UNIVERSITÀ DEGLI STUDI DI VERONA

Department of Biotechnology

Graduate School of Natural Sciences and Engineering

Doctoral program in Biotechnology

Cycle XXIX – year 2014

Engineering Two Human Proteins with β-Trefoil Fold for Therapeutic Applications

S.S.D. BIO/11

Coordinator:

Prof. Massimo Delledonne

Tutor:

Prof. Hugo L. Monaco

Doctoral Student:

María Cecilia González

Abstract

The expression of glycoproteins containing immature truncated O-glycans such as the Thomsen-Friedenreich antigen (Ser/Thr-O-Galβ1–3GalNAc; T-antigen) and the Lewis antigen (sialyl-T-antigen) is a characteristic feature observed on almost all malignant epithelial cells. Those antigens can be recognized by lectins, a group of highly specific carbohydrate-binding proteins whose three-dimensional structure has been studied in our laboratory by X-ray crystallography. BEL β-trefoil is a lectin found in mushrooms that contains three binding sites for the T-antigen, its antiproliferative activity was demonstrated in various human tumor cell lines and it has also been employed for the targeting of antitumor drugs. Unlike other lectins with these properties, BEL β-trefoil presents a structural fold that is also found in human proteins, unlocking the opportunity to use protein engineering tools to design new anticancer therapeutics.

This thesis explores the possibility of modifying existing human proteins to recognize the carbohydrate antigens present on the surface of cancer cells, in order to reduce the potential immunogenicity risk that foreign lectins could have and allowing their future application in drug-delivery targeting. To reach this purpose, two human proteins structurally similar to BEL β-trefoil were modified following different strategies. Human acidic fibroblast growth factor (FGF1) was modified in an attempt to create a new carbohydrate binding site, while a truncated form of human N-acetylgalactosaminyltranferase-6 (GalNAc-T6) was produced to exploit its affinity to N-acetylgalactosamine for this new purpose.

Biophysical methods such as spectrofluorimetry and isothermal titration calorimetry were used to analyze the ability of the engineered proteins to bind the T-antigen monosaccharides. The binding dissociation constant (Kd) of the protein-carbohydrate interaction was determined. The stability of each protein was also studied through their thermodynamic parameters of unfolding using differential scanning calorimetry. Crystallization screenings were set up using a broad variety of precipitants in order to produce crystals to be used to study the three-dimensional structure of the engineered proteins using X-ray diffraction. The crystals that were grown were taken to the European Synchrotron Radiation Facility (ESRF) in Grenoble (France) to carry out the diffraction experiments.

In conclusion, this work provides a new and interesting insight for the production of optimized protein therapeutics applied in drug-delivery methods for cancer treatment. The present biophysical data are the prerequisite for future studies regarding the biological properties of the engineered proteins and clinical parameters for their potential use in medicine.

Table of Contents

Chapter I

Introduction

1.1 Glycosylation Changes in Cancer Development

1.1.1 *Glycoproteins*

Glycosylation is the most frequent and well-known posttranslational modification of proteins. More than half of all human proteins are estimated to be glycosylated, which affects their structure, function, stability, trafficking, and receptor recognition.

The addition of initial sugars to glycoproteins occurs in the endoplasmic reticulum (ER) or on the ER membrane, but they acquire their final carbohydrate pattern during passage through the Golgi apparatus. In this organelle reside multiple glycosyltransferases (GT), glycosidases, and nucleotide sugar transporters that are disposed in a generally ordered manner from the cis-Golgi to the trans-Golgi network. Each Golgi GT transfers a sugar to a specific substrate generated by preceding GTs, and must act at a particular stage in the glycosylation pathway. A sequence motif in one or more of the GTs domains determines its localization along the Golgi network. The expression and arrangement of these transferases, which have overlapping acceptor preferences, and the availability of glycoprotein substrates and activated sugar donors directly affect the final composition and heterogeneity of the oligosaccharide sequences[1].

1.1.2 *Synthesis of O-glycans*

The initiation of O-glycosylation also occurs in the ER, with a subsequent extension in the Golgi complex by the addition of Gal, GlcNAc, sialic acid and fucose to form linear or branched O-GalNAc glycans. The most abundant types of O-glycosylation are found in mucins, a group of highly O-glycosylated proteins which are expressed by mucosal epithelial cells. Mucins play an important role by acting as a protective barrier for cells against mechanical damage and infections, and also initiate intracellular signalling pathways in response to pathogen invasion.

The first step in mucin biosynthesis is covalent binding to the protein of the GalNAc sugar moiety (Tn-antigen), which is a multiple step reaction catalyzed by several polypeptide N-α-acetylgalactosaminyltransferases (GalNAcTs). After this addition, the carbohydrate structure can be further modified by downstream GTs, such as T-synthase (core 1), C3GnT, C2GnT1, and

C2GnT2. But the most common elongation step is the core 1 O-glycan structure (Galβ1–3GalNAcα1-O-Ser/Thr; T-antigen) and this step is catalyzed by a single enzyme, T-synthase, that requires a specific molecular chaperone (COSMC) for folding and activity. The glycan extension usually continues with LacNAc repeats (for cores 2 and 4), blood group antigens, sialylation or fucosylation[2].

Figure 1. Biosynthesis pathways for O-glycan core structures. In the highlights, T-synthase transfers Gal from UDP-Gal to the T-antigen to form the core 1 (T-antigen), it is the key glycosyltransferase for biosynthesis of complex O-glycans.[3]

1.1.3 *Altered O-glycophenotypes of cancer*

An erroneous regulation of O-glycosylation pathways is related to a number of disorders, specially cancer, which is characterized by the overexpression of truncated (Tn, sTn, and T-antigens) and overfucosylated structures (Lewis antigens). In fact, these immature truncated O-glycans are a characteristic feature observed on almost all malignant epithelial cells, being present both early and late in cancer progression and metastasis. Many types of tumors such as colon, breast, pancreatic, lung, cervix and ovarian, exhibit a highly increased expression of altered mucins, and their expression level significantly correlates with cancer prognosis.

The changes in protein O-glycosylation in cancers have been proposed to result from a number of different mechanisms, including altered expression of glycosyltransferases acting in the processing step, somatic mutations or hypermethylation of COSMC, relocation of polypeptide GalNAc-transferases from the Golgi body to the ER, general reorganization of glycosyltransferase topology, and fluctuations in cellular pH. These possibilities are not mutually exclusive and one or other may govern in different situations, however the mechanism and the biological significance of the truncated O-glycophenotype have not been widely understood yet[4].

In the last years, the tumor-associated carbohydrate antigens have been used as prognostic markers in the clinical diagnostic of patients with neoplastic diseases. Nevertheless, recent studies suggest that this utility could be expanded also to that of therapeutic targets, providing an unequivocal label for tumor cells recognition and therefore improving the specificity of cancer treatments[5].

1.2 Lectins

1.2.1 *Lectins as recognition molecules*

In 1888, Peter Hermann Stillmark described a toxin purified from seeds of the castor tree *(Ricinus communis)* called “ricin”, which possesses agglutination activity on red blood cells. Since then, a large number of proteins present ubiquitously in nature have been identified that are capable of interacting with the glycoconjugates present on animal cell surfaces. These proteins were termed “lectins”, a name derived from the Latin word *legere* that means ‘‘to choose’’ or ‘‘to select’’, and refers to their ability to recognize and bind specific carbohydrate structures. Nowadays, this definition was generalized to a broad group of proteins or protein domains of non-immune origin which bind reversibly to specific saccharides through their carbohydrate-binding site. Actually, this binding can be as specific as the interaction between those of antigen-antibody or substrate-enzyme. Lectins may also bind to free sugars or the sugar residues of polysaccharides, glycoproteins, or glycolipids which can be free or bound (as in cell membranes). Due to this property, lectins mediate many biological processes in cells such as recognition, differentiation, signalling, adhesion and migration, apoptosis, immunomodulation, inflammation, and host–pathogen interactions[6].

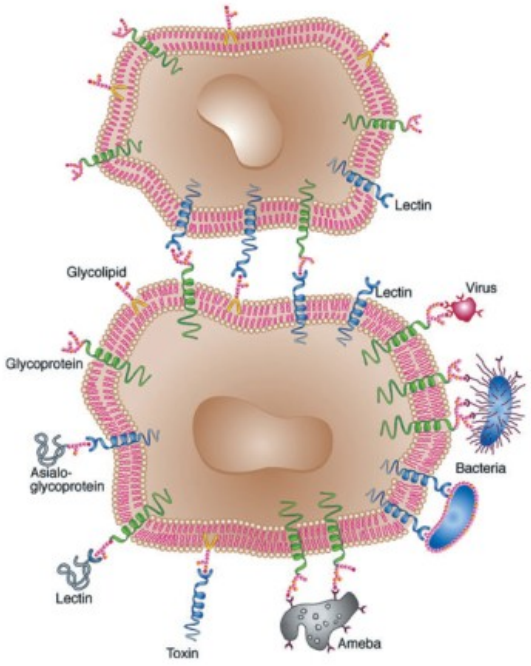


Figure 2. Cell surface lectins-carbohydrates interactions[6].

An interesting fact concerning lectins is that members from different organisms which do not show primary structural homology can have similar preferential binding to the same carbohydrate. However, in the past years, many tertiary structures of lectins from diverse living systems were elucidated. It was then observed that lectins lacking sequence similarity can have related 3D-structures. Even more, lectin-carbohydrate complexes solved by X-ray crystallography suggested that the key interactions responsible for carbohydrate recognition are common, and they are generated by a few amino acid residues designated as the carbohydrate recognition domain (CRD). The CRD confers the lectin a high specificity, that in some cases can even discriminate between anomeric isomers. For instance, the legume lectin concanavalin A (ConA) specifically binds the α-anomer of glucose and mannose, but not the β-anomer of either[7].

The interaction lectin-carbohydrate can be stabilized through hydrogen bonds, metal coordination, van der Waals and hydrophobic interactions. Nonetheless, the numerous hydroxyl groups present on sugars promote the formation of hydrogen bonds with the side chain atoms of some amino acids such as aspartic acid and asparagines, and less commonly, with the main-chain amide hydrogens and carbonyl oxygen. These protein-carbohydrate hydrogen bonds can be both direct and water mediated[8].

1.2.2 *Lectin-mediated drug delivery*

The potential of lectins as drug delivery agents was firstly proposed by Woodley and Naisbettin in 1988, who found that tomato lectin could adhere to the intestinal mucosa in vitro, and suggested its use to target the luminal surface of the small intestine. The use of drug delivery systems (DDS) offers many advantages for the administration of therapeutics. The DDSs not only increase the efficacy of the treatment and limit the side effects on healthy tissues, but also improve drug uptake and internalization. For that reason considerable efforts have been made to exploit lectin-carbohydrate interactions and their induction of endocytosis pathways to target drugs to their specific site of action thus minimizing drug accumulation at non-specific cells[9].

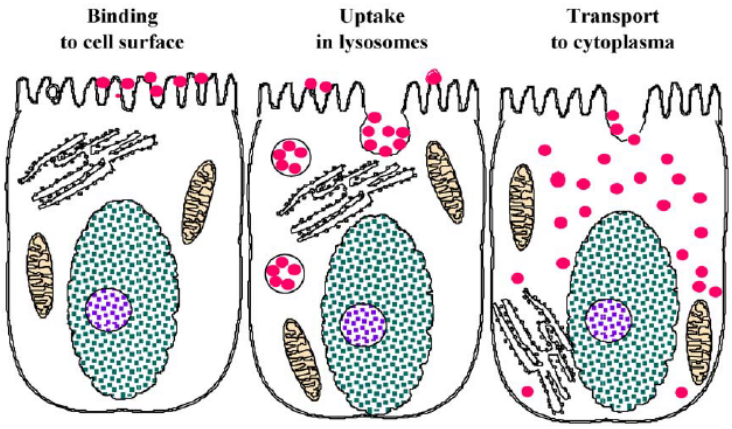


Figure 3. Bio-adhesion and endocytosis of lectins and lectin-conjugates depends on the protein-sugar interactions on cell surfaces[9].

DDS based on lectin targeting may be exploited by two mechanisms: direct and reverse lectin targeting. In the first, the DDS contains sugar moieties that are recognized by endogenous lectins present on the cell surface. The second mechanism utilizes instead exogenous lectins as targeting moieties that carry the DDS to glycoproteins or glycolipids overexpressed on the cell surface. Since the intravenous administration of anti-cancer chemotherapy reagents produces tissue and organ damage due to cytotoxic effects on normal cells, lectin-based DDSs could be greatly beneficial in cancer therapy. It is for this reason that the increased levels of carbohydrate antigens displayed in tumors can be an ideal target in reverse lectin targeting for anti-cancer treatment[7].

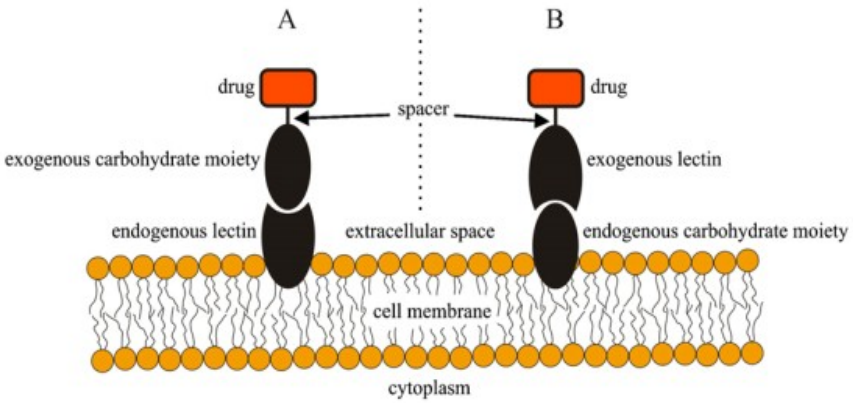


Figure 4. Illustration of direct lectin targeting (left) and reverse lectin targeting (right)[7].

1.2.3 *Anticancer properties of lectins*

A revolution in lectinology took place in the early 60s with the discovery of a plant lectin with the ability to preferentially agglutinate malignant cells in the presence of normal ones. The wheat germ agglutinin (WGA) which was been studied by Aub and collaborators, is able to agglutinate leukemia cells and cells transformed by the tumorigenic polyoma virus more than the relative untransformed cells[6]. This outstanding finding opened the way to research in the field, and since then a multitude of plant and fungal lectins were found to posses anticancer activity. Moreover, this new property offered new perspectives in lectin-based DDSs combining their specific cell recognition and cytotoxic effect.

The potential of WGA in the inhibition of tumor growth *in vitro* was further demonstrated on human pancreatic carcinoma, lymphoma and breast cancer cells. Additionally, WGA induced chromatin condensation, nuclear fragmentation and DNA release, which was consistent with apoptosis. Also the lectin Concanavalin A (ConA), extracted from the jack-bean (*Canavalia ensiformis)*, was demonstrated to induce apoptosis in human melanoma cells, but it also triggered autophagic activities in human hepatoma cells. Another example was found in the lectin-rich Mistletoe (*Viscum album*) extract, which has shown to induce apoptosis in several types of human cancer, but the mechanisms varied with individual cases[10].

The use of several plant lectins as anti-cancer agents has reached the pre-clinical stage. However, experiments on animal models have not yet found significant clinical efficacy in terms of tumor control and survival time for patients.

The discoveries mentioned above suggest that lectins might possess some similar biological activities and anti-tumor mechanisms that are closely correlated with their corresponding molecular structures. This has been the object of study of this laboratory in the last years, in which the three-dimensional structure of several lectins and their complexes with carbohydrates were solved. The following molecules constitute some examples of them.

1.2.3.1 ABL (*Agaricus Bisporus* Lectin)

This lectin, found in the common edible mushroom *Agaricus bisporus,* presents anti-proliferative properties on malignant epithelial cell lines. This effect is not observed on healthy cells, and thus the lectin has been proposed as a potential anti-tumor agent in cancer therapies. ABL possesses the ability to selectively recognize neoplastic cells because of its high affinity binding to the T-antigen expressed on their surface.

The active protein has a molecular mass of 64 kDa and is formed by four monomers. The X-ray diffraction experiments on ABL crystals revealed that the monomer is a single domain structure organized as a β-sandwich with six strands of β-chain in the first sheet and four strands in the second sheet. The first sheet is of the mixed type and the second sheet is antiparallel, both are connected by a helix-loop-helix motif. The resolution of the structure of the complex protein-T-antigen showed the disaccharide bound at the edge of a pocket in each monomer. The carbohydrate interacts with the residues of the chain connecting the two sheets through the loops present between strands B and C and D and E and the β-turn between strands F and G of the second sheet (figure 5). Surprisingly, ABL contains a second carbohydrate-binding site specific for *N*-acetylglucosamine and not related to the T-antigen active site. It is remarkable that the two sites are able to discriminate between the two monomers that differ only in the configuration of a single epimeric hydroxylic group[11].

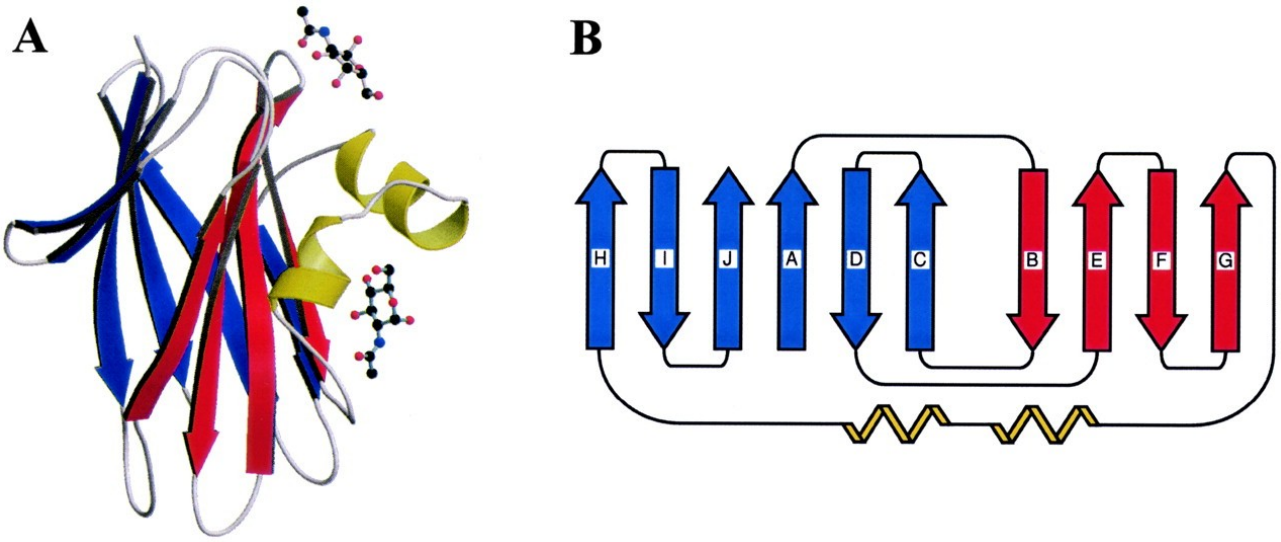


Figure 5-A. A ribbon representation of the ABL monomer. The six-stranded β sheet is shown in blue, the four-stranded sheet is shown in red and the two short helices are yellow. The two carbohydrates are represented as a ball-and-stick model: N-acetylgalactosamine is at the top T-antigen binding site and N-acetylglucosamine at the other site.

