary structure elements around the cysteine residues and this fact clearly indicates that in native LCI the four disulfide bonds are buried and not solvent-accessible. For the III-A intermediate, after 15 min of exchange residues from all secondary structure elements are found protected. However, as expected, residues located in the highly flexible regions are quickly exchanged. The protected residues were Leu10-Gln13 (β1), Gln16-Cys19 and Ile21-Cys22 (β2), Asn35 (β3), Thr60 and Ile63 (β5'). Thus, in III-A both free cysteines (Cys 22 and 58) and the three disulfide bonds are located in protected regions or close to them and therefore are not solvent-accessible or have limited accessibility. The final number of protected residues for native LCI and the intermediate (23 and 13 residues, respectively) is in good agreement with those previously found measuring the global D/H exchange by MALDI-TOF MS (26 and 16 residues, respectively; Arolas et al., 2004), taking into account that Pro residues do not appear in the ^{1}H-^{15}N HSQC spectra and some of them are closely located to the protected residues in both forms. This approximately 40% of decrease in protected residues between native LCI and III-A, mainly localized in the "missing" secondary structure elements of the latter (α1, β4 and β5'), is a reflection of its lower level of conformational packing.

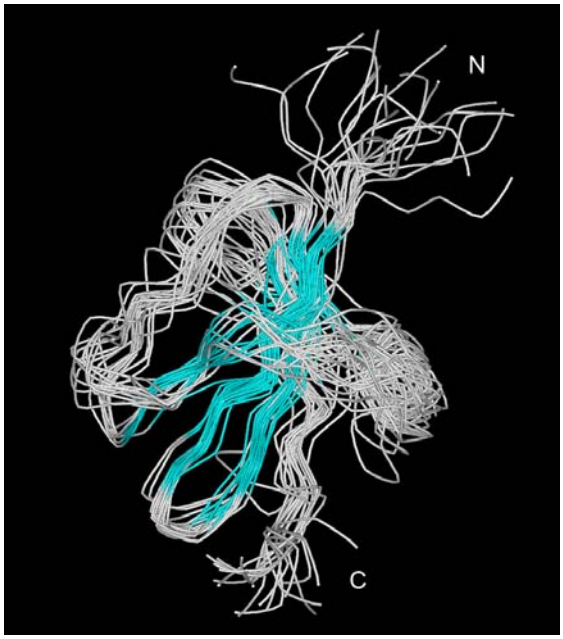


Figure 3.4.4. Stereoview with the superposition of the calculated structures for native LCI. The N- and C-terminal tails are labeled.

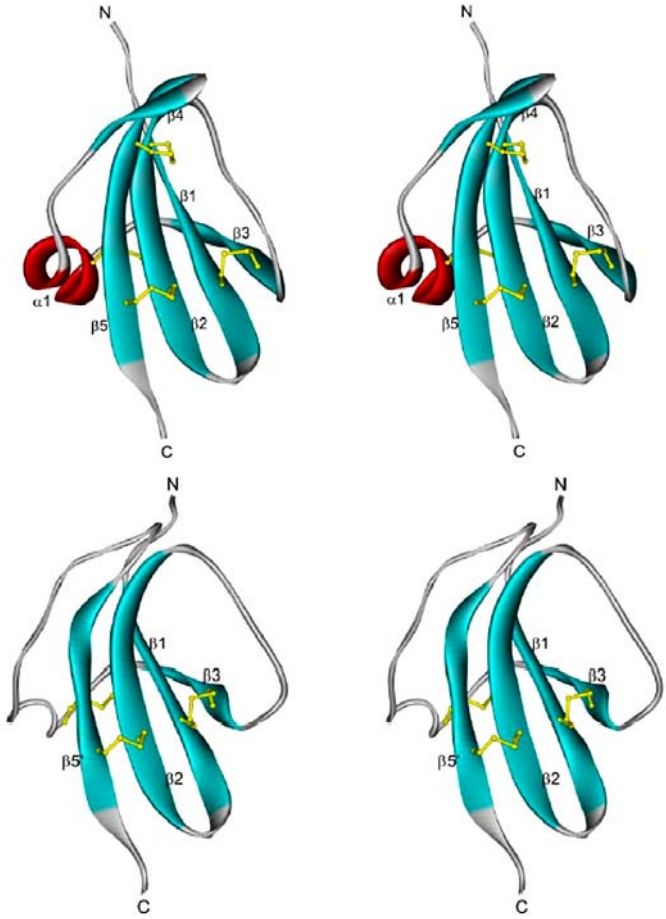


Figure 3.4.5. Stereoview of the structure of native LCI (A) and III-A intermediate (B). Ribbon representation of the calculated structure for native LCI and III-A. β-strands are shown in blue and the α-helix in red. *N* and *C* indicate the location of the N- and C-terminal tails of both proteins. The disulfide bonds are shown in yellow.

3.4.3.5 Conclusion

III-A constitutes the major kinetic trap along the oxidative folding of LCI. Despite the fact that this intermediate lacks a native disulfide bond, we show that it is a highly structured molecule with striking structural similitude to the native state. Comparison of native and intermediate structures allows deciphering why III-A accumulates along the folding reaction: it acts as a *disulfide insecure* intermediate, which protects both their native disulfide bridges and free cysteine residues from rearrangement and direct oxidation, respectively, in the interior of a highly folded protein conformation. Although III-A is a fully functional form that is formed quickly and efficiently along the LCI folding pathway, a conformational search for the formation of the last disulfide bond takes place while losing most of the tertiary structure already gained in the intermediate. The results of this study together with previous stability data (Arolas et al., 2004) demonstrate that the fourth disulfide bond restricts conformational flexibility allowing a net gain in stability and structural specificity to the native form. This assumption makes sense taking into account that LCI is a protease inhibitor from leech saliva evolved to act in blood. In addition, here we show that theoretical approximations based on topological constrains predict accurately the main characteristics of the folding pathway of LCI and those of other proteins such as BPTI or PCI. The overall data provide direct evidence for the importance of native-like interactions between elements of secondary structure in directing the folding of disulfide-rich proteins, which is already assumed in other studies.

4. Summary

The focus of this thesis was on the structural characterization of three groups of proteins: the cytoskeleton related proteins, insulin-like growth factor binding proteins and proteases.

For cytoskeletal proteins, the structures of actin cross-linking protein, filamin (FLN) from *Dictyostelium discoideum,* and the cyclase-associated protein (CAP) were solved by X-ray crystallography and NMR methods, respectively. The structure of the filamin rod domain (repeats 4, 5 and 6) reveals an extended rod configuration of limited flexibility. This structure indicates that a full-length rod domain should be able to cross-link actin filaments over a distance of about 400 Å. The rod domain is built by six repeats, each having an immunoglobulin-like fold; these are interconnected by short linkers and stabilized by salt bridges. The last C-terminal repeats of this domain are involved in dimerization. Homodimerization is essential for the function of filamin. In continuation of the work described in this thesis, structures of larger fragments of filamin are planned to be studied in the future. The structure of an N-terminal domain of the *Dictyostelium* cyclase associated protein (CAP-N) was studied by NMR. This protein is important for downregulation of actin polymerization and cytoskeletal response to cell signaling. The NMR structure appears to be an all-α-helix bundle. Detailed analysis of flexible parts of CAP-N and comparison to X-ray model were performed.

The second aim of this thesis was on the structural characterization of insulin-like growth factor (IGF) binding proteins (IGFBPs). Two structures of the IGF binding fragments of IGFBPs were solved in complex with IGF, with the resolutions of 1.6 Å and 2.5 Å, revealing the presence of an unusual disulphide-ladder subdomain and explaining the mechanism of binding and inhibition of IGF. Additionally, a model of the C-terminal domain of IGFBPs and its interactions with IGF and N-terminal IGFBP was build based crystallographic data, which were not however sufficient to build an atomic resolution ternary complex structure.

Finally, structures of serine protease SplC from *Staphylococcus aureus* and metallo-carboxypeptidases (CPA and CPB) in complex with tick and leech inhibitors were resolved. The SplC serine protease is considered as potential virulence factor of this bacterium. This structure, solved by remote-SAD phasing, represent serine protease fold with an unusual autoinhibition mechanism of blocking the catalytic

histidine by an adjacent flexible loop. Also, the staphostatin A protein structure of this bacterium was analyzed by NMR.

Additionally, a structure of tick carboxypeptidase inhibitor (TCI) in complex with CPA and CPB was studied. The novel structure of TCI reveals interesting double-headed binding mechanism allowing the molecule to adapt to different carboxypeptidases. The structure of CPB is assumed to be the closest analog of the human TAFI carboxypeptidase, a very important target for drug design. Leach carboxypeptidase inhibitor (LCI) folding intermediates in complex with CPA and in free form were also characterized by X-ray crystallography and NMR experiments.

5. Zusammenfassung

Im Mittelpunkt dieser Doktorarbeit stehen die strukturellen Untersuchungen von drei Proteingruppen: die zytoskelettanverwandten Proteine, insulinähnliche wachstumfaktorbindenden Proteine und Proteasen.

Die Strukturen von Aktin bindenden Proteinen, Filamin (FLN) aus *Dictyostelium discoideum* und Adenylyl Cyclase assoziertem Protein (CAP) wurden mit Hilfe von Röntgenkrystallographie (X-ray) und der Kernresonanzspektroskopie (NMR) bestimmt.

