UNIVERSITA' DEGLI STUDI DI VERONA

Department of Biotechnology

Graduate School of Sciences and Engineering

Doctoral Program in Biotechnology

Cycle XXXI – year 2015

Structural studies of a *Tremella fuciformis* lectin: a novel protein with antitumoral properties

S.S.D. BIO/11

Coordinator:

Prof. Matteo Ballottari

Supervisor:

Prof. Matteo Ballottari

Co-Supervisor:

Prof. Hugo L. Monaco

Prof. Roberto A. Steiner

Doctoral Student:

Giulia Glorani

Alla mia famiglia

Abstract

Lectins are carbohydrate-binding proteins of non-immune origin, widely distributed in all the living organisms. They play a role in different biological processes and, due to their carbohydrate specific binding, some lectins have been found to possess antitumoral properties. A particular case is the recognition of the T-antigenic determinant (Galβ1-3GalNAcα) on the malignant cell surfaces, which is present in 90% of human carcinomas. Upon binding to the carbohydrate, some lectins can cause apoptosis, cytotoxicity, inhibition of tumour growth, thus preventing the proliferation of tumoral cells. Considering the fact that these carbohydrates are masked on healthy cells, the highly specific carbohydrate-lectin interaction can be exploited to target only malignant cells.

Although the function and the biological properties of several lectins have been determined, there are still many lectins that remain to be structurally and functionally characterized. As reported in the literature, some *Tremella fuciformis* proteins have been investigated for their potential therapeutical properties and in the light of this, the crude extract of this fungus was examined to assess the presence of lectins.

A lectin of 11 KDa, named TFL, was isolated and purified from the dried fruiting bodies and used for testing several crystal screening conditions. The best crystals were grown in 0.1 M TRIS pH 8.5, 1.5 potassium phosphate dibasic, 1 % DMSO and the final data sets were collected at the ESRF of Grenoble, revealing the three-dimensional structure of the protein. TFL belongs to the space group $P12_11$ and the cell parameters of the crystal are the following: a = 61.62 Å, b = 61.83 Å, c = 67.84 Å, with β = 106.87 °. The protein is a monomer, composed of six β-sheets (from A to F) that arrange to form the so called “β-barrel”. In addition, two α-helixes, named H_1 and H_2 can be recognized in

the structure, which is also stabilized by the presence of two disulphide bridges that connect Cys 54 to Cys 70 and Cys 97 to Cys 100, respectively.

Thermal protein stability was analyzed by means of differential scanning calorimetry, revealing that TFL does not undergo a two-state unfolding process and its denaturation is reversible, a feature that has been rarely observed for lectins before. In addition, chemical and pH-induced unfolding were investigated using fluorescence spectroscopy, highlighting the high stability of TFL in a wide range of conditions. Exploiting isothermal titration calorimetry, more information about sugar-binding were acquired, identifying N-acetylgalactosamine as the best candidate, which has micromolar affinity for TFL. It has also been observed that *Tremella fuciformis* lectin shows no cytotoxicity on malignant and healthy cells and interestingly, it seems to reduce skin malignant cells migration and to have a positive effect in the upregulation of certain genes involved in cancer arresting (SBPP1). In addition, immunomodulatory activity of TFL was analyzed by means of ELISA essay, resulting in the up-regulation of the most important inflammatory markers (IL-6, TNFα), that could be usefully exploited in order to kill intracellular microorganisms and to help the human body to develop a stronger tumoral resistance.

In conclusion, this work provides a new and interesting insight into cancer treatment, being an essential prerequisite for future *in vitro* and *in vivo* experiments. The encouraging results obtained so far, alongside with the fact that TFL has been produced also heterologously in *E.coli*, pave the way for further studies, including the feasibility to perform mutagenesis and to explore the possibility to further encapsulate the protein inside tailor-made nanoparticles, with the aim of reducing the amount of protein used, increasing its therapeutical efficiency, and preventing side effects.