Figure 5-B. A topological diagram of the ABL monomer. B-strands are labelled in the order of their appearance from the N terminus to the C terminus[11].

1.2.3.2 POL (*Pleurotus Ostreatus* Lectin)

Isolated from oyster mushrooms and characterized for the first time by Wang and colleagues (1998), the 3D-structure of this lectin was recently solved in this laboratory and the lectin was shown to belong to a new structural family. The protein is composed of two subunits, each weighing about 40 kDa, and is organized as 2 cylinders formed by 22 β-sheets. The crystal structure at 2 Å resolution showed the monomer formed by two pseudosymmetric domains. The N-terminal domain contains 10 β-strands, while the C-terminal domain contains 12. Both domains are associated face-to-face with the β-sheet radially arranged around a central tunnel[12].

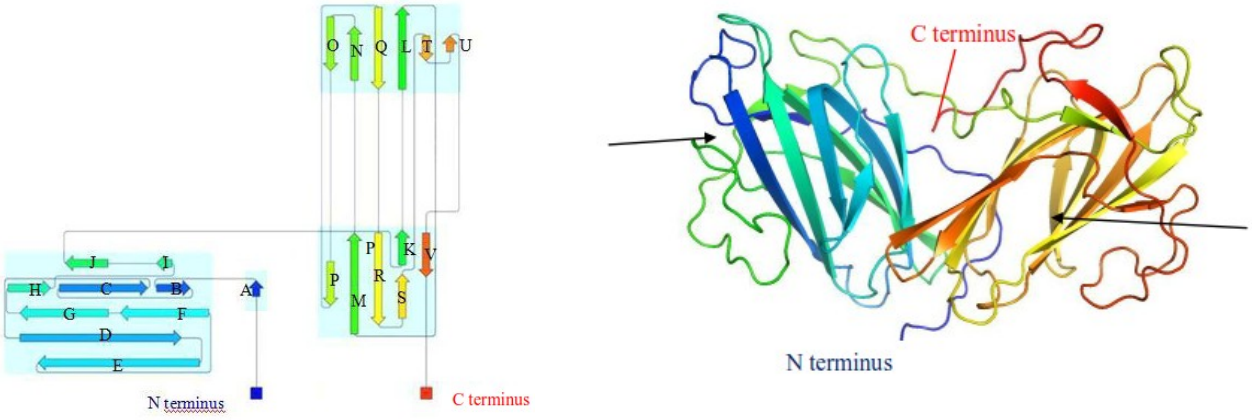


Figure 6-A (left). A topological diagram of POL. β-strands are labelled in the order of their appearance from the N terminus to the C terminus.

Figure 6-B (right). A ribbon representation of POL. The colours used for the β-strands are the same used in the POL topological diagram. Black arrows indicate the cylindrical cavities in the two domains[12].

Initial studies evidenced the high affinity of POL for certain carbohydrates since its hemagglutinating activity was repressed by inulin, lactose, melibiose and D-galactose. The specificity of the lectin for these sugars was further investigated performing studies of the electrophoretic mobility of the lectin in the presence of α-lactose-1-phosphate. In addition, an α-galactosidase activity was measured using p-nitrophenyl-α-galactopyranoside as substrate. Other

carbohydrates that were shown to interact with the lectin using SPR analysis are 2'-fucosyllactose, 2'-fucosyl-N-acetyllactosamine, lactose and galactose. Also recent spectrofluorimetric experiments carried out by our research group with a recombinant POL yielded Kd values in the micromolar range for D-glucose, L-rhamnose, D-galactose and N-acetylgalactosamine.

Further tests performed on human cancer cells from pancreas (MiaPaCa-2) and liver (HepG2) demonstrated the anti-proliferative effect of POL inducing the apoptotic pathway. In addition, the encapsulation in poly lactic-co-glycolic acid (PLGA) nanoparticles presented favourable biophysical properties for their use as therapeutics.

1.2.3.3 BEL β-trefoil: (*Boletus Edulis* Lectin with β-trefoil fold)

After finding the first lectin with antineoplastic properties in the edible mushroom king bolete called BEL[13], additionally studies performed in this laboratory allowed the purification and characterization of a new lectin with similar properties, also in *Boletus edulis,* that was named BEL β-trefoil. This name was given because each protomer in the homodimer folds as a β-trefoil domain.

The apo form of the protein as well as its complex with carbohydrates were examined in detail by X-ray crystallography. The molecule contains 12 β-strands organized in 3 subdomains with similar β–β–β-loop–β (trefoil) units commonly called α, β and γ, starting from the N terminus of the protein. The overall fold has a pseudo threefold symmetry and consists of a six-stranded-barrel capped by a triangular hairpin triplet. Each subdomain of BEL β-trefoil contains a carbohydrate-binding site, consequently the physiological dimer presents six potential binding sites. The X-ray diffraction of the single crystals soaked with lactose, Gal, GalNAc, the T-antigen disaccharide (Galβ1-3GalNAc) and the T-antigen (Ser-Galβ1-3GalNAc) showed that the three binding sites in the molecule can be occupied. The lectin appears to bind to both Gal and GalNAc with the same affinity, and the T-antigen binds to the active site indifferently using one or the other moiety equally well[14].

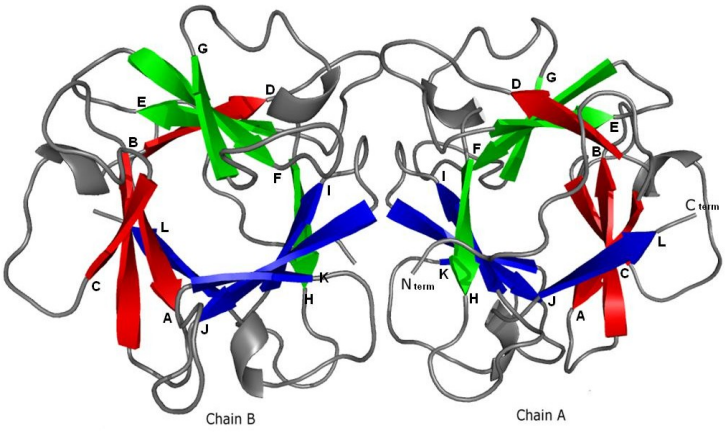


Figure 7. BEL β-trefoil dimer, the view is in approximately the direction of the pseudo three-fold axis. The strands are named from the N terminus of the protein. Subdomain α (red) contains strands L, A, B and C, all antiparallel to each other, subdomain β (green) contains strands D, E, F and G and subdomain γ (blue) strands H, I, J and K[14].

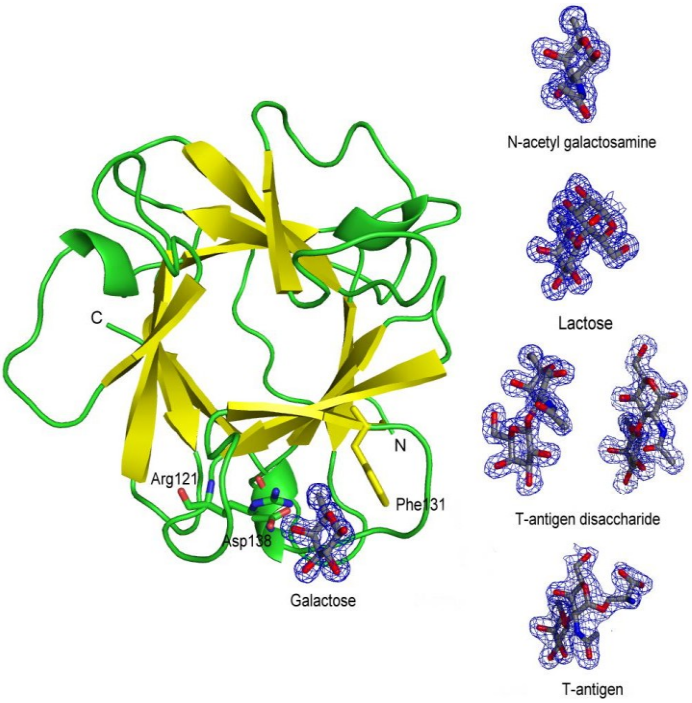


Figure 8. BEL β-trefoil monomer with the electron density of the saccharides that can occupy its binding sites. The carbohydrates are in the orientation they have when bound in site γ and the T-antigen disaccharide is represented again in both the two possible orientations[14].

The antineoplastic activity of BEL β-trefoil was tested on different human tumor cell lines from breast, liver, rectum, pancreas, cervix, brain, colon, skin and lung. The lectin had a dose-dependent antiproliferative effect in all of them, but was found to be stronger on HeLa cervix adenocarcinoma (40 μg/ml of BEL β-trefoil decreased cell viability down to 11% of the control after 48 hours). This effect was not detected on human fibroblast lines, used as a standard of healthy cells, demonstrating the selectivity of the lectin for neoplastic cells. Moreover, the presence of carbohydrates inhibited the antiproliferative activity of the lectin, showing the importance of the sugar moieties on the cell surface for the recognition.

Confocal microscopy demonstrated that BEL β-trefoil enters into the tumor cells, but the mechanism of internalization still remains unclear. Also the mechanism by which the lectin inhibits cell growth has not been elucidated yet, since the results indicated that it does not induce the apoptotic pathway. For this reason, BEL β-trefoil cannot be proposed as an anticancer agent, but it could work very well in drug delivery targeting. To test this ability, the lectin was conjugated to PLGA nanoparticles loaded with α-bisabolol. The active component is a molecule that induces apoptosis in human and murine adenocarcinoma cell lines but not in normal cells. Preliminary results confirmed that the empty nanoparticles had no effect on both neoplastic and normal cells, whereas nanoparticles loaded with α-bisabolol induced apoptosis. However, a noticeable lower selectivity was also observed, which should be improved increasing the lectin concentration though it would be essential to do it without eliciting an immune response.

Although BEL β-trefoil has potent anti-proliferative effects on human cancer cells which confers to it an interesting therapeutic potential as an antineoplastic agent, its safety when used in human patients could not be guaranteed. This is due to its fungal origin, this lectin may induce an immune reaction in vivo, which could have adverse effects on both the carbohydrate recognition function of the lectin and the health of the patients. Many fungal lectins that recognize the T-antigen disaccharide and therefore malignant cells belong to folds that are not found among human proteins but the β-trefoil is present among them. It is for this reason that we decided to explore the possibility to engineer a human protein to create in it similar properties to those that characterize BEL β-trefoil. Finding a human protein structurally similar to the fungal lectin would allow the modification of some essential residues to create in it a T-antigen binding site. On the other hand, if there is a human protein capable of binding to the carbohydrate antigen, it could present similar properties to those found in BEL β-trefoil. Both options would result in a protein that recognizes the T-antigen and therefore malignant cells, but avoiding an immune response.

1.3 Protein Engineering for Therapeutic Applications

Many early attempts to introduce proteins as therapeutic agents failed because the foreign molecules were recognized as non-human and led to an immune response against them. As a result, most proteins used in clinical trials now are primarily human or are "humanized", using protein engineering techniques that allow the reduction of the immunogenicity, enhancement of the affinity to the target and an increase of their half-life in the bloodstream in order to improve their efficacy[15].

This approach permits also to enhance the affinity by mutagenesis, followed by selection of improved variants from the resulting repertoire. Thus, therapeutics candidates with predefined properties in terms of affinity and specificity could be produced. This aim could be reached thanks to the increasing number of protein structures that are being solved. Even with low-resolution structural knowledge, for example, based on homologous family members, could be of value in guiding mutagenic and selection strategies in drug discovery.

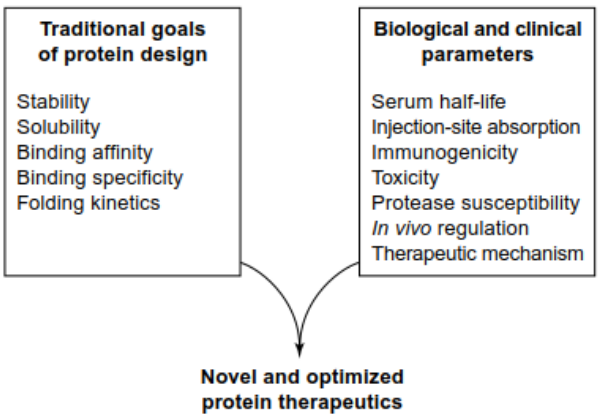


Figure 9. The design of protein therapeutics integrating the traditional goals of protein design (left) and biological and clinical parameters (right)[16].

There are two principal strategies usually used in protein engineering: rational design and molecular evolution. Both strategies allow the generation of novel proteins with modified functions or entirely new properties. The rational design approach involves "site-directed mutagenesis" to introduce specific amino acids into a target gene. This is an effective technique

when the structure and mechanism of the protein of interest are well-known. However, when this information is limited the “evolutionary method” can be used which involves sequentially random mutagenesis and selection of variants[17].

Other methods such as X-ray crystallography to elucidate the resulting protein structures, homology modelling and molecular dynamics simulations are important as they could provide additional key information to protein engineering studies.

1.4 Research Aims

The present work is focused on the study of two human proteins that were customized in order to give them similar carbohydrate-binding properties to those found in BEL β-trefoil. The clinical potential found in the fungal lectin could be reproduced in human proteins making them capable of recognizing carbohydrate antigens over-expressed on tumor cells. In this way, it is possible to lower the risks of incompatibility and enhance safety in the use of proteins carriers in drug-delivery therapies.

The first chapter is dedicated to the modification of acidic Fibroblast Growth Factor, a human protein with high structural homology to BEL β-trefoil. This human protein was engineered by rational design in an attempt to construct a T-antigen binding site. The recombinant protein *wild type* and different mutants were overexpressed in *Escherichia coli* to proceed with their characterization.

The second chapter is centred on the human N-α-acetylgalactosaminyltransferase-6, which contains a catalytic domain and a lectin domain with a binding site to N-acetylgalactosamine (GalNAc), one of the saccharides exposed by cancer cells (Tn-antigen). Only the lectin domain of this enzyme was expressed by cloning the C-terminal portion of the DNA coding sequence and introducing in *Pichia pastoris* for its recombinant production.

The overall aim of this research is to investigate the biophysical and structural properties of the variants generated by protein engineering in order to evaluate whether they may be proposed for tumor cells recognition in cancer treatments.

Chapter II

Creation of a T-antigen Binding Site on Human Acidic Fibroblast Growth Factor

2.1 Background

2.1.1 *Acidic Fibroblast Growth Factor* (FGF1)

Unlike other lectins with the ability to bind the T-antigen, BEL β-trefoil shows structural homology with a human protein, acidic Fibroblast Growth Factor (FGF1)[14]. FGF1 is a member of a large family of growth factors that exhibit a wide variety of biological activities and activate several signalling pathways. FGF1 is a potent mitogen, stimulating DNA synthesis and proliferation in many cell types, thus it plays important roles in angiogenesis, tissue regeneration, and inflammation. As other members of the FGF family, it mediates cellular functions by binding to transmembrane receptors. The cellular receptors for FGF1 are tyrosine kinase receptors (FGFR) which are activated by ligand-induced dimerization requiring heparin as co-receptor. Heparin is a member of the glycosaminoglycan family of carbohydrates that is widely distributed in animal tissues. It consists of a linear chain of a repeated disaccharide unit composed of a 2-O-sulphated iduronic acid and 6-O-sulphated, N-sulphated glucosamine. Signal transduction requires association of FGF1 with its receptor tyrosine kinase (FGFR) and heparin in a specific complex on the cell surface[18].

The three-dimensional structure of FGF1 has been determined by X-ray crystallography, it consists of twelve stranded antiparallel β-sheet with a pseudo three-fold symmetry axis (β-trefoil fold). The monomeric human protein of 16 kDa is composed by 140 amino acids and contains three free cysteine residues located at solvent-inaccessible positions[19]. A cluster of basic residues in the C-terminal region (from Lys112 to Lys128) of the molecule form the heparin binding site[20].

2.1.2 *Similarity with BEL β-trefoil*

Protein structure was much better conserved than sequence during evolution[21]. Therefore, three dimensional structure superposition became a powerful method for detecting homology of distantly-related proteins. The structure database searcher Dali server permits to compare a newly solved structure against the database of known structures in order to detect possible evolutionary relationships[22]. This server allowed the comparison of BEL β-trefoil (PBD entry 4I4O) against those in the Protein Data Bank (PDB). The results summary for the highest scoring superpositions

showed FGF1 as the only human protein with a relevant structural similarity with BEL β-trefoil. The Z-score between the two proteins is 17.9 with a 11% of sequence identity, while the average RSMD is 1.8 Å between the corresponding α-carbons.

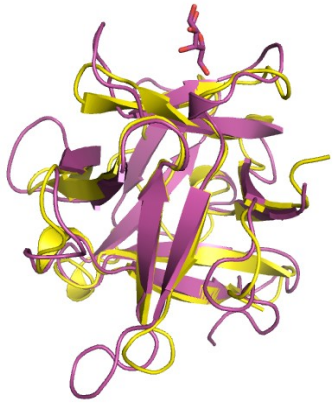


Figure 10. Superposition of FGF1 (yellow) and BEL β-trefoil (violet) tertiary structures.

The crystal structure of BEL β-trefoil soaked with the T-antigen provided a template to determine which residues in FGF1 should be changed to create a binding site for this disaccharide. Each binding site in BEL β-trefoil contains an arginine bound to the O4 of Gal (arginines 27, 74 and 121 at the α, β and γ sites, respectively). Similarly, in every site there is an aspartate (44, 92 and 138) hydrogen bound to the O4 and O6 of the saccharide. Two different amino acids: Gln (45 and 139) and Asn 93, also interact with O6 from Gal. In addition, a hydrophobic cavity is formed by tyrosines 25, 37 and 45 at the α site, Tyr 90, Phe72 and Ile 84 at the β site and tyrosines 119, 136 and Phe 131 at the γ site. Analogous interactions were found for the binding of GalNAc, which indicates that the T-antigen binds to the protein using one or other monosaccharide indifferently[14].

Superposition of the BEL β-trefoil and FGF1 structures suggests that the human protein could be mutated to contain at least one of the binding sites for the T-antigen. Such mutations should create in FGF1 the potential capacity of recognizing tumor cells with less immunogenicity than the fungal protein. To reach this aim, two different strategies were tested to engineer FGF1 employing a rational design. In the first, a form of FGF1 containing all the amino acids needed to create two new sites was produced from a synthetic gene while in the second, FGF1 was modified by iterative amino acid mutations using site-directed mutagenesis.

2.2 Experimental Procedure

2.2.1 *Synthetic gene design of FGF1 full mutant*

Superposition of the models of BEL β-trefoil (4I4O-A) and FGF1 (1RG8-A) was performed using the *Coot* software[23] and it was based on the α-carbons of both chains, defining a one-to-one correspondence of amino acid residues (sequence positions) in the two proteins.

The α site shows the highest structural homology, and the γ site includes the heparin binding site in FGF1, which is essential for its biological activity. These two regions were chosen to be modified in order to introduce two T-antigen binding sites and to suppress the mitogenic activity in FGF1 at the same time.

This requires replacing the following five amino acids in every site:

- site α: Arg24Tyr, Leu26Arg, Thr34Tyr, Asp39Tyr and His41Asp

- site γ: Gly110Tyr, Lys112Arg, Arg122Phe, Gln127Tyr and Ala129Asp

Residues Lys112, Arg122, Gln127 are those directly involved in heparin binding.