Die Struktur von der Filamin Rod Domäne (Einheiten 4, 5 und 6) sagt etwas über die erweiterte Konfiguration mit einer beschränkten Flexibilität aus. Die Struktur deutet darauf hin, dass die volle Länge von der Rod Domäne in der Lage sein soll, die Wechselwirkung über der Distanz von ungefähr 400 Å mit Aktin Filaments durchführen zu können. Die Rod Domäne besteht aus sechs Wiederholungen. Jede der Wiederholungen hat eine immunoglobinähnliche Faltung, die mit der kleineren Verbindung angeschlossen und durch Salzbrücken stabiliziert ist. Die letzte C-terminale Wiederholung von dieser Domäne ist bei der Dimerisierung beteiligt. Die Homodimerisierung wird für die Funktion von Filamin benötigt. Als Fortsetzung, von den hier beschriebenen Untersuchungen, werden demnächst die Strukturen von den längeren Konstrukten von Filamin bearbeitet.

Die Struktur von der N- terminalen Domäne von Adenylyl Cyclase-assoziertem Protein (CAP-N) wurde mit Hilfe NMR studiert. Dieses Protein reguliert Aktins-polymerisation als Antwort auf zellulare Signale. Die NMR Struktur von CAP-N zeigt ein α-helikales Bündel aus 6 antiparallelen Helices. Die detallierten Studien von den flexiblen Fragmenten von CAP-N sowie auch ein Vergleich der NMR Struktur mit der Kristallstrukur sind ebenfalls durchgeführt worden.

Im zweitem Teil der vorliegenden Dissertation wurden die strukturellen Wechselwirkungen von insulinähnlichen wachstumsfaktorbindenden Proteinen (IGFBPs) beschrieben. Es wurden zwei Strukturen von den insulinähnlichen Wachstumfaktoren (IGF's) von IGFBPs im Komplex mit IGF, mit der Auflüsung von 1.6 Å und 2.5 Å, bestimmt. Sowohl die Anwesenheit der ungewönlichen Disulfidleitern als auch des Bindungs- und Inhibitionsmechanismus von IGF sind in der beschriebenen Struktur aufgedeckt. Zusätzlich wurde ein Model von der C-terminale Domäne von IGFPs mit IGF und mit der N-terminalen Domäne von IGFBP als dreifacher Komplex vorgeschlagen.

Zusätzlich wurden die Kristallstrukturen von der Serinprotease SpIC aus *Staphylococcus aureus* und von der Metallocarboxipeptidase (CPA und CPB) im Komplex mit „tick“und „leech“ Inhibitoren bestimmt. Die SpIC Serinprotease wird als eine potenzielle Bösartigkeit der Bakterie *Staphylococcus aureus* angesehen. Die vorgestellte Struktur ist mit SAD (single-wavelenght anomalous dispersion) Phasing gelöst und stellt die Serinprotease Faltung mit einem ungewöhnlichen – Histidin blockierenden Mechanismus dar. Die Struktur von Staphostatin A Protein ist mit der Hilfe von NMR analysiert worden.

Letztlich wurde die Struktur von Tick-carboxipeptidase Inhibitor (TCI) im Komplex mit CPA und CPB untersucht. Diese neuartige Struktur von TCI zeigt einen sehr interessanten Bindungmechanismus, der eine Anpassung von dem Inhibitor an verschiedenen Carboxipeptidasen ermöglicht. Die Struktur von CPB ist übereinstimmend zu menschlichen Carboxipeptidase TAFI und hat dadurch für die Entwicklung von den Wirkstoffen eine sehr grosse Bedeutung. Leach Carboxipeptidasen Inhibitor (LCI) Faltungszwischenstadien sind sowohl im Komplex mit CPA als auch in der ungebundenen Form mit Hilfe von X-ray und NMR analysiert worden.

6. Abbreviations

- 1D — one-dimensional
- 2D — two-dimensional
- Å — Ångstrøm (10^{-10} m)
- aa — amino acid
- ABD — actin binding domain
- CAP — cyclase associated protein
- cDNA — complimentary DNA
- COSY — correlation spectroscopy
- CP — carboxypeptidase
- Da — Dalton (g mol-1)
- dd — *Dictyostelium discoideum*
- FLN — filamin
- g — gravity (9.81 m s^{-2})
- GH — growth hormone
- HSQC — heteronuclear single quantum coherence
- hs — Homo sapiens
- Hz — Hertz
- IGF — insulin-like growth factor
- IGFBP — IGF binding protein
- IGF-IR — IGF receptor type I
- IPTG — isopropyl-β-thiogalactopyranoside
- IR — insulin receptor
- IRS — insulin receptor substrate(s)
- K_D — dissotation constant
- LB — Luria-Broth medium
- LCI — leach carboxypeptidase inhibitor
- MIR — multiple isomorphous replacement
- MW — molecular weight
- NMR — nuclear magnetic resonance
- NOE — nuclear Overhauser effect
- NOESY — nuclear Overhauser enhancement spectroscopy
- OD — optical density

- PAGE polyacrylamide gel electrophoresis
- PEG polyethylene glycol
- PDB Protein Data Bank
- PBS phosphate-buffered saline
- ppm parts per million
- RMS root mean square
- SAD single anomalous dispersion
- SDS sodium dodecyl sulfate
- spl serine protease like
- TCI tick carboxypeptidase inhibitor
- TOCSY total correlation spectroscopy

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

7. References

Abdelnour, A., Arvidson, S., Bremell, T., Ryden, C. and Tarkowski, A. (1993) The accessory gene regulator (*agr*) controls *Staphylococcus aureus* virulence in a murine arthritis model. Infect. Immun. **61,** 3879-3885.

Abrahamson, M. (1994) Cystatins in *Methods in Enzymology* (Barrett, A. J., Ed.) pp 685-700, Academic Press, San Diego.

Ali, O., Cohen, P. and Lee, K.W. (2003) Epidemiology and biology of insulin-like growth factor binding protein-3 (IGFBP-3) as an anti-cancer molecule. Hor. Met. Res. ***35,*** 726-733.

Aloy, P., Catasus, L., Villegas, V., Reverter, D., Vendrell, J., Aviles, F.X. (1998) Comparative analysis of the sequences and three-dimensional models of human procarboxypeptidases A1, A2 and B. Biol Chem **379,** 149-155

Amagai, M., Matsuyoshi, N., Wang, Z.H., Andl, C. and Stanley, J.R. (2000) Toxin in bullous impetigo and *Staphylococcal* scalded-skin syndrome targets desmoglein 1. Nature Med. **6,** 1275 – 1277.

Amodeo, P., Fraternali, F., Lesk, A.M. and Pastore, A. (2001) Modularity and homology: modelling of the titin type I modules and their interfaces. J. Mol. Biol. **31,** 283-296.

Archer, G.L. (1998) *Staphylococcus aureus*: a well-armed pathogen. Clin. Infect. Dis. **26,**1179-81.

Arolas, J.L., Bronsoms, S., Lorenzo, J., Aviles, F.X., Chang, J.Y. and Ventura, S. (2004) Role of kinetic intermediates in the folding of leech carboxypeptidase inhibitor. J. Biol. Chem. ***279,*** 37261-37270.

Arolas, J.L., Lorenzo, J., Rovira, A., Castella, J., Aviles, F.X., Sommerhoff, C.P. (2005) A carboxypeptidase inhibitor from the tick *Rhipicephalus bursa*. Isolation, cDNA cloning, recombinant expression, and characterization. J Biol Chem *in press*

Arron, J.R., Pewzner-Jung, Y., Walsh, M.C., Kobayashi, T. and Choi, Y. (2002) Regulation of the subcellular localization of tumor necrosis factor receptor-associated factor (TRAF)2 by TRAF1 reveals mechanisms of TRAF2 signaling. J. Exp. Med. **196,** 923-34.

Arvidson, S.O. (1983) Extracellular enzymes from *Staphylococcus aureus*. In: Staphylococci and Staphylococcal Infections, Vol. 2, C.S.F. Easmon and C. Adlam, eds. (London, UK: Academic Press, Inc.), pp. 745 – 808.

Arvidson, S., T. Holme, and B. Lindholm. (1973) Studies on extracellular proteolytic enzymes from *Staphylococcus aureus*. I. Purification and characterization of one neutral and one alkaline protease. Biochim. Biophys. Acta **302,**135–148.

Awata, H., Huang, C., Handlogten, M.E., Miller, R.T. (2001) Interaction of the calcium-sensing receptor and filamin, a potential scaffolding protein. J. Biol Chem. **276,** 34871-9.

Aviles, F.X., Vendrell, J., Guasch, A., Coll, M., Huber, R. (1993) Advances in metalloprocarboxypeptidases. Emerging details on the inhibition mechanism and on the activation process. Eur J Biochem **211,** 381-389

Bach, L.A., (1999) The insulin-like growth factor system: basic and clinical aspects. Aust. N.Z. J. Med. ***29,*** 355-361.

Bach, L.A. and Rechler, M.M. (1995) Insulin-like growth factor binding proteins. Diabetes Rev. ***3,*** 38-61.

Bajzar, L., Jain, N., Wang, P., Walker, J.B. (2004) Thrombin activatable fibrinolysis inhibitor: not just an inhibitor of fibrinolysis. Crit Care Med **32,** S320-324

Bajzar L, Manuel R, Nesheim ME (1995) Purification and characterization of TAFI, a thrombin-activable fibrinolysis inhibitor. J Biol Chem **270,** 14477-14484

Barbosa-Pereira, P.J., Segura-Martín, S., Oliva, B., Ferrer-Orta, C., Aviles, F.X., Coll, M., Gomis-Rüth, F.X., Vendrell, J. (2002) Human procarboxypeptidase B: three-dimensional structure and implications for thrombin-activatable fibrinolysis inhibitor (TAFI). J Mol Biol **321,** 537-547

Barrett, A.J., Rawlings, N.D., Davies, M.E., Machleidt, W., Salvesen, G. and Turk, V. (1986) Cysteine proteinase inhibitors of the cystatin superfamily in *Proteinase inhibitors* (Barrett, A. J. and Salvesen, G., Eds.) pp 515-569, Elsevier, Amsterdam.