Sommario

Le lectine sono proteine di origine non-immunogenica, ampiamente diffuse in tutti gli organismi viventi, aventi la capacità di legare, in modo specifico e reversibile, i carboidrati. Esse sono coinvolte in diversi processi biologici e, grazie alla loro abilità di legare zuccheri, alcune di loro posseggono proprietà antitumorali. Nello specifico, il determinante antigenico T (Galβ1-3GalNAcα), esposto soltanto sulla superficie delle cellule tumorali, è associato al 90 % dei carcinomi umani, rappresentando un potenziale target da utilizzare nella terapia antitumorale. A seguito del legame con i carboidrati, le lectine possono causare non solo apoptosi, ma anche citotossicità ed inibizione della crescita del tumore, arrivando sino a prevenirne la proliferazione. Considerando il fatto che questo tipo di residui sono mascherati sulla superficie delle cellule sane, la specifica interazione tra lectina e carboidrati potrebbe essere utilizzata per direzionare i trattamenti chemioterapici soltanto nei confronti delle cellule tumorali, riducendo gli effetti collaterali a carico delle cellule sane.

Nonostante le funzioni e le proprietà biologiche di diverse lectine siano state determinate negli ultimi anni, per molte di loro la struttura tridimensionale è ancora da definire. Come riportato in letteratura, alcune proteine isolate dal fungo *Tremella fuciformis*, sono state analizzate per le loro potenziali proprietà terapeutiche e, in virtù di questo, l'estratto crudo del suddetto fungo è stato esaminato con l'obiettivo di evidenziare la presenza di eventuali lectine.

Una lectina di 11 KDa, rinominata TFL (*Tremella fuciformis* lectin) è stata isolata dai corpi fruttiferi del fungo e purificata, dapprima utilizzando una colonna di scambio anionico e, successivamente, sfruttando il principio di esclusione molecolare. Dopo aver testato diverse condizioni

cristallografiche, i migliori cristalli sono cresciuti in 0.1 M buffer TRIS a pH 8.5, 1.5 potassio fosfato dibasico e 1 % DMSO. I molteplici set di dati raccolti presso il sincrotrone di Grenoble (ESRF) hanno permesso di definirne la struttura tridimensionale, la quale appartiene al gruppo spaziale $P12_11$ ed ha i seguenti parametri di cella: a = 61.62 Å, b = 61.83 Å, c = 67.84 Å, con β = 106.87 °. La lectina è un monomero, caratterizzato dalla presenza di sei foglietti β, la cui organizzazione porta alla formazione di una cavità nota come "β-barrel". In aggiunta alle due α-eliche, denominate H_1 e H_2, la sua struttura è ulteriormente stabilizzata dalla presenza di due ponti disolfuro, che connettono rispettivamente le coppie di residui Cys 50 - Cys 74 e Cys 97 - Cys 100.

La stabilità della TFL, unitamente al suo processo di denaturazione, sono stati analizzati mediante calorimetria a scansione differenziale, localizzando il suddetto meccanismo tra le transizione multi-stato e, ancor più interessante, rivelandone la natura reversibile, una caratteristica raramente osservata all'interno della famiglia delle lectine, prima di oggi.

In aggiunta, le denaturazioni chimica e pH-dipendente sono state analizzante utilizzando la spettroscopia di fluorescenza, rivelando la stabilità del monomero in un ampio range di condizioni.

Il legame della proteina con gli zuccheri è stato analizzato utilizzando la calorimetria isotermica di titolazione, definendo lo zucchero N-acetilgalattosamina come il miglior candidato, avente affinità nell'ordine micro molare.

La lectina isolata dal fungo *Tremella fuciformis* ha evidenziato di possedere un effetto benefico nel ridurre la migrazione delle cellule tumorali e nel regolare l'espressione di alcuni geni coinvolti nell'arresto dello sviluppo delle masse tumorali (SBPP1).

L'attività immunomodulatoria della lectina è stata analizzata attraverso l'utilizzo di un saggio ELISA, evidenziando una iper-regolazione di alcuni dei marcatori infiammatori associati al cancro (IL-6 e TNFα), caratteristica che potrebbe avere un effetto positivo nell'aiutare l'organismo a difendersi dall'attacco della massa tumorale ed a sviluppare una maggiore resistenza.