On the basis of these observations, a mutant form of FGF1 with all these ten amino acids changed was produced (*FGF1 full mut*):

FGF1 WT protein sequence:

FNLPPGNYKKPKLLYCSNGGHFLRILPDGTVDGTRDRSDQHIQLQLSAESVGEVYIKSTETGQYLAMDTDGLLYGSQTPNEECLFLERLEE
NHYNTYISKKHAEKNWFVGLKKNGSCKRGPRTHYGQKAILFLPLPVSSD

FGF1 *full mut* protein sequence:

FNLPPGNYKKPKLLYCSNGGHFL**Y**I**R**PDGTVDG**Y**RDRS**Y**Q**D**IQLQLSAESVGEVYIKSTETGQYLAMDTDGLLYGSQTPNEECLFLERLEE
NHYNTYISKKHAEKNWFV**Y**L**R**KNGSCKRGP**F**THYG**Y**K**D**ILFLPLPVSSD

A synthetic encoded gene, optimized for its expression in *E. coli*, was designed with the software GeneArt Gene Synthesis (ThermoFisher Scientifics). The linear dsDNA fragment of 453 bp contained the ten mutations required and the restriction sites for BamHI, NdeI and XhoI.

GGTGGATCCCATATGTTTAATCTGCCTCCGGGTAATTACAAAAAACCGAAACTGCTGTATTGCAGCAATGGTGGTCATTTT
CTGTATATTCGTCCGGATGGCACCGTTGATGGTTATCGTGATCGTAGCTATCAGGATATTCAGCTGCAACTGAGCGCAGAA
AGCGTTGGTGAAGTTTATATCAAAAGCACCGAAACCGGTCAGTATCTGGCAATGGATACCGATGGTCTGCTGTATGGTAG
CCAGACCCCGAATGAAGAATGTCTGTTTCTGGAACGTCTGGAAGAAAACCACTATAACACCTATATCAGCAAAAAACACGC
CGAAAAAAACTGGTTTGTGTACCTGCGTAAAAATGGTAGCTGTAAACGTGGTCCGTTTACCCATTATGGCTATAAAGATAT
CCTGTTTCTGCCGCTGCCGGTTAGCAGCGATTAACTCGAGGGATCCTGG

Figure 11. Synthetic gene containing residues mutated and restriction sites.

2.2.2 *Cloning of FGF1 full mutant*

The synthetic gene was amplified by PCR using the following primers for its insertion in four different plasmids: pET15b, pGEX-4T-1, pWaldo (pET28-GFP) and pET22b.

The plasmids pET22b and pET15b express a peptide tag of six histidines in tandem with the recombinant protein at its C-terminal and N-terminal, respectively. Histidine-tagged proteins are easy to detect along the expression and purification scheme by western blotting using an anti-his tag antibody. The pWaldo vector, in addition to producing a protein with a His tag at its C-terminal, also adds a fusion protein, GFP. This fusion partner is a folding reporter becoming fluorescent when the upstream protein folds correctly. Similarly, pGEX-4T-1 enables the expression of proteins fused to an N-terminal glutathione S-transferase which acts as a solubility enhancer.

The vectors pET15b, pET22b and pWaldo contain a T7 promoter that regulates the expression of the recombinant protein, whereas pGEX-4T-1 contains a TAC promoter instead. Both promoters are inducible by the lactose analogue IPTG (isopropyl-β-D-thiogalactoside).

pET15b with restriction sites for NdeI and BamHI:

Primer for: 5'-GGTGGATCCCATATGTTTAAT-3'
Primer rev: 5'-AATAATGGATCCATCGCTGCTAACCGG-3'

pGEX-4T using restriction sites for BamHI and XhoI:

Primer for: 5'-GGTGGATCCCATATGTTTAAT-3'
Primer rev: 5'-CCAGGATCCCTCGAGTTAATC-3'

pWaldo with restriction sites for NdeI and BamHI:

Primer for: 5'-GGTGGATCCCATATGTTTAAT-3'

Primer rev: 5'-AATAATGGATCCATCGCTGCTAACCGG-3'

pET22b using restriction sites for NdeI and XhoI:

Primer for: 5'-GGTGGATCCCATATGTTTAAT-3'

Primer rev: 5'-CCAGGATCCCTCGAGTTAATC-3'

Each PCR reaction was run on a 0.8% agarose gel and amplicon bands were excised and extracted using the GenElute Gel Extraction kit (Sigma-Aldrich). DNA fragments and vectors were digested with the appropriate restriction enzymes. Then, plasmids were dephosphorylated with Alkaline Phosphatase (Promega) and isolated from the reaction mix. Finally the DNA fragments were inserted into the corresponding vectors using T4 DNA ligase (Jena Biosciences).

The DNA constructs were further transformed in XL1-Blue, a line of competent *E. coli* cells that are used for routine cloning purposes. Positive selection of transformants was made on LB agar plates containing 100 µg/ml ampicillin (pET15b, pET22b and pGEX-4T-1) or 100 µg/ml kanamycin (pWaldo). The presence of the coding sequence was checked by colony PCR and positive colonies were grown in LB broth at 37°C overnight. Plasmids containing the insert were isolated with Gen Elute plasmid Miniprep kit (Sigma-Aldrich) and conserved at -20°C until use.

2.2.3 *Iterative amino acid mutation on FGF1*

Studies on the structural requirements of FGF1 for receptor binding have reported a double mutant of FGF1 that shows a strong reduced affinity for the receptor, thus inhibiting its activity. Residues Tyr94 and Asn95 in FGF1 are crucial for protein-receptor binding. Tyr94 establishes hydrophobic contacts with Ala167 and Pro169 in the receptor chain, while Asn95 forms a hydrogen bond with the Arg250. The substitution of both residues for alanine causes a strong decrease in receptor binding, blocking the translocation of FGF1 into the cell and suppressing its mitogenic activity[24].

Moreover, considering that in the α site of BEL β-trefoil, there are three essential amino acids (Arg, Asp and Gln) for the interaction with the T-antigen, a new strategy for engineering FGF1 was designed. It consisted in leaving intact the γ site as wild type FGF1 to maintain the affinity for heparin but replacing the amino acids Tyr94 and Asn95 for Ala to render the protein inactive. Then continue mutating one at a time the residues Leu26 for Arg, His41 for Asp and Ile 42 for Gln, to give FGF1 the ability to bind the T-antigen.

2.2.4 *Cloning of FGF1 wild type (FGF1-WT) and site-directed mutagenesis*

The sequence coding *Homo sapiens* FGF1 (MGC:44867 IMAGE:5403677) was amplified by PCR using the following primers and cloned into pET15b using restriction enzymes NdeI and XhoI:

Primer for: 5´-GGTGGTCATATGTTTAATCTGCCTCCAGGG-3'
Primer rev: 5'-GGTGGTCTCGAGTTAATCAGAAGAGACTGG-3´

Cloning into the desired vector and transformation of the DNA construct **FGF1-WT**-pET15b were performed as stated above. Ampicillin was used as selection marker for plasmid DNA isolation.

For site-directed mutagenesis 50 ng of DNA template, 2.5 U/µl of Pfu DNA polymerase, 1X Pfu DNA Polymerase buffer, 0.2 mM DNTPs and 0.3 µM of each primer were used. Each PCR reaction used as template the mutated DNA obtained in the previous reaction, in order to produce accumulated mutations.

Primers used to obtain the double alanine mutant (**FGF1-2M**):

for: 5'-GAGGAGAACCATGCCGCCACCTATATATCC-3'
rev: 5'-GGATATATAGGTGGCGGCATGGTTCTCCTC-3'

Primers used to mutate Leu26 for Arg (**FGF1-3M**):

for: 5´-TTCCTGAGGATCCGTCCGGATGGCACA-3'
rev: 5´-TGTGCCATCCGGACGGATCCTCAGGAA-3'

Primers used to mutate His41 for Asp (**FGF1-4M**):

for: 5´-AGGAGCGACCAGGACATTCAGCTGCAG-3'
rev: 5´-CTGCAGCTGAATGTCCTGGTCGCTCCT-3'

Primers used to mutate Ile42 for Gln (**FGF1-5M**):

for: 5´-GGGACAGGAGCGACCAGGACCAGCAGCTGCAGCTCAGTGCG-3'

rev: 5´-CGCACTGAGCTGCAGCTGCTGGTCCTGGTCGCTCCTGTCCC-3´

Following temperature cycling, the product was treated with *Dpn*I to digest the parental DNA template and to select for mutation-containing synthesized DNA. The nicked vector DNA containing the desired mutations was then transformed into XL1-Blue competent cells. The mutations were confirmed by DNA sequencing.

2.2.5 *Protein expression trials*

Expression constructs were transformed by heat shock in 100 µl of different competent strains of *E. coli* optimized for protein expression: Rosetta, C41, BL21, BL21 pCAM, Origami and C41_pGKJE8.

Protein expression was induced overnight using 0.5 mM IPTG. Optimization of protein induction was made varying expression conditions: time and concentration of induction with IPTG, temperature of incubation, incorporation of additives in growth culture (arginine, sorbitol, sucrose, glucose, glycerol or ethanol), aeration rate, ice shock. The protein expressed on the soluble fraction was detected by Western blotting.

2.2.6 *Expression scale-up and protein purification*

After identifying the *E. coli* strain and the optimal conditions for soluble expression of every protein, a single transformed colony was pre-inoculated in 10 ml LB broth plus 100 µg/ml ampicillin and it was grown at 37°C for 4 hours. The culture was inoculated in 1 litre of the same medium and grown again at 37°C until the optical density measured at 600 nm has reached 0.8 Au. Protein expression was induced by addition of IPTG and incubation overnight.

The growth medium was removed by centrifugation at 10 000xg for 10 minutes and the pellet was resuspended in buffer 1 (see Appendix). Cells were disrupted by addition of 1 mg/ml lysozyme, 1 mM PMSF (phenylmethylsulfonyl fluoride), 1 mM Triton X-100, followed by a sonication step. The lysed insoluble material was removed centrifuging 15 minutes at 10 000xg. The supernatant was

loaded onto a heparin-sepharose column and the protein was desorbed using a linear gradient of NaCl. The chromatographic separation was performed using an ÄKTA PRIME liquid chromatography system.

2.2.7 *Cell proliferation assay*

The effects of FGF1-WT and FGF1-2M on fibroblasts were tested in collaboration with the General Pathology Laboratory in the Department of Medicine at the University of Verona. A WST (see the reagent formula below) assay was set up in order to verify if the double mutation introduced in FGF1 inactivates its proliferation activity. This colorimetric assay analyzes the number of viable cells by the cleavage of tetrazolium salts added to the culture medium. The stable tetrazolium salt WST (slightly red) is cleaved to a soluble formazan (dark red) by a complex cellular mechanism that occurs primarily at the cell surface. This bioreduction is largely dependent on the glycolytic production of NADPH in viable cells. Therefore, the amount of formazan dye formed directly correlates to the number of metabolically active cells in the culture[25].

Cells were grown in 96-well plates to a final density of 20 000 fibroblast/well. Two solutions of 1 mg/ml FGF1-WT and FGF1-2M, previously prepared in sterile PBS buffer, were added to reach a final concentration of 10 µg/ml per well. Samples were incubated with the WST reagent from 0.5 to 1.5 hours. After this incubation period, the formazan dye formed was quantified with a multi-well spectrophotometer. The measured absorbance at 570 nm directly correlates to the number of viable cells.

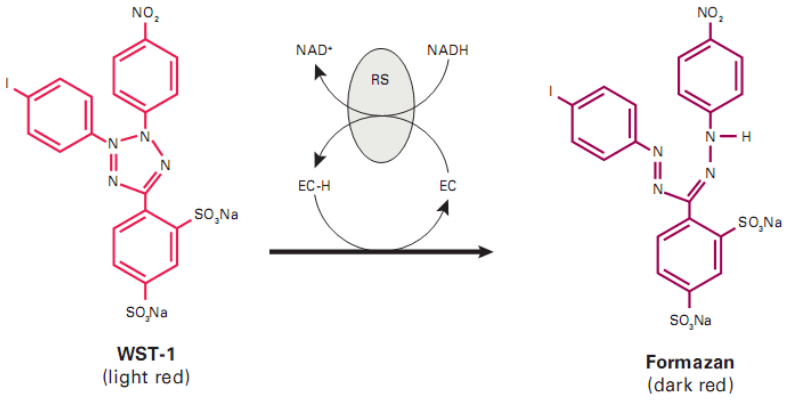


Figure 12. WST-1, in the presence of intermediate electron acceptors (EC), are reduced by succinate- reductase system (RS) using NADH as cofactor.

2.2.8 *Intrinsic fluorescence*

Proteins are naturally fluorescent due to the presence of aromatic amino acids, i.e. tryptophan, tyrosine and phenylalanine. Changes in the emission spectra of these amino acids often occur in response to conformational transitions, subunit association, substrate binding, or denaturation. These processes are usually correlated with collisional quenching and shifts of the emission peak. Hence, the intrinsic fluorescence of a protein is a useful tool for structural studies as for interaction analysis with ligands. In general, in a protein containing the three aromatic amino acids, tryptophan fluorescence dominates tyrosine fluorescence, and both that of phenylalanine, due to a resonance energy transfer. However, for most purposes phenylalanines cannot be considered as useful probes because of their low extinction coefficient and absorption maximum. With an excitation wavelength at 280 nm both tryptophan and tyrosine will be excited[26].

Fluorescence spectra were obtained with a Jasco FP-8200 spectrofluorimeter with all samples at room temperature. A quartz-cuvette was used containing 1 ml of 10 µM of protein sample in buffer 2 (see Appendix). The solutions of galactose and N-acetylgalactosamine, prepared in the same buffer, were added in stepwise manner from 10 µM to 100 µM (final concentrations). Samples were incubated for 30 seconds before excitation at 280 nm. Emission spectra were recorded from 285 to 400 nm, and both excitation and emission width slits were of 5 nm. Background signals due to buffer and ligand absorptions were corrected. Curve fitting was performed in Origin 9.0 (OriginLab). The dissociation constant for one-site binding was calculated using the following equation:

$$\Delta F = \frac{B_{max}.C_{ligand}}{K_d + C_{ligand}}$$

Where ΔF is the quenching of fluorescence intensity, B_{max} is the maximum specific binding, C_{ligand} is the ligand concentration in micro molar units and K_d is the equilibrium dissociation constant, also choosing micromolar as the standard state.

2.2.9 *Protein crystallization*

X-ray crystallography is well known as the main method for the elucidation of three-dimensional structures of macromolecules. This method enables the visualization of protein structures at the atomic level, but requires the growth of large single crystals suitable for X-ray diffraction experiments. The first step is to purify the protein of interest in relatively large quantities (a few milligrams). At this point, the high purity and homogeneity of the sample are crucial for the crystallization to be successful.

The process of crystallization can be divided into two steps: (1) the nucleation process, and (2) crystal growth. With correctly selected conditions, crystal nucleation and growth can occur within the supersaturated regions of the phase diagram (Figure 13). Nucleation is the most difficult problem to overcome because it is a transition by which the molecules dispersed in solution adopt a well-ordered state. Probably this state is achieved after the formation of intermediates, that in last instance associate into small thermodynamically stable clusters with a repeating lattice, known as critical nuclei. Stable nuclei can form within the region of the phase diagram called labile zone. Once crystallization starts, the system moves toward equilibrium until it reaches the metastable zone where the crystal can growth[27].

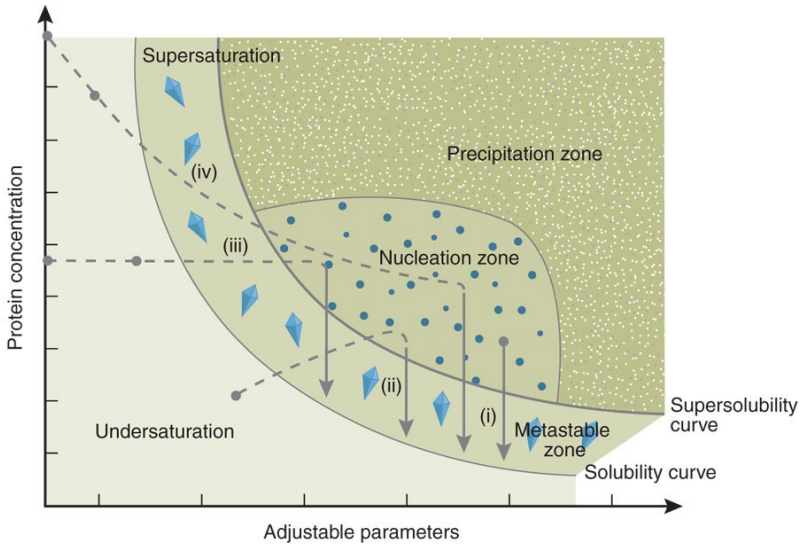


Figure 13. Crystallization phase diagram[28]. The transition from a stable solution to a supersaturated solution can be achieved by adjusting the position of the protein/precipitant mixture. Adjustable parameters include precipitant or additive concentration, pH and temperature.

For crystallization, the protein solution is first brought to a point of supersaturation, this is a non-equilibrium condition in which the concentration of the macromolecule is somewhat above its solubility limit. This point is achieved under specific chemical and physical conditions that combine the right pH, ionic strength, temperature, protein concentration, and the presence of various salts, ligands or additives. These conditions are practically impossible to predict in advance, thus crystallizing a protein is mainly a trial-and-error process of screening and optimization[29].

Crystallization screening was performed by vapor diffusion using the hanging drop technique in 24-well plates. Purified FGF1-WT and FGF1 mutants were dialyzed against buffer 3 (see Appendix) and concentrated to 5-10 mg/ml with Centricon devices (Millipore). Ninety-six different crystallization conditions were set up (Molecular Dimensions Limited) for each protein, combining different buffers and precipitant solutions. The crystallization plates were incubated at both 4°C and 20°C.

For each condition tested, one microlitre of the concentrated protein sample was pipetted onto a previously silanized cover slip, followed by an equal amount of the well solution (reservoir). The cover slip was then inverted over the well and sealed with immersion oil. Since the protein/precipitant mixture in the drop is less concentrated in the precipitant than the reservoir solution, water evaporates from the drop into the reservoir. As a result the concentration of both protein and precipitant in the drop slowly increase and thus crystals may form.

2.2.10 *Differential Scanning Calorimetry (DSC)*

Differential Scanning Calorimetry is a powerful technique for studying thermal transitions in biological systems. Using DSC it is possible to determine the transition midpoint (Tm) of these systems (described as the temperature at which 50% of the molecules are denatured) but also to obtain the thermodynamic parameters associated with these changes. For reversible processes, this technique allows the direct determination of the enthalpy and the heat capacity changes of the system of interest in a single experiment. During a DSC experiment, the temperature on both, sample and reference cells, is raised identically over time. Hence, the difference in thermal power required to maintain them at the same temperature is measured and plotted as a function of temperature. As the temperature increases, the protein sample eventually reaches its

denaturation temperature (Tm) and the denaturation process results in an endothermic peak in the DSC curve. Higher Tm values are representative of a more stable molecule[30].

The thermodynamic stability of FGF1 wild type and variants was compared using a Nano-DSC (TA Instruments). The heating scan rate was set to 0.5°C/min and the experiment was run in the temperature 4-80°C interval. The protein concentration used was 100 µM, and the reference cell was filled with the corresponding buffer. Heat capacity curves were analyzed using the Origin 9.0 software.