Baxter, R.C. (2000) Insulin-like growth factor (IGF)-binding proteins: interactions with IGFs and intrinsic bioactivities. Am. J. Physiol. ***278,*** E967–E976.

Baxter, R.C., Bayne, M.L. and Cascieri, M. A. (1992) Structural determinants for binary and ternary complex formation between insulin-like growth factor-I (IGF-I) and IGF binding protein-3. J. Biol. Chem. ***267,*** 60-65.

Bayne, M.L., Applebaum, J., Chicchi, G.G., Miller, R.E. and Cascieri, M.A. (1990) The roles of tyrosine 24, 31, and 60 in the high affinity binding of insulin-like growth factor-I and the type 1 insulin-like growth factor receptor. J. Biol. Chem. ***265,*** 15648-15652.

Berry, F.B., O'Neill, M.A., Coca-Prados, M., Walter, M.A. (2005) FOXC1 transcriptional regulatory activity is impaired by PBX1 in a filamin A-mediated manner. Mol. Cell Biol. **25,** 1415-24.

Berthold, H., Frorath, B., Scanarini, M., Abney, C.C., Ernst, B. and Northemann, W. (1992) Plasmid pGEX-5T: an alternative system for expression and purification of recombinant proteins. Biotechnol. Lett. ***14,*** 245–250.

Bode, W., Huber, R. (1992) Natural protein proteinase inhibitors and their interaction with proteinases. Eur J Biochem **204,** 433-451.

Bodenhausen, G. and Ruben, D.J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem. Phys. Lett. ***69,*** 185-189.

Bode, W., Huber, R. (2000) Structural basis of the endoproteinase-protein inhibitor interaction. Biochim Biophys Acta **1477,** 241-252.

Boffa, M.B., Bell, R., Stevens, W.K., Nesheim, M.E. (2000) Roles of thermal instability and proteolytic cleavage in regulation of activated thrombin activable fibrinolysis inhibitor. J Biol Chem **275,** 12868-12878.

Boffa, M.B., Wang, W., Bajzar, L., Nesheim, M.E. (1998) Plasma and recombinant thrombin-activable fibrinolysis inhibitor (TAFI) and activated TAFI compared with respect to glycosylation, thrombin/thrombomodulin-dependent activation, thermal stability, and enzymatic properties. J Biol Chem **273,** 2127-2135.

Bork, P., Holm, L. and Sander, C. The immunoglobulin fold**.** (1994) J. Mol. Biol. **242,** 309-320.

Bouma, B.N., Meijers, J.C. (2003) Thrombin-activatable fibrinolysis inhibitor (TAFI, plasma procarboxypeptidase B, procarboxypeptidase R, procarboxypeptidase U). J Thromb Haemost **1,** 1566-1574.

Brink, M., Gerisch, G., Isenberg, G., Noegel, A.A., Segall, J.E., Wallraff, E. and Schleicher, M. (1990) A *Dictyostelium* mutant lacking an F-actin cross-linking protein, the 120-kD gelation factor. J. Cell. Biol. **111,** 1477–1489.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W. et al. (1998) Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D **54,** 905–921.

Buckway, C.K., Wilson, E.M., Ahlsen, M., Bang, P., Oh, Y. and Rosenfeld, R.G. (2001) Mutation of three critical amino acids of the N-terminal domain of IGF-binding protein-3 essential for high affinity IGF binding. J. Clin. Endoc. and Metabol. **86,** 4943-4950.

Bunn, R.C. and Fowlkes, J.L. (2003) Insulin-like growth factor binding protein proteolysis. Trends Endoc. and Metabol. **14,** 176-181.

Butt, A.J., Firth, S.M., King, M.A. and Baxter, R.C. (2000) Insulin-like growth factor-binding protein-3 modulates expression of Bax and Bcl-2 and potentiates p53-independent radiation-induced apoptosis in human breast cancer cells. J. Biol. Chem. ***275,*** 39174-39181.

Carrick, F.E., Wallace, J.C. and Forbes, B.E. (2002) The interaction of Insulin-like growth factors (IGFs) with Insulin-like growth factor binding proteins (IGFBPs): a review. Lett. Pep. Sci. ***8,*** 147–153.

Cascieri, M.A., Chicchi, G.C., Applebaum, J., Hazes, N.S., Green, B.C. and Bayne, M.L. (1988) Mutants of human insulin-like growth factor I with reduced affinity for the type I insulin-like growth factor receptor. Biochemistry, ***27*****,** 3229-3233.

Cavarelli, J., Prevost, G., Bourguet, W., Moulinier, L., Chevrier, B., Delagoutte, B., Bilwes, A., Mourey, L., Rifai, S., Piemont, Y., Moras, D. (1997) The structure of *Staphylococcus aureus* epidermolytic toxin A, an atypic serine protease, at 1.7 A resolution. Structure. **5,** 813-24.

Chan, K. and Spencer, E.M. (1998) Megakaryocytes endocytose insulin-like growth factor (IGF) I and IGF-binding protein-3: a novel mechanism directing them into alpha granules of platelets. Endocrinology, ***139*****,** 559-65.

Chernausek, S.D., Smith, C.E., Duffin, K.L., Busby, W.H., Wright, G. and Clemmons, D.R (1995) Proteolytic cleavage of insulin-like growth factor binding protein 4 (IGFBP-4) Localization of cleavage site to non-homologous region of native IGFBP-4. J. Biol. Chem. **270,** 11377-11382.

Clemmons, D.R. (1999) IGF binding proteins and extracellular matrix. In Rosenfeld, R.G. and Roberts, C.T. (eds), The IGF system. Molecular biology, physiology, and clinical applications. Humana Press, Totowa, pp. 273-279.

Clemmons, D.R. (2001) Use of mutagenesis to probe IGF-binding protein structure/function relationships. Endocr. Rev. **22,** 800-817.

Clore, G.M., Gronenborn, A.M., Nilges, M., Ryan, C.A. (1987) Three-dimensional structure of potato carboxypeptidase inhibitor in solution. A study using nuclear magnetic resonance, distance geometry, and restrained molecular dynamics. Biochemistry. **26,** 8012-8023.

Cohen, P., Clemmons, D.R. and Rosenfeld, R.G. (2000) Does the GH-IGF axis play a role in cancer pathogenesis? Growth Hor. IGF Res. **10,** 297-305.

Collaborative computational project, Number 4. (1994) Acta Crystallogr. D **50,** 760-763.

Condeelis, J., Vahey, M., Carboni, J. M., DeMey, J. and Ogihara, S. (1984) Properties of the 120,000- and 95,000-dalton actin-binding proteins from Dictyostelium discoideum and their possible functions in assembling the cytoplasmic matrix. J. Cell Biol. **99**, 119 s-126 s.

Cooke, R.M., Harvey, T.S. and Campbell, I.D. (1991) Solution structure of human insulin-like growth factor 1: a nuclear magnetic resonance and restrained molecular dynamics study. Biochemistry. **30,** 5484-5491.

Cooper, B.S., Stone, S.P., Kibbler, C.C., Cookson, B.D., Roberts, J.A. Medley, G.F., Duckworth, G.J., Lai, R., Ebrahim, S. (2003) Systematic review of isolation policies in the hospital management of methicillin-resistant *Staphylococcus aureus*: a review of the literature with epidemiological and economic modelling. Health Technol Assess. **39,** 1-194.

Cornilescu, G., Delaglio, F. and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR, **13**, 289–302.

Coulter, S.N., Schwan, W.R., Ng, E.Y., Langhorne, M.H., Ritchie, H.D., Westbrock-Wadman, S., Hufnagle, W.O., Folger, K.R., Bayer, A.S. and Stover, C.K. (1998) *Staphylococcus aureus* genetic loci impacting growth and survival in multiple infection environments, Mol. Microbiol. **30,** 393-404.

Cranmer, S.L., Pikovski, I., Mangin, P., Thompson, P.E., Domagala, T., Frazzetto, M., Salem, H.H., Jackson, S.P. (2005) Identification of a unique filamin A binding region within the cytoplasmic domain of glycoprotein Ibalpha. Biochem, J. **387,** 849-58.

Creighton, T.E. (1986) Disulfide bonds as probes of protein folding pathways. Methods Enzymol. **131**, 83-106.

Creighton, T.E., Darby, N.J. and Kemmink, J. (1996) The roles of partly folded intermediates in protein folding. Faseb J. **10**, 110-118.

Daughaday, W.H., Hall, K., Raben, M.S., Salmon Jr., W.D., Van den Brande, J.L. and Van Wyk, J.J. (1972) Somatomedin: proposed designation for sulphation factor. Nature. **235,** 107.

Daughaday, W. and Salmon Jr, W.D. (1999) Historical perspective: the origins and development of the somatomedin hypothesis. In The IGF system: Molecular biology, physiology and clinical applications. Rosenfeld, R.G. and Roberts, C.T. (eds),. Humana Press, Totowa, New Jersey. 1-15.

Dhalluin, C., Wierusyeski, J. M. and Lippens, G. (1996) An Improved Homonuclear TOCSY Experiment with Minimal Water Saturation, J. Magn. Reson. B **111,** 168-179.

Diederichs, K., Karplus, P.A. (1997) Improved R-factors for diffraction data analysis in macromolecular crystallography. Nat. Struct. Biol. **4,** 269-75.