In conclusione, questo lavoro propone una nuova ed interessante visione nell'ambito del trattamento del cancro, proponendosi come un prerequisito essenziale per i futuri esperimenti *in vivo* ed *in vitro*. Gli incoraggianti risultati ottenuti sino ad ora, assieme al fatto che la TFL è stata prodotta anche in versione ricombinante utilizzando il sistema di espressione batterico *E.coli*, pongono le basi per futuri studi, tra i quali la possibilità di effettuare mutagenesi ed esplorare la potenzialità di incapsulare la proteina all'interno di nanoparticelle, con l'obiettivo di ridurne il quantitativo utilizzato, aumentarne l'efficienza terapeutica e prevenire eventuali effetti indesiderati.

Table of contents

1. Introduction

1.1 Lectin superfamilies

1.1.1 Lectins: a general outlook

Lectins are ubiquitous proteins of non-immune origin, being found in all living organisms. They are involved in a large number of biological and pathological processes: host-pathogen interaction, cellular signalling and tissue metastasis, to name just a few. Moreover, they serve as storage proteins, are fundamental during fungi and plant morphogenesis and development, and take part in their defence processes.

Thanks to their structure, lectins are not only able to specifically and reversibly recognize carbohydrates, but also to promote cell precipitation. Since lectins are often multivalent proteins, upon carbohydrate recognition, they can cause cross-linking between cells and subsequently precipitation, a phenomenon referred to as “agglutination”.

Considering that the blood group system “ABO” is determined by the presence of terminal oligosaccharides on the red blood cell surface, lectins were exploited for a long time to clarify the ABH histocompatibility. In addition, apart from lectins, erythrocyte agglutination can be produced also by viruses or antibodies but unlike them, these proteins have a non-immunogenic origin.

Historically, these proteins were first described by Shapley and Boyd, who have also introduced the term *“lectin”*, derived from the latin word *“legere”*. By introducing this term, meaning *“to choose”* or *“to select”*, they intended to classify lectins as a group of agglutinins from higher plants, some of which showed sugar binding and promoted agglutination.

Although several descriptions and definitions have been proposed, currently to be classified as a lectin, a protein should contain at least one carbohydrate recognition domain (CRD) and should not be an immunoglobulin. According to their overall structure, they can be divided into three main classes:

I. Merolectins: possess only one CRD, which is able to bind a simple sugar but does not show agglutinating or catalytic activity.
II. Hololectins: have two or more identical CRDs that bind to sugars, do not display catalytic activity and can cause agglutination.
III. Chimerolectins: characterized by the presence of one CRD and another domain with a distinct biological activity, such as enzymatic.

According to the number of CRD, lectins may or may not cause erythrocyte agglutination [1].

1.1.2 Concerning the history of lectins

The first information concerning these proteins dates back to 1888, when Peter Stillmark discovered two highly toxic lectins, named *"Ricin"* and *"Abrin"*. The first was isolated from the castor tree seeds (*Ricinus communis*), while the second was extracted from the jequirity bean (*Abrus precatorius*) and both were found to possess hemoagglutinating activity. After becoming commercially available, Paul Elrich started using them as a model antigen for his immunological studies, laying the foundations of modern immunology.

Subsequently, in the 1890s, the specificity of the antibody response was established, together with the transfer of the humoral immunity from mother to offspring, and the phenomenon called

immunological memory. Later on, the first pure lectin, named Concanavalin A, was obtained from *Canavalia ensiformis*, also commonly known as jack bean.

Despite being discovered in 1919, it was only in 1963 that Concanavalin A was demonstrated to agglutinate cells, such as yeasts and erythrocytes.

