UNIVERSITA' DEGLI STUDI DI VERONA

DIPARTIMENTO SCIENTIFICO E TECNOLOGICO
FACOLTA' DI SCIENZE MM. FF. NN.

Dottorato di ricerca in
Biotecnologie Industriali ed Ambientali

CICLO 19°

EXPRESSION AND STRUCTURAL STUDIES OF INSULIN-LIKE GROWTH FACTOR BINDING PROTEINS AND THE C-TERMINAL DOMAIN OF PERLECAN

Coordinatore: Prof. Hugo L. Monaco

Firma _____

Supervisore: Prof. Hugo L. Monaco

Firma _____

Dottorando: Dott. Faggion Beniamino

Firma _____

INDEX

Riassunto

Insulin-like Growth Factors (IGFs), Insulin-like Growth Factor Binding Proteins (IGFBPs) e Insulin-like Growth Factor Binding Proteins-related Proteins (IGFBP-rP).

Le IGF-I e IGF-II sono piccoli ormoni peptidici di 70 e 67 amminoacidi rispettivamente. Essi sono così chiamati a causa della loro alta omologia con l'insulina. Le IGF hanno effetti a livello cellulare promuovendo la crescita e il differenziamento, vengono legate *in vivo* dalle IGFBP e da 2 tipi di recettori (recettore tipo 1 e tipo 2) che mediano i loro effetti. Le IGFBP costituiscono una famiglia di 6 proteine (da IGFBP-1 a -6) le cui dimensioni vanno da 216 a 289 amminoacidi di lunghezza. Dal punto di vista strutturale, esse sono suddivise in tre domini: un dominio ammino-terminale conservato ricco in cisteine, una regione centrale non conservata e un dominio carbossi-terminale conservato. La capacità di legare IGF permette alle IGFBP di regolare la concentrazione di IGF circolanti, di sequestrarle o di concentrarle in prossimità dei tessuti bersaglio. Le IGFBP però possono dare degli effetti biologici anche indipendentemente dal legame con le IGF. E' stato dimostrato, ad esempio, da studi condotti utilizzando frammenti delle IGFBP che questi sono in grado di influenzare l'adesione cellulare attraverso il legame con le integrine. Nel complesso, il sistema delle IGF è coinvolto in numerosi processi nell'organismo e sempre più patologie sono state correlate al suo malfunzionamento.

Successivamente, sono state identificate altre proteine che mostrano omologia per le IGFBP, limitatamente al dominio ammino-terminale, le quali possono legare le IGF con affinità inferiore rispetto alle IGFBP. Si è proposto quindi di chiamare queste proteine IGFBP-related proteins.

In questo lavoro di tesi la IGF-II umana ricombinante è stata espressa in *Escherichia coli* , inizialmente come proteina depositata in corpi d'inclusione e successivamente in forma solubile. La proteina prodotta in corpi d'inclusione è

stata solubilizzata, purificata mediante cromatografia di affinità (tramite una coda di 6 istidine legata alla proteina), ed è stata intrapresa la rinaturazione. La proteina così ottenuta è stata cristallizzata ma i cristalli ottenuti non hanno consentito di determinarne la struttura mediante analisi di diffrazione di raggi X. La proteina in forma solubile è stata ottenuta cambiando vettore di espressione ma la bassa specificità per la matrice di affinità usata ha posto problemi per la sua purificazione.

E' stata poi espressa la IGFBP-3 umana utilizzando il lievito metilotrofico *Pichia pastoris*. La proteina è stata purificata mediante cromatografia a scambio ionico e gel-filtrazione ed è stata tentata la cristallizzazione che però non ha dato risultati.

Inoltre è stata espressa e purificata la IGFBP-rP1 in *Escherichia coli*. La proteina è stata purificata mediante cromatografia di affinità (utilizzando una resina Nickel Sepharose che lega la coda di istidine fusa alla proteina ricombinante), e utilizzata per l'allestimento delle prove di cristallizzazione preliminari. Una delle condizioni testate ha dato dei cristalli che però non possono ancora essere usati per risolvere la struttura tridimensionale.

Dominio LG-3 (Laminin G-like) del perlecano umano.

Il perlecano umano è un eparan-solfato proteoglicano, costituito da un nucleo proteico del peso di 470 kDa legato a oligosaccaridi e catene di eparan solfato che portano il peso complessivo del perlecano fino a 800 kDa. Esso costituisce, insieme a collagene e laminine la membrana basale, cioè una matrice specializzata presente nell'endotelio, nell'epitelio e attorno alle cellule muscolari. Il dominio carbossi-terminale del perlecano (dominio V del peso di 85 kDa) è chiamato endorepellina ed è stata dimostrata la sua capacità di funzionare come fattore anti-angiogenico.

Il dominio LG-3 del perlecano umano è stato espresso come proteina ricombinante in *Escherichia coli* e purificato tramite cromatografia di affinità sfruttando una coda di istidine. La proteina è stata quindi cristallizzata ed è stata determinata la sua struttura tridimensionale mediante analisi di diffrazione di raggi X.

α_1-Microglobulina.

L'α_1-microglobulina è una lipocalina, è costituita da 183 amminoacidi e pesa 26 kDa. L'α_1-microglobulina umana è glicosilata in tre posizioni, due delle quali sono siti di N-glicosilazione e una e un sito di O-glicosilazione. Essa viene sintetizzata nel fegato come proteina di fusione insieme ad un'altra proteina, la bikunina, e viene successivamente tagliata e immessa nel plasma. Una importante funzione della α_1-microglobulina è la sua capacità di modulare negativamente l'attività del sistema immunitario proteggendo i tessuti dalla risposta immunitaria.

L' α_1-microglobulina di topo è stata espressa come proteina ricombinante utilizzando il lievito *Pichia pastoris*. La proteina è stata purificata tramite cromatografia a scambio ionico e gel filtrazione. La proteina purificata è stata quindi utilizzata per l'allestimento delle prove preliminari di cristallizzazione che però non hanno dato risultati.

Abstract

Insulin-like Growth Factors (IGFs), Insulin-like Growth Factor Binding Proteins (IGFBPs) and Insulin-like Growth Factor Binding Proteins-related Proteins (IGFBP-rP).

IGF-I and IGF-II are small hormone peptides of 70 and 67 amino acids respectively and are called insulin-like because of their homology with insulin. IGFs promote cellular growth and differentiation, bind IGFBPs *in vivo* and also two types of receptors (type 1 and type 2 receptor) which mediate their actions. The IGFBPs form a six member family (IGFBP-1 to -6) and are between 216 to 289 amino acids long. They are subdivided into three distinct domains: a conserved amino-terminal cysteine-rich region, a non-conserved midregion, and a conserved carboxy-terminal region. The IGFBPs bind the circulating IGFs thus regulating their concentration, by sequestering them or releasing them in proximity of their target tissues. In addition, IGFBP effects independent of IGF-binding have been demonstrated. Studies with proteolyzed fragments of IGFBPs unable to bind IGFs, have shown their ability to influence cellular adhesion by integrin binding. The IGF system is involved in various processes and many pathological events have been related to its failure.

In addition, other proteins have been identified which have the ability to bind IGFs with lower affinity, compared to IGFBPs. These proteins show homology with IGFBPs in their amino-terminal region and are called IGFBP-related proteins.

During this thesis, human IGF-II was expressed in *Escherichia coli*, in the form of inclusion bodies and in its soluble form. The protein from inclusion bodies was solubilized, purified by affinity chromatography (using a 6 histidine tag), and then refolded. The renatured protein was crystallized but with the data obtained from those crystals it was not possible to determine the three dimensional structure. The soluble form was obtained by using a different expression vector and a 6

histidine tag but purification using the tag was not possible due to the low affinity for the chromatography resin.

Human IGFBP-3 was also expressed using the methylotrophic yeast *Pichia pastoris*. The recombinant protein was purified by ion-exchange chromatography and by gel-filtration but the crystallization trials did not yield good crystals.

Human IGFBP-rP1 was expressed in *Escherichia coli*, purified by affinity chromatography and used for preliminary crystallization trials. One crystallization condition gave small crystals which are still not suitable for X-ray diffraction analysis.

LG-3 (Laminin G like) domain of the human perlecan

Human perlecan is an heparan-sulphate proteoglycan, containing a protein core of 470 kDa with oligosaccharides and heparan sulphate chains attached to it which give a total molecular weight of 800 kDa. Together with collagen and laminin it forms the basement membrane which is a specialized matrix found in epithelium, endothelium and surrounding muscle cells. The carboxy-terminal domain of perlecan (domain V of 85 kDa) is called endorepellin and possesses anti-angiogenic activity

The LG-3 domain of human perlecan was expressed in *Escherichia coli* and purified by affinity chromatography. The purified protein was then crystallized and its molecular structure was determined by X-ray diffraction analysis.

α_1-Microglobulin.

α_1-Microglobulin is a lipocalin of 26 kDa and is 183 amino acids long. Human α_1-microglobulin is glycosylated in three positions: two of these are of the N-linked type and the third one is of the O-linked type. α_1-Microglobulin is synthesized in the liver as a fusion protein with bikunin from which it is

subsequently cleaved and released into plasma. The biological function of α_1-microglobulin is related to its ability to negatively modulate the immune system activity thus preventing tissue damage by immunological responses.

Mouse α_1-microglobulin was expressed as a recombinant protein by using the yeast *Pichia pastoris*. The protein was purified by ion-exchange chromatography and gel filtration and preliminary crystallization trials were prepared which gave no crystals.

Abbreviations:

ALS: Acid Labile Subunit
BM: Basement Membranes
BMGY : Buffered Glycerol-complex Medium
BMMY: Buffered Methanol-complex Medium
BMP-1: Bone Morphogenetic Protein 1
CAM: Chorioallantanoic Membrane
CCP4: Collaborative Computational Project 4
CHO cells: Chinese Hamster Ovary cells
EBV: Epstein-Barr Virus
EBNA: EBV Nuclear Antigen
ECM: Extracellular Matrix
EDTA: ethylen-diaminotetraacetic acid
EGF-like: Epidermal Growth Factor-like
FGF2: Fibroblast Growth Factor-2
FSH: Follicle-Stimulating Hormone
GAG: Glygosaminoglycan
HUMVEC: Human Uterine Microvascular Endothelial Cells
HS: Heparan Sulphate
HSPG: Heparan Sulphate Proteoglycan
IB: Inclusion Bodies
IGF: Insulin-like Growth Factor
IGFBP: Insulin-like Growth Factor Binding Protein
IGFBP-rPs: IGFBP-related Proteins
IGFRI: IGF Receptor type 1
IGFRII: IGF Receptor type 2
IPTG: isopropyl-β-D-1-thiogalactopyranoside
LB: Luria-Bertani
LG: Laminin G-like
M6PR: Mannose-6-phosphate Receptor
MEF: Mouse Embryonic Fibroblast
MMP : Matrix Metalloproteinase
mTLL-1: mammalian Tolloid-like 1
PAPP-A: Pregnancy-associated Plasma Protein-A
PDGF: Platelet-derived Growth Factor
PDK: Perlecan-deficient Keratinocytes
RGD (Arg-Gly-Asp): Integrin-binding Motif
RZPD: Deutsches Ressourcenzentrum fur Genomforschung
SDS-PAGE: Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
TB: Terrific Broth
TBE (Tris Borate EDTA): 45 mM Tris-Borate pH 8.0, 1 mM EDTA
TBS (Tris Buffered Saline): Tris/HCl 20 mM pH 7.5, 0.5 M NaCl, and 0.02% NaN_3
TGF-β: Tumor Growth Factor-β
VEGF: Vascular Endothelial Growth Factor

1. Introduction

1.1 The IGF system

The insulin-like growth factor (IGF) system includes several components IGFs (IGF-I and IGF-II), IGF-binding proteins (IGFBPs), type I and type II IGF receptors, and IGFBP proteases.

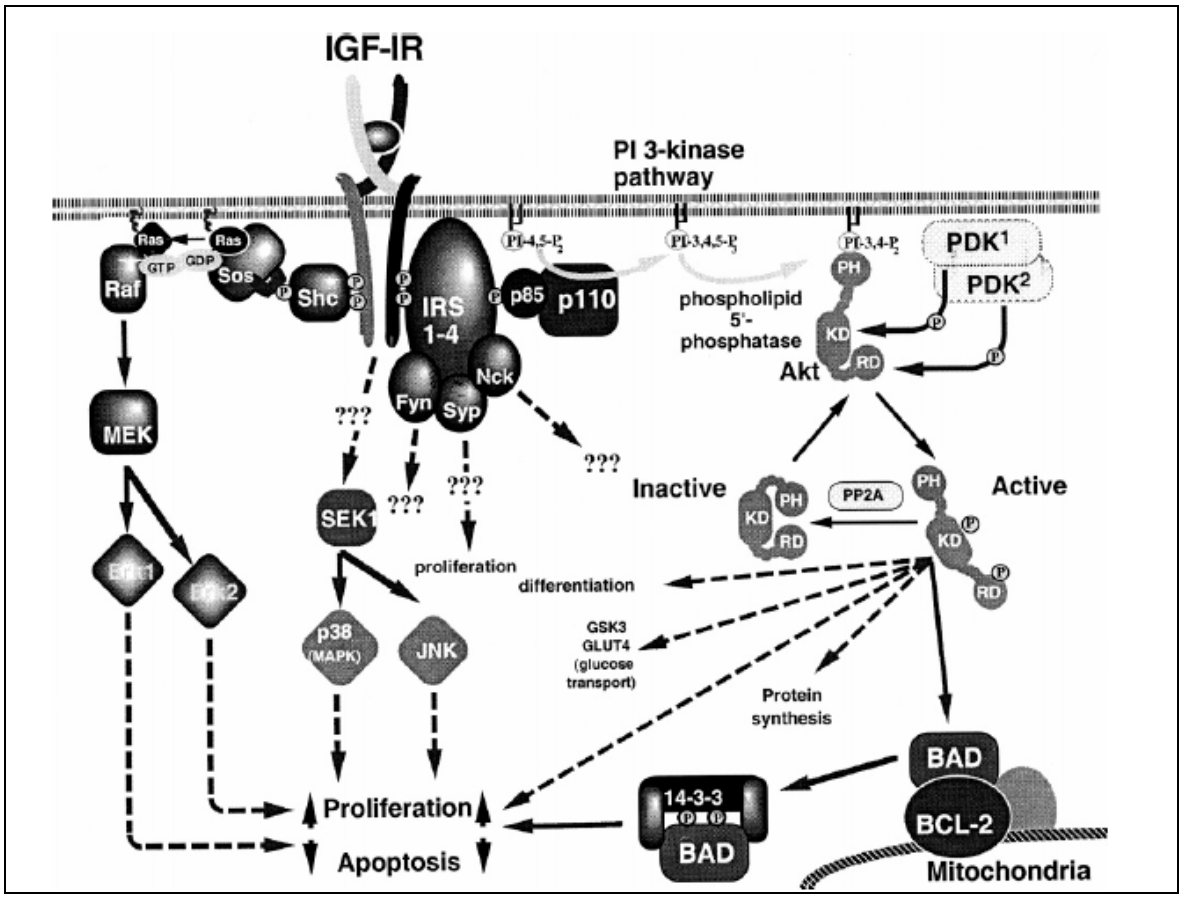


Fig. 1. Representation of IGF-IR and downstream signalling cascade.

Both IGF-I and IGF-II show structural similarity to insulin and are small hormone peptides; they can affect cell growth and metabolism exerting their effects through endocrine actions [1] and through paracrine and autocrine effects on cell proliferation [2]. IGF actions are mediated by specific cell surface receptors termed type I and type II receptors. Type I receptor in particular is responsible for the mitogenic effects of IGFs; while type II receptor interacts with IGF-II only and its

main function is related to IGF-II internalization and degradation. IGFBPs bind *in vivo* circulating IGFs; their functions can be subdivided into IGF-dependent, involving IGF binding and release, and IGF-independent, due to IGFBPs interaction with other cellular elements.

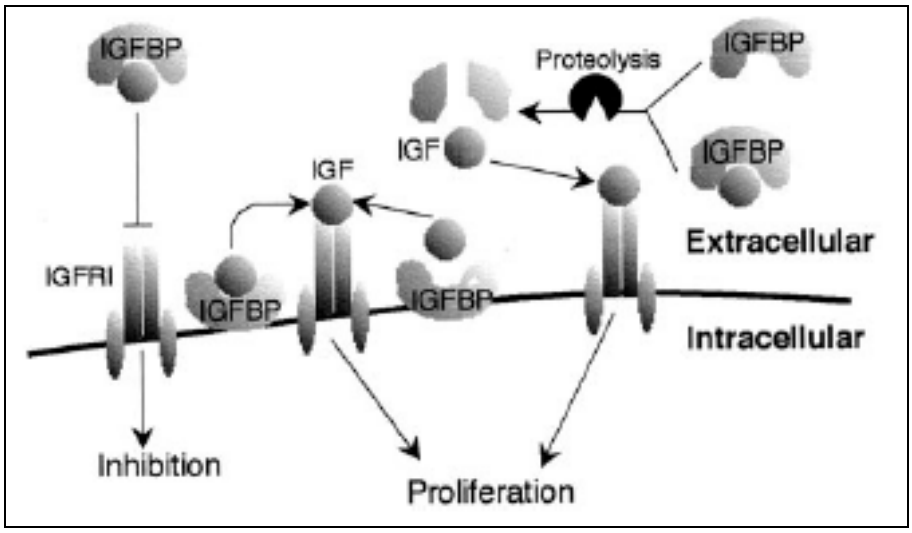


Fig. 2. Schematic representation of IGFBP-dependent actions. IGFBPs are shown to inhibit IGF effects mediated by IGF-I-R. On the contrary, IGFBP proteolytic cleavage releases IGFs thus increasing bioavailability.

Mammalian IGFBPs form a six member family and they have higher affinity for IGFs ($k_d \approx 10^{-10}$ M) than for type I receptors ($k_d \approx 10^{-8}$–10^{-9} M). In addition there is another group of cysteine-rich proteins that show structural and functional similarities to the IGFBPs: the IGFBP-rPs (IGFBP-related proteins).

1.1.2 Insulin-like growth factors (IGFs)

IGF-I and IGF-II are 70 and 67 amino acid peptides with high degree of homology with insulin and with the ability to interact with the insulin receptor. IGF-I and IGF-II are involved in fetal development and can have long term growth-promoting effects as regulators of cell proliferation. The biological functions of IGFs are mediated by 2 receptors, type 1 and type 2, and by the insulin receptor.

Type 1 receptor is structurally similar to the insulin receptor and binds IGF-I with high affinity, IGF-II with slightly lower affinity and insulin with low affinity. Type 2 receptor binds IGF-II with internalization and degradation. The biological effects of IGF-II arise from binding to the type 2 receptor: ligand binding elicits specific tyrosine kinase activity and a protein phosphorylation cascade that may activate other intracellular messenger systems.

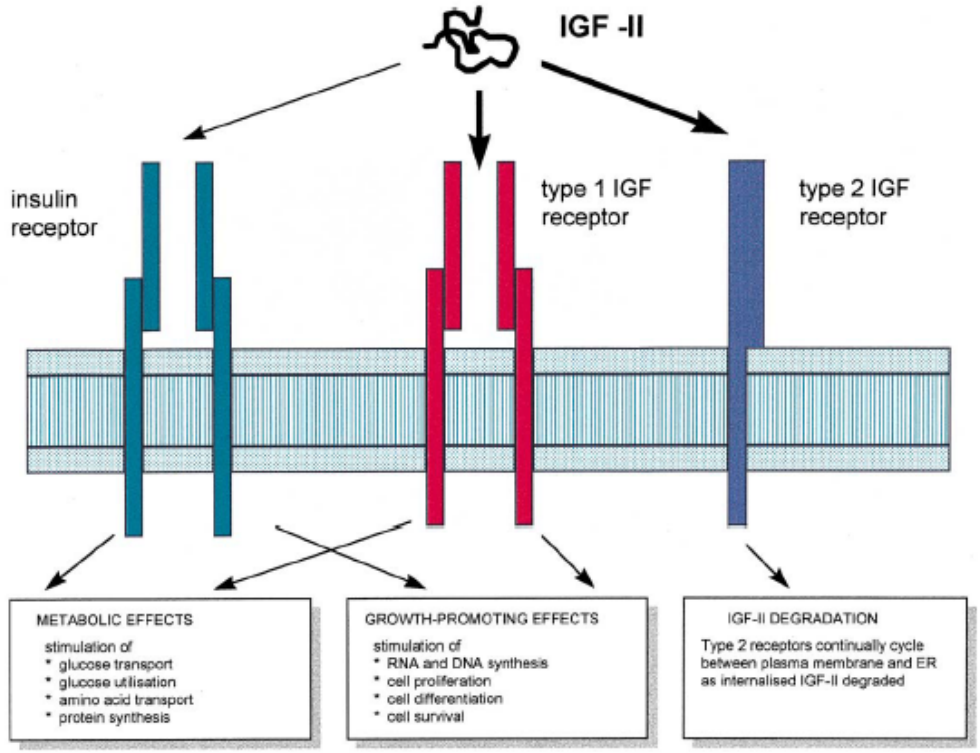


Fig. 3. Representation of IGF-II interaction with receptors.

1.1.3 IGFBPs

All six IGFBPs in their precursor form possess secretory signal peptides ranging from 20 up to 39 residues and the mature proteins are found extracellularly. They are between 216 and 289 amino acids long, corresponding to molecular masses ranging from 22 to 31.3 kDa. All of the IGFBPs contain in their N-terminal domain

six disulfide bonds, with the exception of IGFBP-6 which has five. The N-terminal domain has been shown to contain residues involved in IGF binding. The carboxy-terminal domain has three disulfide bonds in all IGFBPs. In this domain important residues for IGF binding are also present which implies that IGFBPs possess an IGF-binding pocket, shared by both conserved domains. The carboxy-terminal domain contains also an Arg-Gly-Asp (RGD) integrin-binding motif located at residues 221-223 of IGFBP-1 [3] and the chain spanned by residues 265-267 of IGFBP-2; IGFBP-3 and IGFBP-5 contain an 18-residue basic motif with heparin-binding activity.

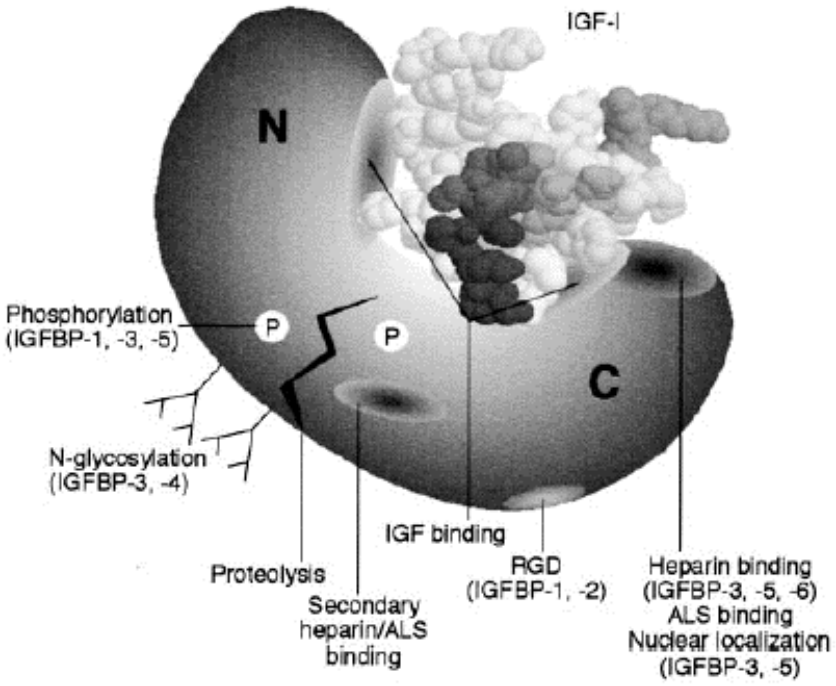


Fig. 4. General features of IGFBPs including IGF binding pocket and posttranslational modifications

The three distinct domains of mammalian IGFBPs are almost equivalent in size:

1. *N-terminal domain.* In mature IGFBPs, the first 80-93 amino acid residues after the signal peptides constitute the N-terminal domain in which there is approximately 58% similarity among the members of the family. In

IGFBP-1 to -5 there are 12 fully conserved cysteines and in IGFBP-6 10 of the 12 cysteines are invariant. Such a high number of cysteines within a small domain gives a highly structured domain with 6 disulfide bridges (5 in the case of IGFBP-6).

2. *Midregion.* The midregion is a portion of the protein ranging in size between 55 to 95 amino acid residues which separates the other two domains: the N-terminal and C-terminal domains. While the other two domains are conserved, the midregion shows a similarity of less than 15% and is therefore unique for the specific protein. Interestingly, the midregion seems the only portion of the protein subject to posttranslational modifications (glycosylation and phosphorylation); in particular, IGFBP-3 and -4 are N-glycosylated, and IGFBP-5 and -6 are O-glycosylated. The asparagine residue on which glycosylation occurs must be part of a consensus sequence Asn-X-Ser/Thr, where X is any amino acid except proline. Three of these consensus sequences have been found in IGFBP-3 and one in IGFBP-4. Even if there are no consensus sequences for predicting O-glycosylation, it has been demonstrated that IGFBP-5 and IGFBP-6 are both O-glycosylated. Another important posttranslational modification is phosphorylation which is used by cells to regulate the activities of numerous intracellular proteins, including proteins involved in signal transduction pathways, in the cell cycle and in gene expression. IGFBP-1, -3, and -5 are modified by phosphorylation predominantly at serine residues found in the midregion.
3. *C-terminal region.* Like the N-terminal domains, the C-terminal domains are highly conserved and contain the remaining 6 cysteines which are strictly conserved.

1.1.4 Modulation of IGF activity by IGFBPs and their proteases

The effects of IGFBP modulation cannot be easily summarized and have been reported by several groups independently; the following is a partial list of the multifunctional roles of IGFBPs.

IGF-I is a regulator of migration, proliferation, and apoptosis in vascular smooth muscle cells; in porcine vascular smooth muscle cells IGFBP-4 and IGFBP-5 expression is differentially regulated by IGF-I [4]. The peptide decreases IGFBP-4 levels by activating an IGFBP-4-specific protease and increases IGFBP-5 levels by stimulating gene expression. In addition, IGF-I induced DNA synthesis in these cells is regulated in opposite fashions by IGFBP-4 and IGFBP-5; IGFBP-4 inhibits whereas IGFBP-5 potentiates IGF-I effects. A calcium-dependent serine protease induced by IGF-I and secreted by smooth muscle cells has been found, which cleaves IGFBP-4 into fragments with low affinity for IGF-I [5]. The major cleavage site is Lys^{120}-His^{121} which generates a 16 kDa amino-terminal fragment. A mutant of IGFBP-4 which is protease-resistant, Asn^{120}-Asn^{121}, has been been shown to have effects on inhibition of DNA synthesis, cell migration, and muscle growth similarly to intact IGFBP-4 [6]. Another IGFBP-4 protease has been purified from human fibroblasts and identified as pregnancy-associated plasma protein-A (PAPP-A) and it specifically cleaves IGFBP-4 on Met^{135}-Lys^{136}. It has been demonstrated that PAPP-A levels increase in injured vascular smooth muscle cell cultures and injured arteries *in vivo* and might stimulate IGF-I release from IGFBP-4 thus showing its role in the repair of arterial injury.

As noted before IGFBP-5 increases the effects of IGF-I. It has been shown that in smooth muscle cells IGFBP-5 possesses the ability to bind to the extracellular matrix (ECM) through a charged region of ten basic amino acids in residues 201-218. In human fibroblasts, ECM-bound IGFBP-5 has a 7-fold loss of IGF-I affinity.

This suggests that the potentiation of IGF-I effects may be due to its release after sequestration and concentration by IGFBP-5. An analogous situation is found in bone cells. IGFBP-5 has a high affinity for extra cellular matrix (ECM) proteins and hydroxyapatite, the mineral constituent of bone, and by specific binding of IGF which is thought to potentiate its proliferative actions on osteoblastic cells [7]. Therefore it has been proposed that the IGFs, sequestered and concentrated by IGFBP-5, stimulate osteoblastic activity during bone remodeling after release during osteoclastic resorption.

1.1.5 IGF-dependent and IGF-independent effects of IGFBPs

In addition to IGF action modulation, IGFBPs have the ability to exert intrinsic bioactivities in the absence or in the presence of IGFs but without triggering IGF receptor type 1 (IGFRI) signalling. There are reports of IGF-independent actions either by intact IGFBPs or by proteolyzed fragments distinct from their effect due to the binding to IGFs.

One of the first pieces of evidence was found for IGFBP-1 which is able to affect cell motility and adhesion. It has been demonstrated that the RGD integrin-binding motif present in the C-terminal domain of IGFBP-1 interacts with $\alpha_5\beta_1$ integrin and is responsible for this effect [8]. While IGFBP-4 and IGFBP-6 are thought to act primarily through IGF modulation, the majority of reports on IGF-independent effects of IGFBPs have focused on IGFBP-3.

A demonstration of these effects has been given by an experiment in which a mouse fibroblast cell line derived from IGFRI-knockout mice showed growth inhibition after IGFBP-3 transfection [9]. This is important evidence, even if the study could not exclude that IGF action could have been mediated by a different receptor. Further demonstration of IGFBP-3 intrinsic activity is given by the

evidence that not only the intact protein is active, but also are its proteolyzed fragments with little or no affinity for IGFs. Nonglycosylated IGFBP-3 digested with plasmin results in a 22- to 25-kDa fragment with 50-fold loss in IGF-I binding affinity and a 16-kDa fragment which does also not bind IGF-I [10]. The 16-kDa fragment has the same effect in inhibiting IGF-I induced DNA synthesis in chicken embryo fibroblasts as intact IGFBP-3. In addition, the 16-kDa fragment (and not intact IGFBP-3) inhibits the mitogenic action of fibroblast growth factor on both wild-type and IGFRI knockout mice. Other experiments support the role of IGFBP-3 amino-terminal fragments as IGF-independent effectors. A semipurified mixture of plasmin digested IGFBP-3 containing a 20-kDa amino-terminal and a carboxy-terminal fragments is able to stimulate glucose uptake by microvessel endothelial cells, while intact IGFBP-3 is not. It is not clear if the 20 kDa fragment is the same as the 16 kDa described above and the mechanism by which the fragments alone can exert the bioactivity still has to be elucidated.

IGFBP-3 has the ability to bind IGFs but also other ligands. It has been shown that in serum the binary complex IGF-IGFBP-3 can bind ALS (Acid Labile Subunit) and hence form a heterotrimeric complex [11]. The formation of this complex is believed to prevent the efflux of the IGFs from the blood vessels thus regulating IGF bioavailability to target tissues. It has also been demonstrated that residues 228-232 of IGFBP-3 play an important role in ALS binding: these residues are found in a 18-residue basic domain involved in cell binding by IGFBP-3 [12]; it is possible that those components which mediate IGFBP-3 binding are involved in ALS dissociation promoting IGF release to target tissues. IGF-independent effects have been also described for IGFBP-5. It has been reported that a 23-kDa IGFBP-5 fragment can stimulate normal osteoblast mitogenesis in the absence of IGF-I [13]. To confirm this observation a recombinant amino-terminal fragment was

produced and found to have the ability to stimulate mitogenesis by low-affinity binding to the osteoblast surface [14].

IGFBPs have also been shown to mediate important cellular actions like cell adhesion and migration. Adhesion to, and release from, ECM (Extra Cellular Matrix) is a complex process which involves cytoskeletal rearrangement, and cell motility, and is also related to tumor invasiveness [15]. Although IGF-dependent effects on cell migration are well-known, it has been subsequently demonstrated that IGFBPs possess this ability in an IGF-independent fashion. Integrins are receptors of cell adhesion which transduce extracellular signals through phosphorylation cascades and through direct connection to cytoskeletal elements. One of the first observations concerning IGF-independent effects of IGFBPs on cell adhesion and motility involves the presence of an RGD integrin binding motif in the carboxy-terminal domain of IGFBP-1 and IGFBP-2. The role of the RGD domain has been elucidated in Chinese hamster ovary (CHO) cells, transfected to express human IGFBP-1, which show increased migration. Furthermore, cells transfected with a mutated form of IGFBP-1 containing WGD in place of the ^{221}RGD motif were unaffected. In addition, the presence of a synthetic peptide containing the RGD sequence blocked the effects of the wild-type IGFBP-1. Affinity chromatography on immobilized IGFBP-1 proved that $\alpha_5\beta_1$ integrin is the interacting partner responsible for cell migration. After the demonstration that IGFBPs exert biological actions upon binding to the cell surface, the question of the role of post-translational modifications in modulating IGFBP effects was raised. IGFBP-3 possesses three N-glycosylation sites: the first two Asn^{89} and Asn^{109} carry 4.5-5 kDa carbohydrate moieties and the third, Asn^{172}, contains either undetectable or a 5 kDa saccharide chain [16]. While IGFBP-1, -5, and -6, present O-glycolsylation, IGFBP-3 and IGFBP-2 appear not to on the basis of their primary structure prediction.

1.1.6 Physiological roles of IGFBP-3

IGFBP-3 is responsible for binding about 75% of circulating IGFs and modulates IGF activity both sequestering IGFs and concentrating and releasing them in the proximity to their target tissues. IGF-independent actions are thought to be initiated by intracellular signalling from a cell surface receptor or by direct nuclear action. For instance, IGFBP-3 has been shown to be capable of translocating into the nucleus from the extracellular compartment in dividing human breast cancer cells. This translocation into the nucleus by β-importin is mediated by its C-terminal nucleus localization signal. Both IGF-dependent and IGF-independent actions of IGFBP-3 are modulated by deactivation from several proteases. A list of potential cleavage sites and their proteases is reported below.

Arg^{97}–Ala^{98}, Lys^{160}–Val^{161}	Plasmin
Arg^{95}–Leu^{96}, Lys^{160}–Val^{161}	Plasmin
Arg^{97}–Ala^{98}, Arg^{206}–Gly^{207}	Thrombin
Arg^{97}–Ala^{98}, Lys^{149}–Lys^{150}, Lys^{150}–Gly^{151},Lys^{154}–Asp^{155}	Serum
Arg^{97}–Ala^{98}, Arg^{132}–Val^{133}, Tyr^{159}–Lys^{160},Phe^{173}–Ser^{174}, Arg^{179}–Glu^{180}	Seminal plasma PSA
Arg^{97}–Ala^{98}, His^{131}–Arg^{132}, Tyr^{159}–Lys^{160}	Urinary PSA
Arg^{97}–Ala^{98}	Cysteine protease
Tyr^{99}–Leu^{100}	MMP-1, MMP-2
Leu^{96}–Arg^{97}, Leu^{141}–His^{142} (minor sites)	
Tyr^{99}–Leu^{100}, Asn^{109}–Ala^{110}	MMP-3
Glu^{176}–Ser^{177}	
Lys^{144}–Ile^{145}	Cathepsin L

These proteases seem to act locally depending on the tissue type. For example, in the quiescent epidermis MMP-19 (matrix metalloproteinase-19) is the major IGFBP-3 degrading enzyme [17], whereas MMP-7 cleaves IGFBP-3 preferentially in tumor tissues increasing IGF bioavailability. It has been demonstrated in fact that co-incubation of the IGF-I/IGFBP-3 complex with MMP-7 restored IGF-I mediated IGF-I-R phosphorylation in cancer cell lines [18]. Another IGFBP-3 activity

modulation event is phosphorylation: some reports indicate that phosphorylated IGFBP-3 shows enhanced IGF-binding affinity, increased nuclear import of IGFBP-3 and binding to components into the nucleus. The physiological function has been investigated also in relation to cancer.