Diekema, D.J., Pfaller, M.A., Schmitz, F.J., Smayevsky, J., Bell, J., Jones, R.N., Beach, M., SENTRY Partcipants Group. (2001) Survey of infections due to *Staphylococcus* species: frequency of occurrence and antimicrobial susceptibility of isolates collected in the United States, Canada, Latin America, Europe, and the Western Pacific region for the SENTRY Antimicrobial Surveillance Program, 1997-1999. Clin Infect Dis. 15;32 Suppl 2:S114-32.

Dill, K. A. and Chan, H. S. (1997. From Levinthal to pathways to funnels. Nat. Struct. Biol. **4**, 10-19.

Djinovic-Carugo, K., Gautel, M., Ylanne, J. and Young, P. (2002) The spectrin repeat: a structural platform for cytoskeletal protein assemblies. FEBS Lett. **513**, 119-123.

Dodatko, T., Fedorov, A.A., Grynberg, M., Patskovsky, Y., Rozwarski, D.A., Jaroszewski, L., Aronoff-Spencer, E., Kondraskina, E., Irving, T., Godzik, A., Almo, S.C. (2004) Crystal structure of the actin binding domain of the cyclase-associated protein. Biochemistry. **43,**10628-41.

Dubin, G. (2002) Extracellular proteases of *staphylococcus* spp. Biol. Chem. **383,** 1075 – 1086.

Eichinger, S., Schonauer, V., Weltermann, A., Minar, E., Bialonczyk, C., Hirschl, M., Schneider, B., Quehenberger, P. and Kyrle, P.A. (2004) Thrombin-activatable fibrinolysis inhibitor and the risk for recurrent venous thromboembolism. Blood **103,** 3773-3776

Emari, T.G. and Gaynes, R.P. (1993) An overview of nosocomial infections, including the role of microbiology labolatory. Clin. Micorbiol. Rev. **6,** 428-442.

Enz, R. (2002) The actin-binding protein Filamin-A interacts with the metabotropic glutamate receptor type 7. FEBS Lett. **514,** 184-8.

Erlanson, D.A., Wells, J.A., Braisted, A.C. (2004) Tethering: fragment-based drug discovery. Annu Rev Biophys Biomol Struct **33,** 199-223

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D. and Kay, L.E. (1994) Backbone Dynamics of a Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by ^{15}N NMR Relaxation, Biochemistry **33,** 5984-6003.

Feng, Y., Walsh, A. (2004) The many faces of filamin: A versatile molecular scaffold for cell motility and signaling. Nat. Cell Biol. **11,** 1034-1037.

Field, J., Vojtek, A., Ballester, R., Bolger, G., Colicelli, J., Ferguson, K., Gerst, J., Kataoka, T., Powers, S., Riggs, M., Rodgers, L., Wieland, I., Wheland, B. and Wigler, M. (1990) Cloning and characterization of CAP, the S. cerevisiae gene encoding the 70 kd adenylyl cyclase-associated protein. Cell. **61**, 319–327.

Filipek, R., Rzychon, M., Oleksy, A., Gruca, M., Dubin, A., Potempa, J. and Bochtler, M. (2003) The staphostatin-staphopian complex: a forward binding inhibitor in complex with its target cysteine protease, J. Biol. Chem. **278,** 40959-40966.

Firth, S.M. and Baxter, R.C. (2002) Cellular actions of the insulin-like growth factor binding proteins. Endocr. Rev. **23,** 824-854

Firth, S.M., Clemmons, D.R. and Baxter, R.C. (2001) Mutagenesis of basic amino acids in the carboxyl-terminal region of insulin-like growth factor binding protein-5 affects acid-labile subunit binding. Endocrinology, **14,** 2147-2150.

La Fortelle, E. de and Bricogne, G. (1997) Maximum-likelihood heavy-atom parameter refinement in MIR and MAD methods. Meth. Enzym. **276,** 472-494.

Fox, J. W. *et al.* (1998) Mutations in filamin 1 prevent migration of cerebral cortical neurons in human periventricular heterotopia. Neuron **21,** 1315–1325.

Franco, R.F., Fagundes, M.G., Meijers, J.C., Reitsma, P.H., Lourenco, D., Morelli, V., Maffei, F.H., Ferrari, I.C., Piccinato, C.E., Silva, W.A. Jr, Zago, M.A. (2001) Identification of polymorphisms in the 5'-untranslated region of the tafi gene: relationship with plasma TAFI levels and risk of venous thrombosis. Haematologica **86,** 510-517

Franzot, G., Sjöblom, B., Gaute, M. and Djinović Carugo K. (2005) The crystal structure of the actin binding domain from alpha-actinin in its closed conformation: structural insight into phospholipid regulation of alpha-actinin. J Mol Biol. **348,**151-65.

Fraternali, F. and Pastore, A. (1999) Modularity and homology: Modelling of the type II module family from titin. J. Mol. Biol. **290,** 581-593.

Fucini, P., Koppel, B., Schleicher, M., Lustig, A., Holak, T.A., Muller, R., Stewart, M., Noegel, A.A. (1999) Molecular architecture of the rod domain of the *Distyostelium* gelation factor (ABP-120*)*. J. Mol. Biol. **291,** 1017-1023.

Fucini, P., McCoy, A.J., Gomez-Ortis, M., Schleicher, M., Noegel, A.A. and Stewart, M. (1997) Crystallization and preliminary X-ray diffraction characterization of a dimerizing fragment of the rod domain of the *Dictyostelium* gelation factor (ABP-120). J. Struct. Biol. **120**, 192-195.

Fucini, P., Renner, C., Herberhold, C., Noegel, A.A. and Holak, T.A. (1997) The repeating segments of the F-actin cross-linking gelation factor (ABP-120) have an immunoglobulin-like fold. Nature Struct. Biol. **4**, 223–230.

Fuentes-Prior, P., Noeske-Jungblut, C., Donner, P., Schleuning, W.D., Huber, R., Bode, W. (1997) Structure of the thrombin complex with triabin, a lipocalin-like exosite-binding inhibitor derived from a triatome bug. Proc Natl Acad Sci USA 94: 11845-11850.

Geerten, V.W. and Bax, A. (1993) Qualitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(HNHA) Coupling Constants in 15N-Enriched Proteins, J. Am. Chem. Soc. **115,** 7772-7777.

Gerst, J.E., Rodgers, L., Riggs, M. and Wigler, M. (1992) SNC1, a yeast homolog of the synaptic vesicle-associated membrane protein/synaptobrevin gene family: genetic interactions with the RAS and CAP genes. Proc. Natl. Acad. Sci. USA, **89**, 4338–4342.

Gill, S.R., Fouts, DE., Archer GL., et al. (2005) Insights on evolution of virulence and resistance from the complete genome analysis of an early methicillin-resistant *Staphylococcus aureus* strain and a biofilm-producing methicillin-resistant *Staphylococcus epidermidis* strain. J Bacteriol.**187,** 2426-38.

Goddard, T.D. and Kneller, D.G., SPARKY 3, University of California, San Francisco.

Goldenberg, D.P. (1992) Native and non-native intermediates in the BPTI folding pathway. Trends Biochem Sci. **17,** 257-61.
Goldsmith, S.C. Pokala, N., Shen, W., Fedorov, A.A., Matsudaira, P., Almo, S.C., (1997) The structure of an actin-crosslinking domain from human fimbrin. Nature Struct. Biol. **4**, 708–712.

Gomis-Rüth, F.X., Companys, V., Qian, Y., Fricker, L.D., Vendrell, J., Aviles, F.X., Coll, M. (1999) Crystal structure of avian carboxypeptidase D domain II: a prototype for the regulatory metallocarboxypeptidase subfamily. EMBO J **18,** 5817-5826.

Gonzalez, C., Neira, J.L., Ventura, S., Bronsoms, S., Rico, M., Aviles,, F.X. (2003) Structure and dynamics of the potato carboxypeptidase inhibitor by 1H and 15N NMR. Proteins **50,** 410-422.

Gorlin, J., Yamin, J., Egan, S., Stewart, M., Stossel, T., Kwiatkowski, D. and Hartwig, J. (1990) Human endothelial actin-binding protein (ABP, nonmuscle filamin): a molecular leaf spring. J. Cell Biol. **111,** 1089–1105.

Gravante, B., Barbuti, A., Milanesi, R., Zappi, I., Viscomi, C., DiFrancesco, D. (2004) Interaction of the pacemaker channel HCN1 with filamin A. J. Biol. Chem. **279,** 43847-53.

Grzesiek, S. and Bax, A. (1992) Improved 3D Triple-Resonance NMR Techniques Applied to a 31 kDa Protein, J. Magn. Reson. **96,** 432-440.

Guasch, A., Coll, M., Aviles, F.X., Huber, R. (1992) Three-dimensional structure of porcine pancreatic procarboxypeptidase A. A comparison of the A and B zymogens and their determinants for inhibition and activation. J Mol Biol **224,** 141-157.

Haley, R.W., Culver, DH., White, J.W., Morgan, W.M., Emori, T.G. (1985) The nationwide nosocomial infection rate. A new need for vital statistics. Am J Epidemiol. **121,**159-67.

Halloran, B.P. and Spencer, E.M. (1988) Dietary phosphorus and 1,25-dihydroxyvitamin D metabolism: influence of insulin-like growth factor I. Endocrinology, **123,** 1225-9.

Hartwig, J. in Protein Profile (ed. Sheterline) P. 711–778 (Academic, London, 1994).

Hartwig, J.H. and DeSisto, M. (1991) The cytoskeleton of the resting human blood platelet: structure of the membrane skeleton and its attachment to actin filaments. J. Cell Biol. **112,** 407-25.