2.3 Results and Discussion

2.3.1 *Creation of expression constructs*

Following the first strategy proposed, the synthetic DNA of FGF1-*full mut* was correctly cloned into four different plasmids. The cloned gene contained the mutations needed to produce two T-antigen binding sites in FGF1. The expression constructs contained different tags and fusion partners to try the expression and purification of the recombinant protein employing several techniques.

In agreement with the second strategy, FGF1 *wild type* cDNA was successfully inserted into the NdeI/XhoI sites in pET15b, and mutations were incorporated one by time performing site-directed mutagenesis. Sequencing of DNA constructs verified that residues Y94 and N95 were mutated to alanines to eliminate the mitogenic activity of FGF1. Positions L26, H41 and I42 were mutated to arginine, aspartic acid and glutamine respectively to create a high affinity site for the T-antigen. Several difficulties were encountered in the introduction of new codons in the DNA sequence, but the addition of DMSO 5% in the PCR mix greatly enhanced the mutagenesis performance due to the inhibition of secondary structures in the DNA.

2.3.2 *Protein expression*

The expression constructs containing the synthetic DNA of FGF1-*full mut* did not show a significant amount of soluble protein in none of the conditions tested. Only the construct of pET15b transformed into the *E. coli* strain C41 produced the protein during expression at 37°C, but as inclusion bodies. Further purification trials were carried out, including denaturation with urea or guanidine and subsequent refolding with solubility enhancers but they turned out to be all unsuccessful. These results evidenced that these mutations introduced in FGF1 cause an irreversible loss of protein folding. In addition, the introduction of all the mutations at once made impossible to know which residues were essential for proper protein folding. In this sense, the second strategy for the engineering of FGF1 offered a possible solution in order to find the right balance between the introduction of a new binding site and the maintenance of protein structure.

Similarly, the constructs containing the cDNA of FGF1-WT and FGF1 mutants into pET15b were also transformed in different *E. coli* strains. Firstly, different temperatures of incubation (15, 20, 28 and 37°C) during induction with IPTG were tested in order to identify the best conditions for soluble protein expression and to proceed to scale protein production. Low temperatures have favored the production of soluble protein in every case, while ice shock before induction became critical for the expression of FGF1 mutants. Different compounds that improve protein solubility were added when poor protein quantities were found in the soluble fraction. Bacterial cultures were supplemented with sterile stock of these additives immediately prior to induction with IPTG. The best conditions for soluble expression of each protein are listed in the table below:

Protein	E. coli strain	Temperature of induction	Additives
WT	BL21	20°C	500 mM IPTG
2M	BL21	20°C	500 mM IPTG, glycerol 4%
3M	BL21	20°C	500 mM IPTG, glycerol 4%, arginine 50 mM
4M	BL21pCAM	15°C	250 mM IPTG, glycerol 5%, sucrose 1 M
5M	C41	15°C	250 mM IPTG, glycerol 5%

Table 1. Best conditions for soluble expression of each FGF1 variant in E. coli.

As can be inferred from the table, as the number of mutations increased, it was more difficult to obtain protein in the soluble fraction. Presumably, secondary structure is affected by the changed residues due to new and different interactions with neighbouring amino acids, then bacteria fail to produce the native-like conformation. The best results were obtained lowering the expression temperature and inductor concentration due to a decrease in the rate of protein synthesis that leads to the production of more soluble protein. Similarly, addition of glycerol and sucrose stabilize native protein structure indirectly by increasing the osmotic pressure which leads to the accumulation of osmoprotectants in the cell[31]. Whereas arginine improves protein stability during expression as it interacts with oppositely charged groups on the protein surface[32].

2.3.3 *Protein purification*

The best conditions for soluble expression of each protein was used to grow transformed cells in one litre of LB culture for protein purification. The presence of the protein of interest in the soluble fraction was checked by SDS-PAGE prior to proceeding with the purification process. After cell disruption, the supernatant was loaded onto a heparin-affinity column and the protein was separated from contaminants and finally eluted performing a NaCl gradient. Protein purity was confirmed by SDS-PAGE and protein concentration was determined spectrophotometrically. All this process was repeated for the purification of FGF1-WT and each mutant.

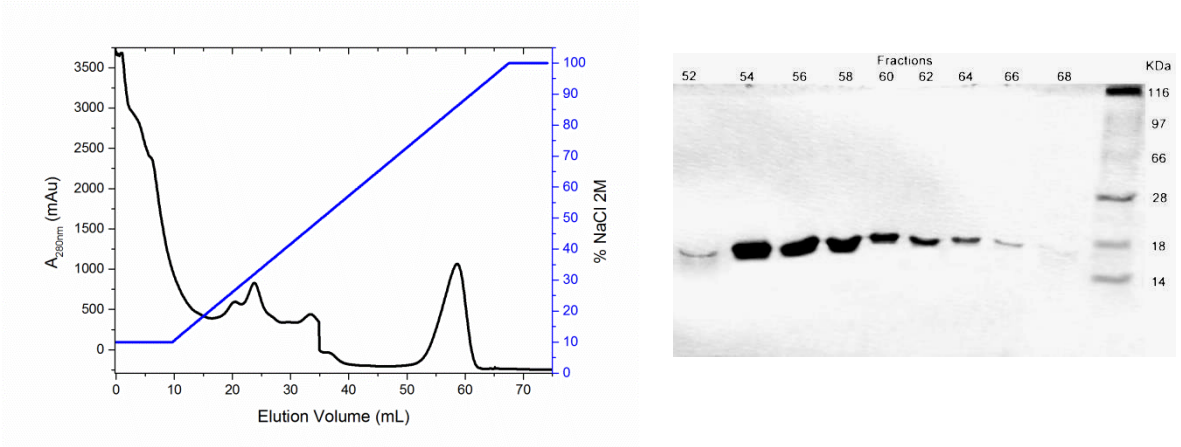
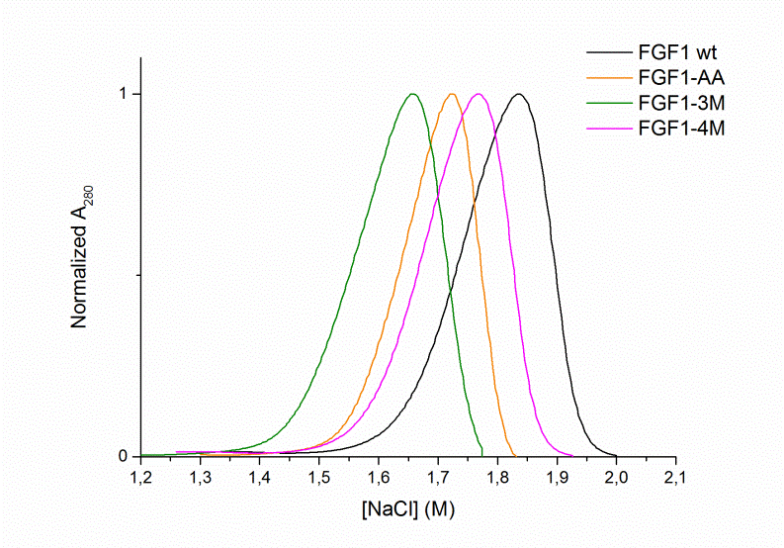


Figure 14. Example of an elution profile from the heparin column (left) and SDS-PAGE gel (right) obtained during the purification of FGF1-WT. The peak containing the protein without contaminants is observed between fractions 52 and 68.

The same elution profiles from the heparin column were obtained for all the FGF1 variants except for FGF1-5M. In this case, the protein seemed not to bind to heparin in the proper form, as its presence was evidenced in almost every fraction collected during the gradient elution. Therefore, it was not possible to isolate the pure protein using the heparin column. Purification trials performed for this protein using histidine-tag binding to a nickel column and an ionic exchange column were unsuccessful. This result suggests that mutation of Isoleucine51, a hydrophobic residue buried into the protein core, when substituted by glutamine, a polar amino acid instead, induces highly destabilizing interactions that lead to the loss of the overall native conformation.

On the contrary, FGF1-WT and its other variants were eluted from the heparin column with no significant differences (Figure 15, left), thus heparin affinity seemed not to be affected with these

mutations. However, protein expression decreased in large scale cultures when mutations were increased, with a concomitant decrease of yield (Figure 15, right). Moreover, FGF1 mutants were more susceptible to undergo degradation and aggregation. Hence, the use of EDTA (ethylenediaminetetraacetic acid) and BME (β-mercaptoethanol) during the experiments became indispensable to avoid protease activity and oxidation of sulfhydryl groups (respectively).



Protein	[NaCl] Elution Peak (M)	Yield (mg/l)
WT	1.85	16.5
2M	1.70	7
3M	1.63	5
4M	1.75	2

Figure 15. Elution peaks of FGF1-WT and its mutants obtained during purification with the heparin-column (left). The table summarizes the concentration of NaCl needed for the desorption of each protein and their final yield (right).

2.3.4 *Mitogenic activity on fibroblasts*

The proliferative activity of FGF1 on fibroblasts was studied in collaboration with the Pathology Laboratory in the Medicine Department of the University of Verona. FGF1-WT and FGF1-2M were added to fibroblast cultures at final concentration of 10 µg/ml. Cell viability was evaluated by a colorimetric assay based on the reduction of the tetrazolium salt WST by mitochondrial dehydrogenase of viable cells to a formazan dye. WST was added to each well and cells were incubated at 37°C for 90 minutes. The absorbance at 570 nm of the samples was measured every 30 minutes using a multiwell-plate reader.

The results obtained for the samples containing the mutated proteins were confronted with those obtained for untreated cells. The effect of FGF1 on fibroblast was evident after 30 minutes of

growth. As expected, the addition of FGF1-WT into the culture stimulated fibroblast proliferation, as cells grew faster than the control group. Under the same conditions, FGF1-2M did not to show this effect. In the wells where the mutant protein was added, cells grew approximately at the same rate as the control. These results indicate that the double mutation made on FGF1 effectively caused a decrease in its mitogenic activity.

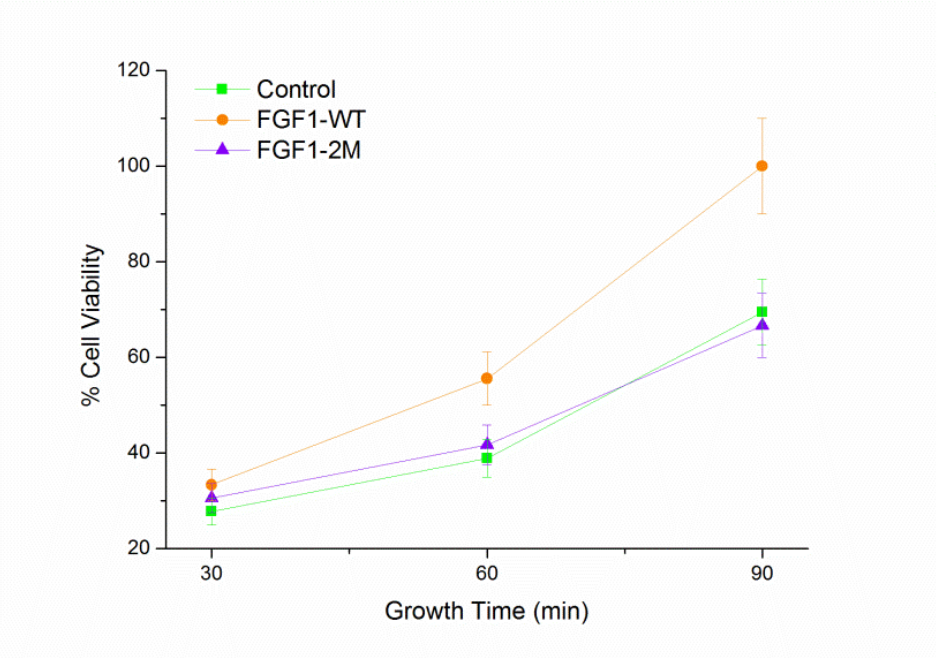


Figure 16. The plot represents the percentage of cell viability (with respect to the control) as a function of growth time.

2.3.5 *Intrinsic fluorescence analysis*

Changes in the structural features of FGF1 and its mutants were studied by spectrofluorimetry. Intrinsic fluorescence emission spectra of these proteins were recorded after excitation at 280 nm. The results showed that, as has already been reported, FGF1 presents an unusual fluorescence emission dominated by tyrosines despite the presence of a tryptophan residue in the protein. Seven of the eight tyrosines contained in the protein are solvent exposed, while its single tryptophan is located in a loop partially inaccessible to solvent. The uncommon emission spectrum would be due to the strong decrease in tryptophan fluorescence caused by the positive side chains of two basic residues (His102 and Lys105) in close proximity to the indole ring stabilizing the ground state[33].

The figure below shows the emission spectra of FGF1 native and its variants. No significant differences were observed between them, suggesting that mutants maintained the overall conformation of the native protein. All spectra exhibited a single peak at 303 nm, characteristic of multiple tyrosine fluorescence. The extended spectral tail in the 320-370 nm region revealed no emission of tryptophan fluorescence.

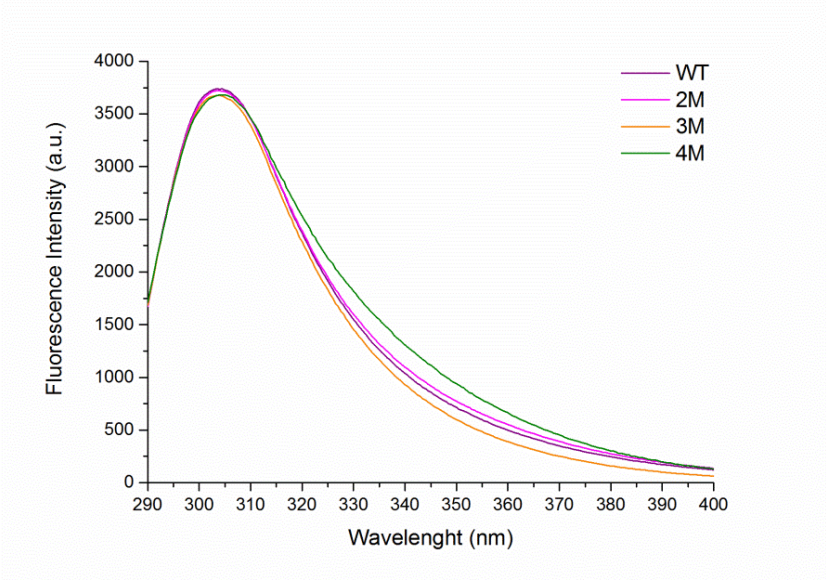


Figure 17. Emission spectra of FGF1-WT and its mutants recorded after excitation at 280 nm.

When proteins interact with other compounds, their intrinsic fluorescence often changes with the concentration of the ligand. Thus, the interactions of FGF1-WT and its mutants with galactose and N-acetylgalactosamine were studied by intrinsic fluorescence quenching. Emission spectra were recorded during titrations after excitation at 280 nm. The table below lists the Kd values calculated from nonlinear regression fitting assuming a single binding site. These data showed that FGF1-WT exhibited a certain affinity for Gal and GalNAc. There is a strong chance that, as acidic molecules, at low ionic strength they were able to bind non-specifically to basic patches on the protein surface. Similarly, FGF1-2M and FGF1-3M presented Kd values not significantly different from those of the wild type. However, the Kd value calculated for FGF1-4M was significantly lower, which means that the inclusion of key amino acids in FGF1 for the binding to these carbohydrates, effectively increased the affinity.

	Kd Gal (µM)	Kd GalNAc (µM)
WT	2.03 ± 0.17	2.82 ± 0.19
2M	3.07 ± 0.33	2.52 ± 0.17
3M	3.20 ± 0.39	2.68 ± 0.34
4M	0.58 ± 0.05	0.47 ± 0.04

Table 2. Dissociation constant values (Kd) calculated from nonlinear regression fitting assuming a single binding site.

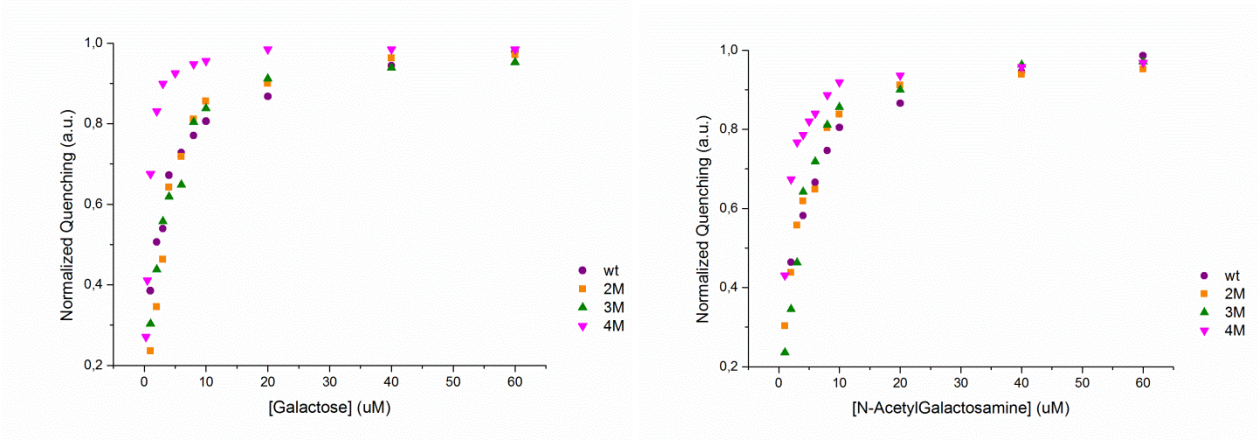


Figure 18. The graphics show the fluorescence quenching for galactose (left) and N-acetylgalactosamine (right) obtained by fluorimetric titrations.

A curious aspect was observed in the emission spectrum of FGF1-4M. As the concentration of the ligands was raised, the 303 nm emission peak strongly decreased and a shoulder near 320-350 nm region began to appear. Deconvolution of the curve evidenced a characteristic tryptophan peak centred at 330 nm, previously quenched and hidden in the long tail of the tyrosines peak (Figure 19).

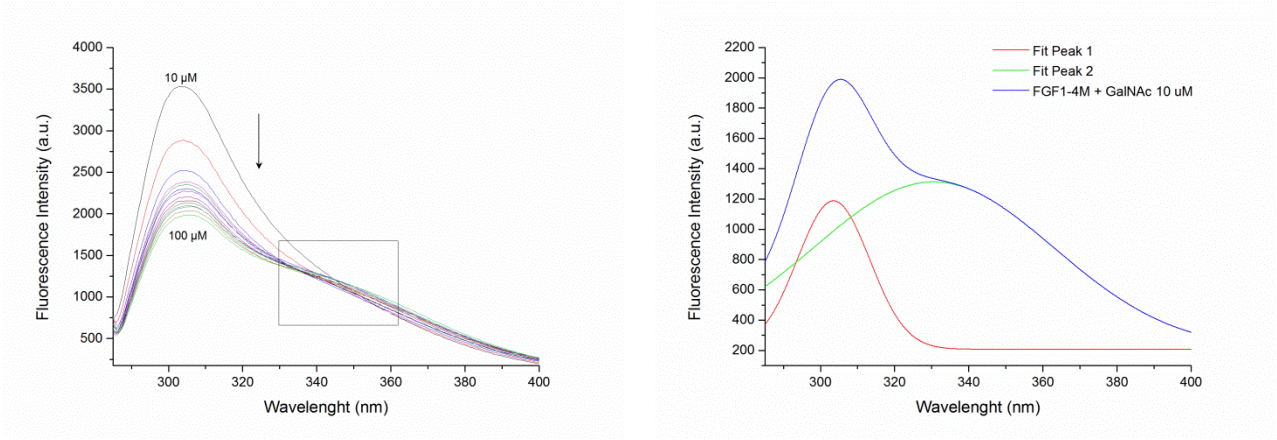


Figure 19. Emission spectra obtained during titrations with GalNAc on FGF1-4M, the shoulder around 330 nm is evidenced (left). Deconvolution of the curve obtained with 10 µM of GalNAc reveals a tryptophan peak (right).