The IGFBP-3 role in breast cancer is controversial since some reports suggest that high levels of IGFBP-3 in serum reduce the relative risk of developing breast cancers, while others indicate the opposite. In some cases, high IGFBP-3 levels in premenopausal women seem to correlate with an increased cancer risk, in contrast with in vitro experiments showing anti proliferative and antiapoptotic activity [19]. It has been suggested that IGFBP-3 could have antiapoptotic or proapoptotic effect in vivo, depending on the cellular environment. These considerations open up the possibility for a therapeutic approach involving IGFBP-3 against some tumoral diseases. Breast cancer cells showed growth inhibition and apoptosis after stable transfection of a construct expressing human IGFBP-3 and recombinant human IGFBP-3 developed by Insmed Incorporated has been subjected to advanced preclinical investigation as a candidate for anticancer therapy.

1.1.7 Physiological functions of IGF-II

IGF-II is a peptide hormone involved in cell differentiation and proliferation with a molecular mass of 6.7 kDa that shows 70% aminoacid sequence identity to IGF-I and 50% to insulin. Several lines of evidence indicate that IGF-II is a growth factor with a role in uterine angiogenesis and fetal growth during pregnancy.

Fetal development is a process which requires the adaptation of uterine blood vessels. Several changes characterize this adaptation: vasodilation; increased

capillary permeability; and angiogenesis, consisting in the growth of new capillaries from pre-existing blood vessels and post-capillary venules. Angiogenesis consists of various steps: proliferation of endothelial cells, degradation of the basal lamina, migration of endothelial cells followed by tube formation. Some reports show that IGF-II/M6PR is responsible for proliferin-induced angiogenesis *in vitro* and *in vivo* [20]; furthermore IGF-II secreted from hepatocellular carcinoma cells possess angiogenic activity both *in vitro* and *in vivo* [21]. These observations led to the hypothesis that IGF-II is involved in angiogenesis during pregnancy. In situ hybridization experiments on human uterine microvascular endothelial cells (HUMVECs) showed cellular localization of IGF-II and IGF-IIR mRNAs in human uterine vessels. The ability of IGF-II to stimulate new vessel formation has been demonstrated by HUMVEC incubation with IGF-II [22]. The addition of 100 ng/ml caused a 2.5-fold increase in the number of outgrown sprouts, an effect comparable to the addition of 10 ng/ml vascular endothelial growth factor (VEGF). Furthermore, HUMVEC proliferation *in vitro* can be induced by physiological concentrations of IGF-II (50-250 ng/ml) to a similar degree as by VEGF. These results suggest a possible involvement of IGF-II in uterine vascular development during pregnancy.

Another role of IGF-II in cell differentiation has been demonstrated in skeletal myoblast cells. Skeletal muscle cells form post-mitotic tubes in a process driven by the coordinated expression of several muscle-specific genes. A family of myogenesis genes (Myo D1, myogenin, myf-5, and herculin/MRF4/myf-6) are, in fact, capable of stimulating non muscle cells, to exhibit muscle-specific characteristics [23]. It has been also demonstrated that higher than physiological insulin concentrations stimulate differentiation of muscle cells. Furthermore lower concentrations of IGF-I and IGF-II cause the same effect and it has been suggested that the effects of insulin are mediated by the IGF-I receptor. It has

been subsequently demonstrated that IGF-II is responsible of muscle cells differentiation with an autocrine action. Muscle cell lines showing a marked difference in the rate of differentiation were incubated and IGF-II mRNA was isolated: this showed an increase of about 1 order of magnitude during differentiation and was present in a 10-fold higher quantity in the cell line with the most rapid differentiation. Similar evidence has been given for the correlation between the amount of protein and rate of differentiation of the two cell lines [24].

1.1.8 IGFBP-rP1

As discussed above it has been shown, on the basis of the primary sequence and of the clustering of invariant cysteines, that IGFBPs can be subdivided in three distinct regions: the conserved N-terminal region, the conserved C-terminal region, and a non conserved midregion. The modularity of those domains was suggested after the observation of the similarity of the C-terminal domain with tyroglobulin type I domains. After the identification of the N-terminal domain of IGFBPs in other cysteine-rich proteins it became clear that there exist proteins structurally related to the IGFBP family and that the N- and the C-terminal domains of the IGFBPs are best thought of as discrete modules of the IGFBPs. A domain can be defined as a region of a protein that can fold in a tertiary structure independent of neighboring sequences. A domain found in two otherwise nonhomologous proteins is considered to be evolutionarily mobile and is therefore termed a module. The N-terminal domain of IGFBPs can be considered a module and is also found in IGFBP-rPs. Primary sequence alignment of IGFBPs and IGFBP-rPs shows that this module is the only common region among the present members of the IGFBPs superfamily. The alignment also suggests that the N-

terminal domain extends up to the last conserved cysteine. The size of the N-terminal domain (including the signal peptide) thus ranges from 70 aminoacids (IGFBP-rP2) to 93 aminoacids (IGFBP-3), with the exception of IGFBP-rP6 which has the longest N-terminal domain, 130 residues long, and contains all the 18 cysteines. It has been demonstrated that in IGFBPs both the N-terminal and C-terminal domains are involved in IGF binding: it has also been shown that loss of the C-terminal domain reduces, but does not abolish the ability of IGFBPs to interact with IGFs. This has led to the hypothesis that IGFBP-rPs could bind IGFs with reduced affinity, compared to IGFBPs. It has been demonstrated that IGFBP-rPs possess approximately 100-fold lower affinity for IGFs, and can thus be considered low-affinity IGF binders in the IGFBPs superfamily. Another feature of IGFBP-rP1 and IGFBP-rP3 consists in the ability to bind insulin with an affinity comparable to that for IGFs. Given the low affinity for IGFs and insulin binding however, IGF-independent actions are considered much more likely than IGF-dependent.

1.1.9 Physiological roles of IGFBP-rP1

Mac25 is the gene coding for IGFBP-rP1 and it was identified as a gene differentially expressed in meningiomas. Normal leptomeningeal cells are present in the membrane surrounding the brain and spinal cord and tumors developing from these cells are called meningiomas. Before the identification of mac25, it was demonstrated that in up to 70% of meningiomas loss of genetic material from chromosome 22 was seen to be more frequent than alterations of other chromosomes [25]. When the locus of mac25 was identified, the gene sequence had not yet been reported but a protein homology search, based on the predicted

protein sequence, showed a significant degree of homology of the N-terminus with the N-termini of the IGFBPs suggesting that mac25 could encode for a new member of the IGFBP family [26]. Three years later another group produced recombinant human-mac25 (rh-mac25) in a baculovirus expression system. Rh-mac25 was proven to be able of IGF-binding by Western Ligand Blot under non-denaturing conditions and by affinity cross-linking. IGF-binding affinity was compared to that of IGFBP-3 and was estimated to be 5-25-fold lower, even if possible overestimation due to affinity cross-linking was not excluded, the authors suggested that affinity could be of 2 or 3 orders of magnitude lower. It was also suggested that the lower affinity, compared to that of IGFBPs could be explained by the reduced similarity of the carboxyl-terminus which is also responsible for IGFs binding. On the other hand, sequence similarity of the N-terminal domain is much more relevant involving 11 conserved cysteines, compared to the 12 found in all IGFBPs, and includes the IGFBP motif (GCGCCXXC). The sequence similarity for IGFBPs found in the N-terminal domain strongly supports the notion that mac25 is another member of the IGFBP family [27].

IGFBP-rP1 has the ability to bind different proteins. It has been recently demonstrated that IGFBPs are able of insulin low-affinity binding; this was excluded until improvements in assay techniques were available. Conversely, it has been demonstrated that IGFBP-rP1has a higher affinity of binding for insulin, compared to all IGFBPs. The C-terminus of IGFBP-rP1 is different from that of IGFBPs due to the loss of 5/6 conserved cysteines. This variation is thought to expose the insulin-binding site and at the same time to reduce IGFs affinity [28]. Thus IGFBP-rP1 can bind IGFs and insulin with similar affinity, but its insulin:IGF binding ratio is 500-fold higher than that of IGFBP-1 to -6. This ability enables IGFBP-rP1 to compete with insulin receptors for insulin binding. The existence of

binding partners other than IGFs implies that IGFBP-rP1 is important for its IGF-independent actions. A “follistatin module” is present in its sequence and has been shown to bind to activin, a member of the tumor growth factor-β (TGF-β) superfamily of growth factors [29]. The first role of activins was found to be their ability to regulate follicle-stimulating hormone (FSH) secretion from the anterior pituitary. They are also able to regulate gonadal functions and are associated with growth modulation of glandular organs. In addition, they are present and regulated during postnatal mammary gland development but they are also necessary for normal mammary cell function. Loss of activin signalling leads in fact to malignant progression as can be seen in high grade breast cancer, where activin and its receptors are decreased in quantity [30].

Type IV collagen is another IGF-independent binding partner. It has been demonstrated by immunostaining techniques that IGFBP-rP1 colocalises with type IV collagen in the vascular basement membrane. This was further confirmed by in vitro experiments showing that IGFBP-rP1 bound with higher affinity to type IV collagen among various extracellular matrix components. Distribution of IGFBP-rP1 has been studied in human biological fluids and tissues using monoclonal antibodies. It was detected in serum, amniotic fluid, urine, and cerebrospinal fluid. In normal human tissues immunohistochemistry studies showed ubiquitous intense staining of peripheral nerves, smooth muscle cells, blood vessel wall cells, gut, bladder, breast, and prostate. It was also demonstrated that IGFBP-rP1 has an important role in the female reproductive system. Gene expression of IGFBP-rP1 showed that the protein is involved in human endometrial receptivity, folliculogenesis, and growth development and regression of the corpus luteum in higher mammals [31]. Endometrial receptivity involves corpus luteum function and progesterone production. IGFBP-rP1 has been shown to increase 35-fold in expression during the receptive phase compared with the pre-receptive phase

which suggests a role of IGFBP-rP1 in endometrial receptivity. IGFBP-rP1 has been studied for its involvement in cancer. It has been demonstrated by subtractive hybridization studies that IGFBP-rP1 is down-regulated in the Hs578T breast cancer cell line. Normal breast tissues show high protein levels while during disease progression IGFBP-rP1 abundance is decreased thus suggesting a potential role as tumor suppressor [32].

1.2 Basement membrane structure

Basement membranes (BM) can be found under epithelial and endothelial sheets, surrounding muscle cells, fat cells, and peripheral nerve axons; they consist in specialized extracellular matrices which form stable sheets through self-assembly mechanisms and many of their components interact with cellular receptors. In response to these interactions cells can be infuenced in their shape, gene expression, cell migration and proliferation and programmed cell death. In addition, BM are important for tissue compartimentalization and act as barriers to cell penetration and filtration. Among the constituents of BM we can name collagens and laminins, proteoglycans, calcium-binding proteins and further structural or adhesive proteins. Two independent networks, one formed from collagen IV and one from some laminin isoforms, constitute the BM. A binding protein, nidogen, connects and stabilizes these two network structures [33].

1.2.1 Perlecan : a basement membrane proteoglycan

Proteoglycans are constituted by a core protein and by one or more glygosaminoglycan (GAG) chains. Most proteoglycans have either chondroitin

suplhate or heparan sulphate chains or both linked to the polypeptide chain. Repeating disaccharides of an amino sugar and an uronic acid constitute chondroitin suplhate and heparan sulphate. In the case of chondroitin suplhate the amino sugar is N-acetylgalactosamine and in heparan sulphate is N-acetylglucosamine [34]. Heparan Sulphate Proteoglycan (HSPG) genes are highly conserved in evolution; their gene products are responsible for several biological functions such as BM homeostasis, modulation of growth factor activity and angiogenesis. They can be expressed in avascular tissues like cartilage, and in musculoskeletal and nervous tissues where they modulate neuronal transmission and eye development. It has been demonstrated that some human congenital diseases, caused by mutations of HSPG genes, show phenotypes similar to those of mutant mice, flies, and worms. There are three well-characterized HSPGs: collagen XVIII, agrin, and perlecan; collagen XVIII is a hybrid collagen-proteoglycan which is part of the multiplexin gene family. Agrin is an abundant constituent of BMs and exerts its funcion at the neuromuscular junction.

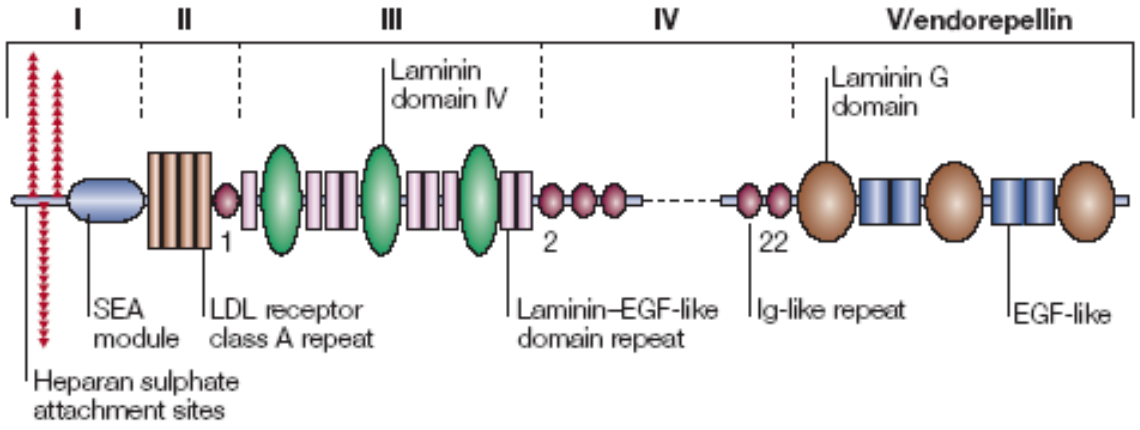


Fig. 5. Schematic representation of the five structural domains of perlecan.

Perlecan is a modular proteoglycan with omology to growth factors and it is one of the largest single-chain polypeptides found in vertebrates and invertebrates.

The protein core of human perlecan weights approximately 470 kDa but it possesses numerous O-linked oligosaccharides and as many as four heparan sulphate (HS) chains and it can reach a molecular mass of over 800 kDa. The five perlecan modules are evolutionarily related to proteins involved in nutrient metabolism, mitogenesis, and adhesion and have probably evolved from ancient ancestors by gene duplication and exon shuffling. Perlecan interactions with its binding partners take place through its five modules and HS chains and involve numerous heparin-binding factors such as fibroblast growth factor-2 (FGF2), vascular endothelial growth factor (VEGF), and platelet-derived growth-factor (PDGF). Perlecan can also be associated with the cell surface through high affinity interactions with integrins [35].

Skin is a body organ with several functions including protection from microorganisms and UV radiation, regulation of body temperature, and is a part of the immune system. Two distinct tissues, separated by a basement membrane (BM) form this organ: epidermis and dermis. The dermis is a collagen rich tissue wich supports and nourishes the epidermis; it is composed of fibroblasts that synthesize and secrete the extracellular matrix (ECM) components [36]. The epidermis contains keratinocytes forming a stratified epithelium made of multiple layers with dinstinct morphology and functions. A characteristic of epidermis is its continuous self-renewal driven by the proliferation of basal cells which are the only epidermis cells capable of proliferation. The kerationocytes differentiate during the translocation from the basal layer to the tissue surface where they die. Epidermal proliferation and differentiation is a process controlled by cell-cell and cell-matrix interactions [37], as well as by a cross-talk through a network of cytokines and growth factors that are secreted from keratinocytes and dermal fibroblasts. HSPG have an effective role in protecting heparin-binding growth

factors from thermal denaturation and proteolytic degradation, and thus they control growth factor bioavailability.

A study [38] has been made using an engineered skin model to determine the role of perlecan in epidermis differentiation. In this model skin equivalents (SE) were used, consisting of keratinocytes cultivated on a air/liquid interphase grown on various substrates which act as dermal equivalents; these equivalents include collagen gels with dermal fibroblasts. This model is different from monolayers in its characteristics showing an architecture and gene profile similar to normal epidermis. The role of perlecan was investigated using targeted disruption of perlecan expression in epidermal keratinocytes and dermal fibroblasts. In the case of dermal fibroblasts, lack of expression of perlecan obtained using perlecan-null cell lines showed no differences in epidermis differentiation. Conversely, epidermal cells transfected with an antisense construct of domain III of the human perlecan gene were unable to develop an organized epidermis. In particular, these perlecan-deficient keratinocytes (PDK) grew in single or double layers with flattened cells. This defective epidermal formation was restored by the exogenous addition of perlecan. It was also demonstrated that perlecan deficiency did not affect the progenitor cells ability to regenerate but PDKs showed an increased rate of apoptosis. In this study the hypothesis that in PDK, perlecan produced by underlying fibroblasts could be effective after accumulation in the basement membrane was also tested. This was proven to be true after 5 or 6 weeks of PDK cultivation, thus demonstrating the essential role of perlecan in the differentiation of epidermis.

1.2.2 Endorepellin: biological activity of the perlecan C-terminal domain

It has been demonstrated that the C-terminal regions of proteoglycans, collagen XVIII and perlecan in particular, possess a specific biological activity. Endostatin, for example, derives from limited proteolysis of collagen XVIII, which is an insoluble BM-embedded molecule. Endostatin is a soluble molecule showing its effects in a paracrine/endocrine manner, is also a heparin-binding protein [42] that can inhibit cell migration and proliferation and induce cell death by apoptosis. Endorepellin is the domain V of perlecan, which weights 85 kDa and contains three laminin-like (LG) domains separated by two sets of epidermal growth factor-like (EGF-like) repeats.

HSPGs act as reservoirs of pro- and anti-angiogenic factors and modulate angiogenesis, along with members of the fibroblast growth factor (FGF) and vascular endothelial growth factor (VEGF) family and their receptors. Angiogenesis is one of the most relevant events in tumor progression and is affected by cell-matrix interactions between the surface of endothelial cells and the tumor-matrix interfaces. A study [39] focusing on the C-terminal domain V of perlecan demonstrated the ability of a 85 kDa fragment, endorepellin, to inhibit several aspects of angiogenesis like endothelial cell migration and collagen induced endothelial tube morphogenesis. The biological activity of recombinant endorepellin was also demonstrated testing VEGF-induced human umbilical vein endothelial cells (HUVEC) migration which resulted completely suppressed in the presence of endorepellin. This effect was also confirmed *in vivo* using the chorioallantanoic membrane (CAM) assay. Endorepellin also blocked the angiogenic process in sponges containg WiDr, a highly tumorigenic colon

carcinoma cell line. In another experiment HUVECs were grown in a collagen matrix showing capillary-like network formation after 4 hours: endorepellin concentrations comparable to those used in the HUVEC migration assay caused a complete block of tube-like structures. The biological action of endorepellin was also tested in the presence of an interacting partner. This partner was found screening a keratinocyte cDNA library in a yeast two-hybrid system using endorepellin as a bait; the protein found, endostatin, was another anti-angiogenic factor derived from collagen type XVIII. Experiments on VEGF-induced HUVEC migration were made in which endostatin was increased, while endorepellin was kept constant; concurrent subministration of endostatin and endorepellin showed an inhibition of their activities with maximum HUVEC migration for a molar ratio of 1:1. ^{125}I-labeled endorepellin was also demonstrated to bind HUVECs which then possess high affinity endorepellin binding sites.

LG3, one of the three Laminin-G like domains of endorepellin, produced by 293-EBNA cells is found together LG3 fragments identical to those found in the urine of patients with end-stage renal disease [40] and in the amniotic fluid of pregnant women with premature rupture of fetal membranes. This observation has suggested that cleavage of endorepellin and production of LG3 could be of biological significance. In a study [41] it has been demonstrated which is the protease responsible for LG3 processing. In these experiments, recombinant endorepellin (85 kDa) produced from 293-EBNA cells often copurified with a 26 kDa fragment. After sequencing residues around the cleavage site it was determined with a database search that two proteins were coherent with the cleavage site: bone morphogenetic protein 1 (BMP-1) and mammalian Tolloid-like 1 (mTLL-1). Incubation of endorepellin with BMP-1 produced a fragment of 60 kDa, corresponding to the N-terminal domain of endorepellin, and the fragment

of 26 kDa. It was also demonstrated that the same process occurs *in vivo* using mouse embryonic fibroblast (MEF) cells from either wild-type or mutant embyos doubly null for the *Bmp-1* and the *Tll-1* genes. While in wild-type cells a 26 kDa fragment was detectable by endorepellin antibody, *Bmp-1/Tll-1* double null failed to show it.

1.3 Heterologous protein expression

The development of molecular biology techniques has enabled the production of recombinant proteins which can be used in place of proteins purified from the organism of interest. This has become the preferred strategy because of the amounts of recombinant protein produced and of the possibility to add fusion tags which are useful for protein purification. *Escherichia coli* has been widely used because it grows rapidly on inexpensive substrates, and because of the availiability of an increasingly large number of cloning vectors and of host strains. The production of a recombinant protein can be achieved by cloning the coding sequence into an expression vector, under control of a promoter which drives protein expression in the presence of an inducer. One of the most used regulatory elements is the lactose utilization (*lac)* operon which is induced by the non-hydrolyzable lactose analog isopropyl-β-D-1-thiogalactopyranoside (IPTG). Even though it has proven useful for the high content of recombinant protein produced, in many cases this was followed by the formation of inclusion bodies. Since protein renaturation often leads to a reduced yield, it is often preferred to modify the cultivation temperatute during expression or to use different combinations of expression vectors or bacterial strains.

Another option for recombinant protein production is given by expression in the methyilotrophic yeast *Pichia pastoris*. A major advantage over bacterial

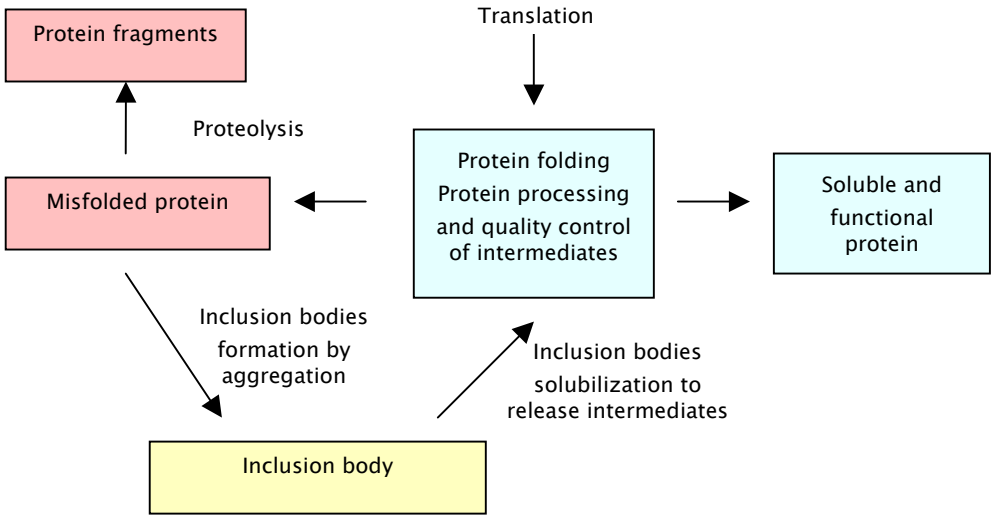
expression systems is that yeast can perform many of the post-translational modifications associated with higher eukaryotes, such as processing of signal sequences, folding, disulfide bridge formation, and N-linked glycosylation. *Pichia pastoris* also offers the possibility to express the recombinant protein either intracellularly or extracellularly. In general, cells are grown initially in a medium containing glycerol as its carbon source: under these conditions cell accumulates but heterologous gene expression is fully repressed. Subsequently after glycerol depletion, methanol is added to the growth medium and protein expression is induced.

1.4 Protein refolding

Protein crystallography has investigated proteins purified from organisms of interest for decades until molecular biology techniques were available that enabled recombinant protein production. One of the hosts most commonly used is *Escherichia coli* because it can be grown fast and it is easy to transform. Even though it is often useful for soluble protein production, in many cases is troublesome due to recombinant protein deposition into inclusion bodies (IB). These aggregates of misfolded protein form intracellularly because of the tendency of the protein to aggregate or when the cell is unable to maintain the expressed protein either in a soluble form or with a correct folding. Aggregate formation can be considered as a dynamic process characterized by addition and removal of partially refolded protein. In order to make protein from inclusion bodies suitable for structural studies it should be released from that misfolded state and solubilized in its native state. Refolding is therefore a process that

brings the protein conformation from unfolded to folded, or native, state. High denaturant concentrations are used to solubilize IBs: under this condition proteins are unfolded, well solvated, and flexible; conversely, after denaturant removal proteins are folded, rigid, and compact. Refolding consists in the transition from the unfolded state, caused by high denaturant concentrations, to the folded state, where the protein collapses into its compact structure; it is commonly recognized that such a drastic process causes protein misfolding and aggregation. A refolding yield will be acceptable only when there is an intermediate denaturant concentration which allows protein molecules to stay in solution and to be flexible enough to reorganize into their native structure. In practice, it is crucial to test how to remove the denaturant and determine the composition of the refolding buffer: the addition of compounds that prevent aggregation and facilitate refolding is widespread.

A flow diagram showing inclusion body formation is presented below.



The preparation of the protein to be refolded from IBs can be subdivided in three tasks: IBs separation and purification, IBs solubilization, and solubilized protein purification. All of the three steps influence the final refolding yield and should be optimized.

After bacterial cell disruption IBs are isolated by centrifugation and removal of the supernatant; IBs are normally purified by resuspending them in buffers containing detergents (Triton X-100 for example) or low urea concentrations (about 1 M); these washing steps should be repeated twice or more times to achieve the complete removal of cellular debris.

IBs are then solubilized with high concentrations of chaotropic agents such as urea or guanidinium chloride. Both of these agents bind to the proteins in a concentration-dependent manner and the concentrations used are 6-8 M urea and 6-7 M guanidinium chloride. Reducing agents should also be included (β-mercapto-ethanol or dithiothreitol) to reduce any uncorrectly formed disulphide bonds.

It has been demonstrated that, once the recombinant protein has been released from IBs, it is critical to include a purification step to remove contaminants which could still be present. Histidine-tagged fusion proteins, for instance, can be purified under denaturant conditions by affinity chromatography. As discussed above, in order to maximize the refolding yield one should test different strategies to remove the denaturant after IBs solubilization. Many approaches are reported in the literature. Although they all possess their own advantages and drawbacks, it is almost impossible to predict which one is the best choice for a particular protein. Some examples are reported in the following list.

- **Dilution:** refolding by dilution is attained by unfolded protein transfer into an appropriate refolding buffer. From the protein's point of view this means the transition from a high denaturant concentration to another

close to zero. This is considered by some to be a severe change and addition of low denaturant concentrations into the refolding buffer is suggested.

- **Reverse dilution**: the refolding buffer is slowly added to the unfolded protein containing concentrated denaturant. Since denaturant and protein concentrations decrease simultaneously, at intermediate denaturant concentration even the protein concentration is significant and this can result in aggregation and precipitation.
- **Dialysis**: consists in the dialysis of the denatured protein against a refolding buffer. It should be noted that during the period of intermediate denaturant concentration the protein is exposed to aggregates and this can severely decrease the final yield.
- **Buffer-exchange by gel filtration**: the unfolded protein sample is loaded onto a gel-filtration chromatography column equilibrated with refolding buffer. An advantage with respect to dialysis is given by the presence of the gel-filtration matrix: protein interaction with the resin, via hydrogen bonding or hydrophobic interactions, may prevent misfolding and aggregation.
- **Solid-phase refolding**: denatured protein is non-covalently bound to a matrix such as Ni-resin or ion-exchange matrix in the presence of a denaturant; this can be useful in immobilizing protein aggregates but can also interfere with refolding because of steric hindrance of the resin.

The refolding protocol is not the only aspect to be tested; different refolding buffer compositions should also be investigated. Although a systematic approach does not exist one can start considering additives which can be classified into two groups: folding enhancers and aggregation suppressors. They are considered

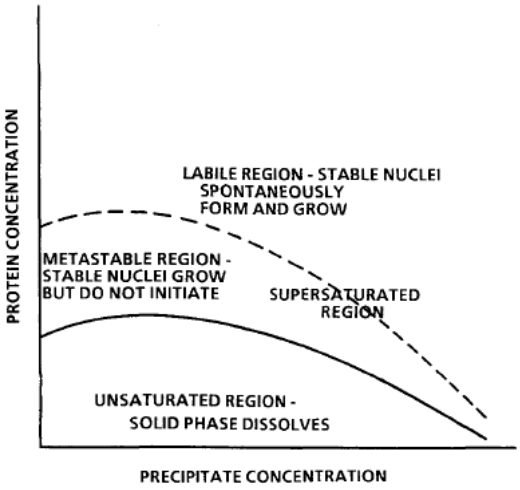
exclusive since the first ones increase protein-protein interactions, while the last ones reduce side-chains interactions. These additives are summarized in the table below.

Classification	Model additive	Effect on protein stability
Folding enhancer	Sucrose Ammonium sulphate	Stabilizer
Aggregation suppressor	Arginine Mild detergent	Neutral
Denaturant	Urea Guanidinium chloride	Destabilizer

1.5 Protein crystallization

The knowledge of the three dimensional structure of a protein has proven to be of key importance for understanding the interactions between different proteins and their roles in the biochemistry of living organisms. Furthermore, the resolution of a protein structure has enabled the development of drugs or specific ligands targeting proteins involved in the biochemical pathways of interest. One of the most powerful techniques to determine a protein structure is X-ray diffraction analysis of protein single crystals. The road to the three dimensional structure of a given protein requires several steps: recombinant protein expression in a host organism, protein purification, crystallization, X-ray diffraction analysis, and diffraction data processing. Recombinant protein production has became easier and easier in the last decades thanks to the development of Molecular Biology tools and the availability of gene clones. Protein purification is also supported by standardized chromatography

techniques and sophisticated equipment. Diffraction analysis and data processing are now extremely fast processes when using synchrotron radiation sources for diffraction intensity collection and using specific softwares designed to run on personal computers. The step which is often considered a bottleneck is protein crystallization: obtaining a protein single crystal can be a difficult task; even more difficult can be growing a single crystal with an ordered structure and of suitable dimensions to collect high-resolution diffraction patterns. Crystallization is still the tricky part due to many reasons: proteins are macromolecules whose properties depend on temperature, pH, ionic strenght, the presence of ligands and other parameters; the process of crystallization is thus unpredictable because of the large number of parameters involved. This makes protein crystallization mostly an empirical procedure [47]. Crystallization kits containing hundreds of conditions are today available which are designed to test parameters like pH, precipitant salts, and organic solvents in a systematic way. However, it is still imperative for the crystallographer to be familiar with the principles underlying crystallization; in fact, it is not unusual to obtain inadequate crystals form commercial crystallization screens which must then be improved adjusting the starting conditions. Crystallization of a macromolecule takes place when a macromolecular solution is brought to a supersaturated state: crystals then appear and constitute a solid phase to which soluble molecules join in order to bring the solution toward a saturated state. A supersaturated solution does not necessarily develop the solid phase in the form of a crystal or precipitate. A phase diagram describes the process of crystal formation as a function of protein and precipitant concentrations.



The metastable region is defined as the condition in which stable nuclei cannot spontaneously form but, if already present, they can grow due to the ordered attachment of soluble molecules. Conversely, in the labile region stable nuclei can form causing protein concentration to decrease until the metastable state is reached. The dynamical nature of crystal formation must be understood to obtain crystals suitable for X-ray diffraction. If the protein solution is brought far away into the supersaturated state then a large number of nuclei will form: those nuclei will grow rapidly but will stop growing too early, due to soluble protein depletion. This will give many crystals too small to be useful for diffraction analysis. One must then determine which are the optimal protein and precipitant concentrations giving less nuclei but larger crystals.

Macromolecular crystals possess different characteristics than crystals made of small molecules; the latter are generally more resistant, can be exposed to dehydration, and diffract X-rays with strong intensity while protein crystals can be easily damaged by temperature variations, dehydration, and diffract less. These differences arise from the structure of the molecular packing and from the overall crystal features. Macromolecular crystals contain about 50% of solvent even though their content can range between 25% up to 90% thus forming large solvent pores freely accessible to salts and small ligands. This aspect of protein

crystals is one of the causes of their weak diffraction power, compared to small molecule crystals. The higher solvent content offers to the protein molecules some freedom of movement and molecules with slightly different orientations diffract less than perfectly ordered ones. In addition, polypeptides orderly packed to form a crystal can present different conformations from each other and this further reduces crystal diffraction quality.

In order to start crystal growth, a limited degree of supersaturation must be reached; this can occur either by increasing the concentation of the precipitating agents or by altering its physical properties like pH. A classical procedure, for example, is to progressively increase the level of saturation of a salt: normally the protein has a tendency to precipitate but can also start to form ordered lattices. Proteins react to precipitant ionic strength in two different ways; first, at very low ionic strength a phenomenon termed “salting-in” occurs in which protein molecules tend to interact among themselves to balance their electrostatic requirements, which were formerly satisfied by ions. In practice, this has been used in the crystallization of some proteins which were dialyzed against distilled water. Second, high ionic strength causes a “salting-out” effect in which both ions and protein molecules compete for their hydration layers and this also drives protein molecules towards the self-aggregation. The dehydration of protein solution can also be achieved by addition to the mother liquor of organic solvents, ethanol or methylpentanediol for example. Polyethyleneglycole is commonly used because of its dehydration ability but also since it forms a network into the mother liquor which promotes phase separation thus excluding protein molecules. Some examples of compounds used for protein crystallization are offered below.

Salt	Organic solvents	Polymers
Ammonium or sodium sulfate	Ethanol	Poly(ethy1ene glycol)
Lithium sulfate	Isopropanol	Jeffamine T
Lithium chloride	1,3-Propanediol	Polyamine
Sodium or ammonium citrate	2-Methyl-2,4-pentanediol	
Sodium or potassium phosphate	Dioxane	
Sodium or potassium or ammonium chloride	Acetone	
Sodium or ammonium acetate	Butanol	
Magnesium or calcium sulfate	Acetonitrile	
Cetyltrimethyl ammonium salts	Dimethyl sulfoxide	
Calcium chloride	2,5-Hexanediol	
Ammonium nitrate	Methanol	
Sodium formate	1,3-Butyrolactone	

The most commonly used technique to reach the supersaturation state is the vapor diffusion method. This method is subdivided in hanging-drop method and in sitting-drop method, but the principle is common for both of them: drops containig a mixture of protein and precipitant of a size ranging from 100 nl up to 10 µl are dispensed on a microscopy glass or on a microbridge. The drop is then closed into a well containing a reservoir of 300 µl of precipitant solution and sealed with mineral oil.

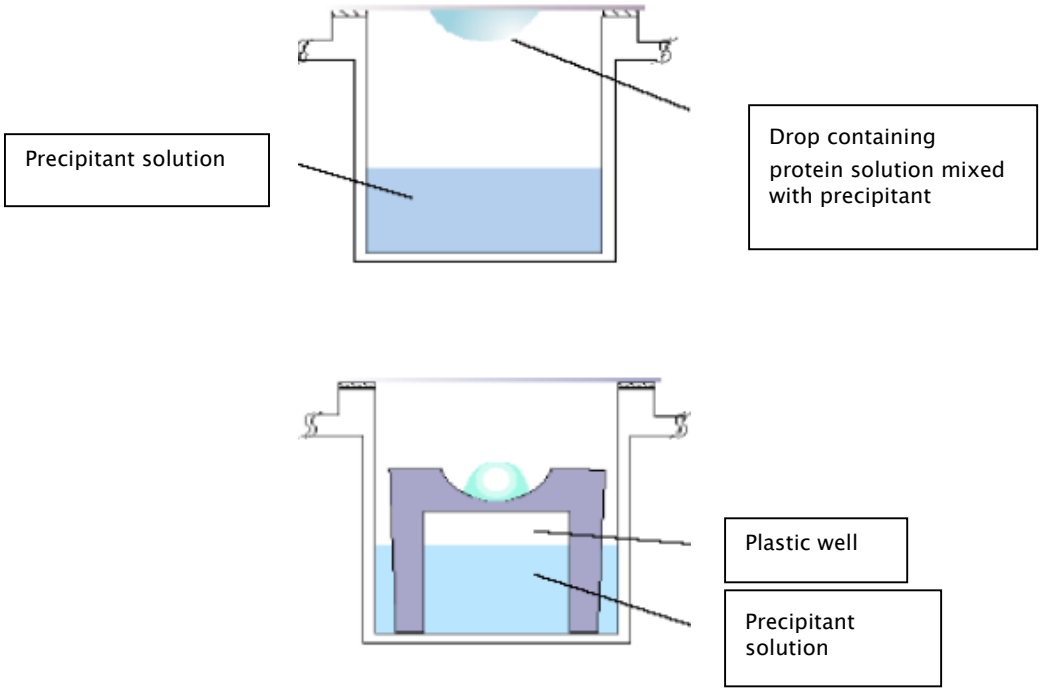


Fig. 6. Representation of hanging drop and sitting drop crystallization methods.