Headey, S.J., Keizer, D.W., Yao, S., Wallace, J.C., Bach, L.A. and Norton, R.S. (2004) Binding site for the C-domain of insulin-like growth factor (IGF) binding protein-6 on IGF-II; implications for inhibition of IGF actions. FEBS Lett. **568,** 19-22.

Himmel, M., van der Ven, P.F.M., Stöcklein W. and Fürst D.O. (2003) The Limits of Promiscuity: Isoform-Specific Dimerization of Filamins. Biochemistry **42,** 430-439.

Hink-Schauer, C., Estebanez-Perpina, E., Wilharm, E., Fuentes-Prior, P., Klinkert, W., Bode, W., Jenne, D.E. (2002) The 2.2-A crystal structure of human pro-granzyme K reveals a rigid zymogen with unusual features. J Biol Chem. **277,** 50923-33.

Hodgson, D.R., May, F.E.B. and Westley, B.R. (1996) Involvement of phenylalanine 23 in the binding of IGF-1 to the insulin and type I IGF receptor. Regul. Pept. **66,** 191-196.

Hofmann, B., Schomburg, D. and Hecht, H. J. (1993) Crystal structure of a thiol proteinase from *Staphylococcus aureus* V8 in the E-64 inhibitor complex, Acta Crystallogr. **49** (Suppl.), 102.

Holak, T.A., Gondol, D., Otlewski, J. and Wilusz, T. (1989) Determination of the complete three-dimensional structure of the trypsin inhibitor from squash seeds in aqueous solution by nuclear magnetic resonance and a combination of distance geometry and dynamical simulated annealing, J. Mol. Biol. **210,** 635-648.

Holmm L,, Sanderm C, (1993) Protein structure comparison by alignment of distance matrices. J. Mol. Biol. **233,** 123-138.

Honig, B. (1999) Protein folding: from the levinthal paradox to structure prediction. J. Mol. Biol. **293**, 283-293.

Hoover, D.M., Rajashankar, K.R., Blumenthal, R., Puri, A., Oppenheim, J.J., Chertov, O., Lubkowski, J. (2000) The structure of human beta-defensin-2 shows evidence of higher order oligomerization. J. Biol. Chem. **275,** 32911-32918.

Hubberstey, A.V. and Mottillo E.P. (2002) Cyclase-associated proteins: CAPacity for linking signal transduction and actin polymerization. FASEB J. **16**, 487–499.

Hutchinson, E.G. and Thornton, J.M. (1990) HERAsa program to draw schematic diagrams of protein secondary structures, Proteins **8,** 203-212.

Hwa, V., Oh, Y., Burren, C.P., Choi, W.K., Graham, D.L., Ingermann, A., Kim, H.S., Lopez-Bermejo, A., Minniti, G., Nagalla, S.R., Pai, K., Spagnoli, A., Vorwerk, P., Wanek, D.L.V., Wilson, E.M., Yamanaka, Y., Yang, D.H. and Rosenfeld, R.G. (1999) The IGF binding protein superfamily. In Rosenfeld, R.G. and Roberts, C.T. (eds), The IGF system. Molecular biology, physiology, and clinical applications. Humana Press, Totowa, pp. 315-327.

Improta, S., Krueger, J.K., Gautel, M., Atkinson. R.A., Lefevre, J.F., Moulton, S., Trewhella, J. and Pastore, A.(1998) The assembly of immunoglobulin-like modules in titin: Implications for muscle elasticity. J. Mol. Biol. **284**, 761-777.

Janssen, K.-P., Eichinger, L., Janmey, P., Noegel, A.A., Schliwa, M., Witke, W. and Schleicher, M. (1996) Viscoelastic properties of F-actin solutions in the presence of normal and mutated actin-binding proteins. Arch. Biochem. Biophys. **325**, 183-189.

Jones, J.L. and Clemmons, D.R. (1995) Insulin-like growth factors and their binding proteins: biological actions. Endocr. Rev. ***16,*** 3-34.

Jung, G.W., Spencer, E.M. and Lue, T.F. (1998) Growth hormone enhances regeneration of nitric oxide synthase-containing penile nerves after cavernous nerve neurotomy in rats. Journal of Urology, **160,** 1899-904.

Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. J. Appl. Cryst. **26,** 795-800.

Kalus, W., Zweckstetter, M., Renner, C., Sanchez, Y., Georgescu, J., Grol, M., Demuth, D., Schumacher, R., Dony, C., Lang, K. and Holak, T.A. (1998) Structure of the IGF binding domain of the insulin-like growth factor-binding protein-5 (IGFBP-5): implications for IGF and IGF-I receptor interactions. EMBO J. **17,** 6558-72.

Kay, L., Ikura, M., Tschudin, R. and Bax, A. (1990) Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins, J. Magn. Reson. **89,** 496-514.

Khandwala, H.M., McCutcheon, I.E., Flyvbjerg, A. and Friend, K.E. (2000) The effects of insulin-like growth factors on tumorigenesis and neoplastic growth. Endocr. Rev. **21**, 215-44.

Kilshtain-Vardi, A., Glick, M., Greenblatt, H.M., Goldblum, A., Shoham, G. (2003) Refined structure of bovine carboxypeptidase A at 1.25 Å resolution. Acta Crystallogr D **59,** 323-333.

Knuth, M., Khaire, N., Kuspa, A., Lu, S.J., Schleicher, M., Noegel, A.A. (2004) A novel partner for Dictyostelium filamin is an alpha-helical developmentally regulated protein. J. Cell Sci. **117,** 5013-5022.

Koradi, R., Billeter, M. and Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. J. Mol. Graph. **14**, 52–55.

Ksiazek, D., Brandstetter, H., Israel, L., Bourenkov, G.P., Katchalova, G., Janssen, K.-P., Bartunik, H. D., Noegel, A.A., Schleicher, M. and Holak, T.A. (2003) Structure, **11,** 1–20.

Kumar, A., Ernst, R.R. and Wüthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton crossrelaxation networks in biological macromolecules. Biochem. Biophys. Res. Commun. **95,** 1-6.

Lamzin, V.S., Wilson, K.S. (1993) Automated refinement of protein models. *Acta Crystallogr*. D **49**, 129-149.

Leahy, D.J., Aukhil, I. and Erickson, H.P. (1996) 2.0 Å crystal structure of a four-domain segment of human fibronectin encompassing the RGD loop and synergy region. Cell **84,** 155-164.

Lebart, M.C., Mejean, C., Casanova, D., Audemard, E., Derancourt, J., Roustan, C. and Benyamin, Y**.** (1994) Characterization of the actin binding site on smooth muscle filamin. J. Biol. Chem. **269**, 4279–4284.

Lebart, M.C., Mejean, C., Roustan, C., Benyamin, Y. (1993) Further characterization of the alpha-actinin-actin interface and comparison with filamin-binding sites on actin. J. Biol. Chem. **268,** 5642-8.

Lehman, W., Craig, R., Kendrick-Jones, J., Sutherland-Smith, AJ. (2004) An open or closed case for the conformation of calponin homology domains on F-actin? J. Muscle Res. Cell Motil. **25,** 351-8.

LeRoith, D. and Helman, L. (2004) The new kid on the block(ade) of the IGF-1 receptor. Cancer Cell, **5,** 201-2.

Leventhal, P.S., Russel, J.W. and Feldman, E.L. (1999) IGFs and the nervous system. In The IGF System. R.G. Rosenfeld and C. Roberts (eds). Humana Press, Totowa, New Jersey.

Lin, R., Karpa, K., Kabbani, N., Goldman-Rakic, P., Levenson, R. (2001) Dopamine D2 and D3 receptors are linked to the actin cytoskeleton via interaction with filamin A. Proc. Natl. Acad. Sci. USA. **98,** 5258-63.

Lindsay JA. and Holden MT. (2004) *Staphylococcus aureus*: superbug, super genome? Trends Microbiol. **12,** 378-85.

Liu, G., Thomas, L., Warren, R.A., Enns, C.A., Cunningham, C.C., Hartwig, J.H., Thomas, G. (1997) Cytoskeletal protein ABP-280 directs the intracellular trafficking of furin and modulates proprotein processing in the endocytic pathway. J. Cell Biol. **139,**1719–1733.

Li, M.-G., Serr, M., Edards, K., Ludmann, S., Yamamotot, D., Tilney, L.G., Field, C.M. and Hays, T.S. (1999) Filamin is required for ring canal assembly and actin organization during *Drosophila* oogenesis. J.Cell Biol. **146,** 1061-1074.

Loddick, S.A., Liu, X.J., Lu, Z.X., Liu, C.L., Behan, D.P., Chalmers, D.C., Foster, A.C., Vale, W.W., Ling, N. and Desouza, E.B. (1998) Displacement of insulin-like growth factors from their binding proteins as a potential treatment for stroke. Proc. Nat. Acad. Sci. USA, **95,** 1894-8.

Lowy, F.D. (1998) *Staphylococcus aureus* infections. N. Engl. J. Med. **339,** 520-32.

Loy, C.J., Sim, K.S. and Yong, E.L. (2003) Filamin-A fragment localizes to the nucleus to regulate androgen receptor and coactivator functions. Proc. Natl. Acad. Sci. USA. **100,** 4562-7.

Mans, B.J., Louw, A.I., Neitz, A.W. (2002) Amino acid sequence and structure modeling of savignin, a thrombin inhibitor from the tick, *Ornithodoros savignyi*. Insect. Biochem. Mol. Biol. **32,** 821-828

Mans, B.J., Neitz, A.W.H. (2004) Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. Insect. Biochem Mol. Biol. **34,** 1-17.

Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: Application to interleukin 1 β. Biochemistry ***28,*** 6150-6156.

Marx, P.F., Hackeng, T.M., Dawson, P.E. Griffin, J.H., Meijers, J.C., Bouma, B.N. (2000) Inactivation of active thrombin-activable fibrinolysis inhibitor takes place by a process that involves conformational instability rather than proteolytic cleavage. J. Biol. Chem. **275,** 12410-12415.

Marx, P.F., Havik, S.R., Marquart, J.A., Bouma, B.N., Meijers, J.C. (2004) Generation and characterization of a highly stable form of activated thrombin-activable fibrinolysis inhibitor. J. Biol. Chem. **279,** 6620-6628.

Massimi, I., Park, E., Rice, K., Muller-Esterl, W., Sauder, D. and McGavin, M. J. (2002) Identification of a novel maturation mechanism and restricted substrate specificity for the SspB cysteine protease of *Staphylococcus aureus*, J. Biol. Chem. **277,** 41770-41777.

Mazerbourg, S., Callebaut, I., Zapf, J., Mohan, S., Overgaard, M. and Monget, P. (2004) Update on IGFBP-4: regulation of IGFBP-4 levels and functions, in vitro and in vivo, Growth Hormone and IGF Research, **14,** 71-84.

McCoy, A.J., Fucini, P., Noegel, A.A., Stewart, M. (1999) Structural basis for dimerization of the *Dictyostelium* gelation factor (ABP120) rod. Nature Struct. Biol. **6,** 836-841.

McGahee, W. and Lowy, F. D. (2000) *Staphylococcal* infections in the intensive care unit, Semin. Respir. Infect. **15,** 308-313.

McGough, A. (1998) F-actin binding proteins. Curr. Opin. Struct. Biol. **8**, 166–176.

McRee, D.E. (1999) XtalView/Xfit - A versatile program for manipulating atomic coordinates and electron density. J. Struc. Biol. **125,** 156-165.

Meng, X., Yuan, Y., Maestas, A., Shen, Z. (2004) Recovery from DNA damage-induced G2 arrest requires actin-binding protein filamin-A/actin-binding protein-280. J. Biol. Chem. **279,** 6098-105.

Mori, S., Abeygunawardana, C., Johnson, M. O. and van Zijl, P. C. M. (1995) Improved senesitivity of HSQC spectra of exchanging protons at short interscan delays using a new fast HSQC (FHSQC) detection scheme that avoids water saturation, J. Magn Reson. B **108,** 94-98.

Morin, C.A. and Hadler, J.L. (2001) Population-based incidence and characteristics of community-onset *Staphylococcus aureus* infections with bacteremia in 4 metropolitan Connecticut areas, 1998. J. Infect. Dis. **184,**1029-34.

Muehlhahn, P., Bernhagen, J., Czisch, M., Georgescu, J., Renner, C., Ross, A., Bucala, R. and Holak, T. A. (1996) NMR characterization of structure, backbone dynamics, and glutathione binding of the human macrophage migration inhibitory factor (MIF), Protein Sci. **5,** 2095-2103.

Myles, T.M., Nishimura, T., Yun, T.H., Nagashima, M., Morser, J., Patterson, A.J., Pearl, R.G., Leung, L.L.K. (2003) Thrombin activatable fibrinolysis inhibitor, a potential regulator of vascular inflammation. J. Biol. Chem. **278,** 51059-51067.

Nagano, T., Morikubo, S., Sato, M. (2004) Filamin A and FILIP (Filamin A-Interacting Protein) regulate cell polarity and motility in neocortical subventricular and intermediate zones during radial migration. J. Neurosci. **24,** 9648-57.

Nagashima, M., Werner, M., Wang, M., Zhao, L., Light, D.R., Pagila, R., Morser, J., Verhallen, P. (2000) An inhibitor of activated thrombin-activatable fibrinolysis inhibitor potentiates tissue-type plasminogen activator-induced thrombolysis in a rabbit jugular vein thrombolysis model. Thromb Res **98,** 333-342.

Nemato, S. and Finkel, T. (2004) Ageing and the mystery at Arles. Nature **429,** 149-152.

Nicastro, G., Franzoni, L., Chiara, C., Mancin, A.C., Giglio, J.R., Spisni, A. (2003) Solution structure of crotamine, a Na^+ channel affecting toxin from *Crotalus durissus terrificus* venom. Eur. J. Biochem. **270,** 1969-1979.

Nilges, M., Clore, G.M. and Gronenborn, A.M. (1988) Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. FEBS Lett. **239,** 129–136.

Nishida, Y., Shima, F., Sen, H., Tanaka, Y., Yanagihara, C., Yamawaki-Kataoka, Y., Kariya, K. and Kataoka, T. (1998) Coiled-coil interaction of N-terminal 36 residues of cyclase-associated protein with adenylyl cyclase is sufficient for its function in Saccharomyces cerevisiae ras pathway. J. Biol. Chem. **273,** 28019–28024.

Noble, W.C. (1998) Skin bacteriology and the role of *Staphylococcus aureus* in infection. Br. J. Dermatol. 139 Suppl **53,** 9-12.

Noegel, A.A., Rapp, S., Lottspeich, F., Schleicher. M. and Stewart, M. (1989) The *Dictyostelium* gelation factor shares a putative actin binding site with alpha-actinins and dystrophin and also has a rod domain containing six 100-residue motifs that appear to have a cross-beta conformation. J. Cell. Biol. **109,** 607-618.

Noegel, A.A., Rivero, F., Albrecht, R., Janssen, K.P., Kohler, J., Parent, C.A. and Schleicher, M. (1999) Assessing the role of the ASP56/CAP homologue of Dictyostelium discoideum and the requirements for subcellular localization. J. Cell Sci. **112**, 3195–3203.

Novick, R. P. (2000) Pathogenicity factors and their regulation in *Gram-Positive Pathogens* (Fischetti, V. A., Novick, R. P., Ferretti, J. J., Portnoy, D. A. and Rood, J. I., Eds.) pp 392-407, American Society for Microbiology, Washington, DC.

Ohta, Y., Suzuki, N., Nakamura, S., Hartwig, J.H., Stossel, T.P. (1999) The small GTPase RalA targets filamin to induce filopodia. Proc. Natl. Acad. Sci. USA. **96,** 2122-8.

Ozanne, D.M., Brady, M.E., Cook, S., Gaughan. L., Neal, D.E., Robson, C.N. (2000) Androgen receptor nuclear translocation is facilitated by the f-actin cross-linking protein filamin. Mol. Endocrinol. **14,** 1618-26.

Payet, L.D., Wang, X.H., Baxter, R.C. and Firth, M. (2003) Amino- and carboxyl-terminal fragments of insulin-like growth factor (IGF) binding protein-3 cooperate to bind IGFs with high affinity and inhibit IGF receptor interactions. Endocrinology, **144,** 2797-806.

Peranen, J., Rikkonen, M., Hyvonen, M. and Kaariainen, L. (1996). T7 vectors with modified T7lac promoter for expression of proteins in Escherichia coli. Anal. Biochem. **236**, 371-373.

Perdue, J.F., Bach L.A., Hashimoto, R., Sakano, K., Fujita-Yamaguchi, Y., Fujiwara, H. and Rechler, M.M. (1994) in: The insulin-like growth factors and their regulatory proteins, Baxter, R.C., Gluckman, P.D. and Rosenfeld, R.G. (eds), Elsevier Science Publishing Co., New York, pp. 67-76.

Perrakis, A., Harkiolaki, M., Wilson, K.S. and Lamzin, V.S. (2001) ARP/wARP and molecular replacement. Acta Cryst. ***D57,*** 1445-50.

Perrakis, A., Morris, R. and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. Nature Struct. Biol. **6,** 458-63.

Piotto, M., Saudek, V. and Sklenar, V. (1992) Gradient-Tailored excitation for single-quantum NMR spectroscopy of aqueous solutions, J. Biomol. NMR **2,** 661-666.

Pollak, M.N., Schernhammer, E.S. and Hankinson, S.E. (2004) Insulin-like growth factors and neoplasia. Nature Rev. Cancer, **4,** 505-18.

Popowicz, G.M., Muller, R., Noegel, A.A., Schleicher, M., Huber, R. and Holak, T.A. (2004) Molecular structure of the rod domain of dictyostelium filamin. J. Mol. Biol. **342,** 1637–1646.

Powell, M.J.D. (1977) Restart Procedures for the Conjugate Gradient Method Math. Program. **12**, 241–254.

Prasad, L., Leduc, Y., Hayakawa, K., Delbaere, L.T. (2004) The structure of a universally employed enzyme: V8 protease from *Staphylococcus aureus*. Acta Crystallogr. D **60,** 256-9.

Pudas, R., Kiema, T-R., Jonathan, P., Butler, G., Stewart, M. and Ylänne, J. (2005) Structural Basis for Vertebrate Filamin Dimerization. Structure **13,** 111–119.

Qin, X.Z., Strong, D.D., Baylink, D.J. and Mohan, S. (1998) Structure-function analysis of the human insulin-like growth factor binding protein-4. J. Biol. Chem. **273,** 23509-16.

Rance, M. (1987) Improved techniques for homonuclear rotating frame and isotropic mixing experiments. J. Magn. Reson. **74,** 557-564.

Reed, S.B., Wesson, C.A., Liou, L.E., Trumble, W.R., Schlievert, P.M., Bohach, G.A. and Bayles, K.W. (2001) Molecular characterization of a novel *Staphylococcus aureus* serine protease operon. Infect. Immun. **69,**1521-1527.