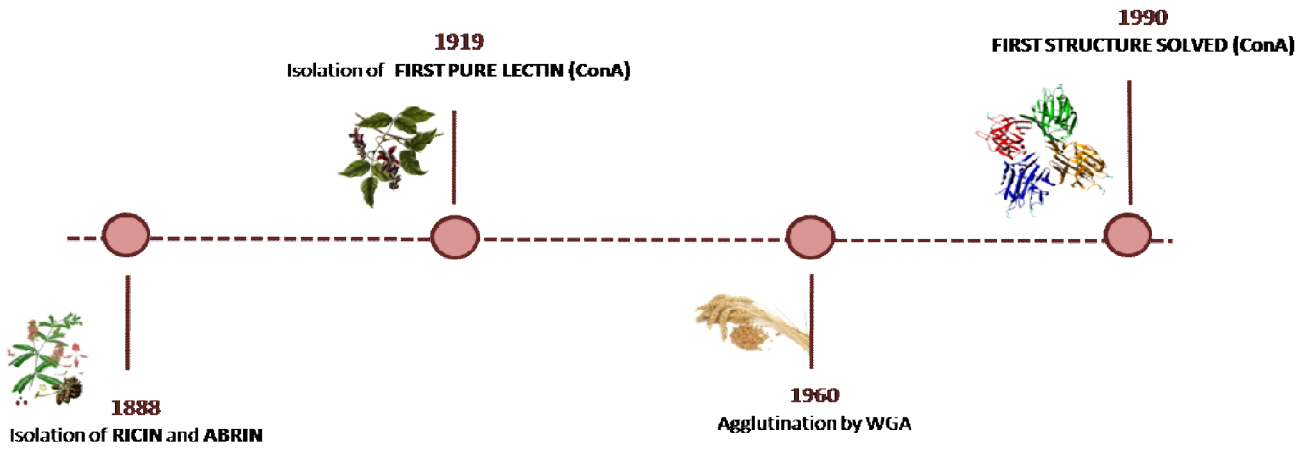


Figure 1: Lectins timeline. From their early discovery (Ricin and Abrin) to the first structure solution by means of X-Ray crystallography (ConA).

In addition, the agglutination process was shown to be dependent upon sucrose binding, highlighting the sharp specificity of lectins.

After the isolation of the first fungal hemagglutinin from the toxic fungus *Amanita phalloides*, the relationship between food toxicity, poisoning, and lectin activity was investigated. However, the following isolation of several lectins from edible mushrooms definitely demonstrated the fact that most of them are not involved in poisoning or in promoting food toxicity.

Going back to the 1940s, a breakthrough in the history of lectins was made by Morgan and Watkins, highlighting the fact that α-N-acetyl-D-galactosamine and α-L-fucose were giving A and H(O) blood group specificity, respectively. The latter study was considered as one of the earlier

evidences for the presence of sugars on cell surfaces, foreseeing also the possibility of exploiting them as identity markers.

Later in the century, the isolation of one phytohemagglutinin from *Phaseolus vulgaris*, also commonly known as red kidney bean, alongside with the discovery of its ability to stimulate lymphocytes to undergo mitosis, put the family of lectins under the spotlight, and emphasized their role in immunology. Interestingly, the mitogenic stimulation upon carbohydrate binding was the demonstration of the biological role of cell surface sugars, suddenly becoming one of the most important tools to analyse several biochemical events occurring during lymphocyte stimulation *in vitro*.

Moreover, the purification of wheat germ agglutinin (WGA) and the demonstration of its ability to selectively agglutinate malignant cells provided a clear evidence that changes in surface carbohydrates were associated with cancer progression and metastasis.

Until the 1970s, although many lectins had been discovered in different organisms, their purification was made difficult by the absence of adequate chromatographic supports. After the introduction of affinity chromatography, their isolation dramatically increased, and they began to be employed not only in the detection, isolation, and characterization of glycoconjugates, but also in the examination of changes on cell surfaces during several pathological processes.

The same decade witnessed also an increasing number of studies concerning lectin molecular properties, ranging from the determination of their sequence, to the elucidation of their three-dimensional structure. Concanavalin A was the first among the lectins whose structure was established, soon followed by Wheat Germ Agglutinin in complex with its ligands.

Since then, the structure of at least 200 lectins has been determined, allowing also the study of their interaction with carbohydrates [2].

The 1970s also witnessed the intensification of studies of the molecular properties of individual lectins, a prerequisite for a deep understanding of their activities at the molecular level. These studies ranged from the determination of the main physicochemical parameters of lectins to the complete amino acid sequencing and elucidation of their 3D structures.