After some time the drop containing the protein/precipitant mixture equilibrates with reservoir through vapor diffusion. Even though the two methods differ only in the size of the drop (100 nl up to 4 µl for hanging-drop and 4 µl up to 10 µl for sitting-drop) they can lead to different results due to the different kinetics of the equilibration process. Another factor affecting crystal growth is temperature and crystallization trials should be made at 4°C and at 20°C since some proteins have been reported to behave differently depending on the temperature chosen; one general principle concerning the choice of the crystallization temperature is that it occurs slowly at colder temperatures than at warmer ones. Crystallization screens test conditions with different pH since it deeply affects the protein behaviour; the charge of a protein, its distribution and the dipole moment of the molecule in fact depend on the ionization state of its side chains. The formation and growth of protein crystals are also influenced by protein concentration and state. As pointed out before, one starts a crystallization trial at a concentration suitable to reach supersaturation (10 mg/ml up to 100 mg/ml are common

starting points) but it should be decreased in case of excessive nucleation. In addition, it has been recognized that the presence of compounds or ligands in the mother liquor can be helpful for crystal formation and growth; some proteins are more likely to crystallize if complexed with a natural ligand because it confers more stability and reduces the dynamical properties of the protein. Although this event is not predictable it is indeed useful to check this option since in many cases an apoprotein and a protein complexed with its ligand can be considered as completely different cases. A major obstacle in crystallizing a protein can be sample heterogeneity. Besides the more obvious sources of contaminations which can be revealed by a SDS-PAGE or a IEF many others can be investigated as suggested from the following list.

1. Presence, absence, or variation in a bound prosthetic group, substrate, coenzyme, inhibitor, or metal ion.
2. Variation in the length or composition of the carbohydrate moiety.
3. Proteolytic modification of the protein during the course of isolation or crystallization.
4. Oxidation of sulfhydryl groups during isolation.
5. Reaction with heavy metal ions during isolation or storage.
6. Presence, absence, or variation in post-translational side chain modifications such as methylation, amidiation, and phosphorylation.
7. Microheterogeneity in the amino or carboxy terminus or modification of termini.
8. Variation in the aggregation or the oligomeric state of the protein due to association/dissociation.
9. Conformational instability due to the dynamic nature of the molecule.

10. Microheterogeneity due to the contribution of multiple but nonidentical genes to the coding of the protein, isozymes.

11. Partial denaturation of sample.

12. Bound lipid, nucleic acid or carbohydrate material, or substances such as detergents used in the isolation.

There are some cases in which the crystals obtained are too small to be useful for diffraction analysis. This may happen for example when too many nuclei appear that cannot grow to an appropriate extent. Seeding techniques have been developed which enable one to use these nuclei and allow further growth. Seeding can be carried out by using both microcrystals or macrocrystals. In the first case one or more nuclei are transferred into the mother liquor and serially diltuted; then small amounts of these dilutions are added to fresh protein-crystallization trials to determine which dilution gave the best extent of nucleation. Macroseeding consists in sequential transfer of a crystal into fresh precipitant solution in order to remove microcrystals adhering its surface and then seeding it into mother liquor with properly saturated protein.

2. Materials and methods

2.1 Human IGF-II: cloning, expression, purification, and crystallization

Cloning – The clone containing the gene coding for human IGF-II was purchased from RZPD (Deutsches Ressourcenzentrum fur Genomforschung). *E. coli* cells containing the plasmid vector pOTB-7 were grown in 3 ml of LB medium overnight at 37°C.

LB medium composition:

1% Tryptone

0.5% Yeast Extract

0.5% Sodium Chloride

Cells were harvested and the plasmid vector was purified using the GenElute Miniprep kit (Quiagen). The gene of interest was then amplified by PCR; the following primers were used:

Primer forward – **GGTGGTGGATCCGCTTACCGCCCCAGTGAG**

Primer reverse – **GGTGGTAAGCTTGGAACCACGCGGAACCAGCTCGGACTTGGCGGGGGTA**

GGATCC – Bam H I restriction site

AAGCTT – Hind III restriction site

GGAACCACGCGGAACCAG – Thrombin cleavage site

A thrombin cleavage site was used for histidine tag removal after protein purification. The histidine tag consists of 6 histidine residues which are added at the end of the recombinant protein and are coded into the plasmid vector used. The histidine tag is useful for protein purification by a Nickel Sepharose affinity column and for recombinant protein detection by Western-blot analysis.

PCR cycle

1- 1 minute at 95°C

2- denaturation: 1 minute at 95°C

3- annealing: 1 minute at 55°C

4- extension: 1 minute at 72°C

5- 15 minutes at 72°C

Steps 2 to 4 repeated 35 times.

The amplified gene was quantified on a 0.8% agarose gel prepared with TBE (45 mM Tris-Borate pH 8.0, 1 mM EDTA) then digested with Bam H I and Hind III restriction enzymes and ligated to a pQE50 expression vector which was previously digested with the same enzymes and dephosphorylated to prevent self-ligation. The amplified sequence to plasmid vector molar ratio used for the ligation reaction was 5:1 and was calculated as follows:

ng (amplified DNA)= Kbp(amplified DNA) * ng(plasmid vector)*molar ratio/Kbp(plasmid vector)

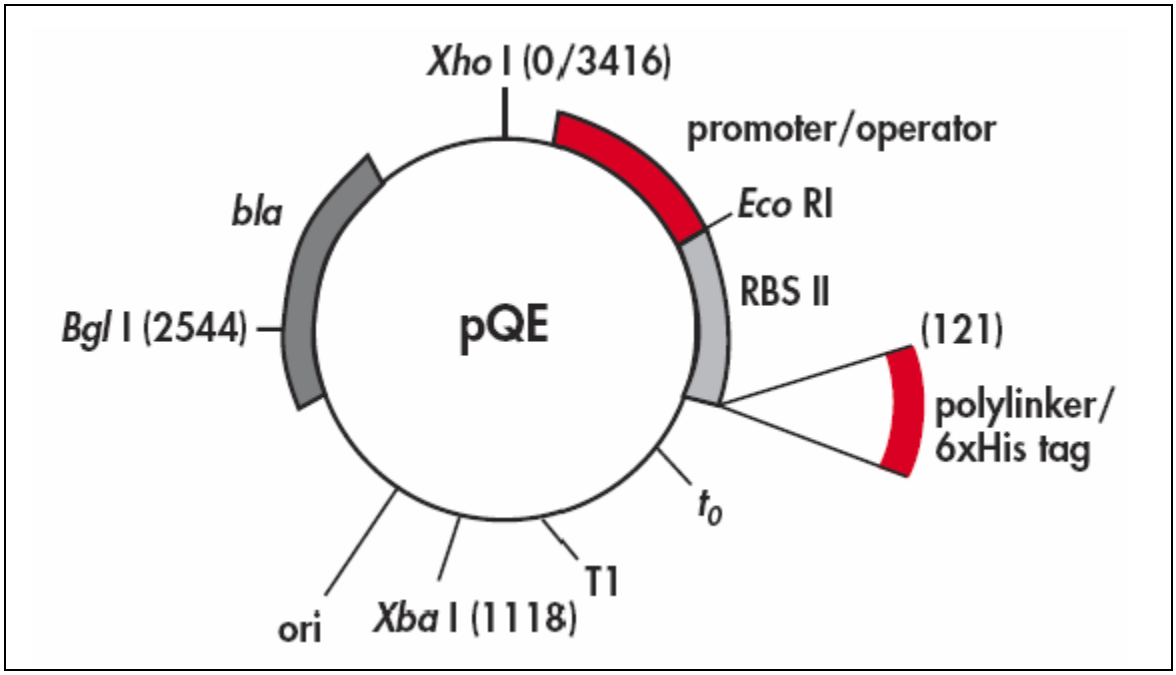


Fig. 8. Schematic representation of a pQE vector.

The construct obtained was then transformed into *E. coli* XL1-Blue strain and cells were plated on a LB-agar plate containing 100 µg/ml ampicillin and grown at 37°C overnight. Bacterial colonies were then checked for transformation by colony PCR and by digestion of the purified plasmid. A positive single colony was transferred into 100 ml of LB liquid medium and grown at 37°C overnight. Plasmid DNA was then recovered and used to transform 100 µl of *E. coli*, Origami strain; the remaining DNA was stored at -80°C. *E. coli* cells were plated and the resulting colonies were tested for protein expression.

Expression - The cultures used for testing protein expression were induced at 16°C, 20°C, and 37°C for 4, 8, and 16 hours with 0.5 mM IPTG. *E. coli* cells were then lysed by sonication. Samples from the total lysate and from the supernatant, after centrifugation, were collected and loaded on SDS-PAGE; protein expression was confirmed by Western-blot analysis using anti-histidine tag monoclonal antibodies (Sigma-aldrich). None of the induction conditions resulted in soluble protein. The largest amount of protein expression was at 37°C overnight induction and all the protein was present in the form of inclusion bodies.

Purification - Three different refolding protocols were tested. 1 L culture was grown and induced with 0.5 mM IPTG when the absorbance at 600 nm reached 0.8, the culture was incubated at 37°C overnight in a shaking incubator. Cells were recovered by centrifugation and inclusion bodies were solubilized. Protein solubilization was obtained as follows:

- After centrifugation *E. coli* cells were resuspended in 30 ml of Tris/HCl 20 mM pH 7.5, 0.5 M NaCl, and 0.02% NaN_3 (TBS) and disrupted by sonication.
- The lysate was then centrifuged and the pellet was washed twice with 15 ml of TBS containing 1% Triton X-100.

- Inclusion bodies were washed with TBS containing 1 M urea.
- Solubilization with TBS containing 8 M urea and 10 mM β-mercaptoethanol.

Crystallization – One of the three refolding protocols gave soluble protein which was used for preliminary crystallization trials.

A new expression vector – Since protein refolding was not practical the expression vector was changed in order to find conditions for soluble protein expression. The gene of interest was amplified by PCR and, after digestion with Bam HI and Hind III restriction enzymes, was ligated to a pET-28 His expression vector, modified to add a histidine tag at the C-terminus. The construct was used for *E. coli* strain XL-1 Blue transformation and the cells were plated on a LB-agar plate containing 50 μg/ml kanamicine. The resulting bacterial colonies were tested by colony PCR and restriction digestion of the purified plasmid. The construct purified from positive colonies was used for the transformation of *E. coli* strains: BL21 (DE3), ER 2566, and Rosetta.

The first expression test was carried out as follows. Colonies from each strain were transferred into 3 ml of LB medium and grown at 37°C until $O.D._{600}$ was about 0.8-1.0 and induced adding 0.25 mM IPTG and incubating overnight at 20°C. After cell lysis by sonication the total lysate and the soluble protein were analyzed by Western-blot. While Rosetta did not express the recombinant protein, both BL21 and ER2566 produced significant amounts of soluble protein. Subsequently BL21 cells were tested for expression levels and induced overnight at 16°C, 28°C, and 37°C and grown both in LB and Terrific Broth (TB) media.

Terrific broth (TB); per litre

12gm tryptone

24gm yeast extract

4ml glycerol (autoclaved).

100ml 0.17M KH_2PO_4(2.31gm/100ml) / 0.72M K_2HPO_4 (12.54gm/100ml); Autoclaved separately.

2.2 Human IGFBP-3: cloning, expression, purification, and crystallization trials

The recombinant protein was produced using the methylotrophic yeast *Pichia pastoris* (Invitrogen).

The *Pichia pastoris* expression kit is a commercial kit which can be purchased from Invitrogen. *Pichia pastoris* is a methylotrophic yeast and has many of the advantages of higher eukaryotic expression systems such as protein processing, protein folding and protein post-translational modification. *Pichia pastoris* shares the same advantages with *Saccharomyces cerevisiae*, another widely used eukaryotic expression system, but expression levels of heterologous proteins are 10-fold to 100-fold higher. *Pichia pastoris* is a methylotrophic yeast capable of using methanol as the sole carbon source. Methanol, which is metabolized in a specialized cell organelle, the peroxisome, is oxidized to formaldehyde by the enzyme alcohol oxidase. This reaction generates hydrogen peroxide which is sequestered into the peroxisome. Alcohol oxidase, which requires molecular oxygen, has a low affinity for O_2 and is produced in large amounts by *Pichia pastoris.* For this reason the promoter used to drive recombinant protein expression in *Pichia* is the same of alcohol oxidase. The gene coding for alcohol oxidase is called AOX1 and the major part of oxidase activity in the yeast is attributable to the product of AOX1 gene. Its expression is strictly regulated and induced by methanol at high levels: its gene product can represent up to 30% of

the total soluble protein in cells induced by methanol. Heterologous protein expression is repressed in the presence of glycerol in the growth medium, even in the presence of methanol. Recombinant protein expression is initiated by complete removal of the growth medium and yeast cells resuspension in liquid medium containing methanol.

Two different options for *Pichia pastoris* recombinant protein expression are available: intracellular or secreted protein expression. With secreted expression the recombinant protein can be found in the growth medium, along with very low levels of native *Pichia* proteins, and this is useful for the purification of the protein of interest. An inconvenience can be the presence of glycosylation sites (Asn-X-Ser/Thr) in the protein's primary sequence which can be recognized by *Pichia*. In this case glycosylation may occur. However, glycosylation is different in *Pichia* as compared to *Saccharomyces*: both of them have a N-linked glycosylation of the high-mannose type but while *Saccharomyces* adds 50-150 mannose residues to the oligosaccharide chains, *Pichia* adds 8-14 residues per side chain. In addition, very low O-linked glycosylation has been reported in *Pichia*.

Plasmid vectors used for extracellular recombinant protein expression are pPICZα A, B, and C. Some features of these vectors are reported below.

Feature	Benefit
5´ *AOX1*	A 942 bp fragment containing the *AOX1* promoter that allows methanol-inducible, high-level expression in *Pichia*. Targets plasmid integration to the *AOX1* locus.
Native *Saccharomyces cerevisiae* □-factor secretion signal	Allows for efficient secretion of most proteins from *Pichia*
Multiple cloning site with 10 unique restriction sites	Allows insertion of the gene into the expression vector

C-terminal *myc* epitope tag Glu-Gln-Lys-Leu-Ile-Ser-Glu- Glu-Asp-Leu-Asn)	Permits detection of the fusion protein by the Anti-*myc* Antibody or Anti-*myc*-HRP Antibody
C-terminal polyhistidine tag	Permits purification of the recombinant fusion protein on metal-chelating resin such as ProBond. In addition, the C-terminal polyhistidine tag is the epitope for the Anti-His(C-term) Antibody and the Anti-His(C-term)-HRP Antibody
AOX1 Transcription Termination (TT)	Native transcription termination and polyadenylation signal from *AOX1* gene (~260 bp) that permits efficient 3´ mRNA processing, including polyadenylation, for increased mRNA stability
TEF1 promoter (GenBank accession numbers D12478, D01130)	Transcription elongation factor 1 gene promoter from *Saccharomyces cerevisiae* that drives expression of the *Sh ble* gene in *Pichia*, conferring Zeocin™ resistance
EM7 (synthetic prokaryotic promoter)	Constitutive promoter that drives expression of the *Sh ble* gene in *E. coli*, conferring Zeocin™ resistance
Sh ble gene (*Streptoalloteichus hindustanus ble* gene)	Zeocin™ resistance gene
CYC1 transcription termination region (GenBank accession number M34014)	3´ end of the *Saccharomyces cerevisiae CYC1* gene that allows efficient 3´ mRNA processing of the *Sh ble* gene for increased stability
pUC origin	Allows replication and maintenance of the plasmid in *E. coli*
Sac I *Pme* I *Bst*X I	Unique restriction sites that permit linearization of the vectors at the *AOX1* locus for efficient integration into the *Pichia* genome

Plasmid vector map is represented below.

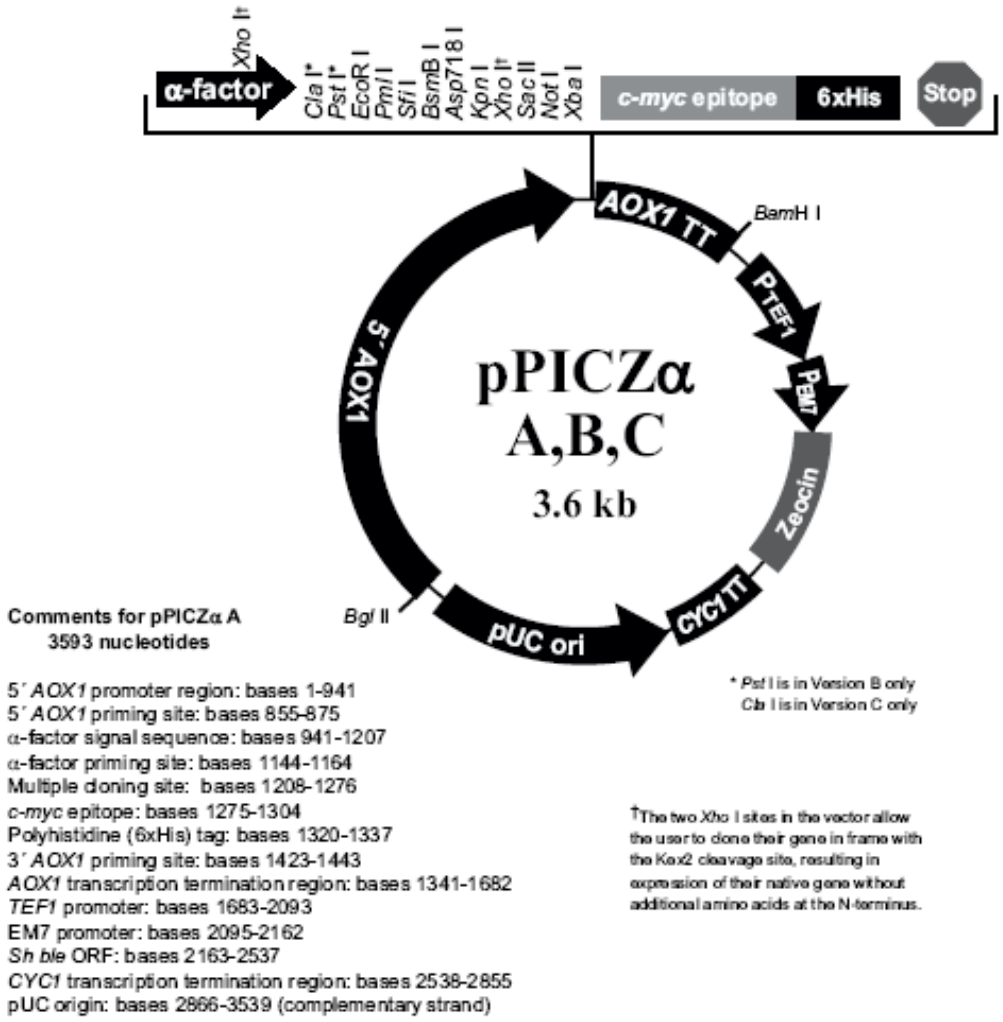


Fig. 7. Plasmid vector map of pPICZα A, B, and C.

Protein expression with *Pichia pastoris* requires a preliminary step of *E. coli* transformation. Brielfly, the construct containing the gene of interest is transformed into *E. coli* XL-1 blue strain cells. Transformant cells are then grown and plasmid vector is purified in amounts suitable for *Pichia* transformation. Yeast cells can be transformed by thermal shock.

Cloning – The human IGFBP-3 clone was purchased from RZPD. The recombinant protein was produced using the methylotrophic yeast *Pichia pastoris* (Invitrogen). *E. coli* cells containing the gene of interest, inserted into a pOTB7 plasmid vector, were grown into 3 ml LB liquid culture at 37°C overnight. Cells were harvested and the plasmid vector was purified using GeneElute Miniprep Plasmid Kit (Quiagen). The gene of interest was then amplified by PCR using the following primers:

Primer forward: GGTGGTGAATTCATGGGCGCGAGCTCGGGGGGCTTG

Primer reverse: GGTGGTTCTAGACTAATGATGATGATGATGATGCTTGCTCTGCATGCTGTAGCA

GAATTC – Eco R I restriction site

TCTAGA – Xba I restriction site

ATGATGATGATGATGATG – 6x Histidine tag

PCR cycle

1- 1 minute at 95°C

2- denaturation: 1 minute at 95°C

3- annealing: 1 minute at 55°C

4- extension: 1 minute at 72°C

5- 15 minutes at 72°C

Steps 2 to 4 repeated 35 times.

The amplified sequence was then quantified on a 0.8% agarose gel and digested with Eco R I and Xba I restriction enzymes and ligated to a pPICZαA plasmid vector (Invitrogen) which was previously digested with the same enzymes and dephosphorylated to prevent self-ligation.

Expression – The pPICZαA vector enables extracellular expression of the recombinant protein which is secreted into the culture medium. The *E. coli* strain XL1 Blue was transformed with the construct obtained and plated on a LB-agar

plate containing 100 µg/ml Zeocin. The resulting colonies were checked for transformation using colony-PCR and digestion with the restriction enzymes. A single positive colony was then transferred into 100 ml LB liquid culture containing 100 µg/ml Zeocin and grown overnight at 37°C. *E.coli* cells were harvested and the plasmid vector was purified using a GeneElute Midiprep Plasmid Kit (Quiagen). 3 µg of purified plasmid were used for *Pichia pastoris* X-33 strain trasformation. Transformant *P. pastoris* cells were plated on a YPDS-Agar plate and after 3 days colonies appeared.

YPDS-agar composition

2% tryptone

1% yeast extract

2% glucose

2% agar

100 µg/ml Zeocin

Four colonies were inoculated into 10 ml Buffered Glycerol-complex Medium (BMGY) liquid cultures and incubated at 28°C on a shaking incubator for 24 hours.

BMGY medium composition

2% tryptone

1% yeast extract

100 mM potassium phosphate pH 6.0

1.34% YNB

$4*10^{-5}$ % Biotin

1% glycerol

Induction was started centrifugating all cultures at 3000 g and resuspending the cells in 10 ml of Buffered Methanol-complex Medium (BMMY).

BMMY medium composition

2% tryptone

1% yeast extract

100 mM potassium phosphate pH 6.0

1.34% YNB

$4*10^{-5}$ % Biotin

1% methanol

Induction was continued for 7 days with a daily addition of 1% methanol. Samples from test cultures were then centrifuged and the supernatant was loaded onto a SDS-PAGE and protein expression was evaluated by Western-blot analysis.

Human IGFBP-3 is susceptible of proteolytic cleavage by metalloproteases in vivo and it has been hypotized that similar proteases could be present in *P. pastoris* cultures. The protocol was then modified in order to prevent protease digestion. The induction medium BMMY was modified with the addition of 25 mM EDTA; 5 mM EDTA was also added daily during the entire induction period. Since EDTA from the medium could not be completely removed and even low concentrations cause Nickel leakage from the affinity resin, ion-exchange chromatography was used to purify the protein.

Purification - After induction the cultures were centrifuged at 8000 RPM for 10 minutes; the supernatant was recovered and washed extensively with Tris/HCl 20 mM pH 7.0, 0.02% NaN_3, and 5 mM EDTA using an Amicon ultrafiltration cell. A DEAE-cellulose syringe column was equilibrated with the same buffer and the protein from the yeast culture was loaded. The column was washed extensively and a step gradient, from 0 M to 0.5 M NaCl was applied. Recombinant IGFBP-3

eluted in fractions containing 150 mM up to 250 mM NaCl. The fractions were pooled, concentrated with an Amicon ultrafiltration cell up to a final volume of 2 ml and loaded onto a Superdex G 75 gel filtration column for a final purification step. After gel filtration the purified protein was stable in the absence of EDTA which indicates that the contaminant proteases were removed.

Crystallization trials – The recombinant protein was concentrated up to 20 mg/ml and used for preliminary crystallization trials.

2.3 Human IGFBP-rP1: cloning, expression, purification, and crystallization trials

Cloning – The clone of human IGFBP-rP1 was purchased from RZPD. *E. coli* cells containing the cloning vector pOTB7 were grown in 3 ml LB culture at 37°C overnight. Cells were harvested and the plasmid DNA was purified using the GeneElute Miniprep Plasmid Kit (Quiagen). The coding sequence was amplified using the following primers:

Primer forward: GGTGGTGGATCCTCCTCTTCGGACACCTGCGGC

Primer reverse: AATAAGCTTGGAACCACGCGGAACCAGTAGCTCGGCACCTTCACC

GGATCC – Bam H I restriction site

AAGCTT – Hind III restriction site

GGAACCACGCGGAACCAG – Thrombin cleavage site

PCR cycle

1- 1 minute at 95°C

2- denaturation: 1 minute at 95°C

3- annealing: 1 minute at 55°C

4- extension: 1 minute at 72°C

5- 15 minutes at 72°C

Steps 2 to 4 repeated 35 times.

The amplified fragment was quantified on a 0.8% agarose gel and digested with Bam H I and Hind III restriction enzymes and then ligated to a pQE50 plasmid vector, which includes a histidine tag added at the C-terminus of the recombinant protein. After ligation, the construct was transformed into *E. coli* XL1 Blue strain and cells were plated on a LB-agar plate containing 100 µg/ml ampicillin. The resulting bacterial colonies were tested by colony-PCR and digestion of the purified plasmid. A single positive colony was inoculated into 3 ml of LB and grown at 37°C overnight. The plasmid DNA was recovered from the cells and used to transform the BL21(DE3) strain of *E. coli*. Transformed cells were tested for protein expression and the production of soluble protein was confirmed by Western-blot analysis.

Expression - Protein expression was checked using 50 ml liquid cultures and incubating at 37°C and 20°C overnight after induction with 0.25 mM IPTG when $O.D._{600}$ reached 0.8-1.0. The optimal conditions for protein expression were 20°C overnight induction which gave a yield of approximately 1.5 mg of purified protein per litre of culture.

Purification - One litre bacterial culture was grown and induced at 20 °C overnight. Cells were harvested and disrupted by sonication. The bacterial lysate containing the recombinant IGFBP-rP1 was loaded onto a Nickel sepharose affinity column and the protein was eluted by applying a linear gradient from 20 mM to 500 mM imidazole. Eluted fractions containing the protein were

concentrated to a final volume of 2 ml and loaded onto a Superdex G 75 column for a final purification step.

Crystallization trials – A preliminary crystallization screen was prepared (Structure Screen 1 and 2, Molecular Dimension Limited) using undigested protein at a concentration of 8 mg/ml but no crystals appeared. The protein eluted from the affinity column showed degradation after 2-3 days and presented 3 distinct bands on SDS-PAGE. Degradation was then prevented by the addition of 25 mM EDTA after the affinity column separation and EDTA removal by gel-filtration was not harmful for the protein since the contaminants have been removed. Protein digestion was carried out on 1 mg/ml IGFBP-rP1 with the addition of 2.5 mM $CaCl_2$ and of 1 unit of thrombin per mg of fusion protein. For digestion the protein was incubated at 20 °C for 16 hours and digestion was then confirmed by Western-blot analysis. The digested protein was loaded onto a gel filtration column to remove thrombin.

2.4 LG3 domain of perlecan: cloning, expression, purification and crystallization

Cloning – The clone containing the coding sequence for the LG3 domain of perlecan was purchased from RZPD. *E. coli* cells (DH10B strain) were inoculated into 3 ml LB liquid medium and incubated overnight at 37°C. Cells were harvested and plasmid DNA was purified using GenElute Plasmid Miniprep kit (Quiagen). The gene of interest was amplified by PCR using the following primers:

Primer forward: **GGTGGTGGATCCCAGTACGGAGCCTATTTC**

Primer reverse: AATAAGCTTGGAACCACGCGGAACCACCGAGGGGCAGGGGCGTGT

GGATCC : Bam H I restriction site

AAGCTT : Hind III restriction site

GGAACCACGCGGAACCAG : thrombin cleavage site

PCR cycle

1- 1 minute at 95°C

2- denaturation: 1 minute at 95°C

3- annealing: 1 minute at 55°C

4- extension: 90 seconds at 72°C

5- 15 minutes at 72°C

Steps 2 to 4 repeated 35 times.

The amplified DNA was quantified on a 0.8% agarose gel and digested with Bam H I and Hind III restriction enzymes and ligated to a pQE50 expression vector, which was digested with the same enzymes and dephosphorylated to prevent self-ligation. An amplified fragment to plasmid vector molar ratio of 5:1 was used. *E. coli* XL1 Blue strain cells were transformed by thermal shock with the construct obtained and plated on a LB-agar plate. Bacterial colonies were checked for transformation by colony-PCR and by restriction enzyme digestion. The plasmid was recovered from the positive colonies and used for *E. coli* BL21 strain and SG strain transformation.

Expression - Transformant colonies were transferred into 10 ml of LB liquid medium containing 100 µg/ml ampicillin and grown at 37°C until absorbance at 600 nm reached 0.8. Induction was started by adding 0.25 mM IPTG and incubating at 20°C overnight on a shaking incubator. Expression levels were

evaluated by SDS-PAGE analysis and optimal conditions found were overnight induction at 20°C with 0.5 mM IPTG.

Purification – Recombinant protein purification was carried out by affinity chromatography using a Nickel Sepharose column. A one litre culture was grown until absorbance at 600 nm was 0.8-1.0 and induction was initiated as described before. Cells were harvested and disrupted by sonication and the recombinant protein was purified by affinity chromatography and by gel-filtration chromatography.

Crystallization trials- The purified protein was concentrated up to 20 mg/ml and was used for a preliminary crystallization trial (Structure Screen 1 and 2, Molecular Dimensions Limited).

3. Results

3.1 Human IGF-II

IGF-II from inclusion bodies – The first refolding protocol tested was refolding by affinity chromatography. The inclusion bodies were solubilized in TBS with 6 M guanidinium chloride and 10 mM β-mercaptoethanol and the solubilized protein was loaded onto a Nickel Sepharose column which was equilibrated with the same buffer.

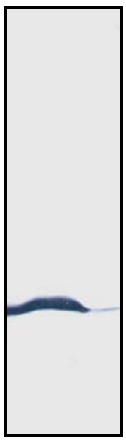


Fig. 10. Western Blot analysis of solubilized IGF-II from inclusion bodies

The affinity column was then washed extensively with the same buffer until absorbance at 280 nm was negligible, indicating that proteins unable to bind to the affinity column had been eluted. Then a gradient was applied against TBS to remove the denaturing agent. The refolded protein was eluted applying a linear gradient from 20 mM to 500 mM imidazole. The eluted protein was then concentrated up to a volume of 2 ml and loaded onto a Superdex G 75 gel filtration column for a final purification step. The protein yield was of 2 mg of refolded protein from 1 litre of bacterial culture. The purified protein was concentrated to 20 mg/ml and used for preliminary crystallization trials (Crystal

Screen 1 and 2, Hampton Research). Condition n. 35 of Crystal Screen 1 gave small crystals after 6 days at 4°C. Diffraction data collection was performed at the European Synchrotron Radiation Facility (ESRF, Grenoble). The crystals are orthorhombic, space group P222, with a=28.22, b=50.61, c=65.25, α=β=γ=90°, and diffract to a resolution of 2.5 Å. The data were processed using the program MOSFLM. After integration and scaling of the data in the CCP4 programs (Collaborative Computational Project 4) an attempt to solve the structure by molecular replacement was made. Molecular replacement is a technique used to determine the orientation and position of a molecule in the unit cell using a previously solved structure as a 'search model'. The search and target molecules must have reasonable sequence identity (> 25 %). The solution of a structure by molecular replacement consists in finding a 'search model', and in determining a rotation and translation function used to fit the 'search model' into the experimental unit cell. The molecular model used as a search probe was a solution structure determined by NMR of IGF-II (PDB ID: 1IGL). Molecular replacement gave very ambiguous rotation and translation function results and no clear solutions indicating that the recombinant protein might have been incorrectly folded.

The second protocol tested was refolding by dilution. After protein solubilization and affinity column loading in TBS with 8 M urea and 10 mM β-mercaptoethanol a linear gradient of imidazole, from 20 mM to 500 mM imidazole in 8 M urea and 10 mM β-mercaptoethanol TBS, was applied. Fractions containing the eluted protein were pooled and rapidly diluted into a refolding buffer. Different compositions were tested; TBS with arginine concentrations ranging from 100 mM up to 500 mM, with a dilution ratio of 1:10, gave soluble protein with a yield of approximately 10% as estimated by SDS-PAGE. Subsequent removal of urea and arginine caused protein precipitation.

The third refolding protocol attempted was dialysis refolding. Denatured protein was eluted from the affinity column as described in the previous protocol. The protein was dialyzed against TBS containing 6 M, 4 M, 2 M, and no urea with an equilibration time of 2 hours for every step. After complete removal of urea a precipitate appeared into the dialysis membrane and no soluble protein could be recovered.

Soluble IGF-II - Recombinant human IGF-II expression levels from *E. coli* cells transformed with the pET-28 His expression vector were evaluated by Western-blot analysis. Optimal induction conditions for soluble protein expression were 28°C induction in Terrific Broth overnight.

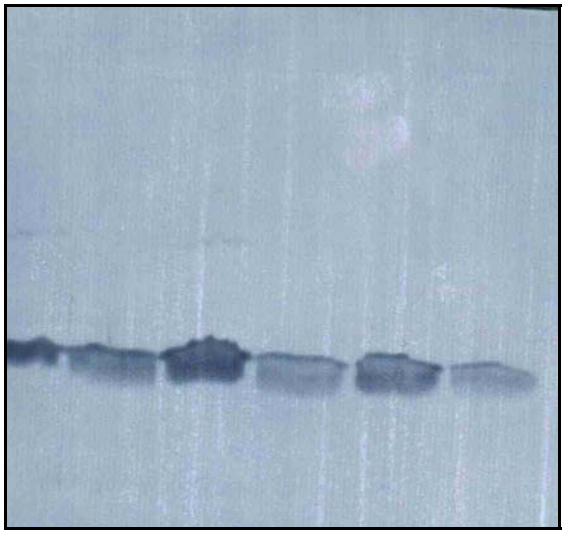


Fig. 11. Western Blot Analysis of soluble IGF-II. Lane 1 and 2: 37°C TB and 37°C LB; lane 3 and 4:28°C TB and 28°C LB; 16°C TB and LB.

Purification - 1 litre of culture was then inoculated with 3 ml of overnight culture of a BL21 positive colony. The soluble protein was applied onto a Nickel-Sepharose affinity column and the retained proteins were eluted with an imidazole linear gradient from 20 mM to 500 mM imidazole. Soluble IGF-II co-eluted with low-affinity *E.coli* proteins.