Rees, D.C. and Lipscomb, W.N. (1982) Refined crystal structure of the potato inhibitor complex of carboxypeptidase A at 2.5 Å resolution. J. Mol. Biol. **160,** 475-498.

Rehm, T., Huber, R. and Holak, T.A. (2002a) Application of NMR in Structural Proteomics: Screening for Proteins Amenable to Structural Analysis. Structure **10,** 1613-1618.

Rehm, T., Mavoungou, Ch., Israel, L., Schleicher, M. and Holak, T.A. (2002b) Sequence-specific (^{1}H, ^{15}N, ^{13}C) resonance assignment of the N-terminal domain of the cyclase-associated protein (CAP) from Dictyostelium discoideum. J. Biomol. NMR **23**, 337–338.

Renner, C., Schlieicher, M., Moroder, L. and Holak, T. A. (2002) Practical aspects of the 2D ^{15}N-{^{1}H}-NOE experiment, J. Biomol. NMR ***23,*** 23-33.

Reverter, D., Fernandez-Catalan, C., Baumgartner, R., Pfander, R., Huber, R., Bode, W., Vendrell, J., Holak, T.A., Aviles, F.X. (2000) Structure of a novel leech carboxypeptidase inhibitor determined free in solution and in complex with human carboxypeptidase A2. Nat. Struct. Biol. **7,** 322-328.

Reverter, D., Ventura, S., Villegas, V., Vendrell, J., Aviles, F.X. (1998) Overexpression of human procarboxypeptidase A2 in *Pichia pastoris* and detailed characterization of its activation pathway. J. Biol.Chem. **273,** 3535-3541.

Reznik, S.E., Fricker, L.D. (2001) Carboxypeptidases from A to Z: implications in embryonic development and Wnt binding. Cell Mol. Life Sci. **58,** 1790-1804.

Rice, K., Peralta, R., Bast, D., Azavedo, J. and McGavin, M. J. (2001) Description of *staphylococcus* serine protease (*ssp*) operon in *Staphylococcus aureus* and nonpolar inactivation of *sspA*encoded serine protease. Infect. Immun. **69,** 159-169.

Rieneck, K., Renneberg, J., Diamant, M., Gutschik, E., Bendtzen, K. (1997) Molecular cloning and expression of a novel *Staphylococcus aureus* antigen. Biochim. Biophys. Acta. **1350,**128-32.

Rinderknecht, E. and, Humbel, R.E. (1976) Amino-terminal sequences of two polypeptides from human serum with nonsuppressible insulin-like and cell-growth promoting activities: evidence for structural homology with insulin B chain. Proc. Natl. Acad. Sci. USA, **73,** 4379-81.

Rinderknecht, E. and, Humbel, R.E. (1978a) The amino acid sequences of human insulin-like growth factor-I and its structural homology with proinsulin. J. Biol. Chem. **253,** 2769-76.

Rinderknecht, E. and Humbel, R.E. (1978b) Primary structure of human insulin-like growth factor II. FEBS Lett. **89,** 283-6.

Rivero, F., Köppel, B., Peracino, B., Bozzaro, S., Siegert, F., Weijer, C.J., Schleicher, M., Albrecht, R. and Noegel, A.A. (1996) The role of the cortical cytoskeleton: F-actin crosslinking proteins protect against osmotic stress, ensure cell size, cell shape and motility, and contribute to phagocytosis and development J. Cell Sci. **109**, 2679–2691.

Roberts, R.R., Scott, R.D. 2nd, Cordell, R., Solomon, S.L., Steele, L., Kampe, L.M., Trick, W.E., Weinstein, R.A. (2003) The use of economic modeling to determine the hospital costs associated with nosocomial infections. Clin. Infect. Dis. **36,** 1424-32.

Robertson, S.P. *et al.* (2003) Localized mutations in the gene encoding the cytoskeletal protein filamin A cause diverse malformations in humans. Nature Genet. **33,** 487–491.

Ruoppolo, M., Freedman, R.B., Pucci, P. and Marino, G. (1996) Glutathione-dependent pathways of refolding of RNase T1 by oxidation and disupfide isomerization: catalysis by protein disulfide isomerase. Biochemistry **35,** 13636-13646.

Rydel, T.J., Ravichandran, K.G., Tulinsky, A., Bode, W., Huber, R., Roitsch, C., Fenton, J.W. II (1990) The structure of a complex of recombinant hirudin and human alpha-thrombin. Science **249,** 277-280.

Rzychon, M., Sabat, A., Kosowska, K., Potempa, J. and Dubin, A. (2003) Staphostatins: an expanding new group of proteinase inhibitors with a unique specificity for the regulation of staphopains, *Staphylococcus* spp. cysteine proteinases, Mol. Microbiol. **49,** 1051-1066.

Sader, H.S., Jones, R.N., Gales, A.C., Silva, J.B., Pignatari, A.C., SENTRY Participants Group (Latin America). (2004) SENTRY antimicrobial surveillance program report: Latin American and Brazilian results for 1997 through 2001. Braz J Infect Dis. **8,** 25-79.

Sakano, K., Enjoh, T., Numata, F., Fujiwara, H., Marumoto, Y., Higashihashi, N., Sato, Y., Perdue, J.F. and Fujita-Yamaguchi Y. (1991) The design, expression, and characterization of human insulin-like growth factor II (IGF-II) mutants specific for either the IGF-II/cation- independent mannose 6-phosphate receptor or IGF-I receptor J. Biol. Chem. **31,** 20626-35.

Salzet, M. (2001) Anticoagulants and inhibitors of platelet aggregation derived from leeches. FEBS Lett. **492,** 187-192.

Sampson, L.J., Leyland, M.L., Dart, C. (2003) Direct interaction between the actin-binding protein filamin-A and the inwardly rectifying potassium channel, Kir2.1. J. Biol. Chem. **278,** 41988-97.

Seck, T., Baron, R., Horne, W.C. (2003) Binding of filamin to the C-terminal tail of the calcitonin receptor controls recycling. J. Biol. Chem. **278,** 10408-16.

Schneider, T.R., Sheldrick, G.M. (2002) Substructure solution with SHELXD, Acta Cryst. D **58,**1772-1779.

Senn, H., Eugster, A., Otting, G., Suter, F. and Wüthrich, K. (1987) 15N-labeled P22 c2 repressor for nuclear magnetic resonance studies of protein-DNA interactions. Eur. Biophys. J. **14**, 301–316.

Schedlich, L.J., O'Han, M.K., Leong, G.M. and Baxter, R.C. (2004) Insulin-like growth factor binding protein-3 prevents retinoid receptor heterodimerization: implications for retinoic acid-sensitivity in human breast cancer cells. Bioch. Bioph. Res. Comm. **314,** 83-8.

Scheraga, H.A., Wedemeyer, W.J. and Welker, E. (2001) Bovine pancreatic ribonuclease A: oxidative and conformational folding studies. Methods Enzymol. **341,** 189-221.

Schlechter, N.L., Russel, S.M., Spencer, E.M. and Nicoll, C.S. (1986) Evidence suggesting that the direct growth-promoting effect of growth hormone on cartiliage in vivo is mediated by local production of somatomedin. Proc. Natl. Acad. Sci. **83,** 7932-4.

Schwaiger, I., Kardinal, A., Schleicher, M., Noegel, A.A., Rief, M. (2004) A mechanical unfolding intermediate in an actin-crosslinking protein. Nat. Struct. Mol. Biol. **11,** 81-5.

Schwaiger, I., Schleicher, M., Noegel, A.A., Rief, M. (2005) The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. EMBO Rep. **6,** 46-51.

Shand, J.H., Beattie, J., Song, H., Phillips, K., Kelly, S.M., Flint, D.J. and Allan, G.J. (2003) Specific amino acid substitutions determine the differential contribution of the N- and C-terminal domains of insulin-like growth factor (IGF)-binding protein-5 in binding IGF-I. J. Biol. Chem. **278,** 17859-66.

Sheen, V. L. *et al.* (2001) Mutations in the X-linked filamin 1 gene cause periventricular nodular heterotopia in males as well as in females. Hum. Mol. Genet. **10,** 1775–1783.

Silveira, A., Schatteman, K., Goossens, F., Moor, E., Scharpe, S., Stromqvist, M., Hendriks, D. and Hamsten, A. (2000) Plasma procarboxypeptidase U in men with symptomatic coronary artery disease. Thromb. Haemost. **84**, 364-368.

Skelton, N.J., Palmer, A.G., Akke, M., Kordel, J., Rance, M. and Chazin, W.J. (1993) Practical Aspects of Two-Dimensional Proton-Detected 15N Spin Relaxation Measurements, J. Magn. Reson. B **102,** 253-264.

Sklenar, V., Piotto, M., Leppik, R. and Saudek, V. (1993) Gradient-Tailored Water Suppression for 1H-15N HSQC Experiments Optimized to Retain Full Sensitivity, J. Magn. Reson. A **102,** 241-245.

So, A.K., Varisco, P.A., Kemkes-Matthes, B., Herkenne-Morard, C., Chobaz-Péclat, V., Gerster, J.C., Busso, N. (2003) Arthritis is linked to local and systemic activation of coagulation and fibrinolysis pathways. J. Thromb. Haemost.**1,** 2510-2515.

Sokol, N. and Cooley, L. (1999) *Drosophila* filamin encoded by the *cheerio* locus is a component of ovarian ring canals. Curr. Biol. **9,** 1221–1230.