In this case, Concanavalin A led the field, being the first lectin whose primary sequence was established. As previously mentioned, this was soon followed by the determination of the structure of WGA as well as of its complexes with its ligands (*N*-acetylneuraminic acid and β-1,4-linked *N*-acetylglucosamine oligomers). During the past few years, the number of lectin primary and 3D structures has increased dramatically, with about 200 of the latter having been elucidated. In addition, many structures of lectin–carbohydrate complexes have been solved [2].

1.1.3 Higher fungi: an abundant source of lectins

In 1910, after the discovery of the first fungal lectin in Amanita muscaria (*Agaricus muscarius*), also known as fly agaric, by Ford, the research on their toxicology suddenly stepped up, proving the absence of a direct link between the poisoning effect of a mushroom and its lectin content [3].

Since the amount of lectins in mushrooms tends to be higher than their content in higher plants, they must play an exceptionally important role not only in growth and morphogenesis, but also in dormancy, morphological changes, and molecular recognition during the early stages of mycorrhization [4].

It is worth mentioning that 82 % of the fungal lectins has been found to be present in higher mushrooms, being equally distributed in the cap, in the stipes and the mycelia, with the remaining remaining 18 % found either in microfungi or in yeasts.

Although their biochemical properties, alongside with their biological activities and localization have already been determined, only a few of them have been thoroughly characterized from the structural point of view, mostly because their primary sequence is often unknown [5-7].

1.1.4 Common fold and association in fungal lectins

In a recent review written by Imberty and colleagues [7], the number of already characterized fungal lectins was estimated to be around 26, including more than one hundred three-dimensional structures, among which 40 % was in complex with ligands.

Although this number represents only 9 % of the total structures of lectins present in the 3D Lectin database (http://glyco3d.cermav.cnrs.fr/search.php?type=lectin), more than ten different folds have been discovered: highlighting the unequivocal relationship between their organization and function. Among them, both classical lectins and CBMs can be recognized, being spread not only in the Fungi domain, but also in Plant, Animal and Bacteria.

Apart from the different folds belonging to the CBM class (β-sandwich and β-trefoil), which will be described later in paragraph 1.5, also galectins, actinoporin-like fold, cyanovirin-N like fold, yeast adhesion fold, and β-propeller domain have been characterized. A description of the main representative for each class is reported below.

Galectin fold

Galectins are a family of conserved β-galactoside-binding proteins widely distributed in vertebrates, invertebrates, and some fungi. Galectins, isolated both from *Coprinopsis cinerea* (CGL2) and *Agrocybe aegerita* (AAL), have been shown to possess the characteristic fold composed

of two parallel 6 stranded antiparallel β-sheets forming a β-sandwich. In particular, one of them is assigned to the binding of β-galactosides, as shown in figure 2. Although most of the members of this family are dimeric, *C. cinerea* lectin assemblies to form a four-leafed clover.

Actinoporin-like fold

The actinoporin-like fold is common among fungal lectins, especially those isolated from the fruiting bodies, of which BEL (*Boletus edulis* lectin), XCL (*Xerocomus chrysenteron* lectin), ABL (*Agaricus bisporus* lectin), and SRL (*Sclerotium rolfsii* lectin) are members. It is composed of two β-sheets made up of six and four β-strands respectively connected by a helix-loop-helix motif, forming a β-sandwich. Lectins belonging to the actinoporin-like fold associate to form dimeric or tetrameric structures, showing two different binding sites for each monomer, usually placed on each side of the helix-loop-helix motif: the primary site is specific for GalNAc, while the secondary binds GlcNAc.

Cyanovirin-N like fold

Besides the actinoporin-like fold, fungal lectins can also associate to form the so called cyanovirin-N like fold (CNVH), being homologous to a small viricidal lectin produced by the cyanobacteria *Nostoc ellipsosporum*. They are mostly monomeric with only one copy of CNVH, such as the *Giberella zeae* lectin (GzCVNH), which consists of two CVNH repeats related by a pseudo-two-fold symmetry axis referred to as domain A and B (type I CNVH).