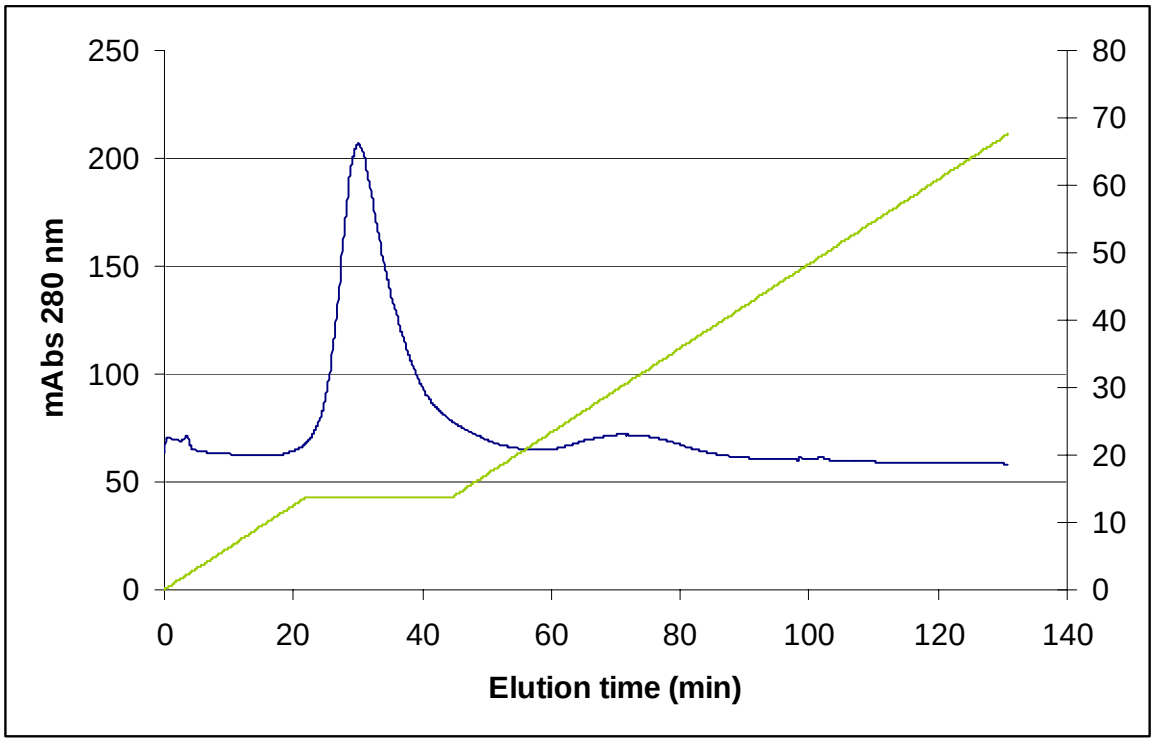


Fig. 12. Elution profile of IGF-II from Nickel Sepharose affinity column. Blue line corresponding to UV absorbance and green line to imidazole gradient in percentage. IGF-II can be found in the first peak, with other low specificity proteins.

3.2 Human IGFBP-3

Recombinant protein produced in BMMY medium without EDTA addition showed strong degradation.

Fig. 13. Western blot showing IGFBP-3 degradation.

The protein produced using the modified protocol was loaded onto a DEAE-cellulose column and fractions containing the recombinant protein were concentrated and purified by gel-filtration chromatography.

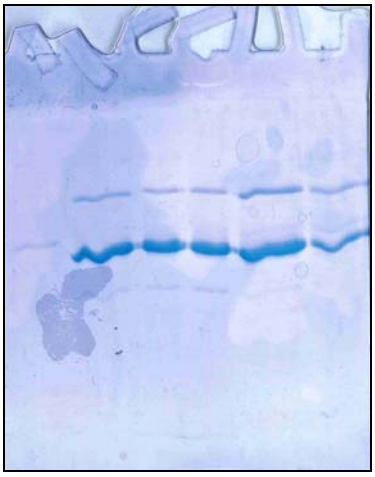


Fig. 14. SDS-PAGE analysis of fractions eluted from gel-filtration chromatography showing recombinant IGFBP-3.

Recombinant human IGFBP-3 from gel-filtration chromatography was dialyzed against 20 mM Tris/HCl pH 7.5, 0.02% NaN_3 and concentrated with Amicon devices up to a final concentration of 20 mg/ml. Preliminary crystallization conditions were then tested (Structure Screen 1 and 2, Molecular Dimensions Limited) but none of the conditions gave crystals. All of the crystallization experiments used the hanging drop vapor diffusion technique with 2 µl drops containing 1 µl of protein solution and 1 µl of precipitant.

Structure screen 1: composition

1	0.02M Calcium chloride dihydrate	0.1M Na Acetate trihydrate pH 4.6	30% v/v 2-methyl-2,4 pentanediol
2	0.2M Ammonium acetate	0.1M Na Acetate trihydrate pH 4.6	30% w/v PEG 4000
3	0.2M Ammonium sulphate	0.1M Na acetate trihydrate pH 4.6	25% w/v PEG 4000
4	None	0.1M Na acetate trihydrate pH 4.6	2.0M Sodium formate
5	None	0.1M Na acetate trihydrate pH 4.6	2.0M Ammonium sulphate
6	None	0.1M Na acetate trihydrate pH 4.6	8% w/v PEG 4000
7	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% w/v PEG 4000

8	0.2M Ammonium acetate	0.1M tri-sodium citrate dihydrate pH 5.6	30% v/v 2-methyl-2,4-pentanediol
9	None	0.1M tri-Sodium citrate dihydrate pH 5.6	20% w/v 2-propanol, 20% w/v PEG 4000
10	None	0.1M Na Citrate pH 5.6	1.0M Ammonium dihydrogen phosphate
11	0.2M Calcium chloride dihydrate	0.1M Na acetate trihydrate pH 4.6	20% v/v 2-propanol
12	None	0.1M Na Cacodylate pH 6.5	1.4M Na acetate trihydrate
13	0.2M tri-sodium citrate dihydrate	0.1M Na Cacodylate pH 6.5	30% v/v 2-propanol
14	0.2M Ammoniium sulphate	0.1M Na Cacodylate pH 6.5	30% w/v PEG 8000
15	0.2M Magnesium acetate tetrahydrate	0.1M Na Cacodylate pH 6.5	20% PEG 8000
16	0.2M Magnesium acetate tetrahydrate	0.1M Na Cacodylate pH 6.5	30% v/v 2-methyl-2,4-pentanediol
17	None	0.1M Imidazole pH 6.5	1.0M Sodium acetate trihydrate
18	0.2M Sodium acetate trihydrate	0.1M Na Cacodylate pH 6.5	30% w/v PEG 8000
19	0.2M Zinc acetate dihydrate	0.1M Na Cacodylate pH 6.5	18% w/v PEG 8000
20	0.2M Calcium acetate hydrate	0.1M Na Cacodylate pH 6.5	18% w/v PEG 8000
21	0.2M tri-sodium citrate dihydrate	0.1M Na Hepes pH 7.5	30% v/v 2-methyl-2,4-pentanediol
22	0.2M Magnesium chloride hexahydrate	0.1M Na Hepes pH 7.5	30% v/v 2-propanol
23	0.2M Calcium chloride dihydrate	0.1M Na Hepes pH 7.5	28% v/v PEG 400
24	0.2M Magnesium chloride hexahydrate	0.1M Na Hepes pH 7.5	30% v/v PEG 400
25	0.2M tri-sodium citrate dihydrate	0.1M Na Hepes pH 7.5	20% v/v 2-propanol
26	None	0.1M Na Hepes pH 7.5	0.8M K, Na tartrate tetrahydrate
27	None	0.1M Na Hepes pH 7.5	1.5M Lithium sulphate monohydrate
28	None	0.1M Na Hepes pH 7.5	0.8M Na dihydrogen phosphate 0.8M K dihydrogen phosphate monohyd.
29	None	0.1M Na Hepes pH 7.5	1.4M tri-Sodium citrate dihydrate
30	None	0.1M Na Hepes pH 7.5	2% v/v PEG 400, 2.0M Amm sulphate
31	None	0.1M Na Hepes pH 7.5	10% v/v 2-propanol, 20% w/v PEG 4000
32	,None	0.1M Tris HCl pH 8.5	2.0M Ammoniium sulphate
33	0.2M Magnesium chloride hexahydrate	0.1M Tris HCl pH 8.5	30% w/v PEG 4000
34	0.2M tri-sodum citrate dihydrate	0.1M Tris HCl pH 8.5	30% v/v PEG 400
35	0.2M Lithium sulphate monohydrate	0.1M Tris HCl pH 8.5	30% w/v PEG 4000
36	0.2M Ammonium acetate	0.1M Tris HCl pH 8.5	30% v/v 2-propanol
37	0.2M Sodium acetate trihydrate	0.1M Tris HCl pH 8.5	30% w/v PEG 4000
38	None	0.1M Tris HCl pH 8.5	8% w/v PEG 8000
39	None	0.1M Tris HCl pH 8.5	2.0M Ammonium dihydrogen phosphate
40	None	None	0.4M K, Na Tartrate tetrahydrate
41	None	None	0.4M Ammonium dihydrogen phosphate
42	0.2M Ammonium sulphate	None	30% w/v PEG 8000
43	0.2M Ammonium sulphate	None	30% w/v PEG 4000
44	None	None	2.0M Ammonium sulphate
45	None	None	4.0M Sodium formate

46	0.05M Potassium dihydrogen phosphate	None	20% w/v PEG 8000
47	None	None	30% w/v PEG 1500
48	None	None	0.2M Magnesium formate
49	1.0M Lithium sulphate monohydrate	None	2% w/v PEG 8000
50	0.5M Lithium sulphate monohydrate	None	15% w/v PEG 8000

Structure Screen 2: composition

1	0.1M Sodium chloride	0.1M Bicine pH 9.0	30% w/v PEG monomethylether 550
2	None	0.1M Bicine pH 9.0	2.0M Magnesium chloride hexahydrate
3	2% w/v Dioxane	0.1M Bicine pH 9.0	10% w/v PEG 20,000
4	0.2M Magnesium chloride hexahydrate	0.1M Tris pH 8.5	3.4M 1,6 Hexanediol
5	None	0.1M Tris pH 8.5	25% v/v tert-Butanol
6	0.01M Nickel chloride hexahydrate	0.1M Tris pH 8.5	1.0M Lithium sulphate
7	1.5M Ammonium sulphate	0.1M Tris pH 8.5	12% v/v Glycerol
8	0.2M Ammonium phosphate monobasic	0.1M Tris pH 8.5	50% v/v MPD
9	None	0.1M Tris pH 8.5	20% v/v Ethanol
10	0.01M Nickel chloride hexahydrate	0.1M Tris pH 8.5	20% w/v PEG monomethylether 2000
11	0.5M Ammonium sulphate	0.1M Hepes pH 7.5	30% v/v MPD
12	None	0.1M Hepes pH 7.5	10% w/v PEG 6000, 5% v/v MPD
13	None	0.1M Hepes pH 7.5	20% v/v Jeffamine M-600
14	0.1M Sodium chloride	0.1M Hepes pH 7.5	1.6M Ammonium sulphate
15	None	0.1M Hepes pH 7.5	2.0M Ammonium formate
16	0.05M Cadmium sulphate octahydrate	0.1M Hepes pH 7.5	1.0M Sodium acetate
17	None	0.1M Hepes pH 7.5	70% v/v MPD
18	None	0.1M Hepes pH 7.5	4.3M Sodium chloride
19	None	0.1M Hepes pH 7.5	10% w/v PEG 8000 8% v/v Ethylene glycol
20	None	0.1M Mes pH 6.5	1.6M Magnesium sulphate heptahydrate
21	0.1M Na phosphate monobasic	0.1M Mes pH 6.5	2.0M Sodium Chloride 0.1M K phosphate monobasic
22	None	0.1M Mes pH 6.5	12% w/v PEG 20,000
23	1.6M Ammonium sulphate	0.1M Mes pH 6.5	10% v/v Dioxane
24	0.05M Cesium chloride	0.1M Mes pH 6.5	30% v/v Jeffamine M-600
25	0.01M Cobalt chloride hexahydrate	0.1M Mes pH 6.5	1.8M Ammonium sulphate
26	0.2M Ammonium sulphate	0.1M Mes pH 6.5	30% w/v PEG monomethylether 5000
27	0.01M Zinc sulphate heptahydrate	0.1M Mes pH 6.5	25% v/v PEG monomethylether 550
28	None	0.1M Hepes pH 7.5	20% w/v PEG 10,000

29	0.2M K/Na Tartrate	0.1M Sodium citrate pH 5.6	2.0M Ammonium sulphate
30	0.5M Ammonium sulphate	0.1M Sodium citrate pH 5.6	1.0M Lithium sulphate
31	0.5M Sodium chloride	0.1M Sodium citrate pH 5.6	4% v/v polyethyleneimine
32	None	0.1M Sodium citrate pH 5.6	35% v/v tert-butanol
33	0.01M Ferric chloride hexahydrate	0.1M Sodium citrate pH 5.6	10% v/v Jeffamine M-600
34	0.01M Manganese chloride tetrahydrate	0.1M Sodium citrate pH 5.6	2.5M 1,6 Hexanediol
35	None	0.1M Sodium acetate pH 4.6	2.0M Sodium chloride
36	0.2M Sodium Chloride	0.1M Sodium acetate pH 4.6	30% v/v MPD
37	0.01M Cobalt Chloride hexahydrate	0.1M Sodium acetate pH 4.6	1.0M 1,6 Hexanediol
38	0.1M Cadmium chloride	0.1M Sodium acetate pH 4.6	30% v/v PEG 400
39	0.2M Ammonium sulphate	0.1M Sodium acetate pH 4.6	30% w/v PEG monomethylether 2000
40	2.0M Sodium Chloride	None	10% w/v PEG 6000
41	0.01M Cetyl trimethyl ammoniumbromide	None	0.5M Sodium chloride 0.1M Magnesium chloride hexahydrate
42	None	None	25% v/v Ethylene glycol
43	None	None	35% v/v Dioxane
44	2.0M Ammonium Sulphate	None	5% v/v Isopropanol
45	None	None	1.0M Imidazole pH 7.0
46	None	None	10% w/v PEG 1000, 10% w/v PEG 8000
47	1.5M Sodium Chloride	None	10% v/v Ethanol
48	None	None	1.6M Sodium citrate pH 6.5

Crystal screening were used to co-crystallize IGFBP-3 both with human and bovine holo-transferrin, which are reported in the literature to bind IGFBP-3, but none of the two crystal screens was successful. Recombinant protein deglycosylation was also investigated: the protein was treated with the enzyme Endoglycosidase F which was added to a molar ratio of 25:1 (recombinant protein to enzyme) and incubated at 37°C for 48 hours. SDS-PAGE analysis of the protein treated with the enzyme showed no deglycosylation compared to the untreated protein, suggesting that no saccharide chains had been removed.

Fig.15. SDS-PAGE analysis of recombinant IGFBP-3 digestion by the enzyme Endo F. First lane, untreated protein. Last lane, protein treated with Endo F. The presence of the enzyme is confirmed by the high molecular weight band in the second lane.

3.3 Human IGFBP-rP1

Purification - Purification of recombinant IGFBP-rP1 was carried out as follows. Cells from a single positive colony were inoculated into 3 ml liquid medium and grown overnight at 37°C. 1 litre culture was inoculated with 3 ml of overnight grown cells and incubated at 37°C until the $O.D._{600}$ reached 0.8-1.0. Protein expression was induced by adding 0.25 mM IPTG and incubating overnight at 20°C. Cells were then centrifuged and resuspended in 30 ml of Tris/HCl 20 mM pH 7.5, 0.5 M NaCl, and 0.02% NaN_3 (TBS) and disrupted by sonication. The soluble fraction of the bacterial lysate was then loaded onto a Nickel-Sepharose affinity column. A linear gradient was applied from 40 mM up to 500 mM imidazole.

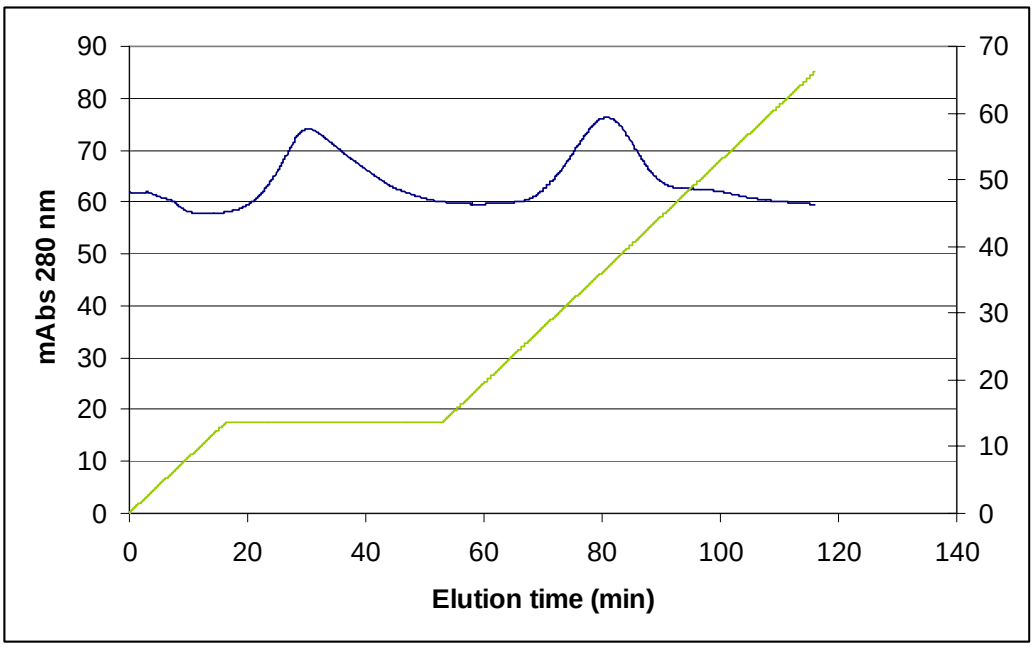


Fig.16.Elution profile of IGFBP-rP1 from the Nickel Sepharose affinity column. The blue line corresponds to UV absorbance and the green line to the imidazole gradient in percentage. IGFBP-rP1 is found in the second peak.

Fractions corresponding to the two peaks were analyzed by SDS-PAGE and fractions containing the recombinant protein were pooled and further purified by gel-filtration chromatography.

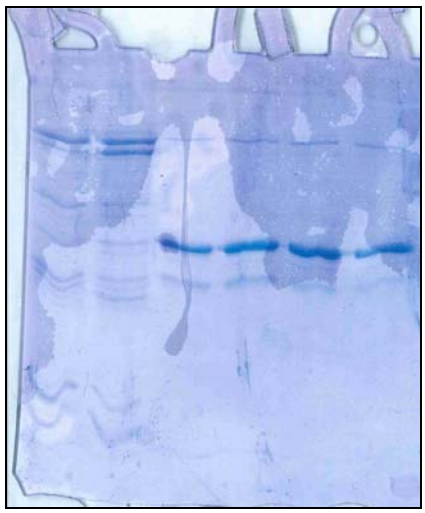


Fig. 17. SDS-PAGE analysis of elution from affinity column of recombinant IGFBP-rP1. The first two lanes correspond to the first peak containing contaminants with low specificity for the Nickel Sepharose matrix. The last four lanes correspond to the second peak and show the recombinant protein eluted from the column.

Crystallization - Fractions containing recombinant IGFBP-rP1 after digestion and gel-filtration were pooled and concentrated to a final concentration of 20 mg/ml and the protein was used for crystal screening. One condition (n.28 Structure Screen 2) gave small crystals after approximately 3 weeks which were not suitable for X-ray diffraction analysis.

Fig. 18. Magnified photograph of crystals of IGFBP-rP1.

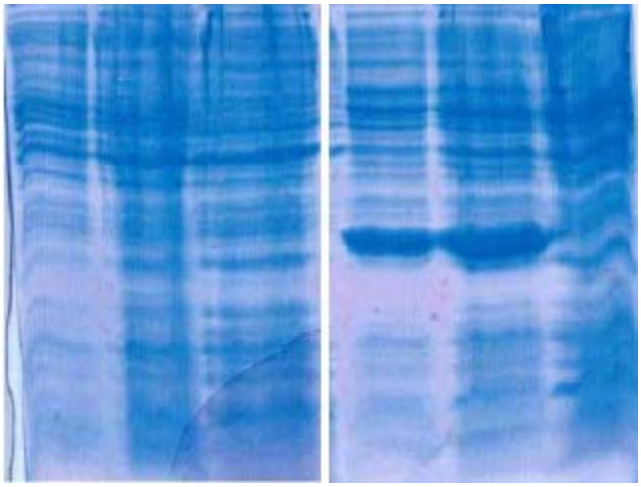
SDS-PAGE analysis of the concentrated protein showed that after 2-3 days there was protein degradation. The major problem of histidine-tag removal was the presence of a potential thrombin cleavage site in the primary sequence located at residue 52, and considering the small difference from thrombin and IGFBP-rP1 in molecular mass, it has been hypotized that gel filtration could not completely remove thrombin from the protein of interest. In order to prevent protein degradation 5 mM EDTA and 1:1000 PMSF (100 mM stock solution) were added

immediately after digestion but some thrombin activity was still present and protein degradation was observed.

Fig. 19. SDS-PAGE analysis showing recombinant IGFBP-rP1 in the first lane. The second lane contains the protein after thrombin digestion.

3.4 LG3, C-terminal domain of perlecan.

Recombinant LG3 expression levels were evaluated by SDS-PAGE analysis, loading samples from bacterial cultures before induction with IPTG and from the total fraction and the soluble fraction after induction.

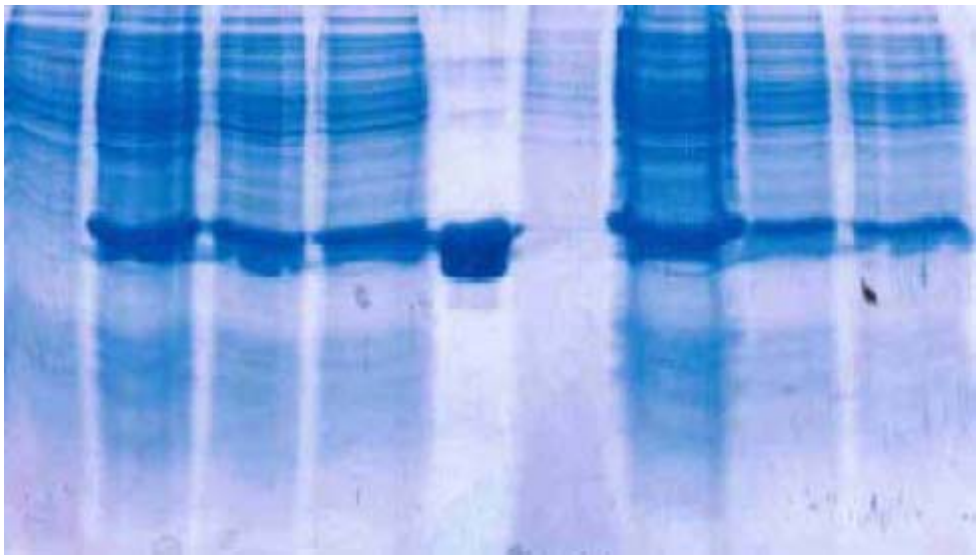


T_0 T S S T T_0

Fig. 20. SDS-PAGE analysis of LG3 expression. T_0 : Before induction ; T : total fraction ; S : soluble fraction. BL21 strain (gel on the left) shows no overexpressed bands. SG strain (gel on the right) shows overexpressed bands on T and S fractions.

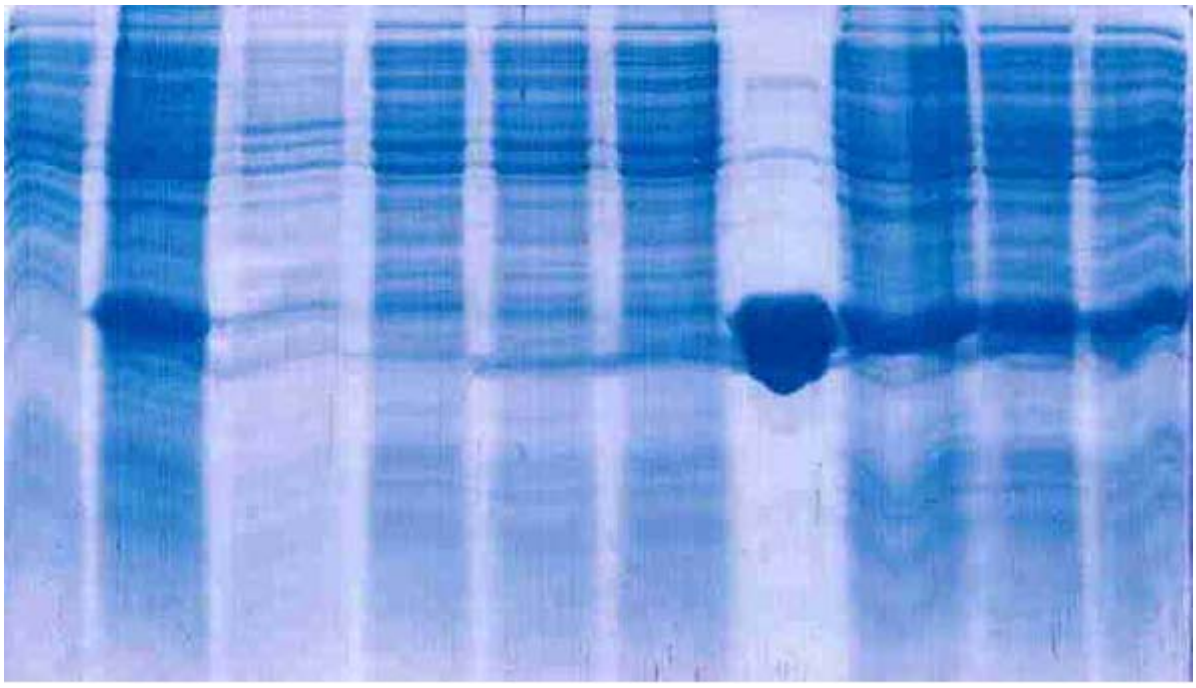
The SG strain showed acceptable levels of protein expression on SDS-PAGE and different induction conditions were tested using 0.25 mM and 0.5 mM IPTG:

- overnight induction at 20°C.
- 8 hours and overnight induction at 25°C.
- 4 hours and 8 hours induction at 37°C.



T_0(20°C) T(20°C) $S_{0.25}$ $S_{0.5}$ marker T_0(25°C) T(25°C) $S_{0.25}$ $S_{0.5}$

Fig. 21. SDS-PAGE analysis of different expression conditions of LG3 domain. T_0(20°C): Bacterial colony before induction; T: total fraction; $S_{0.25}$: soluble fraction after induction with 0.25 mM IPTG; $S_{0.5}$: soluble fraction after induction with 0.5 mM IPTG. 20°C overnight induction on the left. 25°C 8 hours induction on the right.



T_0(37°C) T(37°C) $S_{0.25}$(4 hrs) $S_{0.5}$(4 hrs) $S_{0.25}$(8 hrs) $S_{0.5}$(8 hrs) marker T(25°C) $S_{0.25}$(25°C) $S_{0.5}$(25°C)

Fig. 22. Induction at 37°C for 4 and 8 hours on the left. Induction at 25°C overnight on the right.

SDS-PAGE analysis showed comparable expression levels for induction overnight with 0.5 mM IPTG at 20°C and 25°C.

Purification - Purification of recombinant LG3 was carried out using cells from a 1 L culture induced with 0.5 mM IPTG and incubated at 20°C overnight. Cells were disrupted by sonication and the soluble fraction was applied on a Nickel-Sepharose affinity column.

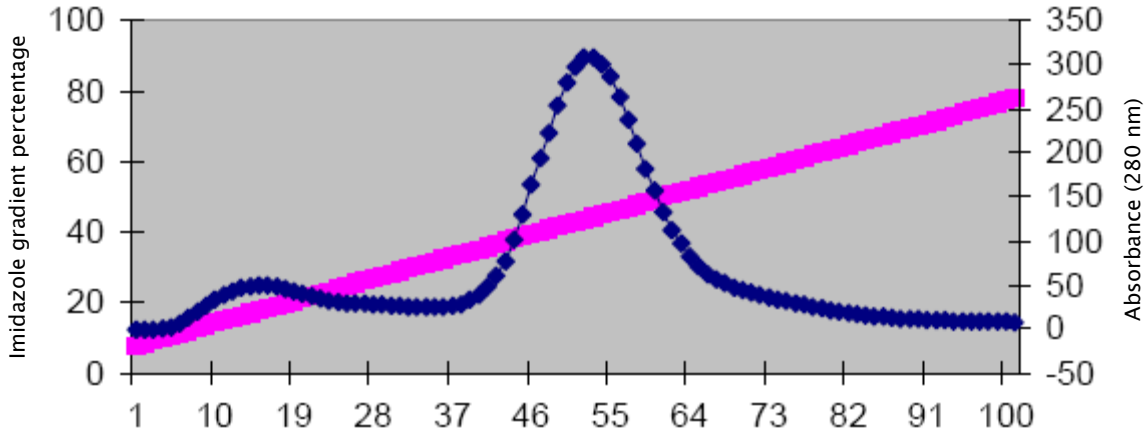


Fig. 23. Elution profile of LG3 from affinity chromatography. Recombinant LG3 eluted in the second peak.

Fractions containing eluted LG3 were pooled and dyalized in an Amicon ultrafitration cell against 20 mM Tris/HCl pH 7.5, 0.15 M NaCl, 2.5 mM $CaCl_2$. The protein of interest was concentrated at 1 mg/ml and 10 units of thrombin were added. Digestion was carried out incubating for 16 hours at 20°C. After digestion the protein was further concentrated up to a final volume of 2 ml and loaded on a Superdex G 75 gel filtration column. Digestion efficiency was evaluated by Western-blot analysis and the digested protein was concentrated to a final concentration of 20 mg/ml.

Crystallization – The recombinant protein was used for preliminary crystallization trials, based on the vapor diffusion method, consisting in 98 hanging drops (Structure Screen 1 and 2, Molecular Dimensions Limited). After one day crystals appeared on condition n. 15 of structure screen 1. The precipitant composition was: 0.2 M Magnesium Acetate, 0.1 Sodium Cacodilate pH 6.5, and 20% PEG 8000. Since crystals grow twinned it was necessary to test similar conditions in order to obtain single crystals.

At first, precipitant pH and PEG 8000 concentration were modified as follows:

- pH 6.5, and 20% PEG 8000
- pH 7.0, and 20% PEG 8000
- pH 6.5, and 15% PEG 8000
- pH 7.0, and 15% PEG 8000

After some days crystals formed in drops containing 15% PEG 8000 but without changing the crystal morphology. The original condition was then modified by adding different concentrations of ethylene glycol, which favors the formation of single crystals. The following conditions, with different concentrations of additives and at different pHs were tested. In every case 5% ethylene glycol was present in the samples.

- 18% PEG 8000 and pH 6.0
- 18% PEG 8000 and pH 6.5
- 18% PEG 8000 and pH 7.5
- 20% PEG 8000 and pH 6.0
- 20% PEG 8000 and pH 6.5
- 20% PEG 8000 and pH 7.5
- 22% PEG 8000 and pH 6.0
- 22% PEG 8000 and pH 6.5

- 22% PEG 8000 and pH 7.5
- 25% PEG 8000 and pH 6.0
- 25% PEG 8000 and pH 6.5
- 25% PEG 8000 and pH 7.5
-

None of the conditions showed significant improvement of crystalline morphology.

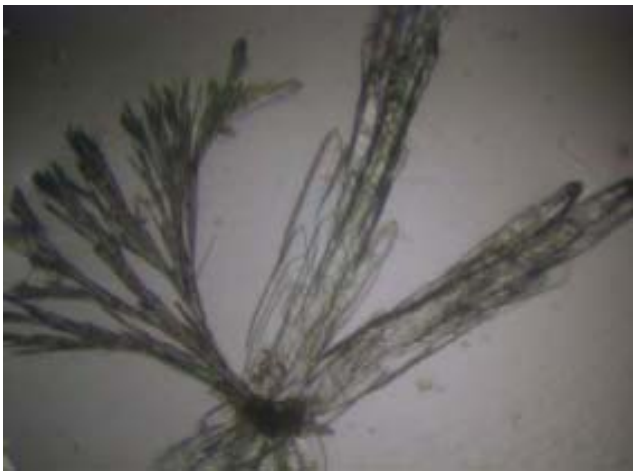


Fig. 24. Microscopy magnification of crystals grown in 0.1 M sodium cacodylate pH 6.0, 0.2 M magnesium acetate, 25% PEG 8000, and 5% ethylen glycole.

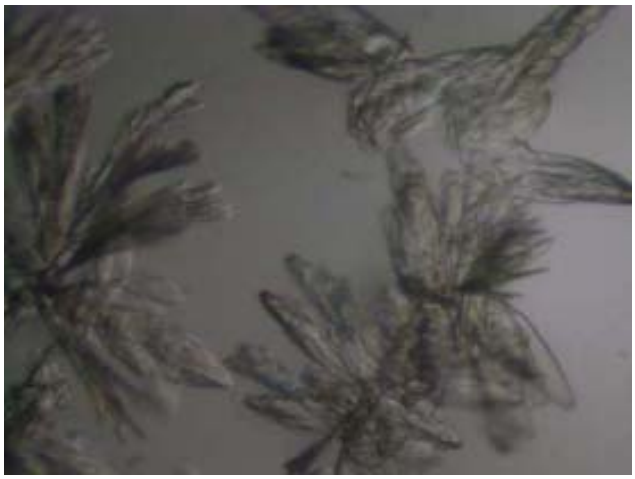


Fig. 25. Microscopy magnification of crystals grown in 0.1 M sodium cacodylate pH 6.5, 0.2 M magnesium acetate, 20% PEG 8000, and 5% ethylen glycole.

Fig. 26. Microscopy magnification of crystals grown in 0.1 M sodium cacodylate pH 7.5, 0.2 M magnesium acetate, 25% PEG 8000, and 5% ethylen glycole.

Ethylene glycol concentrations were also varied: 2%, 4%, 8%, and 20% were included in the original condition.

Fig. 27. Photograph showing crystals grown in the original condition modified by adding 20% ethylen glycole.

An improvement of the crystalline morphology was obtained by using calcium acetate, in place of magnesium acetate.

Fig. 28. Magnification of crystals grown in 0.2 M calcium acetate, 0.1 M sobum cacodylate pH 6.5, 20% PEG 8000, and 20% ethylen glycole.

Crystals grown under these conditions were then used for X-ray data collection.

Data collection – X-ray diffraction measurements were performed at the BM-16 beamline at ESRF (European Synchrotron Radiation Facility; Grenoble, France). Crystals were frozen at -180°C using 20% glycerol as cryoprotectant. 180 frames were collected with a $\Delta\Phi = 1°$. Initial data processing and experimental intensities integration were calculated with MOSFLM.
Initial data processing can be subdivided into three steps:

1. Determining of the crystal orientation, cell parameters and possible space group.
2. Generating the reflection lists and integrating the images.
3. Scaling and merging the resulting data.

The first two steps were carried out using MOSFLM while for the third one the program SCALA was used. After data collection MOLSFLM integrated the collected

images using image files and a command file as input files. The command file included information necessary for MOSFLM such as detector type, detector distance, and X-ray beam wavelength. When integrating a new dataset the first option to be used is autoindexing: MOSLFM then uses the reflections found on the images and generates a list of possible space groups, along with their penalties and calculates an orientation matrix. Once the operator has chosen a possible space group as a solution, diffraction intensities can be integrated. MOSFLM generates an output file containing the integrated intensities and a log file with information and statistics about the calculations performed.