Since tryptophan fluorescence in the properly folded protein is completely quenched, the signal of this residue could indicate a structural rearrangement in FGF1-4M due to carbohydrate binding. The emission peak at 330 nm suggests a tryptophan residue in an apolar environment. Then, a less rigid conformation of the protein could be adopted that moves this residue out of the protein hydrophobic core, where the groups around the indol ring do not suppress its fluorescence emission. This effect was not seen in FGF1-WT or in any of the others mutants and thus it evidenced that insertion of a new residue in position 41 was not only significant for carbohydrate binding but also affected the tryptophan environment.

2.3.6 *Protein crystallization*

Attempts to obtain crystals of FGF1-WT and its mutants were performed to characterize their three-dimensional structure and their affinity for the T-antigen. After purification, proteins were concentrated up to 5-10 mg/ml and crystallization screenings were set up. A total of 96 different crystallization conditions were tested for each protein including different salts, alcohols and organic polymers as precipitants. After two weeks of incubation at 4°C, crystals of FGF1-WT were obtained in the following conditions: i. 10 mM sodium phosphate monobasic / 20 mM ammonium sulphate / PEG 1500 10%, ii. 0.1 M HEPES pH 7.5 / 1.5 M lithium sulphate monohydrate, and iii. 0.1 M sodium acetate trihydrate / 0.2 M sodium chloride.

Since the crystals were too small to perform X-ray experiments, trials in these conditions were repeated using protein samples with different concentrations and incubating sample plates also at 20°C. However, no crystals appeared in any of the conditions tried.

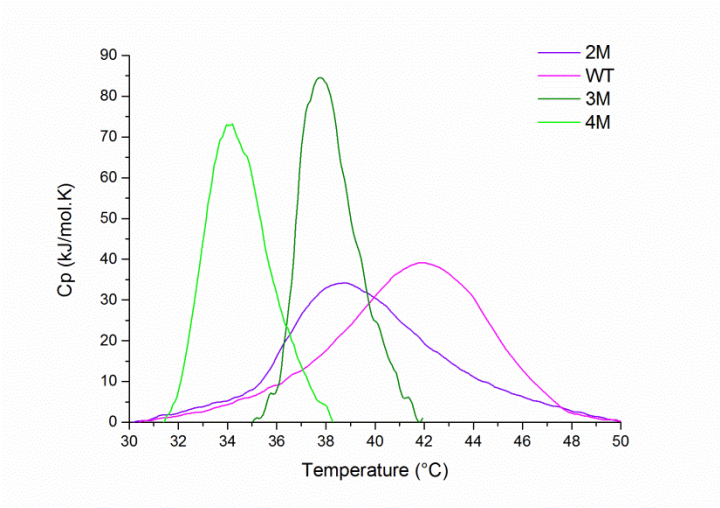
Experiments performed using FGF1 mutants did not produce crystal nuclei. They were more likely to form amorphous precipitates. Then, new tests were made lowering the protein concentration and incubating the sample plates at 4°C in order to move the system more slowly towards the labile region. Still none of the conditions showed positive results.

Furthermore, protein aggregation during concentration and in the crystallization samples was inevitable, even after the addition of reducing agents like dithiotreitol or β-mercaptoethanol.

These results indicated that FGF1 mutants failed to cross the threshold from a disordered supersaturated solution to a more rigid state. This might be because they adopt a less stable structure than native FGF1. Consequently, by increasing the number of molecules in solution, protein-protein contacts are promoted in a disordered manner leading to aggregation.

2.3.7 *Protein thermal stability analysis*

For the rational design of engineered proteins and protein therapeutics it is necessary to study the thermodynamic properties that are correlated to stability. The thermal stability of FGF1-WT and variants was thus analyzed by Differential Scanning Calorimetry (DSC). The protein concentration used was 100 µM and the scan rate was set at 0.5°C per minute. Data were analyzed by subtraction of the reference, normalization to the protein concentration and DSC cell volume, and interpolation of a polynomial baseline. The impact of mutations on the conformational stability of FGF1 was studied by comparing the temperature midpoint (Tm) of each FGF1 variant with that of FGF1-WT (Figure 20).

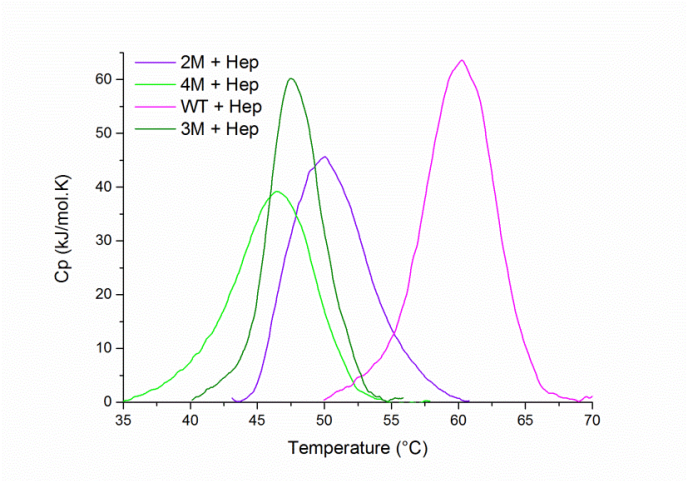


FGF1	Tm (°C)	Tm_{WT} - Tm (°C)
WT	41.97	-
2M	38.78	3.19
3M	37.77	4.20
4M	34.20	7.77

Figure 20. The figure shows the endotherms obtained after the thermal unfolding of the proteins, while the table summarizes the midpoint temperatures observed in each case and the differences with respect to FGF1-WT.

The endotherms show a decrease in Tm as the number of mutations on FGF1 was increased. These results demonstrated that mutations made on FGF1 destabilize the overall protein conformation. The substitution of residues N94 and Y95 by alanines lowered the thermal stability by 3.19°C, while substitution of L26 by an arginine seems to be less significant since it decreased it only 1.01°C with respect to the double mutant. Taken together, these results suggest that the most critical mutation for the maintenance of a native-like conformation was the change of residue H41 by an aspartic acid, destabilizing the protein in 3.57°C with respect to the triple mutant and in a total of 7.77°C with respect to the *wild type*.

There have been several reports concerning improvement of the stability of FGF1 at physiological temperatures by the addition of heparin[33,34]. This polyanion is able to increase long-term stability of the protein and the maintenance of its structure up to 20°C higher. Therefore, changes in the stability of FGF1 mutants due to heparin binding were studied. In this context, the variation of Tm of each heparin-bound protein relative to the apo form was analyzed (Figure 21).



FGF1	T_m (°C)	T_{mHOLO} - T_{mAPO} (°C)
WT	60.23	18.96
2M	50.05	11.27
3M	47.45	9.68
4M	46.45	12.25

Figure 21. The figure shows the endotherms obtained after the thermal unfolding of the protein solutions containing a three-fold excess of heparin, while in the table are shown the midpoint temperatures calculated in each case and the differences observed with respect to midpoint temperatures without the heparin.

Indeed, addition of a three-fold excess of heparin in FGF1-WT solution resulted in an enormous stabilization of the protein with an increase of its Tm of 18.96°C. Analogously, heparin binding to FGF1 mutants produced a more stable protein conformation. This effect was particularly evidenced for FGF1-4M, which increased its Tm in 12.25°C, several degrees above the values of all the other FGF1 mutants.

Although significantly increased stability upon heparin binding was observed, still the denaturation of these proteins occurred near physiological conditions. Unfortunately, the stabilization by heparin did not counteract the destabilization caused by the mutations. Moreover, no signal was observed after down scanning experiments and visible precipitated samples were found after thermal denaturation within the DSC cell. These observations indicated that the denaturation process leads to the irreversible aggregation of the proteins, imposing a problem on their use as a therapeutic agents.

2.4 Concluding Remarks

Lectin-carbohydrate binding is very strong and specific, therefore it is difficult to recreate the same interactions in another protein even in the presence of a high structural similarity between the two. The introduction of all the necessary mutations at the same time leads to the inability to evaluate the balance between the gain of the new function and the stability of the protein, which becomes then possible when performing the mutations by steps. Nevertheless, the current data indicate that designing mutations in FGF1 to create a T-antigen binding site is remarkably complex. Although one FGF1 form with an apparent improvement in the binding constant to the carbohydrate antigen was generated, the introduction of new residues led to the destabilization of the molecule.

Protein-based therapies require high concentration formulations for disease treatment which presents a major problem in using proteins that tend to aggregate. Misfolded proteins have reduced activity and could elicit an immunological response, not being suitable for delivery methods. However, stabilization of therapeutic proteins is generally performed during the development phase using trial-and-error methods. Thus, an understanding of the fundamental mechanisms and regions involved in protein aggregation is important both for stabilizing the protein and for devising strategies to prevent *in vivo* aggregation.

FGF1 contains three free cysteine residues buried into the protein core that can become exposed and form intramolecular thiol adducts which drives to irreversible protein unfolding. Consequently, these buried-free cysteines play a key role in regulating the functional half-life of FGF1, and their mutation can result in a significant improvement in the thermostability and pharmacokinetic properties in protein-based therapeutics.

Recently a cysteine-free form of FGF1 has been described with enhanced thermal stability and *in vivo* increased half-life. By suppressing the cysteine residues buried in the core of the macromolecule it is possible to produce a novel FGF1 with more desirable biophysical properties[35]. This finding could offer a solution to the problems encountered in the use of this protein as a therapeutic agent.

Chapter III

Production of a Truncated Form of GalNAc-T6 for Tn-antigen Binding

3.1 Background

3.1.1 *N-acetylgalactosaminyl-transferases*

The N-acetylgalactosaminyl-transferases (GalNAc-Ts) conform a large family of enzymes present in many species, including humans, that initiates mucin-type *O*-glycosylation. Some evidence indicates that GalNAc-Ts could play an important role in the expression of the immature truncated Tn and STn O-glycans during cancer development[36]. The family includes up to 20 distinct isoforms localized throughout the Golgi apparatus, where they are responsible of the first step in the formation of glycoproteins passing through the secretory pathway. GalNAc-Ts catalyze the addition of a GalNAc moiety from activated donors to the hydroxyl group of serine or threonine residues (GalNAc1-O-Ser/Thr), yielding the Tn-antigen. The different isoforms exhibit selective substrate specificities for these glycopeptides, which may indicate that O-glycosylation involves several successive biosynthetic steps catalyzed by these isoforms. Subsequent elongation, branching and capping of the Tn structure is catalyzed by a large number of different glycosyltransferases in further processing steps[4].

GalNAc-tranferases are type-II transmembrane proteins containing a short cytoplasmic tail, a single transmembrane region, and a stem region followed by two separate structures: one catalytic and one lectin domain. At the N-terminal region, the catalytic domain folds into A type fold, while in the C-terminal region, the lectin domain, possesses a ricin-like (β-trefoil) fold, with three putative carbohydrate-binding sites[37].

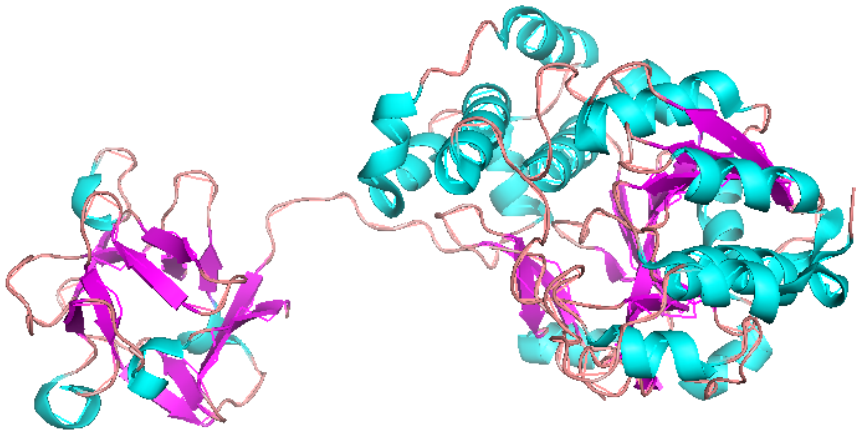


Figure 22. Example of GalNAc-transferases 3D structure (based on PyMOL rendering of PDB 2FFU).

3.1.2 *Isoform GalNAc-T6*

Even if the understanding of the molecular mechanisms for mucin glycosylation changes in human tumors is still limited, one of these changes would be the differential expression of GalNAc-Ts that leads to the production of glycoproteins containing only earlier biosynthetic O-glycan intermediates[38]. Immunohistochemical studies have shown an altered expression of some GalNAc-Ts in the malignant transformation of cells. In fact, isoform GalNAc-T6 was found to be highly over-expressed in human breast cancer cell lines and in the 81% of breast tumors, but not in normal breast epithelium. Due to this finding, GalNAc-T6 was proposed as a specific marker for molecular diagnosis of tumor cells[39].

Furthermore, this member of the GalNAc-tranferases aroused a particular interest as its lectin domain presents certain sequence homology with BEL β-trefoil. The α site of the fungal lectin and residues 94 to 126 of GalNAc-T6 share a sequence identity of 21.2%. This suggests a possible similar mechanism in the carbohydrate binding site where the Tn antigen is recognized.

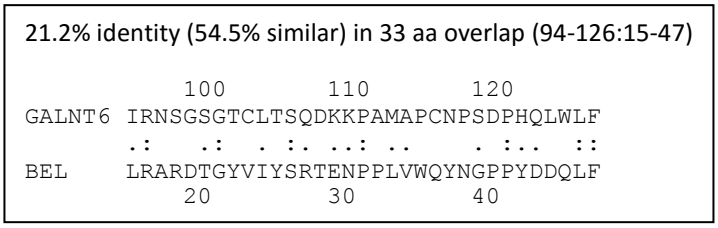


Figure 23. Local sequence homology between GalNAc-T6 and BEL β-trefoil calculated with LALIGN server.

Exploiting the carbohydrate binding site of GalNAc-T6 but suppressing its catalytic activity should result in a good candidate for drug delivery in anti-tumor therapies. The use of a human lectin as a drug carrier would entail great advantages since such protein is expected to have a high affinity for tumor cells and a low or negligible immunogenicity. The production of a truncated form of GalNAc-T6 containing only the lectin domain is thus of utmost interest.

The three-dimensional structure of GalNAc-T6 has not been determined so far, neither has been its substrate specificity. Therefore, the production of a recombinant form containing only the lectin domain can contribute to these two critical points that need to be considered to evaluate its possible value in cancer therapies.

3.2 Materials and Methods

3.2.1 *Designing the expression construct*

The design of the expression construct to be cloned is the most crucial step when it is planned to express just a part of a bigger protein. Usually, only complete domains are expressed and, unless needed for functional reasons, the expressed protein should lack any flexible tails. As domain boundaries can be difficult to define, 3D-structure prediction techniques are useful to derive the necessary information. The homology modelling program Swiss-Model was therefore employed to generate a comparative 3D-structure model of GalNAc-T6 (UniProt Accession N° Q8NCL4).

The 3D-structure data of homologous proteins was used as a guide to define the domain boundaries in order to produce a truncated novel protein containing only the lectin domain of GalNAc-T6 (T6L). The most appropriate template for homology modelling was identified with BLAST[40] and HHBlits[41] against the SWISS-MODEL template library (SMTL version 2016-12-21, PDB release 2016-12-16). The available structures of homologous proteins were referred and models were built based on the target-template alignment using ProMod3[42]. The global and per-residue model quality was assessed using the QMEAN scoring function[43].

3.2.2 *Cloning GalNAc-T6 Lectin domain (T6L) and analysis of transformants in E. coli*

The part of the gene encoding only the lectin domain (D492-V622) was PCR amplified from cDNA of human GalNAc-T6 (MGC:46032 IMAGE:5552059). The pair of primers used were: 5'-GGTGGTGAATTCGACCTGACGCCCACC-3' containing ECORI site (upstream) and 5'-AATAATGTCGACGACAAAGAGCCACAA-3' containing SalI site (downstream). After enzymatic digestion, the PCR fragment was ligated into the respective sites of pPICZαA under the control of the methanol inducible alcohol oxidase 1 promoter (AOX1) and fused in-frame with the α-factor secretion signal peptide. The ligation mixture was transformed into XL1-Blue (*E. coli* strain) and selected on low salt LB medium with 100 µg/ml zeocin. The presence of the appropriate insert was determined by colony PCR and restriction analysis using Xba site. The plasmid pPICZαA–T6L was isolated from the bacterial cells using the plasmid extraction kit (Sigma Aldrich) and then analyzed by DNA sequencing. Then, the isolated plasmid was used to transform *Pichia pastoris* X33 and KM71H competent cells for protein expression.

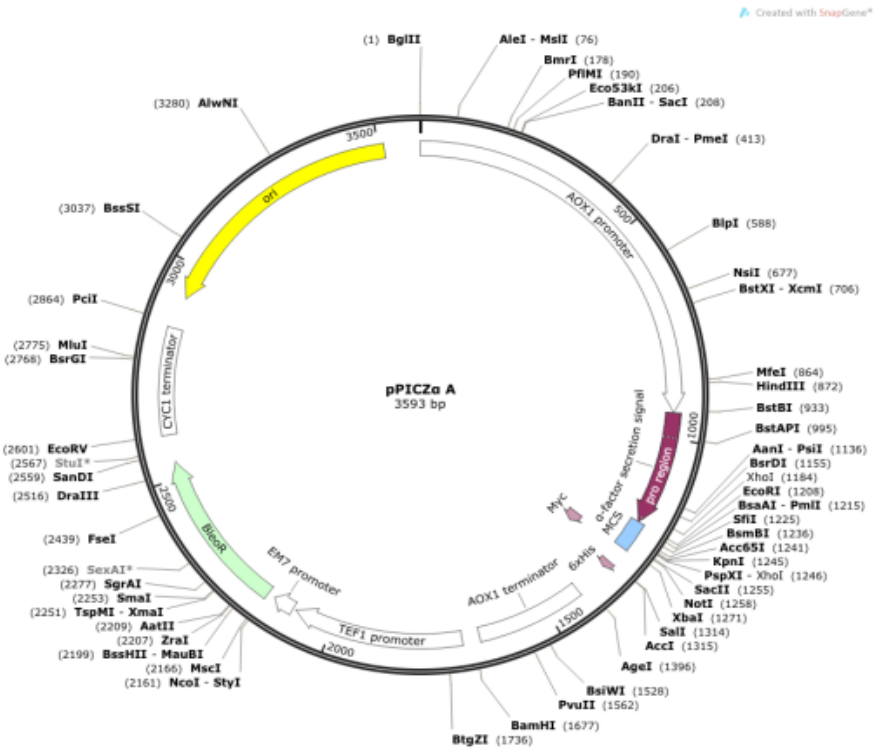


Figure 24. pPICZαA map.

pPICZαA is a 3.6 kb vector used to express and secrete recombinant proteins in *Pichia pastoris*. Recombinant proteins are expressed as fusions to an N-terminal peptide encoding the *Saccharomyces cerevisiae* α-factor secretion signal. The AOX1 promoter regulates methanol-induced expression of the gene of interest. The zeocin resistance gene is useful for the selection in both *E. coli* and *Pichia*. C-terminal peptide containing the c-myc epitope and a polyhistidine (6xHis) tag for detection and purification of a recombinant fusion protein.