Springman, E.B., Dikov, M.M., Serafin, W.E. (1995) Mast cell procarboxypeptidase A. Molecular modeling and biochemical characterization of its processing within secretory granules. J. Biol. Chem. **270,**1300-1307.

Stahlhut, M., van Deurs, B. (2000) Identification of filamin as a novel ligand for caveolin-1: evidence for the organization of caveolin-1-associated membrane domains by the actin cytoskeleton. Mol. Biol. Cell. **11,** 325-37.

Stefanova, M., Meinecke, P., Gal, A., Bolz, H. (2005) A novel 9 bp deletion in the filamin a gene causes an otopalatodigital-spectrum disorder with a variable, intermediate phenotype. Am. J. Med. Genet. A. **132**, 386-90.

Stevenson, V.A. and Theurkauf,W.E. (2000) Actin cytoskeleton: putting a CAP on actin polymerization. Curr. Biol., **10,** R695–697.

Stossel, T.P., Condeelis, J., Cooley, L., Hartwig, J.H., Noegel, A.A., Schleicher, M. and Shapiro, S.S. (2001) Filamins as integrators of cell mechanics and signalling. Nature Rev. Mol. Cell Biol. **2,** 138-145.

Suetsugu, S. and Takenawa, T. (2003) Translocation of N-WASP by Nuclear Localization and Export Signals into the Nucleus Modulates Expression of HSP90. J. Biol. Chem. **278,** 42515-42523.

Suh, D.Y., Hunt, T.K. and Spencer E.M. (1992) Insulin-like growth factor-I reverses the impairment of wound healing induced by corticosteroids in rats. Endocrinology, **131,** 2399-406.

Tabor, S. and Richardson, C.C. (1992) A bacteriophage T7 RNA polymerase/promotor system for controlled exclusive expression of specific genes. Biotechnology. **24**, 280-284.

Takahashi, M., Tezuka, T. and Katunuma, N. (1994) Inhibition of growth and cysteine proteinase activity of *Staphylococcus aureus* V8 by phosphorylated cystatin R in skin cornified envelope, FEBS Lett. **355,** 275-278.

Tigges, U., Koch, B., Wissing, J., Jockusch, B.M., Ziegler, W.H. (2003) The F-actin cross-linking and focal adhesion protein filamin A is a ligand and in vivo substrate for protein kinase C alpha. J. Biol. Chem. **278,** 23561-9.

Torres, A.M., Forbes, B.E., Aplin, S.E., Wallace, J.C., Francis, G.L. and Norton, R.S. (1995) Solution structure of human insulin-like growth factor II. Relationship to receptor and binding protein interactions. J. Mol. Biol. **248,** 385-401.

Torres, A.M., Kuchel, P.W. (2004) The β-defensin-fold family of polypeptides. Toxicon **44,** 581-588

Travis, M.A., van der Flier, A., Kammerer, R.A., Mould, A.P., Sonnenberg, A., Humphries, M.J. (2004) Interaction of filamin A with the integrin beta 7 cytoplasmic domain: role of alternative splicing and phosphorylation. FEBS Lett. **569,** 185-90.

Tu, Y., Wu, S., Shi, X., Chen, K., Wu, C. (2003) Migfilin and Mig-2 link focal adhesions to filamin and the actin cytoskeleton and function in cell shape modulation. Cell. **113,** 37-47.

Tyler, J.M., Anderson, J.M., Branton, D. (1980) Structural comparison of several actin-binding macromolecules. J. Cell Biol. **85,** 489-95.

Vadlamudi, R.K., Li, F., Adam, L., Nguyen, D., Ohta, Y., Stossel, T.P., Kumar, R. (2002) Filamin is essential in actin cytoskeletal assembly mediated by p21-activated kinase 1. Nat. Cell Biol. **4,** 681-90.

van de Locht, A., Lamba, D., Bauer, M., Huber, R., Friedrich, T., Kroger, B., Hoffken, W., Bode, W. (1995) Two heads are better than one: crystal structure of the insect derived double domain Kazal inhibitor rhodniin in complex with thrombin. EMBO J. **14,** 5149-5157.

van de Locht, A., Stubbs, M.T., Bode, W., Friedrich, T., Bollschweiler, C., Höffken, W., Huber, R. (1996) The ornithodorin-thrombin crystal structure a key to the TAP enigma?. EMBO J **15,** 6011-6017.

van den Berg, B., Chung, E.W., Robinson, C.V., Mateo, P.L. and Dobson, C.M. (1999a) The oxidative refolding of hen lysozyme and its catalysis by protein disulfide isomerase. EMBO J. **18**, 4794-4803.

van den Berg, B., Chung, E.W., Robinson, C.V. and Dobson, C.M. (1999b) Characterisation of the dominant oxidative folding intermediate of hen lysozyme. J. Mol. Biol. **290**, 781-796.

van Tilburg, N.H., Rosendaal, F.R., Bertina, R.M. (2000) Thrombin activatable fibrinolysis inhibitor and the risk for deep vein thrombosis. Blood **95,** 2855-2859.

Vendrell, J., Querol, E., Aviles, F.X. (2000) Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. Biochim Biophys. Acta **1477**, 284-298.

Walker, J.B., Hughes, B., James, I., Haddock, P., Kluft, C., Bajzar, L. (2003) Stabilization versus inhibition of TAFIa by competitive inhibitors *in vitro*. J. Biol. Chem. **278,** 8913-8921.

Wang, W., Boffa, M.B., Bajzar, L., Walker, J.B., Nesheim, M.E. (1998) A study of the mechanism of inhibition of fibrinolysis by activated thrombin-activable fibrinolysis inhibitor. J. Biol. Chem. **273,** 27176-27181.

Weber, T. Baumgartner, R. Renner, C. Marahiel, M.A. and Holak, T.A. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure Fold. Des. **8,** *407-418.*

Wedemeyer, W.J., Welker, E., Narayan, M. and Scheraga, H.A. (2000) Disulfide bonds and protein folding. Biochemistry, **39,** 4207-4216.

Wei, A., Alexander, R.S., Duke, J., Ross, H., Rosenfeld, S.A., Chang, C.H. (1998) Unexpected binding mode of tick anticoagulant peptide complexed to bovine factor Xa. J. Mol. Biol. **283,** 147-154.

Weihing, R.R. The filamins: properties and functions. (1985) Can. J. Biochem. Cell Biol. **63,** 397–413.

Weissman, J.S. and Kim, P.S. (1991) Reexamination of the folding of BPTI: predominance of native intermediates. Science **253**, 1386-1393.

Winkler, J., Lunsdorf, H. and Jockusch, B.M. (1997) Flexibility and fine structure of smooth-muscle α-actinin. Eur. J. Biochem. **248,** 193-199.

Wu, C. (2005) Migfilin and its binding partners: from cell biology to human diseases. J. Cell Sci. **118,** 659-64.

Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, Wiley, New York, NY.

Varadi, D.P. and Saqueton, A.C. (1968) Elastase form *Staphylococcus epidermidis*, Nature **218,** 468-470.

Vendrell, J., Querol, E., Aviles, F.X. (2000) Metallocarboxypeptidases and their protein inhibitors. Structure, function and biomedical properties. Biochim. Biophys. Acta **1477,** 284-298.

Vendrell, J., Aviles, F.X., Fricker, L.D. (2004) Metallocarboxypeptidases in the book Handbook of Metalloproteins. Messerschmidt A, Bode W, Cygler M. Wiley and Sons. 167-192

Ventura, S., Villegas, V., Sterner, J., Larson, J., Vendrell, J., Hershberger, C.L., Aviles, F.X. (1999) Mapping the pro-region of carboxypeptidase B by protein engineering. Cloning, overexpression, and mutagenesis of the porcine proenzyme. J. Biol. Chem. **28,** 19925-19933.

Vuister, G.W. and Bax, A. (1993) Quantitative J correlationsa new approach for measuring homonuclear 3-bond J (H(N)H(R)) coupling constants in 15N-enriched proteins, J. Am. Chem. Soc. **115,** 7772-7777.

Yao, S.G., Headey, S.J., Keizer, D.W., Bach, L.A. and Norton, R.S. (2004) C-terminal domain of insulin-like growth factor (IGF) binding protein 6: Conformational exchange and its correlation with IGF-II binding. Biochemistry. **43,** 11187-95.

Yoshida, N., Ogata, T., Tanabe, K., Li, S., Nakazato, M., Kohu, K., Takafuta, T., Shapiro, S., Ohta, Y., Satake, M., Watanabe, T. (2005) Filamin A-bound PEBP2beta/CBFbeta is retained in the cytoplasm and prevented from functioning as a partner of the Runx1 transcription factor. Mol. Cell Biol. **25,** 1003-12.

Zeslawski, W., Beisel, H.G., Kamionka, M., Kalus, W., Engh, R.A., Huber, R., Lang, K. and Holak T.A. (2001) The interaction of insulin-like growth factor-I with the N-terminal domain of IGFBP-5. EMBO J. **20,** 3638-44.

Ziebandt, A.K., Becher, D., Ohlsen, K., Hacker, J., Hecker, M., Engelmann S. (2004) The influence of agr and sigmaB in growth phase dependent regulation of virulence factors in *Staphylococcus aureus*. Proteomics **4,** 3034-47.

Ziebandt, A.K., Weber, H., Rudolph, J., Schmid, R., Hoper, D., Engelmann, S., Hecker, M. (2004) Extracellular proteins of Staphylococcus aureus and the role of SarA and sigma B. Proteomics **1,** 480-93.

Zweckstetter, M. and Holak, T. A. (1999) Robust refocusing of 13C magnetization in multidimensional NMR experiments by adiabatic passage pulses, J. Biomol. NMR **15,** 331-334.