The final dataset statistics are reported below:

Space Group	$P2_1$
Cell Dimensions (Å)	a=43.0229 b=40.9078 c=45.9863 α=90.0° β= 93.5426° γ=90.0°
Resolution Range (Å)	45.883 – 1.60
Observed reflections	63978
Independent reflections	18911
Multiplicity (outer shell)	3.3 (2.3)
R_{sym} (outer shell)	5.4 % (45.5%)
I/σ (outer shell)	8.5 (1.9)
Completeness (outer shell)	91.9 % (63.6%)

The integrated intensities were then processed using SORT, SCALA, and CAD (Collaborative Computational Project 4, CCP4). Electron density maps were calculated using a model of LG3 built on data collected on a conventional X-ray source. The initial model was obtained as follows. Crystals from the same condition (0.2 M calcium acetate, 0.1 M sobum cacodylate pH 6.5, 20% PEG 8000, and 20% ethylen glycol) were collected on a Rigaku RU-300 rotating anode X-ray

source; a crystal was mounted in a glass capillary, data collection was performed at 20°C, and diffraction intensities were recorded on a MarResearch imaging plate. Data processing was done using Automar. The data collection statistics are reported below:

Space group	$P2_1$
Cell Dimensions	a=44.989 b=41.985 c=47.076 α=90.0° β= 94.199° γ=90.0°
Resolution Range (Å)	30–2.4
Observed reflections	18780
Independent reflections	7002
Multiplicity	2.41
R_{sym} (outer shell)	7.27% (42.13%)
I/σ (outer shell)	5.7 (1.1)
Completeness (outer shell)	97.2 % (99.6%)

Experimental phases were determined by molecular replacement. A suitable probe was found using the program MrBump which searched for similar proteins using the automated BLAST sequence similarity search in online databases. The program also modified the model found on the basis of the LG3 sequence with CHSAW and calculated the rotation and translation function with MOLREP or PHASER. This model was then refined and used for refinement with the data collected at the ESRF using REFMAC.

Resolution (Å)	42.954–1.60
Number of used reflections (outer shell)	18449 (842)
R_{crys}(%)(outer shell)	29.36 (36%)
R_{free}(%)(outer shell)	33.61 (49%)

Number of atoms	1410
Solvent molecules	0
Rmsd bond distances	0.009
Rmsd bond angles	1.329

The electron density map obtained after several cycles of manual model building and refinement was well defined for the major part of the protein.

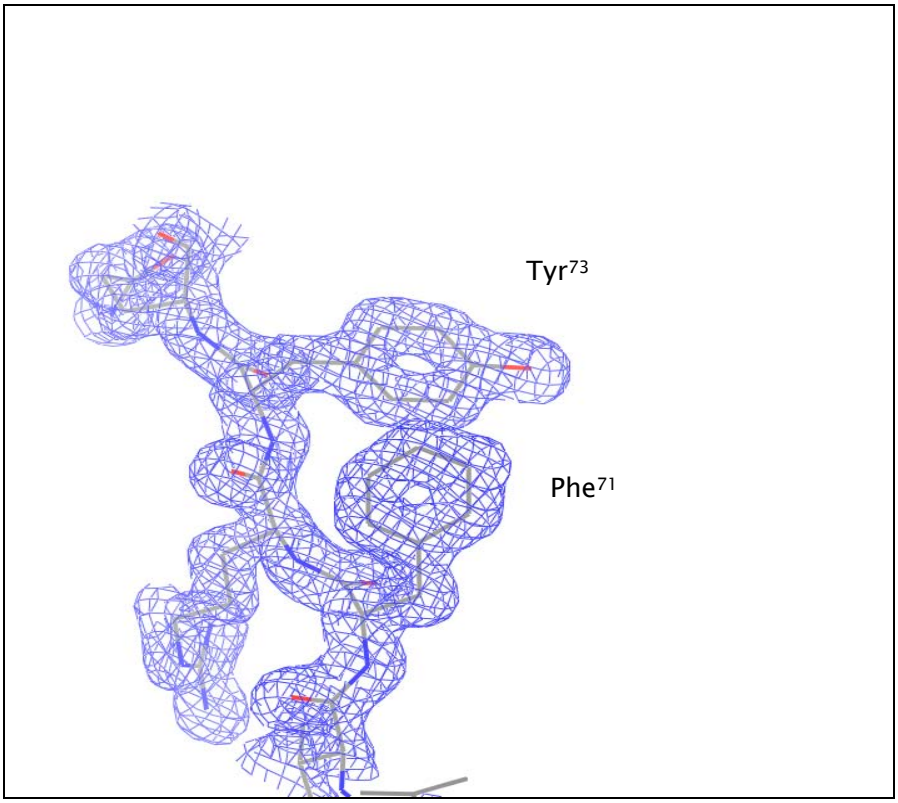


Fig. 29. Electron density map showing Phe[71] and Tyr[73]

A problem which affected both the datasets collected at ESRF and on the conventional X-ray source was represented by the presence of two mobile loops. These loops showed no interpretable electron density and refinement using different models of those regions gave no map improvement.

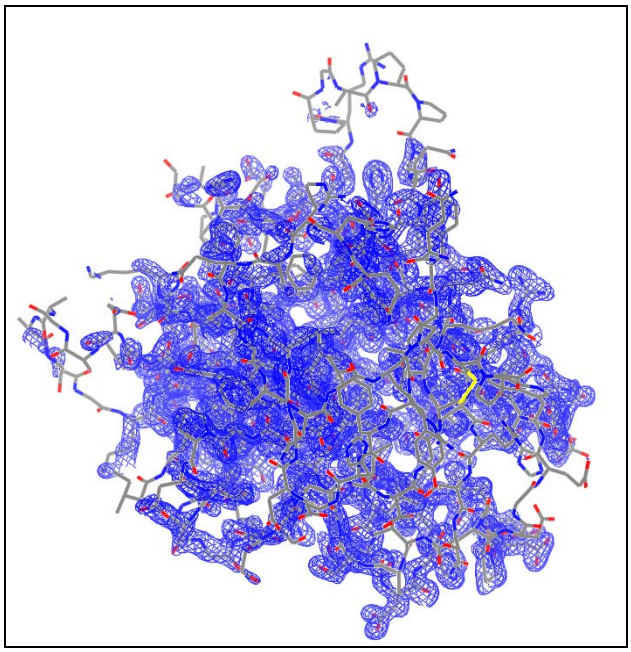


Fig. 30. Electron density map of LG3 with superimposed model. On top and on the left of map are visible two loops with no electron density.

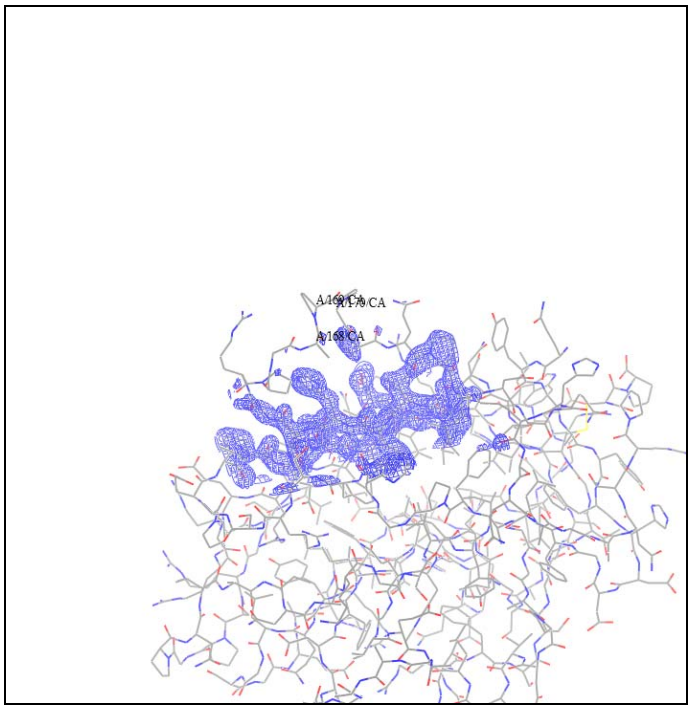


Fig. 31. Picture showing missing electron density of residues 166-171.

Electron density map also showed a disulphide bond between cys^{155} and cys^{189} .

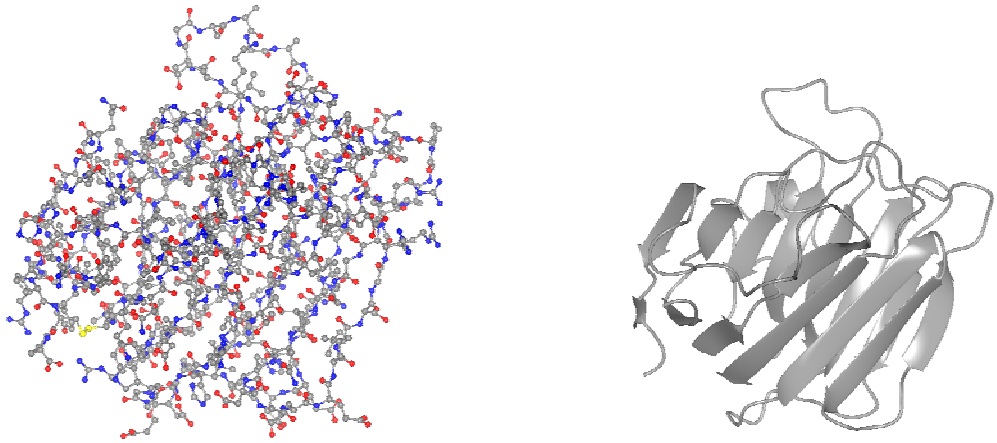


Fig. 32. Pictures showing ball-and-stick model of LG3 with disulphide bond in yellow on the left; ribbon model representation on the right.

4. Conclusions

4.1 Recombinant human IGF-II

Protein refolding in the case of IGF-II was completely unattainable. For all of the protocols described several parameters were varied like temperature, buffer compositions, timing of the various steps. The soluble protein was always barely detectable and decreased due to the precipitation processes. Soluble protein expression using a different plasmid vector was successful but caused other purification problems. The recombinant protein in fact co-elutes along with other proteins with low affinity for the Ni-Sepharose matrix and subsequent separation with ion-exchange and gel-filtration was unsuccessful. Affinity separation was possible using a Ni-Sepharose chromatography column containing recombinant IGFBP-3 attached by means of its histidine tag. It is still to be verified if binding of IGF-II by immobilized IGFBP-3 is specific enough to allow co-elution of the two proteins.

4.2 Recombinant human IGFBP-3

The expression of human IGFBP-3 was straightforward although it required a longer time since *P. pastoris* was used for recombinant protein expression. In addition, it was possible to obtain the stable protein only after the identification of the causes of degradation and after modifying the protocol in order to inhibit metalloprotease activity. Nevertheless, crystallization trials with the purified protein were unsuccessful. One possible explanation could be found in the glycosilation of the secreted protein: human IGFBP-3 in fact possesses three potential N-glycosilation sites and *P. pastoris* has the ability to post-translationally modify secreted proteins. Recombinant protein digestion with the enzyme Endoglycosidase F was carried out to determine the presence of

glycosilation and the digested protein showed no differences on a SDS-PAGE, compared to the untreated protein. This, however does not exclude the possibility that the protein could be glycosilated and that enzyme digestion was ineffective.

4.3 Recombinant human IGFBP-rP1

IGFBP-rP1 was purified in a reproducible manner and the only difficulty encountered was proteolytic cleavage during purification which was solved by adding chelating agents after immobilized metal affinity chromatography. However, the same problem affects the protein when digestion with thrombin is employed; SDS-PAGE analysis in fact showed that in one month old crystallization drops the protein is almost completely degraded. Low concentrations of EDTA (5 mM) were added but this has proven ineffective. It should be verified if higher concentrations of EDTA could prevent protein degradation and if the crystallization process is disturbed by the same chelating agents,

4.4 Recombinant LG3 domain of perlecan

While protein purification and crystallization are completely reproducible and present no difficulties, the problem of mobile loops in the tertiary structure still remains to be solved. One possible solution consists in the co-crystallization with a ligand, or in its binding after crystals have been formed, which could stabilize the mobile loops and then give an interpretable electron density map for those critical zones. Another possibility could be the crystallization in conditions different from the original ones, in order to obtain different intermolecular

interactions which again could stabilize the mobile loops. Unfortunately, up to now none of the two strategies has given results.

5. Appendix

5.1 α_1-Microglobulin: characteristics and physiological roles

α_1-Microglobulin is a lipocalin which was first purified from human urine [43] and was described as a plasma glycoprotein of 26 kDa, consisting of 183 amino acid residues. Several studies following its discovery uncovered its many interesting features. It carries a set of chromophores which confers to the protein a yellow-brown color and an extremely heterogeneous charge; one of these chromophores weights 282 Da and was defined as an unidentified lipocalin ligand. It forms different high molecular weight complexes by covalently associating with many plasma proteins. Furthermore, it is widely distributed in the animal kingdom and its homologues have been described in mammals, birds, amphibians, and fish.

Human α_1-microglobulin is glycosylated in three positions: two of these are of the N-linked type and correspond to the Asn^{17} and to Asn^{96} and the third one is Thr^{5} [44]. Even though the glycosylation state varies greatly among different species, glycosylation at the residue corresponding to the Asn^{96} in humans is conserved in mammals, suggesting that the carbohydrate substitution could be functionally important.

α_1-microglobulin is synthesized in the liver [45] and its cDNA sequence shows that it encodes for α_1-microglobulin and for bikunin, another plasma protein. The mRNA is produced from this precursor gene and it is translated into a 19 residues signal peptide and into the two proteins connected by a tripeptide. After a proteolytic cleavage the two proteins are divided and they leave the hepatocyte. The protein localizes in liver, blood plasma, and kidney: this distribution is correlated with its metabolism. α_1-Microglobulin is in fact produced in the liver, then is secreted into blood, where it can be found in its free form or complexed with various plasma proteins, and in the kidney where it is catabolized. The concentration of α_1-microglobulin in human plasma has been investigated in

order to determine if it can be used as a marker for pathological events. Since α_1-microglobulin often forms high molecular complexes with other plasma proteins its total amounts were evaluated. α_1-Microglobulin levels appear to be stable during various pathological conditions, such as neoplastic diseases, central nervous system disorders, infections, and rheumatoid arthritis. The only fluctuations in the levels of α_1-microglobulin concentration in plasma were always related to impaired liver or kidney functions.

One of the functions established for α_1-microglobulin is its ability to modulate the immune response in vitro. For example, antigen-induced interleukin-2 (IL-2) production by mouse T helper hybridomas is inhibited by human α_1-microglobulin. While α_1-microglobulin normal plasma concentrations partially inhibited IL-2 production, complete inhibition was obtained at 20-50 fold higher concentrations [46]. The role of α_1-microglobulin N-linked saccharide chains on its immunosoppressive effects were investigated by using a crude preparation of α_1-microglobulin saccharide chains that inhibited the antigen-induced lymphocyte proliferation in a manner similar to that of intact α_1-microglobulin. However, α_1-microglobulin expressed in baculovirus-infected insect cells, which produce different N-linked carbohydrates, showed the same immunosoppressive effects. It was then suggested that its immunosoppressive ability could be related to the 282 Da unidentified ligand. A possible biological role of α_1-microglobulin is to protect the tissues by exerting a constant immunological suppression in normal tissues. This effect could be overridden in case of a local immunological response.

5.2 Mouse α_1-microglobulin: cloning, expression, and purification

Cloning – The clone of mouse α_1-microglobulin was purchased from RZPD. The bacterial colony containing the plasmid vector was transferred into 3 ml of LB liquid medium and grown overnight at 37 °C. Cells were harvested and the plasmid vector was purified using the GenElute Plasmid Kit (Quiagen). The coding sequence of mouse α_1-microglobulin was amplified by PCR using the purified plasmid vector as a template and the following primers.

Primer forward: GGTGGTCTCGAGAAAAGAGACCCTGCGTCAACA

Primer reverse: GGTGGTTCTAGACTACCGCCGGGCTCTGGCAAT

Xho I restriction site: CTCGAG

Xba I restriction site: TCTAGA

PCR cycle

1 – 5 minutes at 95°C

2 – 1 minute at 95°C: denaturation

3 – 1 minute at 55°C: annealing

4 – 90 seconds at 72°: extension

5 – 15 minutes at 72°C: final extension

Steps from 2 to 4 were repeated 35 times.

The amplified DNA was quantified on a 0.8% agarose gel and ligated to a pPICZα A expression vector using a coding sequence to plasmid vector molar ratio of 5:1. The construct was used to transform *E. coli* XL1 Blue strain cells which were plated and grown overnight at 37°C. Bacterial colonies were checked for transformation by colony PCR and restriction enzyme digestion. Positive colonies

were inoculated into 100 ml of LB medium containing 100 µg/ml Zeocin™ and grown overnight at 37°C. Cells were harvested and the plasmid vector was purified and used to transform *Pichia pastoris* KM71H strain competent cells.

Expression – Four colonies of *Pichia pastoris* were tested for recombinant protein expression. Each colony was transferred into 10 ml of BMGY liquid medium containing 100 µg/ml Zeocin™ and incubated overnight into a shaking incubator at 28°C. Induction was initiated by centrifugating the four cultures and by resuspending them into 50 ml of BMMY. Induction was maintained by adding 0.5% methanol every 24 hours and expression was daily monitored by loading samples from culture supernatant on SDS-PAGE gels.

Purification – A single positive colony was transferred into 100 ml of Buffered Glycerol-complex Medium (BMGY) containing 100 µg/ml Zeocin™ and grown overnight at 28°C. Cells were collected and resuspended into 500 ml of Buffered Methanol-complex Medium BMMY and induction was carried out for five days. The culture was then centrifugated and the supernatant was dialyzed on an Amicon ultrafiltration cell against 50 mM Tris/HCl pH 7.0, and 0.02% NaN_3. The supernatant was loaded onto a DEAE-cellulose column equilibrated with the same buffer and elution was carried out by applying a linear gradient from 0 M to 0.5 M NaCl.

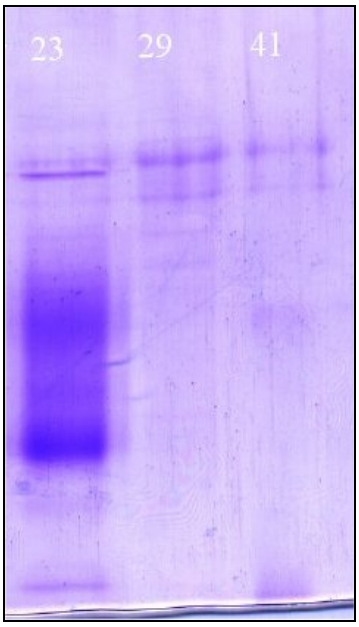


Fig. 33. SDS-PAGE showing fractions from DEAE-cellulose recombinant α_1-microglobulin elution. The recombinant protein is in the first lane.

Fractions containing the recombinant protein were identified by SDS-PAGE, collected, and concentrated up to 2 ml by Amicon ultrafiltration devices. The sample was then loaded onto a Superdex G 75 gel-filtration column for a final purification step.

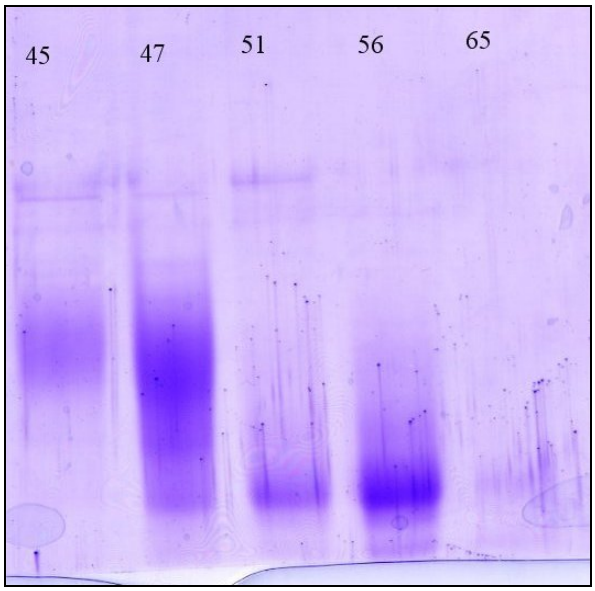


Fig. 34. SDS-PAGE showing fractions from Superdex G 75 elution of α_1-microglobulin. The recombinant protein can be seen in the first two lanes.

Fractions containing the recombinant α_1-microglobulin were collected and quantified by UV absorbance at 280 nm. The final yield was of 8 mg of purified protein per culture litre. The purified protein was concentrated up to 20 mg/ml and preliminary crystallization screens were prepared (Crystal Screen 1 and 2, Hampton Research). The crystal screens were unsuccessful since no crystal formed after 24 months. The protein was then deglycosylated by adding the enzyme PNGase F to a molar ratio of 1:25 (enzyme to α_1-microglobulin). The protein was dialyzed against 50 mM Tris/HCl pH 7.5 in the presence of the enzyme and incubated for 48 hours at 37°C. The protein was then concentrated and loaded onto a Superdex G75 gel filtration column. SDS-PAGE analysis showed a sharp band corresponding to a lower molecular mass as compared to untreated protein thus proving that deglycosylation had been successful. The deglycosylated protein was also concentrated to 20 mg/ml and used in the preliminary crystallization trials but again no crystals appeared.

References

[1] **Rosenfeld R, Wilson D, Lee P, Hintz R** (1986) Insulin-like growth factors I and II in the evaluation of growth retardation. J Pediatr 109:428

[2] **Jones JI, Clemmons DR** (1995) Insulin-like growth factors and their binding proteins: biological actions. Endocrine Reviews 16:3-34

[3] **Jones JI, Gockerman A, Busby WH, Wright G, Clemmons DR** (1993) Insulin-like growth factor binding protein 1 stimulates cell migration and binds to the _5_1 integrin by means of its Arg-Gly-Asp sequence. Proc Natl Acad Sci USA 90:10553-10557

[4] **Duan C, ClemmonsDR** (1998) Differential expression and biological effects of insulin-like growth factor-binding protein-4 and -5 in vascular smooth muscle cells. J Biol Chem 273:16836-16842

[5] **Parker A, Gockerman A, Busby WH, Clemmons DR** (1995) Properties of an insulin-like growth factor-binding protein-4 protease that is secreted by smooth muscle cells. Endocrinology 136:2470-2476

[6] **Zhang M, Smith EP, Kuroda H, Banach W, Chernausek SD, Fagin JA** (2002) Targeted expression of a protease-resistant IGFBP-4 mutant in smooth muscle of transgenic mice results in IGFBP-4 stabilization and smooth muscle hypotrophy. J Biol Chem 277:21285-21290

[7] **Bautista CM, Baylink DJ, Mohan S** (1991) Isolation of a novel insulin-like growth factor (IGF) binding protein from human bone: a potential candidate for fixing IGF-II in human bone. Biochem Biophys Res Commun 176:756-763

[8] **Jones JI, Gockerman A, Busby WH, Wright G, Clemmons DR** (1993) Insulin-like growth factor binding protein 1 stimulates cell migration and binds to the $\alpha_5\beta_1$ integrin by means of its Arg-Gly-Asp sequence. Proc Natl Acad Sci USA 90:10553-10557

[9] **Valentinis B, Bhala A, DeAngelis T, Baserga R, Cohen P** (1995) The human insulin-like growth factor (IGF) binding protein-3 inhibits the growth of fibroblasts with a targeted disruption of the IGF-I receptor gene. Mol Endocrinol 9:361-367

[10] **Lalou C, Lassarre C, Binoux M** (1996) A proteolytic fragment of insulin-like growth factor (IGF) binding protein-3 that fails to bind IGFs inhibits the mitogenic effects of IGF-I and insulin. Endocrinology 137:3206-3212

[11] **Baxter RC, Martin JL** (1989) Structure of the *Mr* 140,000 growth hormone-dependent insulin-like growth factor binding protein complex: determination by reconstitution and affinity-labeling. Proc Natl Acad Sci USA 86:6898-6902

[12] **Firth SM, Ganeshprasad U, Baxter RC** (1998) Structural determinants of ligand and cell surface binding of insulin-like growth factor binding protein-3. J Biol Chem 273:2631-2638

[13] **Andress DL, Birnbaum RS** (1992) Human osteoblast-derived insulin-like growth factor (IGF) binding protein-5 stimulates osteoblast mitogenesis and potentiates IGF action. J Biol Chem 267:22467-22472

[14] **Andress DL, Loop SM, Zapf J, KieferMC** (1993) Carboxy-truncated insulin-like growth factor binding protein-5 stimulates mitogenesis in osteoblast-like cells. Biochem Biophys Res Commun 195:25-30

[15] **Nadav L, Katz BZ** (2001) The molecular effects of oncogenesis on cell-extracellular matrix adhesion. Int J Oncol 19:237-246

[16] **Firth SM, Baxter RC** (1999) Characterisation of recombinant glycosylation variants of insulin-like growth factor binding protein-3. J Endocrinol 160:379-387

[17] **Sadowski T, Dietrich S, Koschinsky F, et al.** (2003) Matrix metalloproteinase 19 regulates insulin-like growth factor-mediated proliferation, migration, and adhesion in human keratinocytes through proteolysis of insulin-like growth factor binding protein-3. Mol Biol Cell 14, 4569-4580

[18] **Miyamoto S, Yano K, Sugimoto S, et al.** (2004) Matrix metalloproteinase-7 facilitates insulin-like growth factor bioavailability through its proteinase activity on insulin-like growth factor binding protein 3. Cancer Res 2004, 64, 665-671.

[19] **Renehan AG, Zwahlen M, Minder C, et al.** (2004) Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. Lancet 2004, 363, 1346-1353.

[20] **Volpert O, Jackson D, Bouck N, Linzer DI** (1996) The insulin-like growth factor II/mannose 6-phosphate receptor is required for proliferin-induced angiogenesis. Endocrinology 137:3871-3876

[21] **Kim KW, Bae SK, Lee OH, Bae MH, Lee MJ, Park BC** (1998) Insulin-like growth factor II induced by hypoxia may contribute to angiogenesis of human hepatocellular carcinoma. Cancer Research 58:348-351

[22] **Herr F, Liang OD, Herrero J, Lang U, Preissner KT, Han V, Zygmunt M** (2003) Possible angiogenic roles of insulin-like growth factor II and its receptors in uterine vascular adaptation to pregnancy. Journal of Clinical endocrinology & metabolism 88:4811-4817

[23] **Olson EN** (1990) MyoD family: a paradigm for development? Genes & Development 4, 1454-1461.

[24] **Florini JR, Magri KA, Ewton DZ, James PL, Grindstaff K, Rotwein PS** (1991) "Spontaneous" differentiation of skeletal myoblasts is dependent upon autocrine sectretion of insulin-like growth factor II. Journal of biological chemistry 286(24):15917-15923.

[25] **Zanckl H, Zang KD** (1980) Correlations between clinical and cytogenetical data in 180 meningiomas. Cancer Genet. Cytogenet. 1:351-356

[26] **Murphy M, Pykett MJ, Harnish P, Zang KD, George DL** (1993) Identification and characterization of genes differentially expressed in meningiomas. Cell Growth & Differentiation 4:715-722

[27] **Oh Y, Nagalla SR, Yamanaka Y, Kim HS, Wilson E, Rosenfeld RG** (1996) Synthesis and characterization of insulin-like growth factor-binding protein (IGFBP)-7. Journal of biological chemistry 271(48):30322-30325

[28] **Yamanaka Y, Wilson EM, Rosenfeld RG** (1997) Inhibition of insulin receptor activation by insulin-like growth factor binding proteins. Journal of biological chemistry 272:30729-30734

[29] **Kato MV** (2000) A secreted tumor-suppressor, mac25, with activin binding activity. Mol med 6:126-135

[30] **Jeruss JS, Sturgis CD, Rademaker AW** (2003) Down-regulation of activin, activin receptors, and Smads in high grade breast cancer. Cancer research 63:3783-3790

[31] **Dominguez F, Avila S, Cervero A** (2003) A combined approach for gene discovery identifies insulin-like growth factor binding protein-related protein I as a new gene implicated in human endometrial receptivity. J Clin Endocrinol Metab 88:1849-1857

[32] **Burger AM, Zhang X, Li H** (1998) Down-regulation of T1A12/mac25, a novel insulin-like growth factor binding protein related gene, is associated with disease progression in breast carcinomas. Oncogene 16:2459-2467

[33] **Timpl R.** (1996) Macromolecular organization of basement membranes, Current Opinion in Cell Biology 8:618-624

[34] **Hassell J, Yamada Y And Arikawa-Hirasawa E** (2003): Role of perlecan in skeletal development and diseases. Glycoconjugate Journal **19**: 263-267

[35] **Iozzo R. V.** (1984) Biosynthesis of heparan sulphate proteoglycan by human colon carcinoma cells and its localization at the cell surface. Journal of Cell Biology 99, 403-417

[36] **O'Toole EA** (2001) Extracellular matrix and keratinocyte migration. Clin Exp Dermatol 26:525-530

[37] **Watt FM** (2002) Role of integrins in regulating epidermal adhesion, growth and differentiation. EMBO J 21:3919-3926

[38] **Iozzo RV et al**. (2006) Targeting perlecan in human keratinocytes reveals novel roles for perlecan in epidermal formation. Journal of Biological Chemistry 281(8):5178-5187

[39] **Mongiat M, Sweeney SM, San Antonio JD, Fu J, and Iozzo RV** (2003) Endorepellin, a novel inhibitor of angiogenesis derived from the C terminus of perlecan. Journal of Biological Chemistry 278(6):4238-4249

[40] **Oda O, Shinzato T, Ohbayashi K, Takai I, Kunimatso M, Maeda K, and Yamanaka N** (1996) Purification and charachterization of perlecan in urine of end-stage renal failure patients. Clin Chim Acta 255:119-132

[41] **Gonzales EM, Reed CC, Bix G, Fu J, Zhang Y, Gopalakrishnan B, Greenspan DS, and Iozzo RV** (2005) BMP-1/Tolloid-like metalloproteases process endorepellin, the angiostatic C-terminal fragment of perlecan. The Journal of Biological Chemistry 280(8):7080-7087

[42] **O'Reilly M. S. et al.** (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth. Cell 88, 277-285

[43] **B. Ekström, P.A. Peterson, I. Berggård** (1975) A urinary and plasma alpha -1-glycoprotein of low molecular weight: isolation and some properties. Biochem. Biophys. Res. Commun. 65: 1427-1433.

[44] **Ekström B, A. Lundblad, S. Svensson** (1981) Structural studies on the carbohydrate portion of human α_1-microglobulin. Eur. J. Biochem. 114:663-666.

[45] **Akerstrom B.** (1983) Synthesis of α_1-microglobulin by Guinea-pig liver. European Journal of Biochemistry 133,235-239

[46] **Wester L, Michaelsson E, Holmdahl R, Olofsson T, Akerstrom B** (1998) Receptor for α_1-microglobulin on T limphocytes: Inhibition of antigen-induced interleukin 2 (IL-2) production. Scandinavian Journal of Immunology 48, 1-7.

[47] **Mc Pherson A** (1990) Current approaches to macromolecular crystallization. European Journal of Biochemistry, 189:1-23

Aknowledgments

I would like to thank:

Prof H.L. Monaco and Dr. Massimiliano Perduca for their helpful support during my PhD thesis.

Dr. Stefano Capaldi, Dr. Maria Elena Carrizo, and all the friends from the Biocrystallography laboratory for their assistance and for the nice time spent.

7. Papers

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 33, Issue of August 19, pp. 29812–29819, 2005
Printed in U.S.A.

Structure and Properties of the C-terminal Domain of Insulin-like Growth Factor-binding Protein-1 Isolated from Human Amniotic Fluid*

Received for publication, April 20, 2005, and in revised form, June 21, 2005
Published, JBC Papers in Press, June 22, 2005, DOI 10.1074/jbc.M504304200

Alberto Sala‡§, Stefano Capaldi§¶, Monica Campagnoli‡, Beniamino Faggion¶, Sara Labò‡, Massimiliano Perduca¶, Assunta Romano‡, Maria E. Carrizo¶, Maurizia Valli‡, Livia Visai‡, Lorenzo Minchiotti‡, Monica Galliano‡‖, and Hugo L. Monaco¶**

From the ‡Department of Biochemistry "A. Castellani," University of Pavia, via Taramelli 3b, 27100 Pavia, Italy and the ¶Biocrystallography Laboratory, Department of Science and Technology, University of Verona, Ca Vignal 1, strada Le Grazie 15, 37134 Verona, Italy

Insulin-like growth factor (IGF)-binding protein-1 (IGFBP-1) regulates the activity of the insulin-like growth factors in early pregnancy and is, thus, thought to play a key role at the fetal-maternal interface. The C-terminal domain of IGFBP-1 and three isoforms of the intact protein were isolated from human amniotic fluid, and sequencing of the four N-terminal polypeptide chains showed them to be highly pure. The addition of both intact IGFBP-1 and its C-terminal fragment to cultured fibroblasts has a similar stimulating effect on cell migration, and therefore, the domain has a biological activity on its own. The three-dimensional structure of the C-terminal domain was determined by x-ray crystallography to 1.8 Å resolution. The fragment folds as a thyroglobulin type I domain and was found to bind the Fe^{2+} ion in the crystals through the only histidine residue present in the polypeptide chain. Iron (II) decreases the binding of intact IGFBP-1 and the C-terminal domain to IGF-II, suggesting that the metal binding site is close to or part of the surface of interaction of the two molecules.

Insulin-like growth factor-binding protein-1 is a member of a family comprising six secreted proteins (designated IGFBP-1[1] to IGFBP-6) that can modulate upon binding the availability and thus the biological effects of insulin-like growth factors I and II (IGF-I and IGF-II). Both inhibition and enhancement of the hormone action have been described, and several mechanisms, including binding of IGFBPs to the extracellular matrix, phosphorylation, and proteolysis, have been shown to modulate their affinity for IGFs. Moreover, an increasing number of reports indicate that IGFBPs can regulate cellular functions independently of their ability to interact with IGFs (1, 2).

Each IGFBP polypeptide chain, ranging in length from 216 to 289 amino acids, may be divided into three distinct domains of approximately equal size. The N- and C-terminal portions exhibit a high primary sequence identity across the six IGFBPs and contain spatially conserved cysteine residues that form intra-domain disulfide bonds. The central domain is the least conserved region and in some of the proteins contains post-translational modifications and proteolytic cleavage sites. On the basis of their sequence homology it is assumed that IGFBPs share a common overall fold and have very similar IGF binding pockets. The N-and C-terminal domains are known to be involved in IGF binding (3), and recently the x-ray structure of the ternary complex of the two domains of IGFBP-4 and IGF-I has been reported (4). In this complex the C-terminal domain was partially disordered, and therefore, a detailed model of this part of the molecule could not be produced. The structure of the N-terminal domain of IGFBP-5, both isolated (NMR data (5)) and complexed to IGF-I (x-ray data (6)), is also known, and NMR spectroscopy has been used to produce a detailed model of the C-terminal domain of IGFBP-6 (7) and to study its interactions with IGF-II (8), but there is no x-ray model available of the C-terminal domain of any member of this protein family, although the C-terminal fragment of IGFBP-4 was crystallized recently (9). Proteolytic cleavage at the mid-region between the two domains of the protein is considered the predominant mechanism for IGF release from all IGFBPs (10), but several studies indicate that the resulting N- and C-terminal fragments still retain the ability to inhibit IGF activity (11). Additionally, recent reports suggest that proteolysis of IGFBPs results in fragments with potential functional properties that differ from those of the intact protein (9, 11).

IGFBP-1 is normally expressed in a tissue-specific manner in the liver, kidney, decidualized endometrium, and luteinizing granulosa cells (12). It is the predominant IGFBP in amniotic fluid (13), a major IGF-binding protein in fetal plasma, and its concentration is increased in maternal circulation during pregnancy. Several studies indicate that the protein plays a relevant role in embryonic growth as a local modulator of IGFs bioavailability at the maternal-fetal interface (14, 15). Moreover, the up-regulation of IGFBP-1 was demonstrated in breast human tumors and was associated with the malignant transformation of breast tissue (16). In addition, IGFBP-1 contains an Arg-Gly-Asp tripeptide that can bind to the recognition site of $\alpha_5\beta_1$ integrin, leading to the stimulation of migration in various cell species (17–20). The mature polypeptide chain of IGFBP-1 is composed of 234 amino acids (the unprocessed precursor is 259 amino acids long) and is phosphorylated at

* This work was supported by grants from the Italian Ministry of Education and Scientific Research (Fondo per gli Investiment della Ricerca di Base and Progetti di Ricerca di Interesse Nazionale).