3.2.3 *Preparation of competent Pichia cells*

Two different phenotypes of *Pichia pastoris* were used to evaluate which of them would be better for protein expression. X33 is a *wild-type* strain that has the ability to metabolize methanol as the sole carbon source (Mut^+ phenotype: Methanol utilization plus). Conversely, KM71H has a mutant aox1 locus generating a Mut^S phenotype ("Methanol utilization slow"), caused by the loss of alcohol oxidase activity encoded by the AOX1 gene.

X33 and KM71H cells were grown overnight in 5 ml of YPD (see Appendix) in a 30°C shaking incubator. The next day, cultures were diluted to an A_{600} of 0.15–0.20 in 50 ml YPD using 200 ml sterile flasks. Yeast were grown to an A_{600} of 0.8–1.0 in a 30°C shaking incubator (approximately 4 hours based on a generation time of 100-120 minutes). Cultures were centrifuged at 500×g for 5 minutes at room temperature and supernatants were discarded. Pellets were resuspended in 9 ml of ice-cold BEDS solution (see Appendix) supplemented with 1 ml 1.0 M dithiothreitol (DTT). Cell suspensions were incubated for 5 minutes at 100 rpm in the 30°C shaking incubator. Cultures were centrifuged again at 500×g for 5 minutes at room temperature and cells were resuspended in 1 ml of BEDS solution without DTT. Aliquots were made into individual 1.5 ml tubes and placed at −80°C until needed.

3.2.4 *Transformation of P. pastoris and selection of high-level expression colonies*

The DNA construct pPICZαA–T6L was digested with SacI for 2 hours at 37°C and purification of the linearized plasmid was performed using a Gel Extraction Kit (Sigma Aldrich). Four microlitre (about 40 ng) of linearized DNA were mixed with 40 µl of cells in an electroporation cuvette and incubated for 2 minutes on ice. Recovery medium (0.5 ml sorbitol 1M + 0.5 ml YP) was immediately added after a pulse shock at 1.5 kV. Cells were incubated for 3 hours before plating on YPDS agar supplemented with 100 µg/ml zeocin. Transformants were grown after 4 days at 28°C. Further selection of multicopy integrants was carried out plating transformants on media containing increasing concentrations of zeocin (100, 200, 300, 400 and 500 µg/ml).

3.2.5 *Determining the Mut phenotype*

Transformation of X33 or KM71H with the linearized DNA construct favour single crossover recombination at the AOX1 locus. For the *wild-type* strain X33, recombination in the 5´AOX1 region will yield Mut^+ transformants, but there is a chance that recombination occurs in the 3´AOX1 region also, disrupting the wild-type AOX1 gene and creating Mut^S transformants. While transformation of the mutant strain KM71H will yield only Mut^S transformants. Both types of phenotypes could be useful for protein expression, therefore to confirm expected phenotypes, a test on MD and MM plates (see Appendix) was carried out. Each transformant was picked and streaked on both media. Plates were analyzed after incubation for two days at 28°C.

3.2.6 *PCR analysis of Pichia integrants*

Direct PCR screening of *Pichia* clones was performed to determine if the gene of interest was integrated into the *Pichia* genome. In this method the cells are lysed by a combined treatment that includes enzyme digestion, freezing and heating. The genomic DNA can be used directly as a PCR template.

One single colony was picked and resuspended in 10 µl of water. A 5 U/µl lyticase solution was added and the digestion mix was incubated at 37°C for 10 minutes. After incubation, the sample was frozen at -80°C for another 10 minutes. Then, the PCR mix was prepared as follows: buffer TAQ Polymerase 1X, 1 mM DNTPs, 1 µM of both primers, 0.75 mM $MgCl_2$, 5 U/µl TAQ Polymerase, and sterile water up to 20 µl.

The polymerase reaction was set up using the following parameters:

1° initialization	95°C	3 minutes
2° denaturation	95°C	30 seconds
3° annealing	55°C	30 seconds
4° elongation	72°C	90 seconds
5° final elongation	72°C	5 minutes

The steps 2°, 3° and 4° were repeated for 35 cycles. After temperature cycling, 6 µl of loading buffer were added to samples and they were analyzed by agarose gel electrophoresis.

3.2.7 *Expression of the recombinant protein in Pichia strains*

To determine the optimal method and expression conditions to produce T6L in *Pichia pastoris*, small-scale expression conditions were tested on both strains, X33 and KM71H. A single colony was inoculated in 5 ml of BMGY (see Appendix) plus zeocin 100 µg/ml. It was grown at 30°C in a shaker incubator until the culture reached an OD_{600} = 2-6 (approximately 20 hours). Then, cells were harvested at 3000xg for 5 minutes and the pellet was resuspended to an OD_{600}=1 in BMMY medium (see Appendix) to induce expression. Cultures were returned to the incubator to continue growth for another 96 hours. To maintain induction, 100% methanol was added to cultures to a final concentration of 0.5% every 24 hours. During the induction period, 1 ml of culture was taken

at different times and the supernatant was separated from the cells by centrifugation at maximum speed. Samples were conserved at -20°C until the secreted expression was analyzed by western blot.

3.2.8 *Expression scale-up and protein purification*

Fresh colonies were grown from glycerol stocks on YPD-agar supplemented with 100 µg/ml zeocin for 48 hours. Next, they were inoculated in 50 ml of BMGY and grown for another 24 hours in a shaker incubator set at 30°C. This 50 ml culture was used to inoculate 250 ml of BMGY in a 3 litre flask and yeasts were grown again at 30°C until OD_{600} = 2-6 (approximately 24 hours). Cells were harvested by centrifuging at 3000xg in sterile bottles and resuspended to an OD_{600}=1 in BMMY medium (see Appendix) to start induction. The culture was incubated for 96 hours at 28°C with vigorous shaking. One hundred percent methanol was added daily to a final concentration of 0.5% to maintain induction. After expression, cells were harvested by centrifugation at 5000xg for 30 minutes and the supernatant was first cleared using a 0.22 µm filter and then saved to proceed with protein purification.

In order to concentrate the secreted proteins, 90% ammonium sulphate w/v was added to the supernatant and proteins were precipitated overnight at 4°C. The protein fraction was recovered by centrifugation at 17 000xg for one hour. The pellet was resuspended in buffer 4 (see Appendix) and ammonium sulphate was removed by exhaustive dialysis against the same buffer.

The sample was loaded onto a CM-cellulose column (GE Healthcare) previously equilibrated in buffer 4. T6L was separated from contaminants in a step elution with NaCl. Firstly, the column was washed with buffer 4 plus 100 mM NaCl until the absorbance at 280 nm was negligible. Secondly, T6L was eluted with buffer 4 plus 500 mM NaCl. Protein purity was confirmed by SDS-PAGE.

For crystallization purposes, further purification steps were carried out to obtain a highly homogeneous solution. First, the fraction containing T6L was loaded onto a Lipidex column (Sigma Aldrich) to eliminate hydrophobic compounds. The protein was eluted from this column using buffer 4 and monitoring the absorbance at 280 nm. Then, the eluted fraction was concentrated by ultrafiltration with a Centricon device (Mllipore) up to 2 ml, and loaded onto Sephacryl S-100 HR

size exclusion column (GE Healthcare). A FPLC system (Äkta Prime) was used for automatic elution in buffer 4 at 0.6 ml/min flow rate and collecting 2 ml per fraction.

An alternative large-scale expression was carried out growing X33 cells as described above but performing methanol induction with the minimal buffered medium BMM (see Appendix) at two different temperatures, 20°C and 25°C. In this case, the secreted protein was purified employing only the CM-cellulose column.

3.2.9 *Intrinsic fluorescence*

The fluorescence emission of aromatic residues is sensitive to their chemical environment and can provide qualitative information on the changes in protein conformation and precise information on protein-ligand interactions[26]. This method was useful to analyze local structure changes in T6L after additions of N-acetilgalactosamine (GalNAc) and to calculate the dissociation constant of this binding event.

For spectrofluorimetric experiments, one ml of 1 µM T6L in buffer 5 (see Appendix) was incubated at room temperature in a quartz-cuvette. A solution of GalNAc prepared in the same buffer was added to a final concentration from 0.1 to 30 µM. After incubation for 30 seconds, samples were excited at 280 nm and emissions were measured from 290 to 400 nm. All spectra were obtained using a Jasco FP-8200 spectrofluorimeter, with both excitation and emission slits of 5 nm. The recorded spectra were corrected for buffer emission and data were plotted using the Origin 9.0 (OriginLab) software for curve-fit analysis. One-site binding was considered to calculate the dissociation constant using the Michaelis-Menten equation:

$$\Delta F = \frac{B_{max} . C_{ligand}}{K_d + C_{ligand}}$$

Where ΔF is the quenching of fluorescence intensity, B_{max} is the maximum specific binding, C_{ligand} is the ligand concentration in micro molar units and K_d is the equilibrium dissociation constant.

3.2.10 *Isothermal Titration Calorimetry (ITC)*

Isothermal titration calorimetry (ITC) is a useful technique to study biological interactions in solution. Its application in protein-ligand interactions allows the precise determination of affinity, thermodynamic parameters and the stochiometry of the binding reaction. In a ITC experiment, the ligand contained in the injection syringe is gradually titrated into the protein solution placed in the ITC cell, under isothermal conditions. As injections proceed, the heat released or absorbed upon binding reaction can be measured. When all the binding sites in the protein become saturated, the heat signal approaches zero. Fitting of the isotherm leads to the calculation of enthalpy (ΔH), entropy (ΔS), binding constant (Kd), and stoichiometry of binding (n) from one single experiment. For accurate measurements both reactants must be in the same buffer and at known concentrations[44].

The experiments were performed at 25°C with a stirring speed of 250 rpm using a Nano- ITC (TA Instruments). The sample cell was filled with 50 µM of T6L solution in buffer 4. The same buffer was placed in the reference cell, while the syringe was loaded with a solution of 1 mM GalNAc. After setting up the titration experiment, a series of 2 µl of the carbohydrate solution were injected into the protein solution with intervals of 300 seconds. The values for the thermodynamic parameters were calculated from data analysis with the NanoAnalyze Software (TA Instruments). The independent fitting model was used to determine the binding constant (Ka) and the variation of enthalpy associated with binding (ΔH). The Ka and ΔH values were then used to calculate the variation in free energy (ΔG) and entropy (ΔS) using the following classical equations:

$$\Delta G^\circ = -RT \ln Ka \quad (1)$$

$$\Delta G = \Delta H - T\Delta S \quad (2)$$

The dissociation constants were obtained from the fitted Ka using:

$$Kd = 1/Ka \quad (3)$$

3.2.11 *Differential Scanning Calorimetry (DSC)*

Differential scanning calorimetry was used to measure the heat capacity (Cp) of T6L solutions. Since changes in the heat capacity are associated to folding/unfolding transitions, the protein was thermally unfolded while heating at a constant rate, giving rise to an endothermic peak in the Cp vs temperature curve. From the maximum of the peak the transition midpoint temperature (Tm), was determined which can be correlated to protein stability[30].

The protein was concentrated up to 100 µM in buffer 5 and 300 µl of this solution were used to fill the sample cell, while the reference cell was loaded with the same buffer. A Nano-DSC (TA Instruments) was used to perform a heating scan from 10 to 100°C at 0.25°C/min scan rate. After running the experiments, the data were analyzed with the NanoAnalyze Software (TA Instruments).

3.2.12 *Protein crystallization*

The production of high-quality crystals is the major obstacle in determining the three-dimensional structure of macromolecules by X-ray crystallography. Optimal conditions for growing high-quality protein crystals are extremely difficult to predict and usually must be empirically determined[29]. As the study of the tridimensional structure of T6L requires single crystals suitable for X-ray experiments, a total of 192 conditions were tested including three different types of precipitants: inorganic salts, organic solvents and polyethylene glycol.

The purified protein was concentrated to 10 mg/ml in buffer 4 and used to set up the crystallization screenings. These trials were performed using the precipitant solutions contained in Structure Screens 1 and 2 (Molecular Dimensions Limited) and MD-21 and MD-25 (MemSysTM). The screenings were set up manually, using 24-well plates (MultiwellTM, Falcon®) and adopting the hanging drop technique. A droplet (2 µl) containing the protein/precipitant mixture was deposited onto a glass cover slip, and then closed to allow equilibration with the reservoir containing 300 µl of the same precipitant solution. The screenings were incubated at 20°C for several weeks, during which crystal formation was controlled using an optical microscope.

3.2.13 *X-ray data collection*

The aim of macromolecular X-ray crystallography is to describe the three-dimensional structure of a protein at the atomic level. To reach this purpose, the protein crystal is irradiated with a finely focused monochromatic beam of X-rays producing a diffraction pattern of regularly spaced spots. The intensities of these reflections are directly correlated with the structure factor amplitudes of the diffracted beam. But as the structure factor phases cannot be directly calculated, several methods available can be applied to produce a three-dimensional electron density map of the molecules in the crystal[45].

The diffraction experiments of the T6L crystals were carried out at beamlines ID-30B and BM-30A of the European Synchrotron Radiation Facility (ESRF, Grenoble, France) using Pilatus3 6M and ADSC Quantum 315r detectors (respectively). The advantage of the use of synchrotron radiation is the production of a high photon flux, making possible high-resolution X-ray diffraction experiments.

For X-ray exposure, crystals were picked up from wells using a loop, then soaked for a few seconds into a cryoprotectant solution made up by mixing 70% of reservoir solution with 30% glycerol, and finally frozen at 100 K. Once on the beamline, crystals were mounted in a goniostat using an automatic sample changer and cooled in a cryogenic nitrogen-gas stream that prevents crystal damage. The data collected was indexed and integrated with iMOSFLM and then reduced using Scala (both from the CCP4 package software).

3.3 Results and Discussion

3.3.1 *Designing and cloning the expression construct*

To initiate the evaluation of GalNAc-T6 as a potential candidate as a therapeutic agent, it was necessary to produce a recombinant form of the protein containing only the lectin domain (T6L). As the 3D-structure of the protein has not been determined, homology modelling was performed with the SWISS MODEL server, which allowed the identification of the secondary structure elements. Using BLAST analysis, GalNAc-T2 was identified as the most appropriate template to perform homology modelling, a sequence identity of 43.66% was calculated (PDB entry 4d0z.3.A, resolution=2.20 Å, R value=0.23).

Considering that sequence conservation (especially of the hydrophobic residues) is higher in the structured domains (β-strands and α-helix), the isoform GalNAc-T2 was used as search probe. This allowed to detect the zone of the sequence where there were less risks of compromising global conformation when producing a truncated protein.

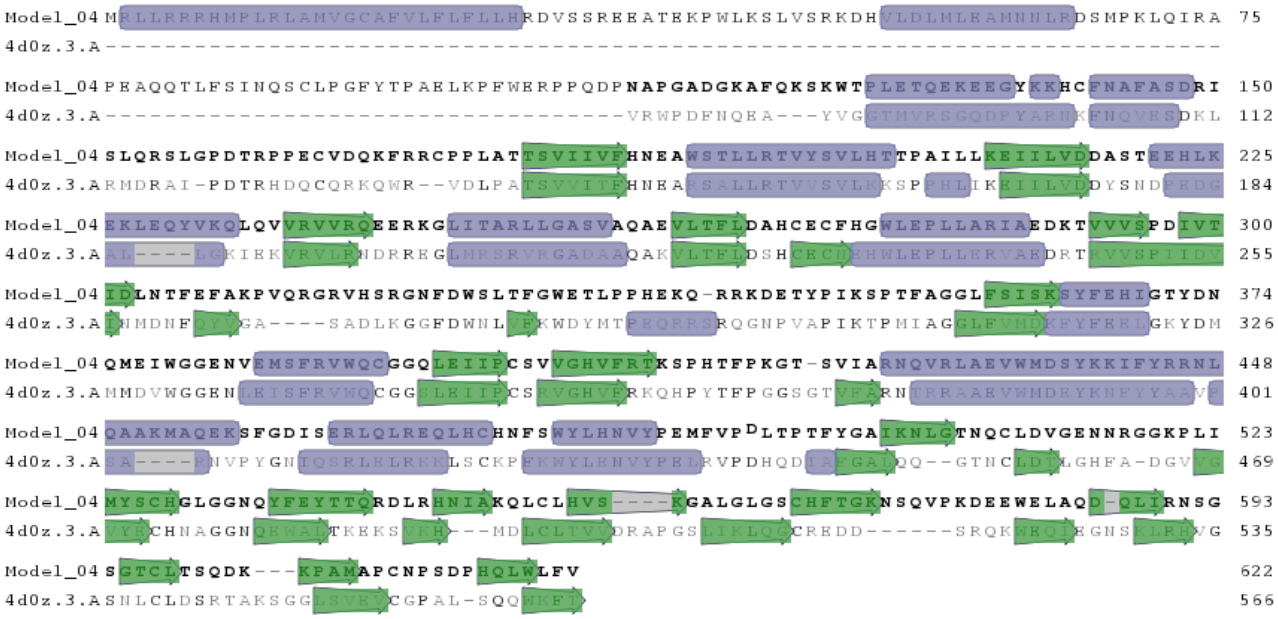


Figure 25. Homology modelling employing GalNAc-T2 as search probe. Secondary structure is represent in blue (α-helix) and in green (β-strands). Bold letters indicate random coils.

Through the prediction of secondary structures, it was possible to identify a potential linker region situated between the catalytic and the lectin domains containing residues Pro486 to Ala500. Inside this area, residue Asp492 is apparently far enough from the structured domains, then this site was chosen to truncate the protein.

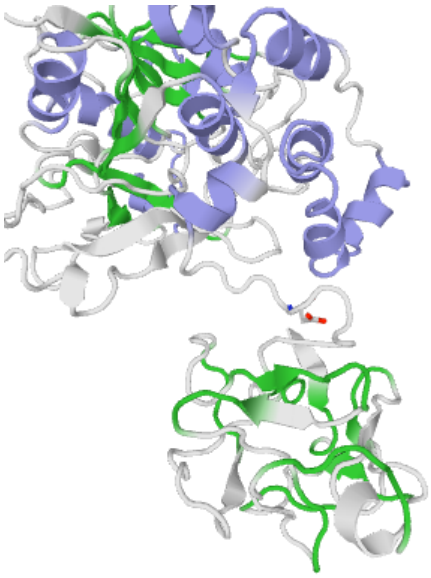


Figure 26. GalNAc-T6 3D structure model (residue Asp 492 is evidenced).

The coding sequence of the lectin domain (D492-V622) was cloned into the pPICZαA vector for its expression in *Pichia pastoris*. After transformation into XL1-Blue (*E. coli* strain), ten colonies grown in LB-zeocin were randomly selected to control proper insertion into the plasmid by colony PCR. The DNA fragment was found amplified in three of them, and further analysis with the restriction enzyme XbaI confirmed the presence of the gene. The construct pPICZαA–T6L was isolated from the bacteria and full sequence match was identified by DNA sequencing.