The atomic coordinates and structure factors (code 1ZT3 and 1ZT5) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

§ Both authors contributed equally to this paper.
‖ To whom correspondence may be addressed. Tel.: 39-0382-987-724; Fax: 39-0382-423-108; E-mail: galliano@unipv.it.
** To whom correspondence may be addressed. Tel.: 39-045-8027-903; Fax: 39-045-8027-929; E-mail monaco@sci.univr.it.

[1] The abbreviations used are: IGFBP-1, insulin-like growth factor (IGF)-binding protein-1; ELISA, enzyme-linked immunosorbent assay; PBS, phosphate-buffered saline; SFM, serum-free Dulbecco's modified Eagle's medium; r.m.s.d., root mean square deviation.

three serine residues, 126, 144, and 194 (we will follow the sequence numbering of the unprocessed precursor) (21). Pregnancy-associated proteolysis of several other binding proteins has been described (22), and recent studies have focused on the cleavage of IGFBP-1 by matrix metalloproteases present in amniotic fluid and conditioned medium from decidualized endometrial cells (23, 24)

In this study we describe the isolation of the C-terminal domain of IGFBP-1 from human amniotic fluid and its structural characterization by x-ray crystallography to 1.8 Å resolution. We also show that the C-terminal fragment, which contains the Arg-Gly-Asp tripeptide, retains the capability of the intact protein to stimulate cell migration and that it binds the ferrous ion through the only histidine residue present in the chain.

MATERIALS AND METHODS

Protein Purification and Electrophoretic Analysis—Human amniotic fluid was obtained from discarded amniocentesis samples collected in the weeks 16–18. The pooled fluid (3 liters) was saturated to 90% with ammonium sulfate, and the precipitated proteins, after centrifugation at 10,000 × *g* for 1 h, were dissolved in 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, dialyzed against the same buffer, and fractioned by gel filtration on a 4 × 100-cm Sephadex G100 column. Proteins eluting after the albumin peak were pooled, equilibrated in 6.25 mM Bis-Tris-propane, pH 7.5, and separated on a 2.6 × 30-cm high performance Q-Sepharose column equilibrated with the same buffer (buffer A) and connected to an Åkta Purifier system (Amersham Biosciences). The elution was performed with a 375-min linear gradient from 0 to 100% of 6.25 mM Bis-Tris-propane, pH 9.5, 350 mM NaCl (buffer B) and monitored at 280 nm. The fractions containing the C-terminal fragment of IGFBP-1 were pooled, concentrated, and submitted to gel filtration on a Superdex G75 column (Amersham Biosciences) equilibrated and eluted with 20 mM Tris-HCl buffer, pH 8.0, 150 mM NaCl. Intact IGFBP-1 was isolated by gel filtration on a Superdex G75 column followed by ion exchange chromatography on a Mono Q HR 5/5 column (Amersham Biosciences) equilibrated against 20 mM Tris-HCl, pH 8.0, and resolved with a gradient from 0 to 0.5 M NaCl in the same buffer. The homogeneity of the proteins was checked by SDS-PAGE analysis under reducing and non-reducing conditions and by N-terminal sequence determination in a Hewlett-Packard model G 1000 A sequencer (Centro Grandi Strumenti, University of Pavia). Isoelectric focusing was performed using the immobilized pH gradient system on a laboratory-made polyacrylamide gel cast on GelBond with a non-linear 4–10 pH gradient (25) obtained with acrylamido buffer solutions (Fluka). The pI values were determined by comparison with marker proteins. Analytical polyacrylamide slab gel electrophoresis in the presence of SDS was performed as reported by Laemmli (26) using a Mini PROTEAN II Cell (Bio-Rad). For dephosphorylation 5-μg aliquots of each of the three isoforms of IGFBP-1 and of the C-terminal fragment in 0.05 M Tris HCl, pH 8.5, containing 0.5 mM $MgCl_2$ were exposed to 200 units of alkaline phosphatase (Sigma) at 37 °C for 1 h. The reaction mixtures and samples containing the untreated proteins were diluted 1:1 with the 8 M urea solution containing 0.5% carrier ampholites and submitted to analytical isoelectric focusing.

Antibody Production and Western Blot Analysis—For polyclonal antibody production against IGFBP-1, BALB/c mice were injected intraperitoneally five times at 1-week intervals with 100 μg of the purified IGFBP-1. The antigen was emulsified with an equal volume of complete Freund's adjuvant for the first immunization followed by four injections in incomplete adjuvant. The mice were bled, and the sera were tested for reactivity to the purified IGFBP-1 and its C-terminal fragment using ELISA and Western blot. The specific IgGs were purified by affinity chromatography on protein A-Sepharose columns according to the manufacturer's recommendations (Amersham Biosciences). Antibody titers were assayed either by ELISA or immunoblotting. For Western blot analyses electrophoreses were run on 17% SDS-PAGE, and the proteins were transferred by electroblotting to Immobilon-P polyvinylidene difluoride membranes (Millipore) using a Mini Protean II apparatus (Bio-Rad). The membrane was blocked with 5% w/v skim milk and probed with the mouse antiserum diluted 1:1000. Immunoreactive spots were detected with horseradish peroxidase-conjugated anti-mouse immunoglobulin and developed by the enhanced chemiluminescence method (ECL system, Amersham Biosciences).

Analysis of Cell Migration—Normal human skin fibroblasts (PromoCell Gmbh, Germany) were used for cell culture experiments. Cells were maintained in Dulbecco's modified Eagle's medium supplemented with 10% (v/v) fetal calf serum, 100 units/ml penicillin, 1 mg/ml streptomycin (complete medium) at 37 °C in a 5% CO_2 humidified atmosphere. Purified IGFBP-1 and its C-terminal fragment were dissolved in 0.05 M Tris-HCl, pH 7.4 containing 0.15 M NaCl and diluted with serum-free Dulbecco's modified Eagle's medium at the indicated concentrations. Cells were plated in 35-mm Petri dishes and allowed to grow to confluency. The confluent monolayers were "wounded" by scraping cells in the center of the culture dishes using a rubber policeman to generate an acellular area (27). After repeated rinses with PBS, the monolayers were examined and photographed under an inverted microscope and incubated for 48 h in serum-free Dulbecco's modified Eagle's medium (SFM) with and without 1 μg/ml IGFBP-1 or its C-terminal fragment. At the end of the incubation the cells were fixed in methanol for 15 min at −20 °C, examined, and photographed. The plates were previously marked to ensure that the same area was recorded.

Cell migration was also evaluated in Transwell inserts (8-μm pore size, Nunc, Denmark); 5×10^4 fibroblasts were seeded into the upper wells in 500 μl of SFM alone or with 1 μg/ml IGFBP-1 or its C-terminal fragment. The lower wells contained 500 μl of SFM. Cells were allowed to migrate through the filters for 48 h; non-migrating cells were removed from the upper surface using a cotton swab, and the cells that had migrated to the bottom surface of the inserts were fixed and stained with the Diff-Quick staining kit (Medion Diagnostic GmbH, Switzerland) according to the manufacturer's instructions. Migration was quantified by counting the stained cells in 10 non-overlapping fields using a light microscope fitted with a grid eyepiece. The data are the means ± S.D. of triplicate wells from at least two different protein preparations.

For the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay cells (20,000 cells/well) were plated on 96-well plates and grown for 48 h. 3-(4,5-Dimethylthyazol-2-yl)-2,5 diphenyltetrazolium bromide) (Sigma) was added to 1 mg/ml for 4 h followed by extraction as reported (20).

Production of Recombinant IGF-II—The cDNA coding for mature IGF-II (residues 25–91) was amplified by PCR using specific primers designed to introduce BamHI and HindIII restriction sites and a sequence coding for a thrombin cleavage site at the end of the protein chain. The fragment was digested and ligated in the pQE50 vector (Qiagen), which expresses a C-terminal His_6-tagged fusion protein. *Escherichia coli* BL21 cells, transformed with the IGF-II-pQE50 construct, were grown at 37 °C in LB medium, and protein expression was induced overnight by the addition of 0.5 mM isopropyl 1-thio-β-D-galactopyranoside. The cells were harvested and sonicated. Inclusion bodies containing the recombinant IGF-II were washed twice with 20 mM Tris, pH 7.5, 0.5 M NaCl containing 0.1% Triton X-100, and the protein was solubilized in the same buffer containing 6 M guanidine HCl and 20 mM reduced glutathione. The solution was clarified by centrifugation, and the supernatant was applied to a Ni^{2+}- Sepharose column equilibrated in the same buffer. On-column refolding was carried out with a linear 6–0 M guanidine HCl gradient in 20 mM Tris, pH 7.5, 0.5 M NaCl containing 10 mM imidazole. The recombinant His_6-tagged protein was eluted with a linear 10 mM–0.5 M imidazole gradient. The fractions containing the recombinant IGF-II were collected, concentrated, and further purified by gel filtration chromatography on a Superdex G75 column. The final purity was checked by SDS-PAGE and immunoblotting with an anti-His_6 horseradish peroxidase-conjugated monoclonal antibody (Sigma). The protein was concentrated to 1 mg/ml, dialyzed against saline phosphate buffer (PBS), and stored at - 80 °C.

ELISA Assays—To examine the binding of IGFBP-1 and its C-terminal domain to IGF-II by ELISA, microtiter wells were coated for 90 min at 37 °C with 100 μl of 1 μg/ml IGF-II in 50 mM sodium carbonate, pH 9.5. After washing in PBST (PBS containing 0.1% v/v Tween 20), the wells were blocked at 4 °C overnight with 200 μl of 2% (w/v) bovine serum albumin in PBS. For the binding assay, 1 μg of IGFBP-1 and 0.5 μg of its C-terminal fragment in 100 μl of 0.1 mM sodium acetate, pH 5.0, were added to the coated wells and incubated for 60 min at room temperature.

For the inhibition assay, IGFBP-1 and its C-terminal fragment were incubated for 1 h at room temperature with 100 μM $FeCl_2$ dissolved in 0.1 mM sodium acetate, pH 5.0, before addition to the coated wells. Dose dependence was tested by incubating the intact protein with different concentrations of $FeCl_2$. The plates were then extensively washed with PBST and incubated for 90 min with 1 μg of mouse anti-IGFBP-1 IgG dissolved in PBS with 2% bovine serum albumin. After washing 5 times

with PBST, the microtiter wells were incubated for 1 h with rabbit anti-mouse IgG conjugated to horseradish peroxidase (1:1000 dilution; Dako, Gostrup, Denmark). Finally, the conjugated enzyme was treated with *o*-phenylethylenediamine dihydrochloride (Sigma), and the absorbance at 492 nm was monitored with a micro-plate reader (Bio-Rad).

Crystallization and X-ray Data Collection—Screening for crystallization conditions was performed with the hanging drop method at 4 and 20 °C using Hampton Research Screens and mixing 1 μl of the protein solution and the same volume of precipitating solution and equilibrating *versus* 0.3 ml in the reservoir. Larger diffraction quality crystals could be obtained by using bigger volumes with the sitting drop method. The crystals grow by mixing equal volumes of 35% dioxane, and the protein dissolved in 0.02 M Tris-HCl, pH 7.5, at a concentration of 20 mg/ml. The crystals are orthorhombic, space group $P2_12_12$, with unit cell parameters a = 38.54 Å, b = 60.39 Å, and c = 31.24 Å and contain one C-terminal domain of IGFBP-1 in the asymmetric unit.

The diffraction data were collected from crystals frozen at 100 K after a brief immersion in a mixture of 70% of the mother liquor and 30% glycerol. The data for the native crystals and the two heavy atom derivatives were obtained using copper Kα radiation from a Rigaku RU-300-rotating anode x-ray generator with a Mar345 imaging plate area detector. The data for the co-crystal with $FeSO_4$ were collected at the XRD1 beamline of the Elettra synchrotron in Trieste (λ = 1.00 Å). The two heavy atom derivatives were prepared by overnight soaking of a crystal in mother liquor with the addition of the two compounds at a final concentration of about 1 mM. The data were indexed, integrated, and reduced using the programs MOSFLM (28) and AUTOMAR and Scala (29). The diffraction data statistics are summarized in Table I.

Structure Determination and Refinement—Initial phases to 2.3 Å resolution were determined by multiple isomorphous replacement with the two heavy atom derivatives. The two platinum sites were located in a difference Patterson map (30) and refined using the program MLPHARE (29). The single isomorphous replacement phases were used to locate the most significant mercury site in the difference Fourier map. These two major sites (one site from each of the two derivatives) were used as input for the program autoSHARP (31) that was used to locate the minor sites of the two derivatives and for density modification and final phasing. The electron density map thus produced was of excellent quality and could be readily interpreted. The initial model was built in the high quality map at 2.3 Å resolution using the program O (32) and was subsequently refined with the program REFMAC (33). During the process of refinement and model building, the quality of the model was controlled with the program PROCHECK (34). Solvent molecules were added to the model in the final stages of refinement according to hydrogen-bond criteria and only if their B factors refined to reasonable values and if they improved the *R*free. The final model contains 642 non-hydrogen protein atoms and has very reasonable geometry (see Table I), with 87.5% of the residues in the most favored regions of the Ramachandran plot and the remaining 12.5% in the additionally allowed region. The iron in the co-crystal was modeled into a difference Fourier map phased by the refined, unliganded structure, and the model was refined with REFMAC using the same criteria followed in the refinement of the apoprotein. The final statistics of the co-crystals are also given in Table I.

RESULTS

Isolation of IGFBP-1 C-terminal Fragment and Intact IGFBP-1 from Human Amniotic Fluid—The purification protocol used in the present work was developed with the aim of isolating a large number of components from human amniotic fluid. Proteins were thus precipitated with up to 90% ammonium sulfate concentration, and after a preliminary fractionation by gel filtration chromatography, the low molecular mass components were resolved by anion exchange chromatography and further purified by gel filtration (data not shown). The fractions showing the presence of a single band, when checked by SDS gel electrophoresis, were submitted to N-terminal sequence determination. The first 15 residues of a homogeneous 11-kDa polypeptide chain were VTNIKKWKEPCRIEL, and a search in the SWISSPROT data bank revealed that this sequence was identical to an internal amino acid sequence of IGFBP-1 starting from residue 166 (the numbering corresponds to the full-length precursor, SWISSPROT accession number P08833). The apparent molecular mass of the fragment accounts for the entire C-terminal domain (spanning residues 176–259) and the last portion of the central domain (residues 166–175). The recovery of the fragment was estimated in 1.5 mg/liter of amniotic fluid, using the extinction coefficient calculated from the amino acid composition. Three components showing an apparent molecular mass of 30 kDa were identified as isoforms of the intact IGFBP-1 on the basis of their unique N-terminal sequence, APWQCAPCSAEKLAL (residues 26–40 in the SWISSPROT numbering). It is known that during pregnancy decidual cells produce the non-phosphorylated as well as the differently phosphorylated IGFBP-1 forms characteristic of human pregnancy plasma (23, 35). The anion exchange chromatography retention times of the isoforms correlate with their more acidic pI values, and both acquire the same net charge of the non-phosphorylated protein after treatment with alkaline phosphatase. These results show that the IGFBP-1 isoforms present in human amniotic fluid are differently phosphorylated forms of the protein. The C-terminal fragment was submitted to analytic isoelectric focusing before and after the alkaline phosphatase treatment, and its mobility shift upon removal of the phosphate group revealed that it is phosphorylated as well. The pI values measured for the unmodified intact protein and its C-terminal fragment were 5.2 and 7.2, respectively, in agreement with the theoretical values calculated on the basis of their amino acid composition. The intact IGFBP-1 was used to produce polyclonal antibodies in mouse that were found to react with both the intact protein and the C-terminal region. A fresh sample of amniotic fluid was then resolved by SDS gel electrophoresis and, after electroblotting, probed with the anti-IGFBP-1 antibodies that recognized a single band corresponding to the molecular mass of the intact protein. This result shows that the cleavage producing the C-terminal fragment occurred in the amniotic fluid either during the long-term storage or during the purification procedure. Although the isolation of the C-terminal portion of IGFBP-1 from a natural source has not been reported before, amniotic fluid is known to contain several proteases belonging to different classes that can cleave IGFBPs.

The IGFBP-1 C-terminal Fragment Promotes Cell Migration—IGFBP-1 stimulates cell migration in an IGF-independent manner through the binding of its RGD motif to $\alpha_5\beta_1$ integrin (17, 18, 19). Because the C-terminal fragment contains the RGD sequence, its possible effect on cell migration was evaluated. In these experiments human fibroblasts were used because of their ability to migrate in response to different stimuli and because they do not express IGFBP-1 as verified by immunoblotting of concentrated conditioned medium (data not shown). Migration was assessed by the ability of cells to move into an empty area created by a rubber policeman. The cells, after wounds, were allowed to grow for 48 h in SFM containing 1 μg/ml IGFBP-1 or its C-terminal fragment, and images were taken at 0 and 48 h. Fig. 1*A* shows that in the control plate the cells grow to almost confluency slightly beyond the wounding line, whereas fibroblasts incubated with both the intact protein and the C-terminal fragment are still sparse but have migrated in the denuded area far beyond the wounding line.

Cell migration was confirmed and quantified in a Transwell system using the same protein concentrations and incubation time as in the wounding assay. IGFBP-1 and the C-terminal fragment showed, respectively, a 233 ± 26 and 186 ± 27% increase over the control of the cells migrated to the underside of the filters. The percentage of migrating cells as the average of triplicate wells from different protein preparations is reported in Fig. 1*B*.

IGFBP-1 or the C-terminal fragment does not significantly affect cell proliferation as evaluated using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide as a marker of cellu-

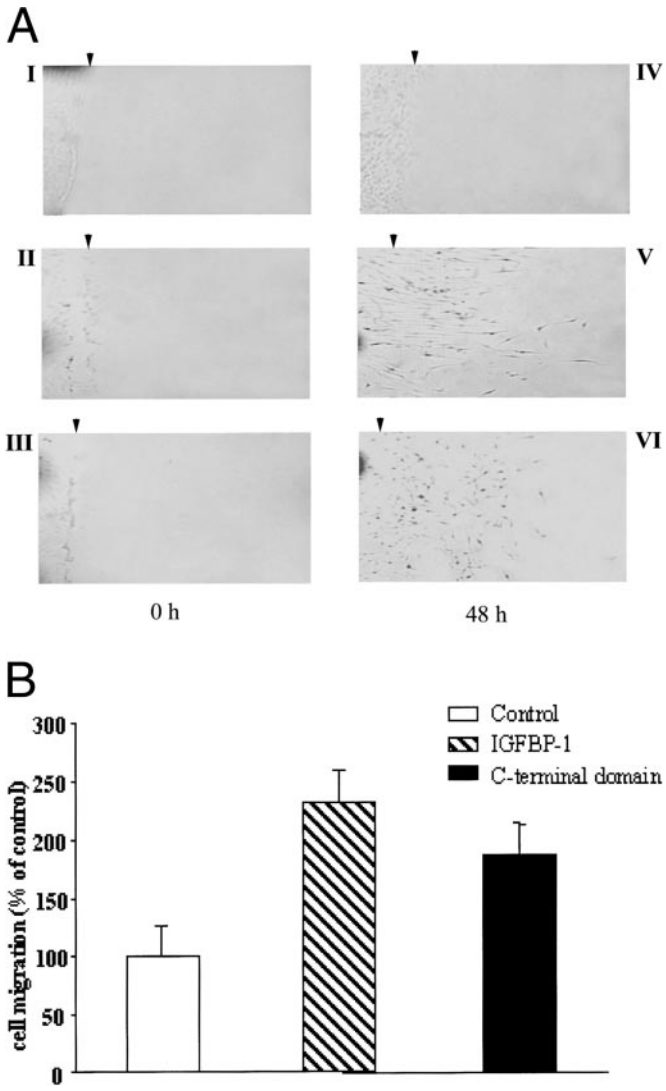


FIG. 1. **IGFBP-1 and its C-terminal fragment induce cell migration.** *A*, confluent fibroblast cells were wounded with a rubber policeman (time 0 h *panels I–III*) and allowed to migrate into the denuded area for 48 h (time 48 h *panels IV–VI*) in either SFM alone (*panel IV*) or containing 1 μg/ml IGFBP-1 (*panel V*) or 1 μg/ml C-terminal fragment (*panel VI*). The *dark spot* on the left side of each photograph marked the Petri dishes to ensure that the same area was recorded. The *arrowheads* indicate the wound edge. *B*, 5×10^4 fibroblasts were seeded into the upper wells of Transwell inserts (8-μm pore size membrane) and allowed to migrate in the presence of either SFM alone or with 1 μg/ml IGFBP-1 or its C-terminal fragment for 48 h. Migrated cells to the underside of the membranes were stained with Diff-Quick and visually counted. The *graph* represents the mean ± S.D. of triplicate wells from at least two different protein preparations

larity after a 48 h culture (data not shown), and therefore, the cell migration observed is not a secondary effect of cell proliferation. These results show that the C-terminal fragment of IGFBP-1 can stimulate cell migration and, therefore, that it retains the effect of the intact protein.

X-ray Structure of the C-terminal Fragment of IGFBP-1—The 2.3 Å experimentally phased map did not show electron density for the first 6 amino acids of the polypeptide chain and for the last 8 amino acids. The polypeptide chain was built starting with Trp-172 up to Cys-251. The side chains of the first two amino acids (Trp-172 and Lys-173) were not well defined in the maps. The final model, refined to 1.8 Å resolution, corresponds to 80 amino acids, 642 protein atoms, 1 dioxane molecule, and 42 water molecules. The conventional R factor is 22.4%, and the free R factor, calculated with 10% of the reflections, is 27.5% (Table I). The R factors and r.m.s.d. values of Table I were calculated with the program REFMAC (33). The stereochemical quality of the protein model was assessed with the program PROCHECK (34); 87.5% of the residues are in the most favorable region of the Ramachandran plot, and the remaining 12.5% are in the additionally allowed region.

The first element of secondary structure of the domain is a β-turn spanning residues 172–175, the second is an α-helix, which spans 19 residues, from Pro-175 to Thr-193. The next 4 residues in the sequence, Ser-194, Gly-195, Asp-196, and Asp-197, are disordered in the maps. Ser-194 was phosphorylated in the protein used to prepare the crystals that were used in all the x-ray diffraction experiments, but no clear electron density for the phosphate group was present in the maps. The rest of the domain contains four very short strands of anti-parallel β-sheet and 5 additional β turns. The 4 strands span the following residues: 200–201, 219–220, 215–217, and 228–230. A total of 10 residues are, thus, found in β strands, and 24 are found in the 6 β turns. A γ turn of the inverse type is formed by residues Ile-245–Arg-246–Gly-247, where 2 of the 3 residues are part of the RGD motif that binds to integrins. The three disulfide bridges are formed by Cys-176 and Cys-206, by Cys-217 and Cys-228, and by Cys-230 and Cys-251, the last residue that is clearly defined in the maps (36). The electron density of the three bridges is very clear in all the maps.

Fig. 2*A* is a schematic representation of the C-terminal domain of IGFBP-1. The α-helix is *light blue*, and the four short strands of β structure are represented as *arrows*. The three disulfide bridges and the phosphorylated serine (Ser-194) are also represented in the figure. Notice the position of the RGD motif (Arg-246–Gly-247–Asp-248), represented in the figure as *ball and stick models*, on the opposite side of the α-helix. The figure shows also the position of the Fe^{2+} binding site (see next paragraph). Fig. 2*B* is a stereoview of the Cα chain trace of the molecule in approximately the same orientation.

The IGFBP-1 C-terminal Fragment Binds the Ferrous Ion—Crystals of the C-terminal domain of IGFBP-1 were soaked in 10 mM solutions of the following metal compounds: $CaCl_2$, $FeSO_4$, $FeCl_3$, $CoCl_2$, $NiSO_4$, and $MnCl_2$. X-ray data were collected, and difference Fourier maps were calculated and examined. The crystals soaked in $FeCl_3$ were found not to diffract at all, whereas the crystals soaked in the other metals did not show any significant extra electron density with only one exception, $FeSO_4$, which presented a very high peak clearly visible with a cutoff of 7 σ in the $F_{obs} - F_c$ map and present in a position of the molecule where it could be easily interpreted in chemical terms. Fig. 3 shows the electron density of the Fe^{2+} ion in the co-crystals of the C-terminal domain of IGFBP-1 in both an $F_{obs} - F_c$ and a $2F_{obs} - F_c$ maps. The red density of the $F_{obs} - F_c$ map is contoured at a 7-σ level, and the blue electron density of the $2F_{obs} - F_c$ map is contoured at a 1.5-σ level. The *diagram* on the *right hand* of the figure represents the species coordinated to the bivalent ion. The Fe^{2+} ion is coordinated tetrahedral to a nitrogen of the ring of the only histidine present in the polypeptide chain, His-213, and to the oxygen of the side chain of Ser-214. The other two positions are occupied by 2 water molecules that are also bound to the polypeptide chain nitrogens of residues 207 and 215. Of the 4 coordination positions 2 involve side chains, those of His-213 and Ser-214, whereas the other two are dependent on two polypeptide chain nitrogens.

Iron (II) Reduces the Binding of IGFBP-1 and Its C-terminal Domain to IGF-II—The hypothesis that the binding of intact IGFBP-1 and its C-terminal domain to IGF-II, the major growth factor in amniotic fluid, might be affected by the presence of ferrous ion was tested by using an ELISA assay. The

TABLE I
Crystal structure determination statistics

The values in parentheses refer to the highest resolution shell, *i.e.* for the data collected 1.90–1.80 Å for the native data set, 2.42–2.30 Å for the heavy atom derivatives, and 1.88–1.82 for the co-crystal with Fe(II). The highest resolution shells used in the refinements are 1.90–1.80 Å for the apoprotein and 1.88–1.82 Å for the co-crystals with $FeSO_4$.

Data set	Native	K_2PtCl_4	$C_2H_5HgOHPO_2$	Native + $FeSO_4$
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P2_12_12$
a (Å)	38.54	38.92	38.63	38.21
b (Å)	60.39	60.40	60.71	59.60
c (Å)	31.24	31.42	31.40	31.06
Resolution range (Å)	20.1-1.8	32.7-2.3	32.6-2.3	32.1-1.8
Observed reflections	44,453	12,950	24,374	23,571
Independent reflections	7,168	3,557	3,580	6,773
*R*sym (%)	4.1 (27.4)	6.8 (29.4)	6.7 (17.3)	4.4 (8.5)
I/σ	14.2 (2.8)	9.0 (3.4)	10.7 (4.4)	8.9 (5.6)
Completeness (%)	99.5 (98.1)	99.5 (99.5)	99.9 (99.2)	97.0 (98.3)
Sites		2	2	
*R*cullis (acentric/centric)		0.718/0.696	0.682/0.622	
Phasing power (acentric/centric)		1.331/1.275	1.519/1.627	
Reflections in refinement	6,801			6,267
*R*cryst (%)	22.36 (32.1)			21.64 (22.8)
*R*free (%) (test set 10%)	27.50 (27.4)			27.18 (43.3)
Protein atoms	642			642
Water molecules	42			61
r.m.s.d. on bond lengths (Å)	0.008			0.010
r.m.s.d. on bond angles (°)	1.082			1.570
Planar groups (Å)	0.004			0.005
Chiral volume deviation ($Å^3$)	0.073			0.086
Average B factor ($Å^2$)	23.59			24.30
Protein atoms	23.19			23.42
Solvent atoms	29.74			33.06

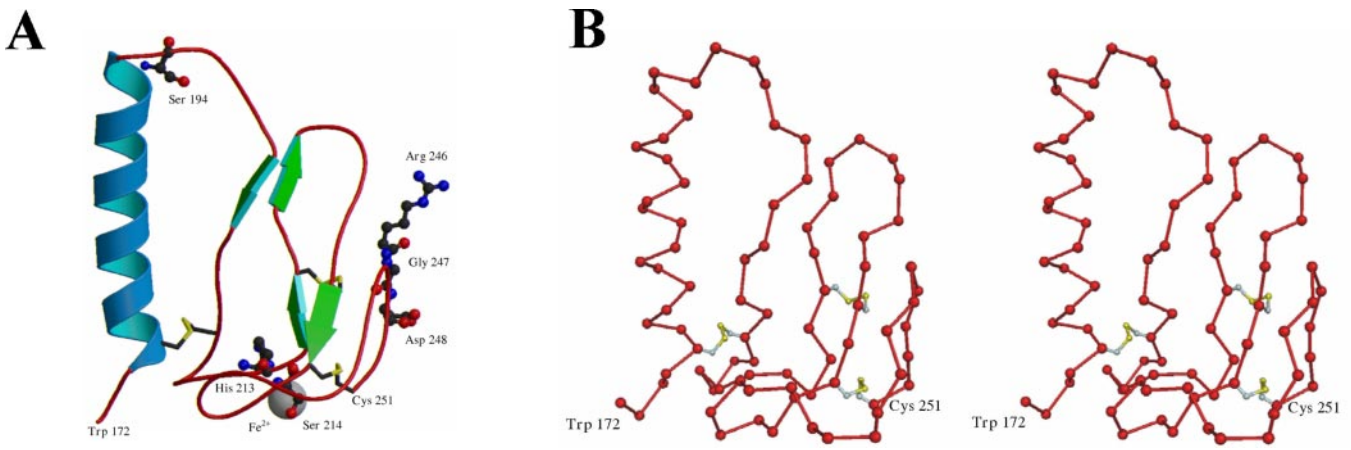


FIG. 2. **Overall structure and folding of the C-terminal domain of IGFBP-1.** *A*, *ribbon representation* of the C-terminal domain of IGFBP-1. The α- helix is *light blue*, the four short strands of β structure are *green*, and coil is *red*. The three disulfide bridges are shown in *yellow*. The diagram shows the phosphorylated serine (Ser-194, disordered in the electron density maps), and the RGD motif (Arg-246–Gly-247–Asp-248) as *ball and stick models*. The Fe^{2+} ion is represented as a *gray sphere*. The figure was prepared using the program MOLSCRIPT (49). *B*, stereoview of the Cα chain trace of the domain in approximately the same orientation. This figure was prepared using Dino (www.dino3d.org).

assays were performed in sodium acetate buffer instead of PBS to avoid the precipitation of ferrous phosphate and to maintain a slight acidic pH, which prevents oxidation of the metal ion. The binding experiments showed that the intact protein binds to IGF-II with higher affinity when compared with the C-terminal domain, a result in agreement with the data reported for other IGFBPs (4). In addition, incubation with ferrous ion reduced the binding of both intact IGFBP-1 and its C-terminal fragment with the growth factor, and this inhibition process was abolished by the addition of EDTA (Fig. 4*A*). The inhibition curve was determined by incubating the intact protein with increasing amounts of ferrous chloride, and the fitting shown in Fig. 4*B* was obtained assuming that the effect of the ferrous ion is competitive. A similar analysis was not performed with the C-terminal fragment because of its lower affinity for IGF-II. These results suggest that the residues involved in the binding to the ferrous ion are in a region of IGFBP-1 at or near the area of interaction with IGF-II.

DISCUSSION

This work describes the structural properties of the C-terminal domain of IGFBP-1 and shows that it maintains the effect of the intact protein on cell motility. The fragment was isolated from human amniotic fluid by conventional chromatographic procedures, and our results indicate that the degradation of IGFBP-1 occurred during the storage or the processing of the fluid that, when examined immediately after sample collection, contains only the intact protein. The polypeptide chain has a unique N-terminal sequence, and therefore, it originates from a specific proteolytic cleavage of the entire protein. IGFBP proteolysis is a well recognized mechanism that controls IGF bioavailability and, although little is known about the susceptibility of IGFBP-1 to enzymatic cleavage *in vivo*, the protein has been recognized as a potential physiological substrate for matrix metalloproteases (24) such as MMP-11 (stromelysin-3) (37) and MMP-26 (matrilysin 2) (38). Both enzymes cleave *in vitro*

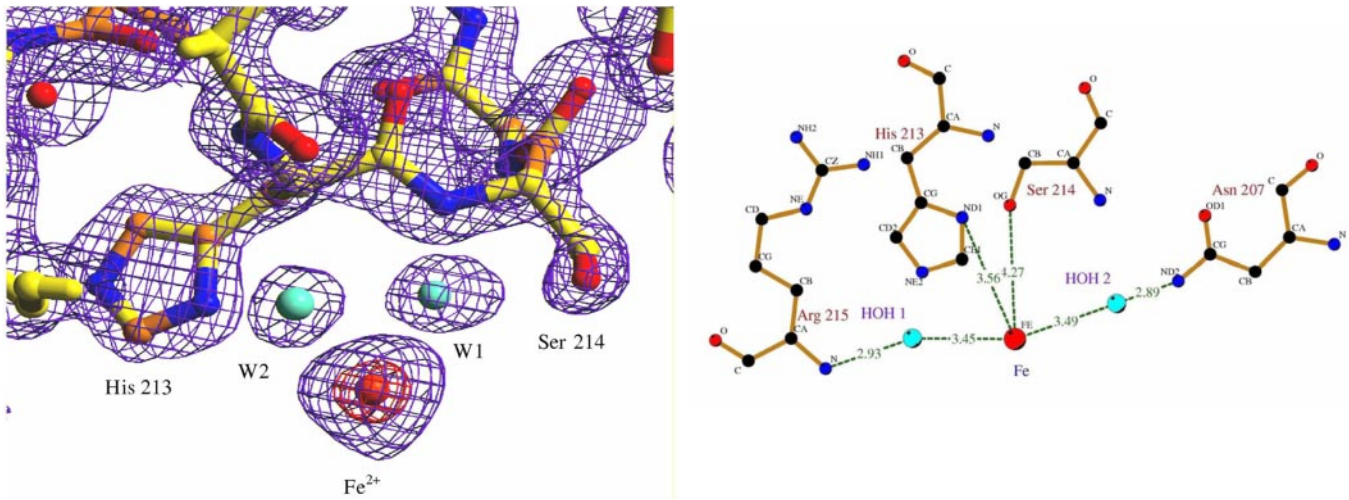


FIG. 3. **Iron ion binding to the C-terminal domain of IGFBP-1.** *Left*, electron density of the Fe^{2+} ion in the co-crystals of the C-terminal domain of IGFBP-1. The *red density* of the $F_{obs} - F_c$ map is contoured at a 7-σ level, the *blue electron density* of the $2F_{obs} - F_c$ map is contoured at a 1.5-σ level. The figure was prepared using the programs XtalView (50) and Raster3D (51). The *right diagram*, representing the groups coordinated to the Fe^{2+} ion, was prepared using the program LIGPLOT (52).

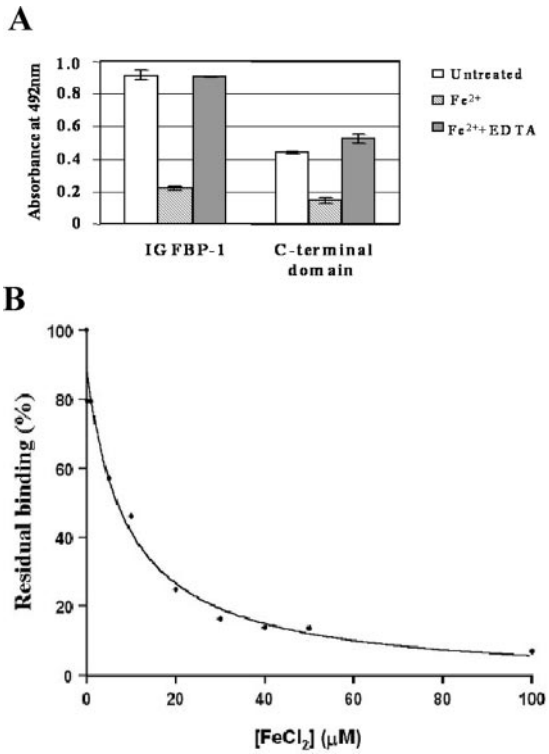


FIG. 4. **Inhibition of IGFBP-1 and of its C-terminal domain binding to IGF-II induced by ferrous ions.** *A*, binding of intact IGFBP-1 and its C-terminal fragment to IGF-II by ELISA. The inhibition was assessed after preincubating the protein with 100 μM $FeCl_2$; the addition of 100 μM EDTA restores the binding. The data shown represent the means and S.D. from duplicated ELISA assays and are representative of two separate experiments. The *y* axis shows the absorbance at 492 nm determined with a micro-plate reader. *B*, inhibition of IGFBP-1 binding to IGF-II determined by ELISA. The protein was preincubated with increasing amounts of $FeCl_2$, and the *y* axis shows the residual binding. The binding of IGFBP-1 to IGF-II in the absence of ferrous ion was set equal to 100%. The fit was obtained considering the process induced by the metal competitive and represents the duplicated mean and S.D. of two independent measurements.

the polypeptide chain of IGFBP-1 between His-165 and Val-166, thus producing a fragment with the same N-terminal sequence that we have found. The observation that the fragment represents a major component of the fluid whereas only small amounts of intact IGFBP-1 are obtained can be ascribed to long term storage of the amniotic fluid in the absence of protease inhibitors. The lack of comparable amounts of the 16-kDa N-terminal domain is likely due to the intrinsic instability of this portion of the molecule. On the contrary, the stability of the C-terminal region suggests that it is a completely very well structured domain of the whole protein, a property that has also been reported for other IGFBPs (11) and is confirmed by our x-ray diffraction studies.

IGFBP-1 stimulates cell migration, and several reports have established that this effect is caused by the binding of its RGD motif to the specific domain of $\alpha_5\beta_1$ integrin present on cell surfaces and have suggested that this interaction results in an intracellular signaling event (20). The requirement of this sequence for a migration promoting effect, initially shown with Chinese hamster ovary (18) and vascular smooth muscle cells (19), has also been demonstrated for placental trophoblast cells (20). In all these models, however, it is clear that not only an intact RGD motif but also some other regions of IGFBP-1 are necessary for binding. Our results indicate that the stable C-terminal domain has the same effect on cell motility as the intact molecule and suggest that proteolytic cleavage of the protein yielding a polypeptide chain with a biological activity on its own might be relevant for the regulatory role of IGFBP-1 on trophoblast migration and invasion.

The C-terminal domain of all the IGFBPs has been recognized on the basis of the disulfide bonding pattern (36) as a member of the thyroglobulin type-I domain (39), which is found in a number of proteins with very diverse function and in different organisms (40, 41). The thyroglobulin type I-fold was first described in the major histocompatibility complex class II-associated p41 Ii fragments bound to cathepsin L (42). More recently, the structure of the C-terminal domain of IGFBP-6 was determined by NMR spectroscopy (7).

The most important element of secondary structure of the C-terminal domain of IGFBP-1 is, as in the other thyroglobulin domains, an α-helix, which in this case spans residues 175–193, *i.e.* it is somewhat longer than in the other two proteins. The α-helix of the p41 Ii fragment is 9 amino acids long (Thr-195—His-203), and that of the C-terminal domain of IGFBP-6 is 15 amino acids long (Pro-162—Thr-176). Using the program LSQKAB (43), we have superimposed the two sets of coordinates of these thyroglobulin domains (PDB entry codes 1ICF and 1RMJ) to the coordinates of the C-terminal domain of IGFBP-1. Fig. 5*A* is a stereo diagram that shows the three protein

FIG. 5. **Comparison of the C-terminal domain of IGFBP-1 model and two related structures.** *A*, stereo diagram of the three models. The coordinates were superimposed by using the program LSQKAB (43). The model of C-IGFBP-1 is represented in *red* in the figure. The other two structures are (*a*) The NMR structure of C-IGFBP-6, represented in *blue* (Ref. 7, PDB code 1RMJ; the coordinates used for the NMR structure were the first set listed in the PDB file, and the first 27 amino acids were removed from the file) and (*b*) the x-ray structure of the human major histocompatibility complex class II-associated p41 Ii fragment represented in *green* (Ref. 42, PDB code 1ICF). *B*, r.m.s.d. between α-carbon atoms (in Å) of the NMR structure of the C-terminal domain of IGFBP-6 (*blue*), the x-ray structure of the human major histocompatibility complex class II-associated p41 Ii fragment (*green*), and the model of the C-terminal domain of IGFBP-1.

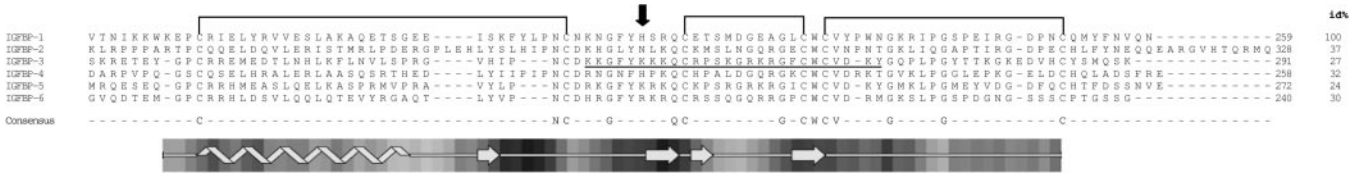


FIG. 6. **Sequence comparison of the C-terminal domain of IGFBP-1 and the other 5 members of the human IGFBP family.** The six sequences of the entire proteins were aligned using the program ClustalW (47), but only the C-terminal domains are represented in the figure. The *column* on the *right hand* gives the identity percentage of each sequence in the figure and that of the C-terminal domain of IGFBP-1; the disulfide bridges are indicated only for IGFBP-1. The *last line* has the amino acids identical in all the C-terminal domains of the group, whereas the *bottom strip* represents the elements of secondary structure of IGFBP-1. An *arrow* indicates the position of His-213, and the metal binding domain of IGFBP-3 is *underlined* and in *bold*.

structures superimposed. Notice that although the essential features of the thyroglobulin type-I domain are preserved in the three models, there is a significant variability in some details like for example the α-helix length and the exact position of the loops defined by the disulfide bonds. The r.m.s.d. in Å of the equivalent α carbons of these two structures and the C-terminal domain of IGFBP-1 are represented in Fig. 5*B* as a function of the amino acid position using the numbering of IGFBP-1.

Ferrous and other divalent ions, Ni^{2+}, Co^{2+}, Zn^{2+}, Mg^{2+}, and Mn^{2+} were found to bind to the C-terminal domain of IGFBP-3, and a 14-amino acid-long metal binding synthetic polypeptide was shown to trigger apoptotic effects as efficiently as intact IGFBP-3 (44). In addition, IGFBP-3 was found to interact with high affinity with the iron-binding proteins transferrin (45) and lactoferrin (46). The hypothesis that the C-terminal domain of IGFBP-1 might bind metal ions as well was tested by immersing crystals of the fragment in sufficiently concentrated solutions of the metals. A total of six ions were tested: Fe^{2+}, Fe^{3+}, Ni^{2+}, Co^{2+}, Ca^{2+}, and Mn^{2+}. Ferric ion was found to completely abolish the diffraction pattern of the crystals, whereas ferrous ion was found to bind in a single very well defined position. Fig. 6 shows the sequences of the C-terminal domains of the six IGFBPs aligned using the program ClustalW (47). The position of His-213 is indicated in the figure with an *arrow*, and the sequence of the metal binding domain of IGFBP-3 is in *bold* and *underlined*. Notice that the domain contains the amino acids aligned with those identified by us as relevant for the binding of Fe^{2+} to the C-terminal domain of IGFBP-1, and therefore, it may be suggested that metal binding in the two proteins could involve equivalent amino acids. The relevant histidine is present only in IGFBP-4, but there is no equivalent to Ser-214 in any of the other members of this protein family, and there is currently no evidence supporting the binding of IGFBP-4 to any metal ion.

On the basis of sequence homology with other IGFBPs, the residues of the IGFBP-1 C-terminal domain that bind the iron ion are located in the region of the protein involved in the binding to IGFs. Photoaffinity-labeling experiments with IGFBP-2 have identified as belonging to the IGF binding site amino acid residues 212–227 but also amino acid residues 266–287 (48). The latter contain Asn-271, which is in a position equivalent to that of His-213 in IGFBP-1 (see Fig. 6). Therefore, our observation that iron (II) reduces the binding of intact IGFBP-1 and of its C-terminal domain to IGF-II is in agreement with a model proposed for another member of this protein family. The fact that metal binding to full-length IGFBP-3 is also inhibited by the presence of IGF-I (44) gives further support to this proposal, but the exact physiological role of metal binding to these proteins will undoubtedly require further work. Taken together, our results indicate that the regulatory action of IGFBP-1 on IGFs and on their functions is more complex than perceived and may involve aspects that had not been identified so far.

Acknowledgments—We thank the staff of Sincrotrone Elettra for assistance during data collection, Rossella Greco, Department of General Biology and Medical Genetics, University of Pavia for supplying the amniotic fluid, Patrizia Arcidiaco, Centro Grandi Strumenti, University of Pavia for sequencing, and Angelo Gallanti, Department of Biochemistry, University of Pavia for cell harvesting.

REFERENCES

1. Baxter, R. C. (2000) *Am. J. Physiol. Endocrinol. Metab.* **278,** 967–976
2. Firth, S. M. and Baxter, R. C. (2002) *Endocr. Rev.* **23,** 824–854

3. Clemmons, D. R. (1998) *Mol. Cell. Endocrinol.* **140,** 19–24
4. Siwanowicz, I., Popowicz, G. M., Wisniewska, M., Huber, R., Kuenkele, K. P., Lang, K., Engh, R. A., and Holak, T. A. (2005) *Structure* **13,** 155–167
5. 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) *EMBO. J.* **17,** 6558–6572
6. Zeslawski, W., Beisel, H. G., Kamionka, M., Kalus, W., Engh, R. A., Huber, R., Lang, K., and Holak, T. A. (2001) *EMBO J.* **20,** 3638–3644
7. Headey, S. J., Keizer, D. W., Yao, S., Brasier, G., Kantharidis, P., Bach, L. A., and Norton, R. S. (2004) *Mol. Endocrinol.* **18,** 2740–2750
8. Yao, S., Headey, S. J., Keizer, D. W., Bach, L. A., and Norton, R. S. (2004) *Biochemistry* **43,** 11187–11195
9. Fernandez-Tornero, C., Lozano, R. M., Rivas, G., Jimenez, M. A., Standker, L., Diaz-Gonzalez, D., Forssmann, W. G., Cuevas, P., Romero, A., and Gimenez-Gallego, G. (2005) *J. Biol. Chem.* **280,** 18899–18907
10. Bunn, R. C. and Fowlkes, J. L. (2003) *Trends Endocrinol. Metab.* **14,** 176–181
11. Mark, S., Kubler, B., Honing, S., Oesterreicher, S., John, H., Braulke, T., Forssmann, W. G., and Standker, L. (2005) *Biochemistry* **44,** 3644–3652
12. Lee, P. D., Giudice, L. C., Conover, C. A., and Powell, D. R. (1997) *Proc. Soc. Exp. Biol. Med.* **216,** 319–357
13. Drop. S. L., Kortleve, D. J., and Guyda, H. J. (1984) *J. Clin. Endocrinol. Metab.* **59,** 899–907
14. Han, V. K., Bassett, N., Walton, J., and Challis, J. R. (1996) *J. Clin. Endocrinol. Metab.* **81,** 2680–2693
15. Kajimura, S., Aida, K., and Duan, C. (2005) *Proc. Natl. Acad. Sci. U. S. A.* **102,** 1240–1245
16. Pekonen, F., Nyman, T., Ilvesmaki, V., and Partanen, S. (1992) *Cancer Res.* **52,** 5204–5207
17. Irving, J. A. and Lala, P. K. (1995) *Exp. Cell Res.* **217,** 419–427
18. Jones, J. I., Gockerman, A., Busby, W. H., Jr., Wright, G., and Clemmons, D. R. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 10553–10557
19. Gockerman, A., Prevette, T., Jones, J. I., and Clemmons, D. R. (1995) *Endocrinology* **136,** 4168–4173
20. Gleeson, L. M., Chakraborty, C., McKinnon, T., and Lala, P. K. (2001) *J. Clin. Endocrinol. Metab.* **86,** 2484–2493
21. Jones, J. I., Busby, W. H., Jr., Wright, G., Smith, C. E., Kimack, N. M., and Clemmons, D.R. (1993) *J. Biol. Chem.* **268,** 1125–1131
22. Davies, S. C., Holly, J. M., Coulson, V. J., Cotterill, A. M., Abdulla, A. F., Whittaker, P. G., Chard, T., and Wass, J. A. (1991) *Clin. Endocrinol.* **34,** 501–506
23. Gibson, J. M., Aplin, J. D., White, A., and Westwood, M. (2001) *Mol. Hum. Reprod.* **7,** 79–87
24. Coppock, H. A., White, A., Aplin, J. D., and Westwood, M. (2004) *Biol. Reprod.* **71,** 438–443
25. Gianazza, E., Giacon, P., Sahlin, B., and Righetti, P. G. (1985) *Electrophoresis.* **6,** 53–56
26. Laemmli, U. K. (1970) *Nature* **227,** 680–685
27. Chen, P., Gupta, K., and Wells, A. (1994) *J. Cell Biol.* **124,** 547–555
28. Leslie, A. G. W. (1992) *Joint CCP4/ESF-EACMB Newslett. Protein Crystallogr.* **26,** pp. 27–33
29. Collaborative Computational Project Number 4 (1994) *Acta Crystallogr. Sect. D* **50,** 760–767
30. Sheldrick, G. M. (1991) in *Isomorphous Replacement and Anomalous Scattering: Proceedings of the CCP4 Study Weekend, January 25–26, 1991* (Wolf, W., Evans, P. R., and Leslie, A. G. W., eds) pp. 23–28, SERC Daresbury Laboratory, Warrington, UK
31. Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. and Paciorek, W. (2003) *Acta. Crystallogr. D. Biol. Crystallogr.* **59,** 2023–2030
32. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47,** 110–119
33. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. D. Biol. Crystallogr.* **53,** 240–255
34. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26,** 283–291
35. Westwood, M., Gibson, J. M., Davies, A. J., Young, R. J., and White, A. (1994) *J. Clin. Endocrinol. Metab.* **79,** 1735–1741
36. Forbes, B. E., Turner, D., Hodge, S. J., McNeil, K. A., Forsberg, G., and Wallace, J. C. (1998) *J. Biol. Chem.* **273,** 4647–4652
37. Manes, S., Mira, E., Barbacid, M. M., Cipres, A., Fernandez-Resa, P., Buesa, J. M., Merida, I., Aracil, M., Marquez, G., and Martinez-A, C. (1997) *J. Biol. Chem.* **272,** 25706–25712
38. Park, H. I., Turk, B. E., Gerkema, F. E., Cantley, L. C., and Sang, Q. X. (2002) *J. Biol. Chem.* **277,** 35168–35175
39. Malthiery, Y., and Lissitzky, S. (1987) *Eur. J. Biochem.* **165,** 491–498
40. Molina, F., Bouanani, M., Pau, B., and Granier, C. (1996) *Eur. J. Biochem.* **240,** 125–133
41. Galesa, K., Pain, R., Jongsma, M. A., Turk, V., and Lenarcic, B. (2003) *FEBS Lett.* **539,** 120–124
42. Guncar, G., Pungercic, G., Klemencic, I., Turk, V., and Turk, D. (1999) *EMBO. J.* **18,** 793–803
43. Kabsch, W. (1978) *Acta Crystallogr. A* **32,** 922–923
44. Singh, B., Charkowicz, D., and Mascarenhas, D. (2004) *J. Biol. Chem.* **279,** 477–487
45. Weinzimer, S. A., Gibson, T. B., Collett-Solberg, P. F., Khare, A., Liu, B., and Cohen, P. (2001) *J. Clin. Endocrinol. Metab.* **86,** 1806–1813
46. Baumrucker, C. R., Gibson, C. A., and Schanbacher, F. L. (2003) *Domest. Anim. Endocrinol.* **24,** 287–303
47. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680
48. Horney, M. J., Evangelista, C. A., and Rosenzweig, S. A. (2001) *J. Biol. Chem.* **276,** 2880–2889
49. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
50. McRee, D. E. (1999) *J. Struct. Biol.* **125,** 156–165
51. Merritt, E. A., and Bacon, D. J. (1997) *Methods Enzymol.* **277,** 505–524
52. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) *Protein Eng.* **8,** 127–134

Available online at www.sciencedirect.com
ScienceDirect
Journal of Structural Biology xxx (2006) xxx–xxx
Journal of Structural Biology
www.elsevier.com/locate/yjsbi

Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin ☆

Stefano Capaldi [a], Massimiliano Perduca [a], Beniamino Faggion [a], Maria E. Carrizo [a], Aldo Tava [b], Laura Ragona [c], Hugo L. Monaco [a,*]

[a] *Biocrystallography Laboratory, Department of Science and Technology, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy*
[b] *C.R.A.-Istituto Sperimentale Colture Foraggere, Viale Piacenza 29, 26900 Lodi, Italy*
[c] *Laboratorio NMR, Istituto per lo Studio delle Macromolecole, CNR, Via Bassini 15, 20133 Milano, Italy*

Received 10 July 2006; received in revised form 10 October 2006; accepted 10 October 2006

Abstract

The structure of the ternary complex of the anticarcinogenic Bowman–Birk protease inhibitor purified from snail medic (*Medicago scutellata*) seeds (MSTI) and two molecules of bovine trypsin has been solved by X-ray diffraction analysis of single crystals to a resolution of 2.0 Å. This is the highest resolution model of a ternary complex of this type currently available. The two binding loops of the MSTI differ in only one amino acid and have in both cases an arginine in position P1. The distances between the residues of the inhibitor at the binding interface and the trypsin side chains that recognize them are almost identical in the two sites. When compared to the NMR model of the uncomplexed MSTI, the inhibitor in the functional assembly with trypsin shows the largest differences in the two P2′ residues. Compared with the similar ternary complex of the soybean trypsin inhibitor, this model shows very small differences in the polypeptide chain of the trypsin binding sites and its largest difference in the area between Asp 26 and His 32 of the MSTI which in the soybean inhibitor has an extra Leu inserted in position 29.

Keywords: *Medicago scutellata*; Protease inhibitor; Bowman–Birk inhibitor; Trypsin; Ternary complex; X-ray structure

1. Introduction

The Bowman–Birk protease inhibitors (BBIs) are relatively small proteins of between 60 and 90 amino acids that are found in the seeds of legumes and other plants and that contain seven conserved disulfide bridges that stabilize their active conformations (Laskowski and Kato, 1980; Mello et al., 2003; Odani and Ikenaka, 1973). Characteristic of their sequence is the presence of two tandem homology regions that can bind independently, simultaneously and, in general, with different specificity to two different protease active sites. The amino acids present in the primary contact region of the inhibitor with the enzyme are usually labeled P1, P2, P3 on one side of the scissile bond of the inhibitor and P1′, P2′ and P3′ on the other. The P1 residue of the inhibitor and the corresponding S1 enzyme pocket together with those very few adjacent residues are primarily responsible for the recognition phenomenon. Thus the presence of an Arg or Lys at P1 recognizes the active site of trypsin, the presence of a hydrophobic residue, Leu, Phe, Trp or Tyr that of chymotrypsin while the presence of an Ala is selective for the active site of elastase. The interaction inhibitor-enzyme has been very extensively studied and reviewed (Bode and Huber, 2000; Qi et al., 2005) and several three-dimensional structures of the isolated inhibitors have been determined by X-ray diffraction and NMR spectroscopy. The list of the known three-dimen-

☆ The atomic coordinates and structure factors (code 2ILN) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/) .

* Corresponding author. Fax: +39 045 8027 929.
E-mail address: monaco@sci.univr.it (H.L. Monaco).

1047-8477/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2006.10.017

sional structures of isolated inhibitors includes the following: tracy soybean (Chen et al., 1992), peanut (Suzuki et al., 1993), pea (Li de la Sierra et al., 1999), soybean (Werner and Wemmer, 1992; Voss et al., 1996) and, more recently, sunflower (Luckett et al., 1999) and also snail medic (Catalano et al., 2003) seeds. Several X-ray structures of inhibitor–protease complexes are also known. The first reported complex was that of the azuki bean inhibitor–trypsin complex (Tsunogae et al., 1986), followed by the mung bean inhibitor–trypsin complex (Lin et al., 1993) and, more recently, the soybean inhibitor–trypsin complex (Koepke et al., 2000), barley inhibitor–trypsin complex (Park et al., 2004) and mung bean inhibitor–dengue virus NS3 protease complex (Krishna Murthy et al., 2000). The resolution achieved in some of these structures was moderate as well as the level of completeness of the data in the highest resolution shells. The discovery that Bowman–Birk inhibitors have anticarcinogenic effects on many different cell lines (Yavelow et al., 1985; Billings and Habres, 1992; Zhang et al., 1999; Miyata et al., 2000; Meyskens, 2001; Kennedy and Wan, 2002; Kennedy, 1998a,b) and radioprotective effects on normal tissue (Dittmann et al., 2003; Dittmann et al., 2000) arouse considerable interest as these molecules begun to be viewed as potential new useful agents in cancer therapy. Much effort is currently being devoted to the study of the possible mechanisms of these actions that are thought, at least in some cases, to be linked to the protease inhibition activity of these molecules (Billings et al., 1988; Ware et al., 1997, 1999).

Snail medic (*Medicago scutellata*) is an important pasture cultivar in Australia, taxonomically related to the more widely diffused alfalfa (*M. sativa*), and with a very high seed content of proteinase inhibitor. The inhibitor was isolated and sequenced (Ceciliani et al., 1997) and its three-dimensional structure in solution determined (Catalano et al., 2003). At the same time it was shown that it possesses anticarcinogenic activity against human breast and cervical cancer cells (Lanza et al., 2004) and the cDNA coding for it was isolated and its functional expression in response to biotic and abiotic stress studied (Balestrazzi et al., 2004).

In this paper, we describe the three-dimensional structure of the ternary complex of the *M. scutellata* seed protease inhibitor and two molecules of bovine trypsin determined by X-ray diffraction to 2.0 Å resolution. The structure of the inhibitor in the complex is compared with that in solution determined by NMR and with the similar soybean inhibitor in complex with trypsin.

2. Materials and methods

2.1. MSTI purification, complex formation and crystallization

The procedures followed to purify MSTI [1] from the seeds of the cultivar are described in the literature (Ceciliani et al., 1997). The protein sample used in the crystallization experiments showed one band in both SDS–PAGE and analytical isoelectric focusing (pI = 7.9, Molecular Mass = 6925, Ceciliani et al., 1997). The purified inhibitor was stored freeze dryed at −20 °C. A 1:2 stoichiometric complex with bovine trypsin (from Sigma–Aldrich, used without further purification) was prepared by mixing the two components in 20 mM Tris–HCl, pH 7.4. The properties of this complex have been extensively described by Ceciliani et al., 1997. A solution of the complex at an approximate concentration of 30 mg/ml was used for the initial screen of crystallization conditions. Hampton Research Screens were used at 20 °C with the hanging-drop method, mixing 1 µl of the protein solution with the same volume of the precipitating solution, and equilibrating versus a volume of 0.3 ml of the latter in the reservoir. The conditions yielding very small crystals were later refined and the sitting-drop method with larger volumes was also tested until crystals that were large enough for data collection were obtained. The best crystals of the MSTI–bovine trypsin complex were grown by mixing equal volumes (5 µl) of the protein solution and 1.5 M lithium sulfate, 0.1 M Hepes, pH 7.5 and equilibrating against 0.5 ml of the precipitant solution in the reservoir. They grow to a size of about 100 × 50 × 20 µm in approximately 20 days at 20 °C.

2.2. Data collection, structure solution and refinement

The crystals of the complex were indexed in the tetragonal space group $P4_1$, and contain one ternary complex (a molecule of MSTI and two molecules of bovine trypsin) in the asymmetric unit (see Table 1). Indexing the data in the space group P422 (one trypsin molecule and half of the inhibitor molecule in the asymmetric unit) significantly worsened all the statistics increasing the Rsym to 11.0%. Thus it was decided not to consider the pseudomolecular twofold axis in the complex as present in the crystals. The data for structure solution and refinement were collected at the XRD1 diffraction beamline of the Elettra Synchrotron in Trieste ($\lambda = 1.00$ Å), at 100 K after a brief soaking in a mixture of 80% of the mother liquor and 20% glycerol. The data were indexed, integrated and reduced using the programs MOSFLM (Leslie, 1992) and Scala (Collaborative Computational Project Number 4, 1994). The diffraction data statistics are summarized in Table 1.

The structure of the complex was solved using the CCP4 suite of programs for crystallographic computing. The initial phases were calculated by the molecular replacement method as implemented in the program MOLREP (Vagin and Teplyakov, 2000) with the coordinates of the highly similar (66% homology and 58% identity) soybean trypsin inhibitor complex (Protein data bank, accession code 1D6R) as the search probe. The 3.0 Å resolution rotation function presented two very high distinct peaks related by a twofold non-crystallographic axis with an $R_f = 142.1$ and 138.2, while the third peak had an $R_f = 56.8$. The solution of the translation function corresponding to the first peak, cal-

[1] *Abbreviations*: MSTI, *Medicago scutellata* trypsin inhibitor, PI, protease inhibitor; BBI, Bowman–Birk inhibitor; rms, root-mean-square.

Please cite this article in press as: Capaldi, S. et al., Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin, J. Struct. Biol. (2006), doi:10.1016/j.jsb.2006.10.017

Table 1
Data collection and refinement statistics

Data collection statistics	
Space group	$P4_1$
Unit-cell parameters (Å)	$a = 54.47$ $b = 54.47$ $c = 180.08$
Observed reflections	80,296
Independent reflections	33,730
Redundancy	2.4 (2.4)[a]
R_{sym} (%)[b]	9.3 (28.2)
Multiplicity weighted R_{merge}	12.1 (35.3)
$<I/\sigma(I)>$	4.0 (2.4)
Mosaicity	0.65°
Overall completeness (%)	95.8 (97.5)
Refinement statistics	
Resolution range (Å)	27.0–2.00 (2.05–2.00)
Reflections in working set	31,996 (2380)
Reflections in test set	1680 (115)
$R_{cryst.}$(%)[c]	23.2 (29.6)
R_{free} (%)[d]	28.7 (36.6)
Protein atoms	3670
Water molecules	376
r.m.s.d on bond lengths (Å)[e]	0.009
r.m.s.d on bond angles (°)	1.356
Planar groups (Å)	0.004
Chiral volume dev. (Å^3)	0.097
Average B factor (Å^2)	29.67
Trypsin molecule A	28.82 (rmsd = 0.41)
Trypsin molecule B	29.29 (rmsd = 0.43)
Inhibitor	38.68 (rmsd = 0.56)
Solvent atoms	33.86 (lowest value = 15.1, highest value = 55.5)

[a] The values in parentheses refer to the highest resolution shell, 2.05–2.00 Å.

[b] $R_{sym} = \Sigma h \Sigma i \mid Iih - <Ih> \mid \Sigma h \Sigma i <Ih>$ where $<Ih>$ is the mean intensity of the i observations of reflection h.

[c] R-factor $= \sum \| F_{obs} | - | F_{calc} \| / \sum | F_{obs} |$, where $|F_{obs}|$ and $|F_{calc}|$ are the observed and calculated structure factor amplitudes, respectively. Summation includes all reflections used in the refinement.

[d] $R_{free} = \sum \| F_{obs} | - | F_{calc} \| / \sum | F_{obs} |$, evaluated for a randomly chosen subset of 5% of the diffraction data not included in the refinement.

[e] Root mean square deviation from ideal values.

culated with data to 3.0 Å resolution was also very clear and had a correlation coefficient of 37.7 and an R factor of 54.1%. The second peak had a correlation coefficient of 30.4. By superimposing one of the trypsin molecules of the biological assembly of the search probe to the solution of the translation function, the ternary complex in the asymmetric unit was generated. The overall correlation value of this solution was 84.7 and its R factor 36.1% The three protein molecules were initially rigid body refined, the electron density examined and the proper side chains of the inhibitor introduced. The entire complex was then conventionally refined using the program REFMAC (Murshudov et al., 1997). In the subsequent cycles of manual model building and refinement, the model was modified using the program XtalView (McRee, 1999). Solvent molecules were added to the model in the final stages of refinement according to hydrogen-bond criteria and only if their B factors refined to reasonable values (lower than 60.0) and if they improved the R free.

3. Results and discussion

3.1. Overall structure of the complex and crystal packing

The structure of the complex was solved by molecular replacement and refined to a resolution of 2.0 Å. Statistics for the experimental data and final refined model are summarized in Table 1. The final model of the ternary complex contains the trypsin sequences from Ile 16 to Asn 245. The numbering that will be used for the trypsin residues in this and the following sections will be the chymotrypsin numbering scheme. The first seven as well as the last two amino acids of the inhibitor are also disordered, in other words the clearly defined polypeptide chain of the inhibitor runs from Cys 8 to Ser 60; Cys 61 is not well ordered in the maps. It is worthwhile mentioning that the N and C terminal regions are also disordered in the NMR structure of the uncomplexed inhibitor. The model contains 3670 protein atoms, and 376 water molecules. The conventional R factor is 23.2% and the free R factor, calculated with 5% of the reflections, 28.7% (Table 1). The R factors and r.m.s deviations of Table 1 were calculated with the program REFMAC (Murshudov et al., 1997). The stereochemical quality of the protein model was assessed with the program PROCHECK (Laskowski et al., 1993). 86.2% of the residues are in the most favorable region of the Ramachandran plot and the remaining 13.8% in the additionally allowed region. The average B factors of the inhibitor where somewhat higher than those of the trypsin molecules and increased towards both the N and C termini of the molecule.

Fig. 1(A) is a cartoon representation of the ternary complex with the approximate twofold axis of the molecular assembly vertical in the plane of the figure. Fig. 1(B) is a stereo diagram in which the view is looking down the approximate twofold axis of the biological assembly. All the intercomplex contacts in the crystal involve trypsin molecules which explains why this crystal form is so very closely related to that of the similar complex of the soybean inhibitor. The main contacts of the two trypsin molecules in the assembly are established with the inhibitor but there are other relevant contacts, the most important being the ND2 of Asn 92 of one trypsin molecule (A in our notation) with the peptide carbonyl of Gly 169 of the other molecule (B) and the ND2 of Asn 94 of molecule A and the carbonyl of Pro 168 of molecule B. The symmetric contact of the ND2 of Asn 92 of molecule B and the peptide carbonyl of Gly 169 of molecule A is also present but Asn 94 of molecule B is in contact with the OE1 of Gln 170 of molecule A and not with Pro 168.

The main packing contacts in the crystal, established between two trypsin molecules with the same label (A–A or B–B) are between Glu 80, Gln 81, Phe 82 and Ile 83 of one molecule and Ser 166 and Ser 170 of the other molecule, between Ser 110 of one molecule and Ser 164 of the other and between Ser 113 of one molecule and Ser 167 and Leu 185 of the other molecule. Interestingly, the contacts appear to be tighter between the two A molecules and somewhat looser

Please cite this article in press as: Capaldi, S. et al., Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin, J. Struct. Biol. (2006), doi:10.1016/j.jsb.2006.10.017

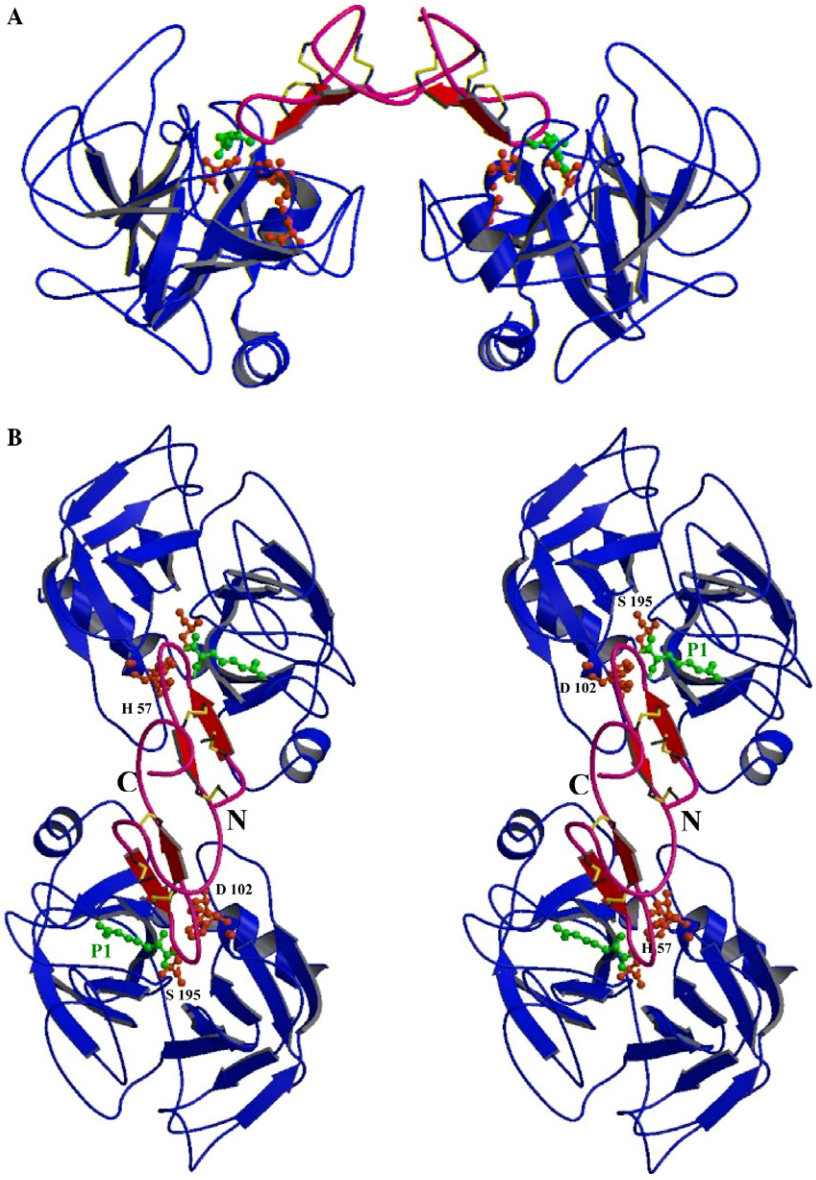


Fig. 1. Overall structure of the ternary complex of the *Medicago scutellata* trypsin inhibitor and the two trypsin molecules. (A) Ribbon representation of the complex (space group $P4_1$). The inhibitor molecule is represented red and the two trypsin molecules are blue. The approximate twofold symmetry axis of the molecular assembly is vertical and runs through the center in the plane of the figure. (B) Stereo diagram of the ternary complex viewed looking down the approximate twofold axis. The side chains of the P1 and P1′ residues of the inhibitor and the S1 residues of the trypsin molecules are represented as ball and stick models. The figure was prepared using the program MOLSCRIPT (Kraulis, 1991). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

between the two B molecules. The A–B contacts in the crystal are more extensive. They are established between Gly 38 (A) and Trp 237 (B) and Pro 92 (B), between Tyr 39 (A) and Ser 244 (B) and Asn 245 (B), between Tyr 59 (A) and Asn 72 (B), Asn 74 (B) and Asp 153 (B), between Arg 66 (A), and Pro 92 (B), between Ser 86 (A) and Lys 87 (A) and Tyr 151 (B) and between Ser 88 (A), Ile 89 (A) and Val 90 (A) and Asn 74 (B). Only slightly different distances are found exchanging in the description the roles of A and B although in some cases neighbouring amino acids are involved in the contacts.