GALNAC-T6
>sp|Q8NCL4|1-622
MRLLRRRHMPLRLAMVGCAFVLFLFLLHRDVSSREEATEKPWLKSLVSRKDHVLDLMLEAMNNLRDSMPKLQIRAPEAQQTLFSINQSCLPGFYTPAELK
PFWERPPQDPNAPGADGKAFQKSKWTPLETQEKEEGYKKHCFNAFASDRISLQRSLGPDTRPPECVDQKFRRCPPLATTSVIIVFHNEAWSTLLRTVYSVL
HTTPAILLKEIILVDDASTEEHLKEKLEQYVKQLQVVRVVRQEERKGLITARLLGASVAQAEVLTFLDAHCECFHGWLEPLLARIAEDKTVVVSPDIVTIDLNTF
EFAKPVQRGRVHSRGNFDWSLTFGWETLPPHEKQRRKDETYPIKSPTFAGGLFSISKSYFEHIGTYDNQMEIWGGENVEMSFRVWQCGGQLEIIPCSVVG
HVFRTKSPHTFPKGTSVIARNQVRLAEVWMDSYKKIFYRRNLQAAKMAQEKSFGDISERLQLREQLHCHNFSWYLHNVYPEMFVPDLTPTFYGAIKNLGT
NQCLDVGENNRGGKPLIMYSCHGLGGNQYFEYTTQRDLRHNIAKQLCLHVSKGALGLGSCHFTGKNSQVPKDEEWELAQDQLIRNSGSGTCLTSQDKKP
AMAPCNPSDPHQLWLFV

Figure 27. Primary structure of GalNAc-T6 (sequence of the lectin domain is indicated in blue).

3.3.2 *Selection of high level expression colonies of Pichia pastoris*

The construct pPICZαA–T6L was introduced by electroporation into X33 and KM71H competent cells. After incubation during 4 days, several transformants grew on YPDS-zeocin (100 µg/ml) plates. Since recombination can occur in many different ways that can effect expression, 16 colonies of each strain were randomly selected to perform further testing. Only 9 colonies were able to growth on zeocin highest concentration (500 µg/ml) and they were selected since they should contain multiple copies of the gene of interest. Finally, the genomic DNA of these colonies was used as a template to run a direct PCR screening of the clones, confirming that in all of them the T6L gene was integrated into the *Pichia* genome.

Different *Pichia* phenotypes were evident after 48 hours of incubation depending on the carbon source present in the culture plates. KM71H mutant colonies plated on medium containing methanol (MM) presented a retarded growth with respect to those plated on culture medium containing glucose (MD). This observation confirmed the reduced ability of these mutant cells to metabolize methanol as the sole carbon source due to the disruption of the AOX1 gene. Conversely, wild type X33 colonies grew at similar rates when plated on both media.

The selected transformants were purified on minimal plates to ensure pure clonal isolates before testing for expression. The importance of finding colonies with multiple copies of the gene inserted is that a higher dose of the gene could favour higher levels of protein expression. Since it is not possible to know in advance which strain would express the protein better, both phenotypes must be tested.

3.3.3 *Protein expression*

Small-scale protein expression assays were carried out with the nine selected colonies of each *Pichia* strain. Cells were growth on buffered medium containing glycerol (BMGY) to generate a large quantity of biomass prior to methanol induction. Next, protein expression was induced on buffered methanol medium (BMMY) at OD_{600}=1. Induction was maintained for 96 hours with daily additions of methanol (up to 0.5% final concentration) into the culture. A Western blot revealed which colonies presented an enhanced protein expression.

KM71H colonies did not show any protein production in the first 24 hours of induction, while only one sample did after 48 hours. On the other hand, the protein was identified on two X33 colonies from the beginning of the induction period, and its expression was still increased on the second day.

After 72 hours of induction, a large quantity of protein expressed was detected in one sample of X33, which was found even enhanced at 96 hours (Figure 28, lanes 1 and 2). Thus, glycerol stocks of this colony were prepared and conserved at -80°C for its future use on larger scale protein expression. Unfortunately, poor quantities of protein were identified on the KM71H culture (Figure 28, lanes 5 and 6).

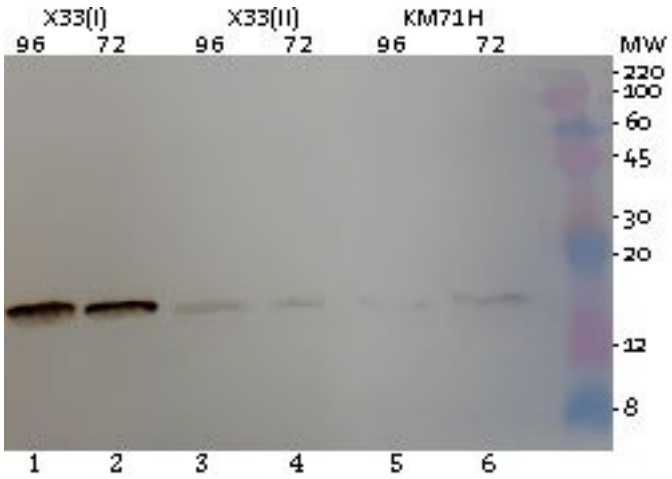


Figure 28. Western blot membrane with samples taken after 72 and 96 hours of induction with methanol.

Pichia pastoris colonies with the quicker methanol utilization X33 appeared to achieve a high expression level of the recombinant protein T6L. The full-length protein was detected at least up to 96 hours of incubation at 28°C, suggesting that the buffered medium BMMY chosen for growth and induction prevents the action of proteases with success.

3.3.4 *Expression scale-up and protein purification*

Protein production was optimized for larger scale cultures that allowed the purification of recombinant T6L. X33 cells secreted the protein into the culture broth using methanol to induce

AOX1 promoter for 96 hours. Good aeration and a fixed temperature of 28°C were critical parameters during induction for optimal protein production.

The secreted protein was recovered from the supernatant by precipitation using 90% w/v ammonium sulphate. Purification was performed with an CM-cellulose column, from where T6L was eluted with 500 mM NaCl. The contaminants having a negative surface charge did not bind to the cation exchange resin, and this became a key step not only for the elimination of other proteins secreted by *Pichia*, but also to remove AOX (alcohol oxidase) crystalloids present in the supernatant. If those had not been completely removed, then they would have made it difficult to use techniques such as protein concentration, dialysis and measurements of UV absorbance.

Since protein crystallization requires a highly homogeneous sample, T6L was further purified employing hydrophobic and size exclusion chromatography. Finally, protein purity was checked by SDS-PAGE (Figure 29). The yield for T6L was about 2 mg starting from one litre of yeast culture.

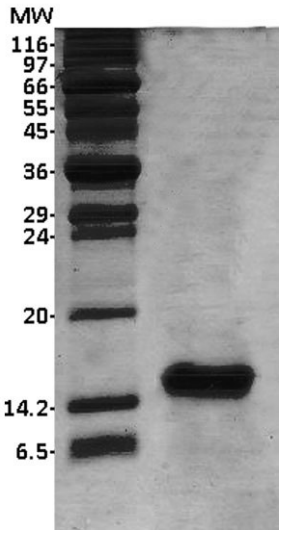


Figure 29. SDS-PAGE gel of T6L after elution from chromatographic columns.

From every purification step, the protein was obtained together with non identified pigments, giving a dark brown colour to the preparation. This is a common inconvenient problem encountered when working with secreted proteins in yeast cultures. Thus, additional expression assays were performed to work against this effect.

As the presence of the yeast extract and peptone into the culture medium could produce these coloured substances, protein expression was induced using buffered minimal medium (BMMH)

and lower temperatures (20 and 25°C). Although both alternatives showed a reduced production of these pigments, protein yield decreased strongly (~0.2 mg per litre). Presumably, those substances were able to interact non-specifically on the protein surface, stabilizing the protein when secreted into the culture medium.

The following experiments were performed using T6L obtained under induction at 20°C on BMMH broth. An exception was made for crystallization assays, in which the protein was prepared using the first method described as higher amounts of samples were needed.

3.3.5 *Intrinsic fluorescence analysis*

The intrinsic fluorescence emission provides an effective probe of local structural changes in the protein upon ligand interaction. Therefore, this technique was used to analyze the binding between T6L and GalNAc. The protein contains two tryptophans, four tyrosines and four phenylalanines, the residues capable of exhibit fluorescence emission.

After excitation at 280 nm only Trp and Tyr in the molecule are excited, and the emission spectra of T6L exhibited a single peak at 340 nm (Figure 30). This result indicates that it is mostly the tryptophan residues in the protein that contribute to the total emission, and the position of the maximum value of the curve allowed to infer that they should not be buried into the protein core, but semi-exposed to the solvent.

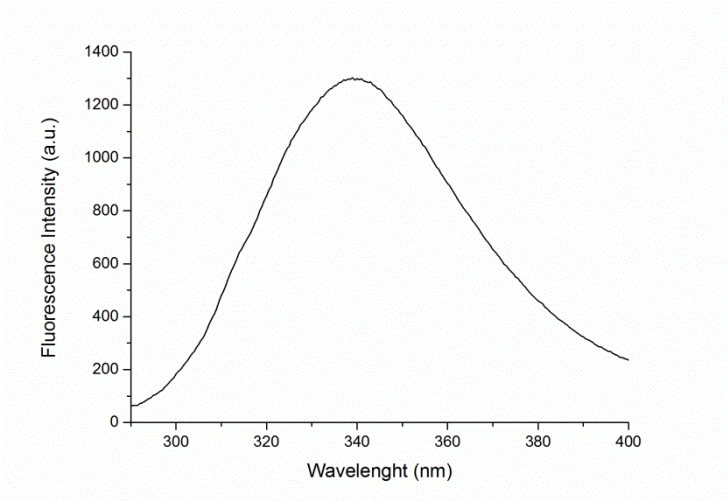


Figure 30. Fluorescence emission spectra of apo-T6L after excitation at 280 nm.

The sensitivity of tryptophan residues to their local environment was used to analyze the affinity of T6L for GalNAc. Several injections of 1 mM GalNAc were added to T6L 1 µM to a carbohydrate final concentration of 30 µM in the solution. After excitation at 280 nm, fluorescence emission spectra were acquired and the quenching curves were evaluated using a non-linear regression function. Fitting of the data assuming a single binding site model gave Kd values of 0.91 ± 0.14 µM, confirming the presence of a high affinity for this carbohydrate in T6L.

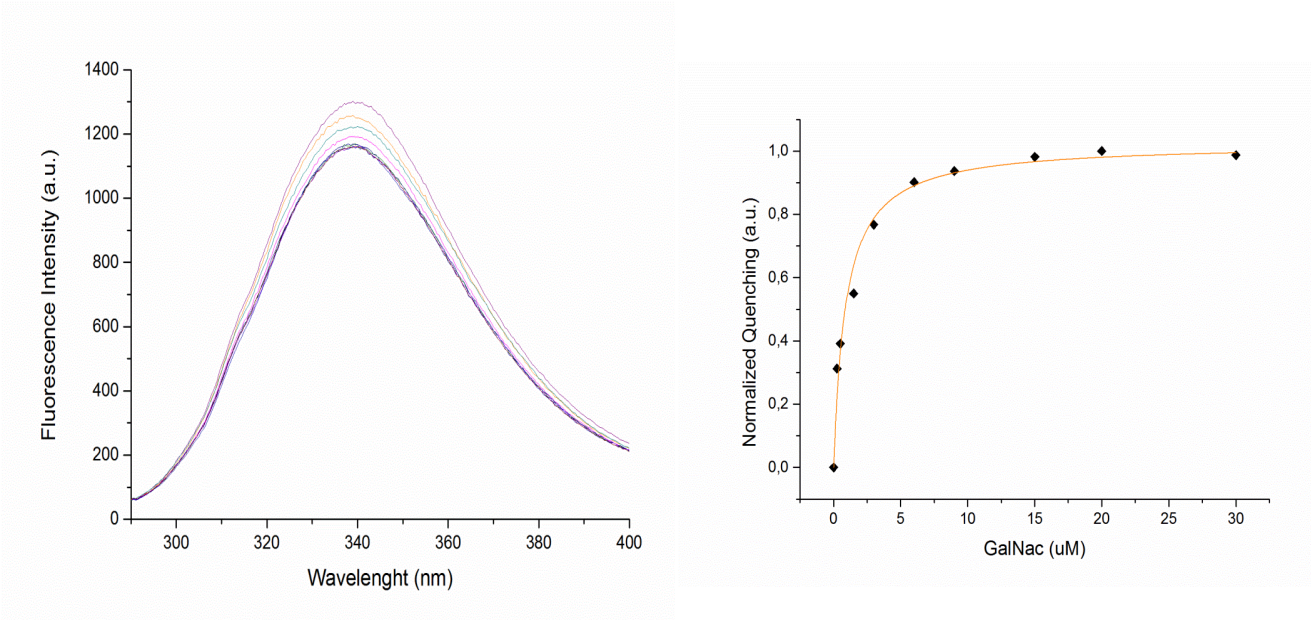


Figure 31. Fluorescence quenching curves (left) and non-linear fitting using a single binding site (right).

3.3.6 *Thermodynamic characterization of carbohydrate binding*

Isothermal Titration Calorimetry (ITC) is a biophysical technique used to study interactions between biomolecules in solution and in this case it was used to assess the carbohydrate binding properties of T6L. The ITC experiments were performed by the titration of a 1 mM GalNAc solution into a 50 µM T6L solution at discrete time intervals and under isothermal conditions. The signal measured was the heat released upon macromolecule-ligand interaction. The area of each heat peak was then integrated and plotted versus the molar ratio ligand-macromolecule and the resulting isotherms were fitted to a binding model from which the interaction parameters were derived.

Representative ITC curves and the derived isotherm for the binding of GalNAc to T6L are shown in Figure 32, and a summary of the thermodynamic parameters obtained by using an independent

binding fitting model is listed in Table 3. Titration of GalNAc into T6L resulted in a binding isotherm that indicated the presence of only one binding site with a dissociation constant (Kd) of 1.25 μM, a value very close to that obtained from the fluorimetric experiments. The binding of GalNAc appeared to be an exothermic process with a small negative enthalpy value (-4.4 kJ/mol) and a high overall positive change in entropy (98.23 $J.mol^{-1}K^{-1}$). Therefore, the interactions stabilized with the carbohydrate would stabilize the protein conformation causing the formation of energetically favourable hydrogen bonds (negative ΔH), but it would also promote hydrophobic interactions as the process is accompanied by a large positive ΔS, which can be associated with the release of water molecules from the binding site.

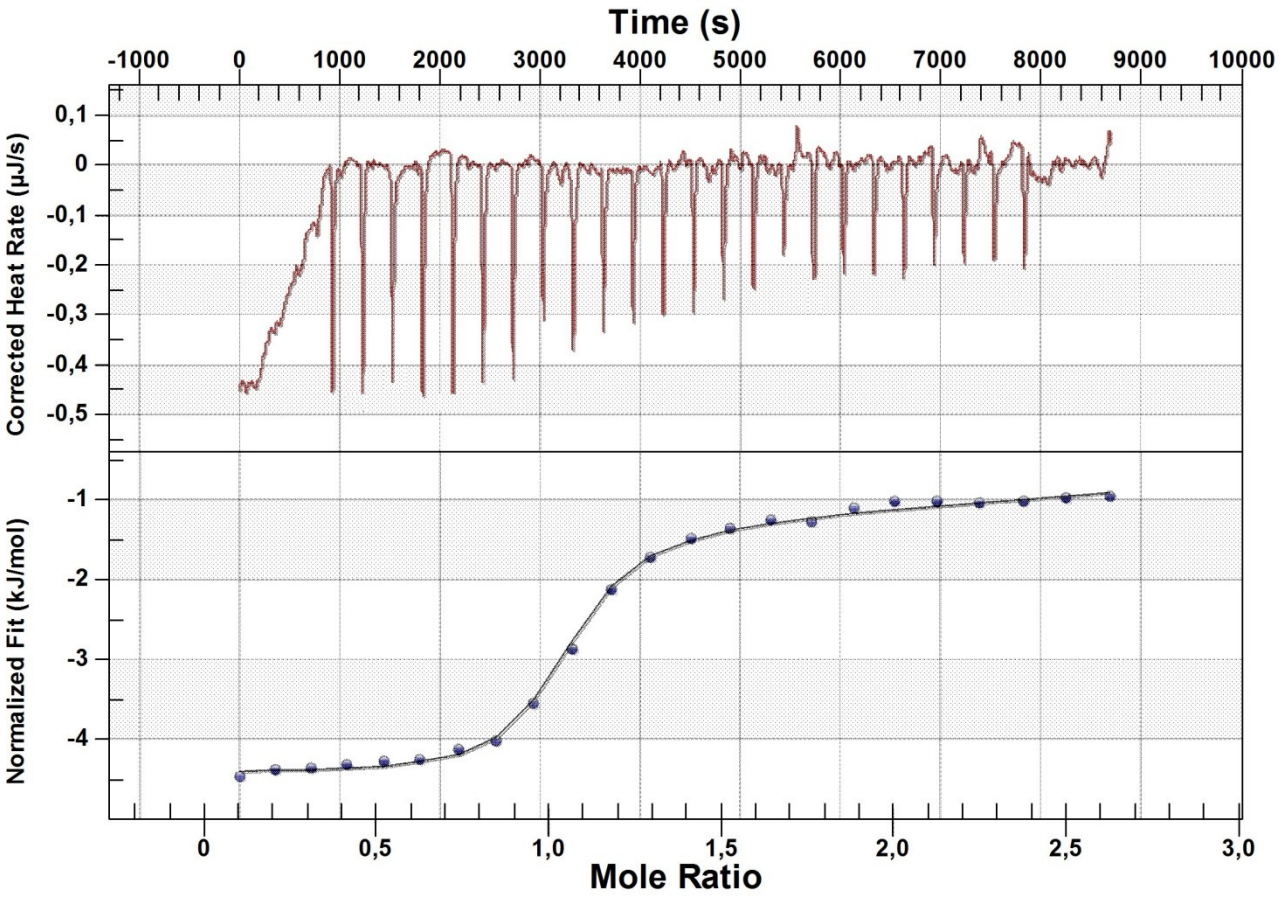


Figure 32. ITC analysis of the GalNAc binding to T6L. Raw data minus baseline (top) and fitted data (bottom) are displayed.

Variable	Value
Ka (μM^{-1})	0.8
n	0.92
ΔH (kJ/mol)	-4.41
Kd (μM)	1.25
ΔS (J/mol·K)	98.23
ΔG (kJ/mol)	-33.69

Table 3. Affinity constant and thermodynamic parameters for the binding of GalNAc to T6L. Data were fitted using the independent-set of sites model provided by the NanoAnalyze Software (TA).

3.3.7 *Protein thermal stability analysis*

The thermal stability of T6L was evaluated performing Differential Scanning Calorimetry (DSC). This technique measures the change in the heat capacity (Cp) of the protein solution when heated at a constant rate. Changes in the heat capacity are originated by the disruption of the forces stabilizing the native protein structure, and during protein unfolding an endothermic peak comes up in the Cp vs temperature curve. The maximum value represent the temperature midpoint (Tm).

The thermal scanning was set from 10 to 100°C, with an increase of 0.25°C/min. The temperature midpoint (Tm) was determined at 58.1°C. The reverse cooling scanning did not show an equivalent exothermic peak, indicating an irreversible process of unfolding. However, no sample aggregate was found within the cell after the experiments. These results suggests that T6L would have desirable biophysical properties under physiological temperatures.