The solvent accessible surface of trypsin molecule A, calculated with the program MSMS (Sanner et al., 1995), is 7609.7 Å^2, that of molecule B 7557.4 Å^2 and that of the inhibitor 3205.1 Å^2. The solvent accessible surface of the ternary complex is 16406.9 Å^2. Thus the area that becomes buried when the ternary complex is formed is 1965.3 Å^2. The same type of analysis for the interaction of the inhibitor with each trypsin molecule reveals that the surface buried upon complex formation is 929.9 Å^2 with trypsin molecule A and 913.4 Å^2 with trypsin molecule B. As expected, these two values are almost identical.

3.2. Structure of the MSTI

The covalent structure of the MSTI is shown in Fig. 2(A) in which the conserved residues before and after the P1–P1′

Please cite this article in press as: Capaldi, S. et al., Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin, J. Struct. Biol. (2006), doi:10.1016/j.jsb.2006.10.017

sites of the two proteinase binding regions are coloured in the same way. The presence of an Arg in both P1 sites explains the lack of antichymotryptic activity of this inhibitor (Ceciliani et al., 1997). The inhibitor presents two antiparallel beta sheets, the first with two and the second with three strands. The first strand of the first sheet is formed by Cys 22 and Gln 23, at the end of the first proteinase-binding region while the second strand of this sheet is formed by Ile 53 and Thr 54. The second sheet contains three strands, the second and the third are formed by Cys 38, Leu 39 and Cys 40 and by Cys 48, Arg 49 and Cys 50. These two strands define a beta hairpin (Sibanda et al., 1989) that contains the loop which is the second proteinase-binding region of the inhibitor. The first strand of this second sheet is formed by two amino acids, Val 27 and Arg 28. Five beta turns are formed by amino acids 9–12, 17–20, 32–35, 35–38 and 43–46. Of the six disulfide bridges present in the model of the molecule (Cys 61 was not included in the model and therefore the Cys 61–Cys 8 bridge is not present) four are right handed while those formed by Cys 12 and 57 and by Cys 31 and 38 are left handed (Hutchinson and Thornton, 1995). Fig. 2(B) shows the molecular surface of the inhibitor in which the polar residues are represented red while the hydrophobic residues are green. Fig. 2(C) shows the amino acid conservation in six inhibitors of this protein family.

3.3. Inhibitor–trypsin interactions

Figs. 3(A) and (B) are stereo representations of the electron density, contoured at a 1.5 σ level in the 2Fobs-Fc maps, and ball and stick models of the two areas of interaction of the inhibitor with the two trypsin molecules while Fig. 3(C) is a ribbon diagram and molecular surface representation of the entire inhibitor molecule and the areas of interaction of the trypsin molecules. Table 2 lists the shortest distances, i.e the polar interactions in the two contact areas of the inhibitor with the enzyme molecules (A and B, side by side). As evidenced from the table, the distances in the contacts of the two binding loops are very similar. In addition to these interactions, there are also several relevant hydrophobic contacts. The most important involve the inhibitor loops Cys 12–Ile 18 and Thr 35–Phe 44 and the trypsin chains Asp 189–Ser 195 and Ser 214–Ser 217, Phe 41, His 57, Asn 97–Leu 99, Tyr 151 and Gln 175. As shown in Fig. 2(A) and (B) the two P1 side chains of the arginines of the inhibitor (16 and 42) are totally embedded into the S1 pocket of the enzyme molecules. It is worth mentioning that water molecules were found coordinated to Arg 16 and Ser 17, the P1 and P1′ residues of the first inhibitory loop and Arg 42 and Ser 43, the P1 and P1′ residues of the other inhibitory loop. The mode of interaction of inhibitor and trypsin molecules is similar to that observed for other Bowman–Birk inhibitors.

3.4. Comparison of the inhibitor structure free in solution and in the crystal bound to bovine trypsin

The NMR structure of the uncomplexed *M. scutellata* inhibitor in solution has been determined at pH 5.6 and 27 °C (Catalano et al., 2003). In that study two triple stranded beta sheets were defined and the residues that form them were the following: Phe 11–Thr 15 (A), Gln 21–

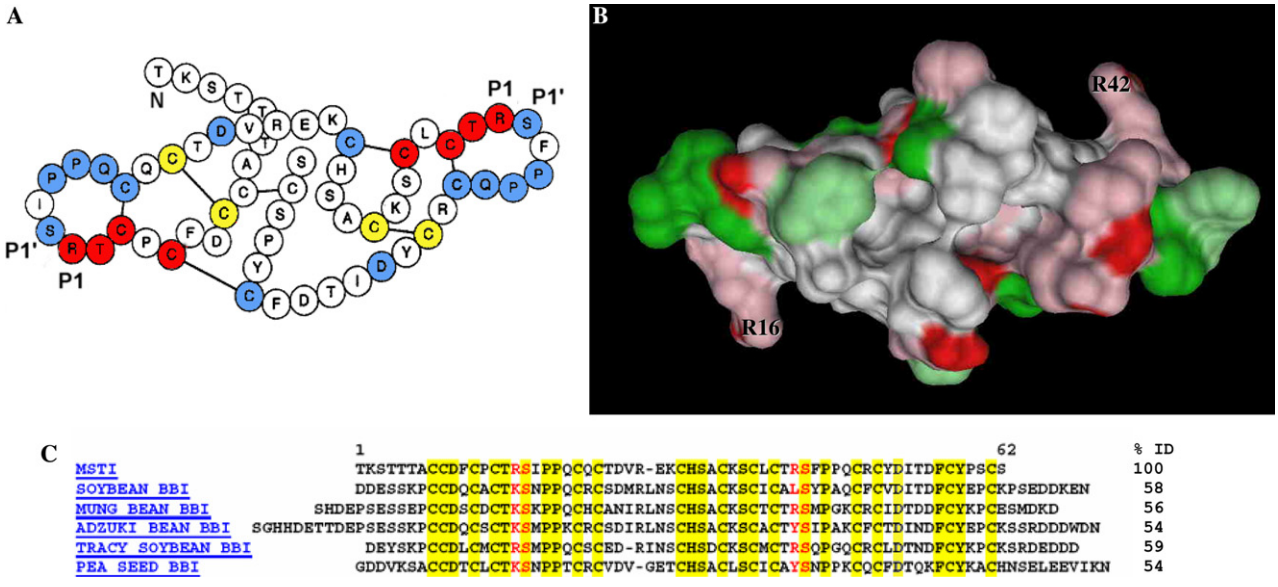


Fig. 2. The *Medicago scutellata* trypsin inhibitor (A) Covalent structure and disulfide bridges of the MSTI. The conserved residues at the two inhibitory sites are represented with the same colour. P1 and P1′ are indicated in the figure. (B) Buried surface area of the inhibitor viewed down the approximate twofold axis. R16 and R42 are labeled; red was used for the polar and green for the hydrophobic residues. (C) Sequence comparison of MSTI and other Bowman–Birk inhibitors. The sequences were aligned using the program CLUSTAL W (Thompson et al., 1994). The residues conserved in all the sequences are colored yellow and those equivalent to Arg16, Arg 42, Ser 17 and Ser 43 red. (b) Figure 2B was prepared with the program Dino (http://www.dino3d.org). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Please cite this article in press as: Capaldi, S. et al., Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin, J. Struct. Biol. (2006), doi:10.1016/j.jsb.2006.10.017

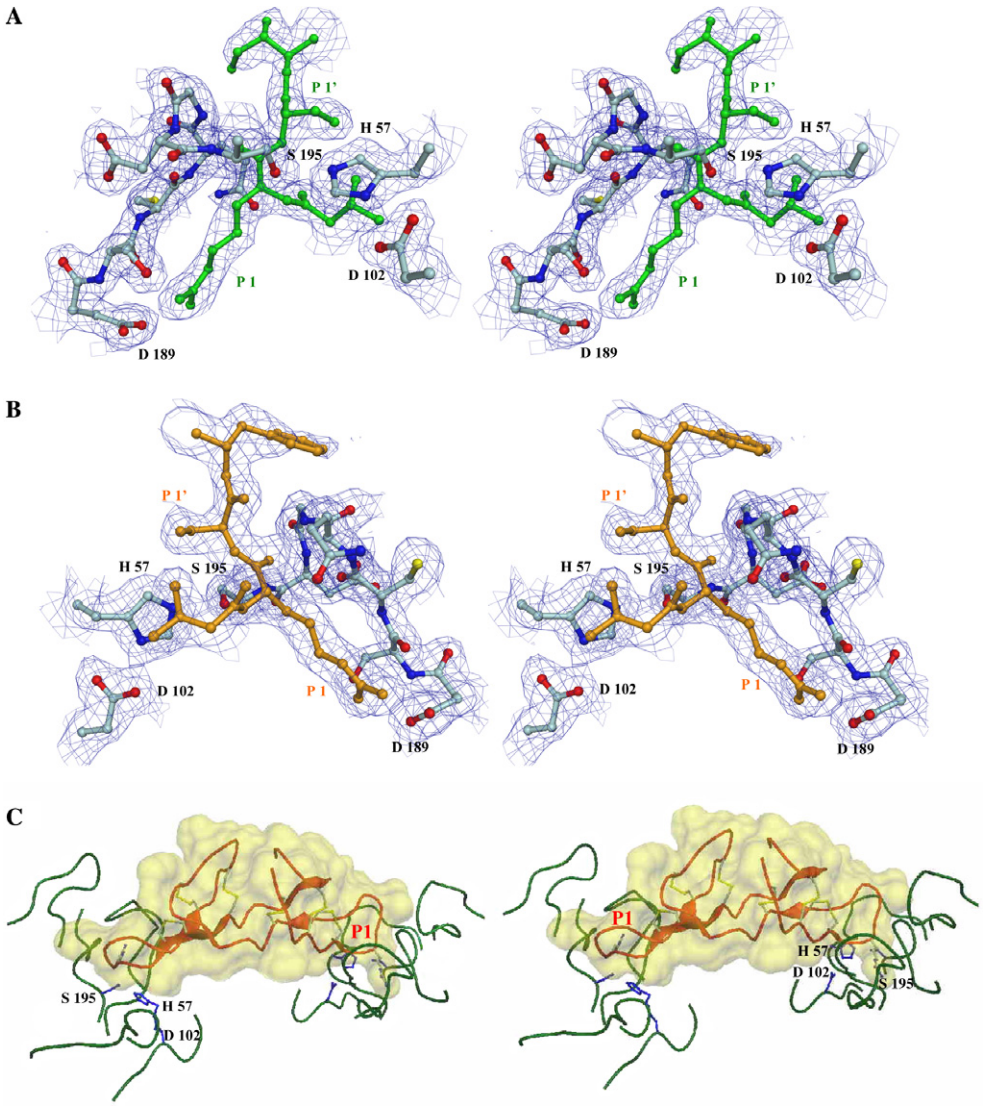


Fig. 3. Interaction of the *Medicago scutellata* trypsin inhibitor and the two trypsin molecules. (A) Electron density of the area of interaction and ball and stick model that fits the density at the interface between the inhibitor (green) and trypsin molecule A. (B) The interface between inhibitor (yellow) and molecule B. The blue electron density of the 2Fobs-Fc maps was contoured at a 1.5 σ level. (C) Stereoscopic representation of the inhibitor molecule and the areas of interaction with the trypsin molecules. A ribbon diagram is shown within the molecular surface of the inhibitor. The figure was prepared using the visualization program Dino (http://www.dino3d.org). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Thr 25 (B), Ile 53–Asp 55 (F), Val 27–Glu 29 (C), Ser 37–Thr41 (D), and Gln 47–Tyr 51 (E). These assignments were made using chemical shifts (Wishart and Sykes, 1994). We have used the program PROMOTIF (Hutchinson and Thornton, 1995) to define the structural elements of our model of the inhibitor bound to trypsin. In our model, there are only five strands of beta structure defined and they are all shorter than those identified in the NMR structure. Residues 9–12 are defined by the program as a type VIII beta turn and the other 5 strands are the following: Cys 22–Gln 23 (B′), Ile 53–Thr 54 (F′), Val 27–Arg 28 (C′), Cys 38–Cys 40 (D′) and Cys 48–Cys 50 (E′). These slight differences in the definitions are due to the different criteria used in secondary structure assignment. If the lowest energy NMR structure is analysed with PROMOTIF, the program recognizes as beta sheet three strands that overlap exactly with C′, D′ and E′ of the crystal structure of the inhibitor complexed with trypsin.

Using the program LSQKAB (Kabsch, 1978), the lowest energy NMR model and the X-ray coordinates of the inhibitor were superimposed and the distances between equivalent alpha carbons were calculated. Fig. 4 shows the two structures superimposed. The two largest differences found correspond to the P2′ residues of the two inhibitory loops. It is worthwhile mentioning that the electron density of the side chains of the two residues in the P2′ positions are the least clearly defined in the binding loops in the crystal (see Fig. 3A and B). Thus the most important differences between the inhibitor free in solution and complexed with the two enzyme molecules are found where expected, in the central residue of the loop that interacts with the enzyme active sites in the complex. The third and fourth peaks, sig-

Please cite this article in press as: Capaldi, S. et al., Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin, J. Struct. Biol. (2006), doi:10.1016/j.jsb.2006.10.017

Table 2
Distances between the inhibitor and the two trypsin molecule residues in contact at the enzyme active sites (labelled A and B)

Main contacts between the inhibitor and trypsin residues									
Inhibitor molecule	Atom	Trypsin residue	Atom	Distance (Å)	Inhibitor molecule	Atom	Trypsin residue	Atom	Distance (Å)
Cys 12	O	Ser 217 (A)	OG	3.16	Cys 38	O	Ser 217 (B)	OG	3.74
Cys 14	N	Gly 216 (A)	O	2.90	Cys 40	N	Gly 216 (B)	O	2.91
Cys 14	O	Gly 216 (A)	O	3.13	Cys 40	O	Gly 216 (B)	O	3.09
Cys 14	O	Gly 216 (A)	N	3.18	Cys 40	O	Gly 216 (B)	N	3.02
Thr 15	O	Gln 192 (A)	OE1	3.08	Thr 41	O	Gln 192 (B)	OE1	3.20
Arg 16	N	Ser 195 (A)	OG	2.97	Arg 42	N	Ser 195 (B)	OG	2.92
Arg 16	N	Ser 214 (A)	O	3.10	Arg 42	N	Ser 214 (B)	O	3.00
Arg 16	NH1	Asp 189 (A)	OD1	2.63	Arg 42	NH1	Asp 189 (B)	OD1	2.63
Arg 16	NH1	Gly 219 (A)	O	2.76	Arg 42	NH1	Gly 219 (B)	O	3.13
Arg 16	NH2	Asp 189 (A)	OD2	3.13	Arg 42	NH2	Asp 189 (B)	OD2	2.95
Arg 16	NH2	Asp 189 (A)	OD1	3.21	Arg 42	NH2	Asp 189 (B)	OD1	3.47
Arg 16	NH2	Ser 190 (A)	OG	2.96	Arg 42	NH2	Ser 190 (B)	OG	2.90
Arg 16	C	Ser 195 (A)	OG	2.61	Arg 42	C	Ser 195 (B)	OG	2.75
Arg 16	O	Gly 193 (A)	N	2.82	Arg 42	O	Gly 193 (B)	N	2.66
Arg 16	O	Ser 195 (A)	OG	2.76	Arg 42	O	Ser 195 (B)	OG	3.07
Arg 16	O	Asp 194 (A)	N	3.21	Arg 42	O	Asp 194 (B)	N	3.22
Arg 16	O	Ser 195 (A)	N	2.87	Arg 42	O	Ser 195 (B)	N	2.91
Ser 17	OG	His 57 (A)	NE2	3.77	Ser 43	OG	His 57 (B)	NE2	3.99
Ser 17	N	Ser 195 (A)	OG	3.01	Ser 43	N	Ser 195 (B)	OG	3.07
Ile 18	N	Phe 41 (A)	O	3.00	Phe 44	N	Phe 41 (B)	O	3.16
Gln 21	NE2	His 57 (A)	O	2.87	Gln 47	NE2	His 57 (B)	O	3.13
Gln 23	NE2	Asn 97 (A)	O	2.86	Tyr 51	OH	Ser 217 (B)	OG	3.01

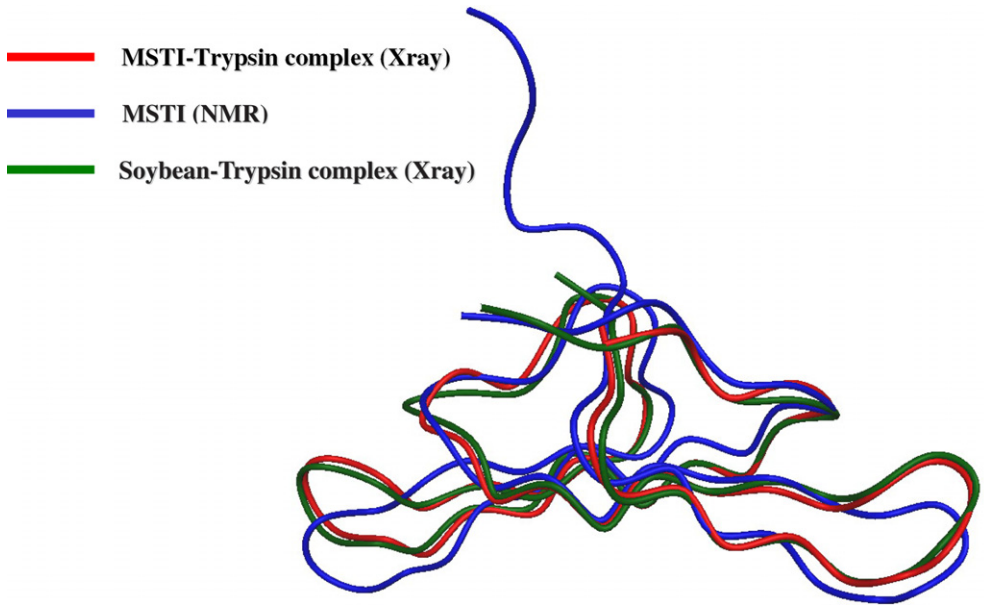


Fig. 4. Comparison of the model of the MSTI in the complex with trypsin (this study, red), the NMR structure of the uncomplexed inhibitor (blue, PDB code 1MVZ) and the X-ray structure of the soybean inhibitor bound to trypsin (green, PDB code 1D6R). The models were superimposed using the program LSQKAB (Kabsch, 1978). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

nificantly lower than the first two, correspond to His 32 and Asp 52, respectively. His 32 is in the first position of a beta turn while Asp 52 is one of the two amino acids found between the third strand of the second sheet and the second strand of the first sheet. It is worthwhile mentioning that the differences in the second binding loop are somewhat larger, in agreement with the proposal that this area has a larger conformational flexibility.

The NMR data showed that the two hydroxyl groups of the threonines in position P2, Thr 15 and Thr 41, form hydrogen bonds with the amide groups of the residues in position P5′, Gln 21 and Gln 47, and that the second bond appeared to be more mobile. In our model, the distance between the OG of both Thr 15 and Thr 41 is at about 4 Å from the NE2 of the two Glns and the OG of Thr 15 is at 2.98 Å from the N of Gln 21 while the distance between the OG of Thr 41 and the N of Gln 47 is 2.97 Å. Thus in the complexed structure of the inhibitor these two hydrogen bonds appear to be virtually identical. In the NMR structure of the MSTI, three aromatic residues: His 32, Phe 56

Please cite this article in press as: Capaldi, S. et al., Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin, J. Struct. Biol. (2006), doi:10.1016/j.jsb.2006.10.017

and Tyr 58 are found in a cluster with their ring planes orthogonal to each other, a conformation that is common to other protease inhibitors. Although in the X-ray model of the complex, the main polypeptide chain of the inhibitor in these areas is very well defined, the side chains are not and therefore a detailed comparison of the two models is not very informative. In particular, the flexibility of Phe 56 and Tyr 58 may be correlated to the observation that in solution the second reactive loop appears to be more flexible than the first.

3.5. Comparison of the MSTI complex with those of other Bowman–Birk inhibitors

The X–ray structure of several similar complexes of other Bowman–Birk inhibitors and trypsin has been determined. In particular the soybean (Koepke et al., 2000), azuki bean (Tsunogae et al., 1986) and mung bean (Lin et al., 1993) complexes are very closely related to the complex described here. We have not compared our structure with the complex of the barley seed inhibitor with porcine trypsin, solved at 2.2 Å resolution, because the inhibitor is 125 amino acid long an folds into two different domains (Park et al., 2004). In general our X-ray data compares rather well with those published before; the resolution is slightly better and the completeness is good. Of the three complexes mentioned above, the one that resembles most closely to the MSTI is the soybean trypsin inhibitor complex. The inhibitors have 58% sequence identity (only the tracy soybean inhibitor has a slightly higher homology but the structure of its complex is not available) and all the functionally important residues are conserved with the exception of the P1 residue in the second loop. In this position a Leu is present in the soybean inhibitor and an Arg in MSTI, which confers to the first the antichymotryptic activity, absent in MSTI. If the program PROMOTIF is used to identify the secondary structure elements of the soybean inhibitor in the complex, it recognizes five strands of beta sheet, Cys 12–Cys 14 (A′′), Cys 22–Cys 24 (B′′), Met 27–Leu 29 (C′′), Cys 39–Cys 41 (D′′) and Gln 48–Cys 51 (E′′). Equivalent strands are found for the last four elements of secondary structure in the MSTI, the first strand (A′′) is missing in MSTI and the last strand in our model, labeled F′ is missing from the soybean inhibitor structure.

Using the program LSQKAB (Kabsch, 1978), the model of the soybean inhibitor complexed with two trypsin molecules (PDB code 1D6R, green in Fig. 4) was superimposed to the model of the MSTI complex. The largest differences in the two models are present in the region between Val 26 and His 32 of the MSTI polypeptide chain i.e. in the region were Leu is inserted in the soybean inhibitor chain.

The polypeptide chain of both active site loops of the soybean inhibitor show small differences with that of the MSTI although the second loop, that has a Leu in position P1, is specific for chymotrypsin. Therefore, the different specificity is found to be due, as expected, to differences in the side chains of the two loops. It has been proposed that the specificity of the trypsin loop (the first in the sequence of the soybean inhibitor) is very strict while the antichymotryptic loop has a broader specificity. Indeed it has been shown that the soybean inhibitor can form a ternary complex with two trypsin molecules both in solution and in the crystals. The contact area of the first inhibitory loop in the soybean inhibitor is predominantly hydrophilic while that of the second loop is hydrophobic. In MSTI both types of contacts are present in both sites as shown in Fig. 2(B) that has the hydrophobic side chains coloured green and those of the polar residues red while intermediate levels of polarity are coded with white and grey. The cancer chemopreventive properties of the Bowman–Birk inhibitors have been associated to their antichymotryptic activity and it has been suggested that it is due to their inhibition of chymotrypsin-like proteases (Kennedy, 1998a). Since MSTI has been shown to possess anticarcinogenic effects (Lanza et al., 2004) in the absence of antichymotryptic activity it would appear that this notion will have to be revised.

Acknowledgments

We are grateful to the staff of the Elettra synchrotron in Trieste for assistance during data collection. We thank Henriette Molinari and Lucia Zetta for critical reviewing of the manuscript. This work was supported by a CNR (Consiglio Nazionale delle Ricerche, National Research Council) grant “Progetto Agenzia 2000” and a FIRB grant (Fondo per gli Investimenti della Ricerca di Base, Italian Ministry of the Universities and Scientific Research).

References

Balestrazzi, A., Confalonieri, M., Odoardi, M., Ressegotti, V., Allegro, G., Tava, A., Carbonera, D., 2004. A trypsin inhibitor cDNA from a novel source, snail medic (*Medicago scutellata* L.): cloning and functional expression in response to wounding, herbivore, jasmonic and salicylic acid. Plant Sci. 167, 337–346.

Billings, P.C., Habres, J.M., 1992. A growth-regulated protease activity that is inhibited by the anticarcinogenic Bowman–Birk protease inhibitor. Proc. Natl. Acad. Sci. USA 89, 3120–3124.

Billings, P.C., St Clair, W., Owen, A.J., Kennedy, A.R., 1988. Potential intracellular target proteins of the anticarcinogenic Bowman Birk protease inhibitor identified by affinity chromatography. Cancer Res. 48, 1798–1802.

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

Catalano, M., Ragona, L., Molinari, H., Tava, A., Zetta, L., 2003. Anticarcinogenic Bowman Birk inhibitor isolated from snail medic seeds (*Medicago scutellata*): solution structure and analysis of self-association behavior. Biochemistry 42, 2836–2846.

Ceciliani, F., Tava, A., Iori, R., Mortarino, M., Odoardi, M., Ronchi, S., 1997. A trypsin inhibitor from snail medic seeds active against pest proteases. Phytochemistry 44, 393–398.

Chen, P., Rose, J., Love, R., Wei, C.H., an Wang, B.C., 1992. Reactive sites of an anticarcinogenic Bowman–Birk proteinase inhibitor are similar to other trypsin inhibitors. J. Biol. Chem. 267, 1990–1994.

Collaborative Computational Project Number 4 (1994). Acta Cryst. D50, 760–767.

Dittmann, K.H., Dikomey, E., Mayer, C., Rodemann, H.P., 2000. The Bowman–Birk protease inhibitor enhances clonogenic cell survival of

ionizing radiation-treated nucleotide excision repair-competent cells but not of xeroderma pigmentosum cells. Int. J. Radiat. Biol. 76, 223–229.

Dittmann, K.H., Mayer, C., Rodemann, H.P., 2003. Radioprotection of normal tissue to improve radiotherapy: the effect of the Bowman Birk protease inhibitor. Curr. Med. Chem. Anticancer Agents 3, 360–363.

Hutchinson, E.G., Thornton, J.M., 1995. PROMOTIF—a program to identify and analyze structural motifs in proteins. Protein Sci. 5, 212–220.

Kabsch, W., 1978. A solution for the best rotation to relate two sets of vectors. Acta Cryst. A 32, 922–923.

Kennedy, A.R., 1998a. Chemopreventive agents: protease inhibitors. Pharmacol. Ther. 78, 167–209.

Kennedy, A.R., 1998b. The Bowman–Birk inhibitor from soybeans as an anticarcinogenic agent. Am. J. Clin. Nutr. 68, 1406S–1412S.

Kennedy, A.R., Wan, X.S., 2002. Effects of the Bowman–Birk inhibitor on growth, invasion, and clonogenic survival of human prostate epithelial cells and prostate cancer cells. Prostate 50, 125–133.

Koepke, J., Ermler, U., Warkentin, E., Wenzl, G., Flecker, P., 2000. Crystal structure of cancer chemopreventive Bowman.Birk inhibitor in ternary complex with bovine trypsin. J. Mol. Biol. 298, 477–491.

Kraulis, P.J., 1991. Molscript: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24, 946–950.

Krishna Murthy, H.M., Judge, K., DeLucas, L., Padmanabhan, R., 2000. Crystal structure of Dengue virus NS3 protease in complex with a Bowman–Birk inhibitor: implications for flaviviral polyprotein processing and drug design. J. Mol. Biol. 301, 759–767.

Lanza, A., Tava, A., Catalano, M., Ragona, L., Singuaroli, I., Robustelli della Cuna, F.S., Robustelli della Cuna, G., 2004. Effects of the *Medicago scutellata* trypsin inhibitor (MsTI) on cisplatin-induced cytotoxicity in human breast and cervical cancer cells. Anticancer Res. 24, 227–233.

Laskowski Jr., M., Kato, I., 1980. Protein inhibitors of proteinases. Annu. Rev. Biochem. 49, 593–626.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. Procheck: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.

Leslie, A.G.W., 1992. Recent changes to the MOSFLM package for processing film and image plate data Jnt. CCP4/ESF-EACMB Newslett. Protein Crystallogr., 26.

Li de la Sierra, I., Quillien, L., Flecker, P., Gueguen, J., Brunie, S., 1999. Dimeric crystal structure of a Bowman–Birk protease inhibitor from pea seeds. J. Mol. Biol. 285, 1195–1207.

Lin, G., Bode, W., Huber, R., Chi, C., Engh, R.A., 1993. The 0.25 nm X-ray structure of the Bowman–Birk type inhibitor from mung bean in ternary complex with porcine trypsin. Eur. J. Biochem. 212, 549–555.

Luckett, S., Garcia, R.S., Barker, J.J., Konarev, A.V., Shewry, P.R., Clarke, A.R., Brady, R.L., 1999. High-resolution structure of a potent, cyclic proteinase inhibitor from sunflower seeds. J. Mol. Biol. 290, 525–533.

McRee, D.E., 1999. XtalView/Xfit—A versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. 125, 156–165.

Mello, M.O., Tanaka, A.S., Silva-Filho, M.C., 2003. Molecular evolution of Bowman–Birk type proteinase inhibitors in flowering plants. Mol. Phylogenet. Evol. 27, 103–112.

Meyskens, F.L., 2001. Development of Bowman–Birk inhibitor for chemoprevention of oral head and neck cancer. Ann. N.Y. Acad. Sci. 952, 116–123.

Miyata, S., Koshikawa, N., Yasumitsu, H., Miyazaki, K., 2000. Trypsin stimulates integrin alpha(5)beta(1)-dependent adhesion to fibronectin and proliferation of human gastric carcinoma cells through activation of proteinase-activated receptor-2. J. Biol. Chem. 275, 4592–4598.

Murshudov, G.N., Vagin, A.A., Dodson, E.J., 1997. Refinement of macromolecular structures by the maximum likelihood method. Acta Crystallogr., D Biol. Crystallogr. 53, 240–255.

Odani, S., Ikenaka, T., 1973. Studies on soybean trypsin inhibitors. 8. Disulfide bridges in soybean Bowman–Birk proteinase inhibitor. J. Biochem. (Tokyo) 74, 697–715.

Park, E.Y., Kim, J.A., Kim, H.W., Kim, Y.S., Song, H.K., 2004. Crystal structure of the Bowman–Birk inhibitor from barley seeds in ternary complex with porcine trypsin. J. Mol. Biol. 343, 173–186.

Qi, R.F., Song, Z.W., Chi, C.W., 2005. Structural features and molecular evolution of Bowman–Birk protease inhibitors and their potential application. Acta Biochim. Biophys. Sin. 37, 283–292.

Sanner, M.F., Olson, A.J., and Spehner, J.C. (1995). Fast and robust computation of molecular surfaces. Presented at the 11th Symposium on Computational Geometry held in Vancouver, BC, Canada.

Sibanda, B.L., Blundell, T.L., Thornton, J.M., 1989. Conformation of beta-hairpins in protein structures. A systematic classification with applications to modelling by homology, electron density fitting and protein engineering. J. Mol. Biol. 206, 759–777.

Suzuki, A., Yamane, T., Ashida, T., Norioka, S., Hara, S., Ikenazka, T., 1993. Crystallographic refinement of Bowman–Birk type protease inhibitor A-II from peanut (*Arachis hypogaea*) at 2.3 Å resolution. J. Mol. Biol. 234, 722–734.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl. Acids Res. 22, 4673–4680.

Tsunogae, Y., Tanaka, I., Yamane, T., Kikkawa, J., Ashida, T., Ishikawa, C., Watanabe, K., Nakamura, S., Takahashi, K., 1986. Structure of the trypsin-binding domain of Bowman–Birk type protease inhibitor and its interaction with trypsin. J. Biochem. 100, 1637–1646.

Vagin, A., Teplyakov, A., 2000. An approach to multi-copy search in molecular replacement. Acta Crystallogr. D Biol. Crystallogr. 56, 1622–1624.

Voss, R.H., Ermler, U., Essen, L.O., Wenzl, G., Kim, Y.M., Flecker, P., 1996. Crystal structure of the bifunctional soybean Bowman–Birk inhibitor at 0.28-nm resolution. Structural peculiarities in a folded protein conformation. Eur. J. Biochem. 242, 122–131.

Ware, J.H., Wan, X.S., Rubin, H., Schechter, N.M., Kennedy, A.R., 1997. Soybean Bowman–Birk protease inhibitor is a highly effective inhibitor of human mast cell chymase. Arch. Biochem. Biophys. 344, 133–138.

Ware, J.H., Wan, X.S., Kennedy, A.R., 1999. Bowman–Birk inhibitor suppresses production of superoxide anion radicals in differentiated HL-60 cells. Nutr. Cancer. 33, 174–177.

Werner, M.H., Wemmer, D.E., 1992. Three-dimensional structure of soybean trypsin/chymotrypsin Bowman–Birk inhibitor in solution. Biochemistry 31, 999–1010.

Wishart, D.S., Sykes, B.D., 1994. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J. Biomol. NMR 4, 171–180.

Yavelow, J., Collins, M., Birk, Y., Troll, W., Kennedy, A.R., 1985. Nanomolar concentrations of Bowman–Birk soybean protease inhibitor suppress x-ray-induced transformation in vitro. Proc. Natl. Acad. Sci. USA 82, 5395–5399.

Zhang, L., Wan, X.S., Donahue, J.J., Ware, J.H., Kennedy, A.R., 1999. Effects of the Bowman–Birk inhibitor on clonogenic survival and cisplatin- or radiation-induced cytotoxicity in human breast, cervical, and head and neck cancer cells. Nutr. Cancer 33, 165–173.

Please cite this article in press as: Capaldi, S. et al., Crystal structure of the anticarcinogenic Bowman–Birk inhibitor from snail medic (*Medicago scutellata*) seeds complexed with bovine trypsin, J. Struct. Biol. (2006), doi:10.1016/j.jsb.2006.10.017

8. Meeting Communications

Structural characterization of two bile acid binding proteins in zebrafish (Danio rerio)

Authors [a]S. Capaldi, G. Saccomani, B. Faggion, M. Guariento and H.L. Monaco
E-mail: faggion@sci.univr.it

The family of the fatty acid-binding proteins (FABPs) includes a large number of intracellular molecules with the same fold, a beta barrel with 10 strands of antiparallel beta sheet plus two very short alpha helices. The bile acid-binding proteins (BABPs) belong also to this group and therefore share the same fold and present sequence similarity. Their function is not known but it is generally assumed that it is related to that of solubilization, storage and transport of lipophilic ligands in the cell [1]. In the zebrafish (*Danio rerio*) two different BABPs are present: the liver type, L-BABP, (formerly called live basic fatty acid-binding protein (Lb-FABP) and the intestinal type, I-BABP, highly homologous to the mammalian ileal BABPs (or gastrotropins) [2].
The two recombinant proteins were expressed in E.coli and crystallized in the presence of cholic acid and the two structures were solved by X-ray diffraction methods. Our study shows that while the intestinal BABP binds two cholate molecules per protein monomer in its internal cavity, the liver BABP binds only one bile acid, which is different from the stoichiometry of chicken liver BABP [3]. Our poster will present a comparison of the binding sites of the two proteins and a site directed mutagenesis study aimed at dissecting the causes of the different stoichiometries.

[1] Banaszak, L., Winter, N., Xu, Z., Bernlohr, D.A., Cowan, S. and Jones, T.A. (1994) Lipid-binding proteins: a family of fatty acid and retinoid transport proteins, *Advances in Protein Chemistry 45*, 89-151.
[2] Gantz, I., Nothwehr, S.F., Lucey, M., Sacchettini, J.C., DelValle, J., Banaszak, L.J., Naud, M., Gordon, J.I. and Yamada, T. (1989) Gastrotropin: not an enterooxyntin but a member of a family of cytoplasmic hydrophobic ligand binding proteins, *J. Biol. Chem. 264*, 20248-20254.
[3] Nichesola.D., Perduca, M., Capaldi, S., Carrizo, M.E., Righetti, P.G. and Monaco, H.L. (2004) Crystal structure of chicken liver basic fatty acid-binding protein complexed with cholic acid, *Biochemistry 43*, 14072-14079.