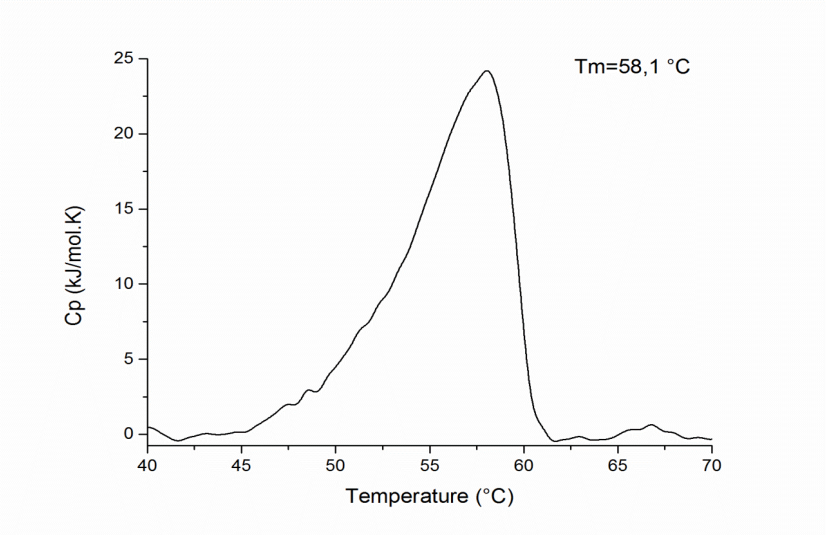


Figure 33. Endotherm obtained after performing the heating scan on T6L solution.

3.3.8 *X-ray diffraction data analysis*

For the study of the three-dimensional structure of T6L, purified protein was concentrated to about 10 mg/ml and used to set up the crystallization screenings using different precipitants. The first crystal nuclei appeared after 48 hours of incubation in many of the conditions screened. Although polyethylene glycol (PEG) seemed to be the optimal precipitant, not every condition tested with it contained crystals suitable for X-ray experiments. In many drops only microcrystals were found, hence screening conditions were probably appropriate but the concentration of the protein and/or the precipitant was only slightly too high and the experiment remained blocked in the nucleation phase.

The crystals chosen to perform X-ray diffraction experiments were grown in a week at 20°C in the following conditions:

A. 30% PEG 1500 / etilenglycol 1%

B. 0.1 M sodium chloride / 0.1 M lithium sulphate / 0.1 M MES pH 6.5 / 12% PEG 4000

C. 0.1 M magnesium chloride / 0.1 M ADA buffer pH 6.5 / 12% w/v PEG 6000

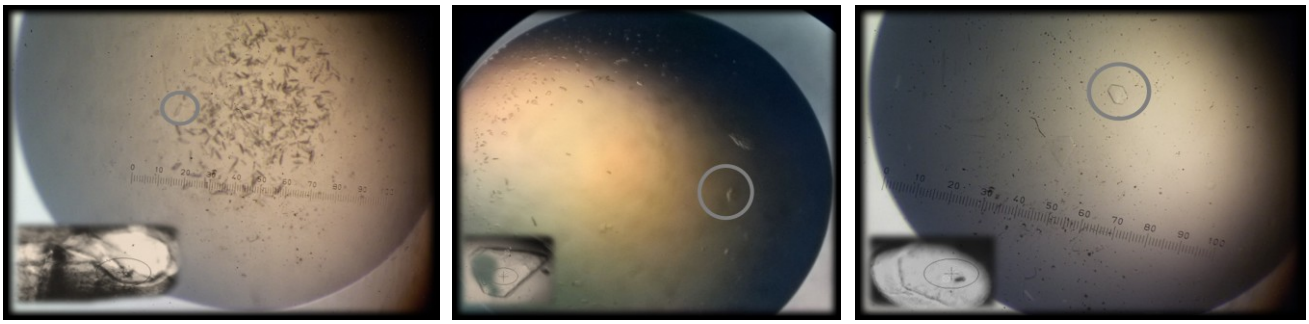


Figure 34. Crystals of T6L obtained in conditions A (left), B (middle) and C (right).

Different crystal forms can be more or less well ordered and hence give diffraction patterns of different quality. Samples obtained in conditions A and B gave a low number of reflections due to crystal decay caused by the radiation. The frames were processed using iMOSFLM but it was not possible to record high resolution reflections that could not be measured due to crystal damage. Moreover, crystals were formed in a hexagonal plane shape, giving reflections in one direction but a few or none when they were rotated 90 degrees. The experimental setup was modified in order to counteract this problem. Although the use of a higher transmittance, longer exposure times and small ranges of oscillation benefited data quality, they led to incomplete high resolution data due to crystal disintegration. In addition, the overlapping of diffraction points resulted in a small signal-to-noise ratio.

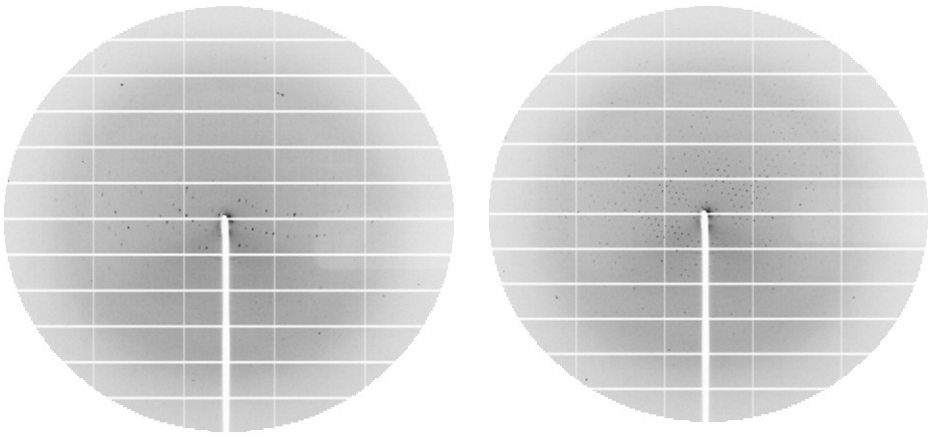


Figure 35. Example of two images obtained during the diffraction experiment of T6L crystals, both differ by an angle of 90 degrees.

The maximum resolution attained was 3.5 Å and 567633 reflections were collected with crystals obtained in sample C. Diffraction data were analyzed with the CCP4 package software and parameters of the unit cell were compatible with space group C2.

Space group	C2 (monoclinic system)	
Average unit cell	a=86.32 b=151.02 c=91.16	α=90.00 β=101.68 γ=90.00
Resolution limit	3.50-50.00	
Rmerge	0.452	
Total number of observations	7515	
Total number unique reflections	5249	

Table 4. Data collection and indexing statistics of T6L.

Additional problems such as high anisotropic mosaicity and twinning did not allow resolution of the T6L structure using these data. As small but well-diffracting crystals grew, several optimization tests were developed to produce larger crystals and to improve the quality of the diffraction data. No other crystals suitable for X-ray diffraction experiments have been obtained thus far. Therefore, in the near future it is planned to improve the purification protocol by increasing the purity of the sample, since it is the critical point affecting the crystallization process. Obtaining better crystals not only would yield a detailed structure of T6L but it would also make possible the 3D characterization of GalNAc binding.

3.4 Conclusions and Future Directions

The presented protein-engineering strategy opens the way to the use of human lectins for therapeutic applications. Nonetheless, further studies are required to determine if the role of GalNAc-T6 in protein glycosylation could be exploited for delivery of nanoparticles loaded with antitumor drugs.

A recombinant human protein containing only the lectin domain of GalNAc-T6 was produced and its biophysical properties regarding carbohydrate binding and thermal stability were characterized. Its high affinity to N-acetylgalactosamine was demonstrated by spectrofluorimetry and isothermal titration calorimetry, and both techniques suggested that only one of the three putative binding sites in the molecule would be active. This result may be confirmed by future crystallographic studies, while i*n vivo* assays will determine whether tumor cells exposing this carbohydrate (i.e. Tn antigen) can be recognized. Moreover, using differential scanning calorimetry it was observed that the truncated protein is stable under physiological temperatures, which makes it a feasible candidate to be proposed for drug-delivery therapies.

The present results encourage additional research to extend our observations. Future studies on tumor cells using the truncated form of GalNAc-T6 conjugated with drug-loaded nanoparticles would determine the potential clinical value of this engineered protein.

Chapter IV

Appendix

4.1 Media Compositions

4.1.1 *Escherichia coli recipes*

Luria-Bertani (LB) medium	*1% w/v peptone* *0.5% w/v yeast extract* *0.5% w/v sodium chloride*

4.1.2 *Pichia pastoris recipes*

YP medium	*1% w/v Yeast extract* *2% w/v Peptone*
YPD medium	*YP medium* *2% w/v Dextrose (glucose)*
YPDS medium	*YPD medium* *0.1 M Sorbitol*
MD medium *(Minimal medium containing Glucose)*	*1.34% yeast nitrogen base without amino acids* *4×10^{-5}% biotin* *2% Dextrose*
MM medium *(Minimal medium containing Methanol)*	*1.34% yeast nitrogen base without amino acids* *4×10^{-5}% biotin* *0.5% methanol*
BMGY medium *(Buffered complex medium containing glycerol)*	*1% yeast extract* *2% peptone* *100 mM potassium phosphate pH 6.0* *1.34% yeast nitrogen base without amino acids* *4 × 10-5% biotin* *1% glycerol*
BMM medium *(Buffered Minimal Medium containing Methanol)*	*100 mM potassium phosphate pH 6.0* *1.34% yeast nitrogen base without amino acids* *4×10^{-5}% biotin* *0.5% methanol*
BMMY medium *(Buffered complex medium containing Methanol)*	BMM medium *1% yeast extract* *2% peptone*

*For plates 1.5% *w/v* bacteriological agar was added

BEDS solution	*10 mM bicine-NaOH pH 8.3* *3% (v/v) ethyleneglycol* *5% (v/v) dimethylsulfoxide* *1 M sorbitol*

4.2 Buffer compositions

4.2.1 *FGF1 recipes*

Buffer 1	*10 mM ammonium sulphate* *20 mM sodium phosphate monobasic* *2mM ethylenediaminetetraacetic acid* *2 mM β-mercaptoethanol*
Buffer 2	*20 mM sodium phosphate monobasic*
Buffer 3	*50 mM HEPES* *10 mM ammonium sulphate* *20 mM sodium phosphate monobasic* *0.5 mM ethylenediaminetetraacetic acid* *5 mM β-mercaptoethanol* *0.5 M sodium chloride*
PBS buffer	*1.37 M sodium chloride* *2.7 mM potassium chloride* *10 mM sodium phosphate dibasic* *1.8 mM potassium phosphate monobasic*

4.2.2 *T6L recipes*

Buffer 4	*20 mM MES pH 6.0* *5 mM β-mercaptoethanol* *5 mM ethylenediaminetetraacetic acid* *0.2 M sodium chloride*
Buffer 5	*20 mM sodium phosphate monobasic* *2 mM β-mercaptoethanol*

Chapter V

References

1. Stanley P. Golgi glycosylation. *Cold Spring Harb. Perspect. Biol.* **3,** 1–13 (2011).

2. Brockhausen I. Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep.* **7,** 599–604 (2006).

3. Ju T & Cummings RD. Core 1 β-3-Galactosyltransferase (C1GalT1, T-Synthase) and Its Specific Molecular Chaperone Cosmc (C1GalT1C1). Handbook of Glycosyltransferases and Related Genes. *Springer, Japan,* 149–169 (2014).

4. Radhakrishnan P, Dabelsteen S, Madsen F, Francavilla C, Kopp KL, Steentoft B, Vakhrushev SY, Olsen JV, Hansen L, Bennett EP, Woetmann A, Yin G, Chen L, Song H, Bak M, Hlady RA, Peters SL, Opavsky R, Thode C, Qvortrup K, Schjoldager K, Clausen H, Hollingsworth MA, and Wandall HH. Immature truncated O-glycophenotype of cancer directly induces oncogenic features. *Proc. Natl. Acad. Sci. U. S. A.* **111,** 4066-75 (2014).

5. Stowell SR, Ju T & Cummings RD. Protein glycosylation in cancer. *Annu. Rev. Pathol.* **10,** 473–510 (2015).

6. Sharon N & Lis H. History of lectins: From hemagglutinins to biological recognition molecules. *Glycobiology* **14,** 53–62 (2004).

7. Ghazarian H, Idoni B & Oppenheimer SB. A glycobiology review: Carbohydrates, lectins and implications in cancer therapeutics. *Acta Histochem.* **113,** 236–247 (2011).

8. Weis W & Drickamer K. Structural Basis of Lectine-Carbohydrate Recognition. *Annu. Rev. Biochem.* **65,** 441–473 (1996).

9. Bies C, Lehr CM & Woodley JF. Lectin-mediated drug targeting: History and applications. *Adv. Drug Deliv. Rev.* **56,** 425–435 (2004).

10. Liu B, Bian H & Bao J. Plant lectins: Potential antineoplastic drugs from bench to clinic. *Cancer Lett.* **287,** 1–12 (2010).

11. Carrizo ME, Irazoqui FJ, Lardone RD, Nores GA, Curtino JA, Capaldi S, Perduca M & Monaco HL. The antineoplastic lectin of the common edible mushroom (Agaricus bisporus) has two binding sites, each specific for a different configuration at a single epimeric hydroxyl. *J. Biol. Chem.* **280,** 10614–10623 (2005).

12. Destefanis L. Structural studies of Pleurotus ostreatus Lectin (POL), a fungal protein of medical interest. (Ph.D. Thesis, University of Verona, 2015).

13. Bovi M, Carrizo ME, Capaldi S, Perduca M, Chiarelli LR, Galliano M & Monaco HL. Structure of a lectin with antitumoral properties in king bolete (Boletus edulis) mushrooms. *Glycobiology* **21,** 1000–1009 (2011).

14. Bovi M, Cenci L, Perduca M, Capaldi S, Carrizo ME, Civiero L, Chiarelli LR, Galliano M & Monaco HL. BEL β-trefoil: A novel lectin with antineoplastic properties in king bolete (Boletus edulis) mushrooms. *Glycobiology* **23,** 578–592 (2013).

15. Carter P. Introduction to current and future protein therapeutics: A protein engineering perspective. *Exp. Cell Res.* **317,** 1261–1269 (2011).

16. Lazar GA, Marshall SA, Plecs JJ, Mayo SL & Desjarlais JR. Designing proteins for therapeutic applications. *Curr. Opin. Struct. Biol.* **13,** 513–518 (2003).

17. Marshall SA, Lazar GA, Chirino AJ & Desjarlais JR. Rational design and engineering of therapeutic proteins. **8,** 212–221 (2010).

18. Ornitz DM & Itoh N. The fibroblast growth factor signaling pathway. *Wiley Interdiscip. Rev. Dev. Biol.* **4,** 215–266 (2015).

19. Blaber M, DiSalvo J & Thomas KA. X-ray crystal structure of human acidic fibroblast growth factor. *Biochemistry* **35,** 2086–94 (1996).

20. Pellegrini L. Role of heparan sulphate in fibroblast growth factor signalling: A structural view. *Curr. Opin. Struct. Biol.* **11,** 629–634 (2001).

21. Chothia C & Lesk AM. The relation between the divergence of sequence and structure in proteins. *EMBO J.* **5,** 823–6 (1986).

22. Holm L & Rosenstrom P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38,** W545-9 (2010).

23. Emsley P, Lohkamp B, Scott WG & Cowtan K. Features and development of "*Coot*". *Acta Crystallogr. Sect. D* **66,** 486–501 (2010).

24. Zakrzewska M, Krowarsch D, Wiedlocha A, Olsnes S & Otlewski J. Structural requirements of FGF-1 for receptor binding and translocation into cells. *Biochemistry* **45,** 15338–15348 (2006).

25. Berridge M, Tan A, McCoy K & Wang R. The biochemical and cellular basis of cell proliferation assays that use tetrazolium salts. *Biochemica* **4**, 4–9 (1996).

26. Lakowicz JR. Protein fluorescence. Principles of fluorescence spectroscopy. *Springer, New York, USA, 3rd edition,* 529-575 (2006).

27. McPherson A. Introduction to protein crystallization. *Methods* **34,** 254–265 (2004).

28. Chayen NE. Turning protein crystallisation from an art into a science. *Curr. Opin. Struct. Biol.* **14,** 577–583 (2004).

29. McPherson A. Current approaches to macromolecular crystallization. *Eur. J. Biochem.* **189,** 1–23 (1990).

30. Bruylants G, Wouters J & Michaux C. Differential scanning calorimetry in life science: thermodynamics, stability, molecular recognition and application in drug design. *Curr. Med. Chem.* **12,** 2011–2020 (2005).

31. Georgiou G & Valax P. Expression of correctly folded proteins in Escherichia coli. *Curr. Opin. Biotechnol.* **7,** 190–197 (1996).

32. Golovanov AP, Hautbergue GM, Wilson SA & Lian L. A simple method for improving protein solubility and long-term stability. *J. Am. Chem. Soc.* **126,** 8933–9 (2004).

33. Copeland RA, Ji H, Halfpenny AJ, Williams RW, Thompson KC, Herber WK, Thomas KA, Bruner MW, Ryan JA, Marquis-Omer D, Sanyal G, Sitrin RD, Yamazaki S & Middaugh CR. The structure of human acidic fibroblast growth factor and its interaction with heparin. *Arch. Biochem. Biophys.* **289,** 53–61 (1991).

34. Blaber SI, Culajay JF, Khurana A & Blaber M. Reversible thermal denaturation of human FGF-1 induced by low concentrations of guanidine hydrochloride. *Biophys. J.* **77,** 470–477 (1999).

35. Xia X, Kumru OS, Blaber SI, Middaugh CR, Li L, Ornitz DM, Sutherland MA, Tenorio CA & Blaber M. Engineering a Cysteine-Free Form of Human Fibroblast Growth Factor-1 for "Second Generation" Therapeutic Application. *J. Pharm. Sci.* **105,** 1444–53 (2016).

36. Munkley J, Mills IG & Elliott DJ. The role of glycans in the development and progression of

prostate cancer. *Nat Rev Urol* **13,** 324–333 (2016).

37. Pedersen JW, Bennett EP, Schjoldager KTBG, Meldal M, Holmér AP, Blixt O, Cló E, Levery SB, Clausen H & Wandall HH. Lectin domains of polypeptide GalNAc transferases exhibit glycopeptide binding specificity. *J. Biol. Chem.* **286,** 32684–32696 (2011).

38. Hanisch FG. O-glycosylation of the mucin type. *Biol. Chem.* **382,** 143–149 (2001).

39. Berois N, Mazal D, Ubillos L, Trajtenberg F, Nicolas A, Sastre-Garau X, Magdelenat H & Osinaga E. UDP-N-acetyl-D-galactosamine: Polypepticle N-acetylgalactosaminyltransferase-6 as a new immunohistochemical breast cancer marker. *J. Histochem. Cytochem.* **54,** 317–328 (2006).

40. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W & Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).

41. Remmert M, Biegert A, Hauser A & Soding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat. Methods* **9,** 173–175 (2011).

42. Guex N & Peitsch MC. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18,** 2714–2723 (1997).

43. Benkert P, Biasini M & Schwede T. Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics* **27,** 343–350 (2011).

44. Duff MR, Grubbs J & Howell EE. Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity. J. Vis. Exp. **55**, e2796, DOI : 10.3791/2796 (2011).

45. Smyth MS & Martin JH. X Ray Crystallography. *J. Clin. Pathol. Mol. Pathol.* **53,** 8–14 (2000).