Although carbohydrate binding can take place in both domains, often only one is functional. After the isolation of the *Magnaporte oryzae* lectin, the existence of a second sub-family of CNVHs was demonstrated, and it was called type III CNVH, characterized by displaying a LysM domain

between the two CNVH repeats. The importance of the LysM domain is due to the fact that it is able to selectively recognize chitin oligomers, differently from the CNVH part that is responsible for oligomannoside binding.

Yeast adhesin fold

Several yeast adhesins, especially those involved in flocculation and epithelial adhesion, present a N-terminus lectin-like domain with a β-sandwich that resembles the PA14 domain from the anthrax toxin protective antigen. Differently from the previously described fold, the yeast adhesin requires the presence of a calcium ion as cofactor to accomplish carbohydrate binding in a surface loop, which is bound in a surface loop through a D*cis*D motif. In certain cases, such as Flo5 from *Saccharomyces cerevisiae*, there is also a secondary binding site composed of five short β-strands that are responsible for the discrimination between various sugars.

β-propeller fold

The only lectin from a human fungal pathogen whose structure has been determined is AFL1 from *Aspergillus fumigates*, which adopts a 6-bladed β-propeller fold. In particular, each blade is composed of four antiparallel β-strands that arrange around a pseudo-six-fold axis and the β-propeller is closed as a zipper through the hydrogen bonds. The AFL1, as well as for the AAL (*Aleuria aurantia* lectin) is dimeric and possesses the ability to bind fucose at the interfaces between two blades.

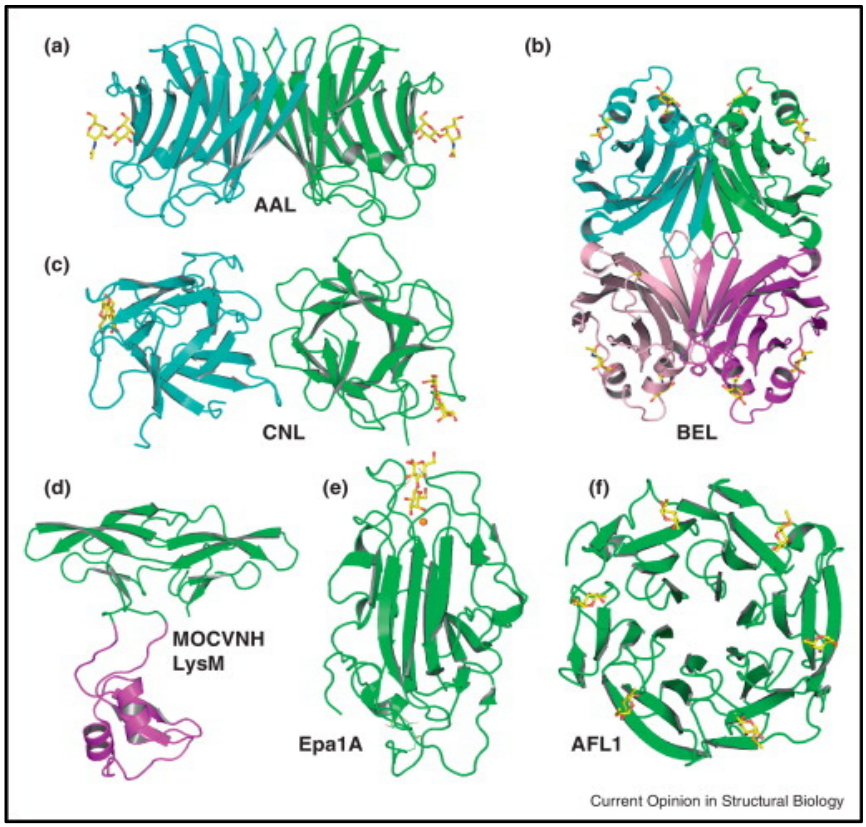


Figure 2: Graphical representation of the most common folds of fungal lectins, coloured to identify monomers. Carbohydrates are represented in yellow. From a to f: galectin fold (AAL), actinoporin-like fold (BEL), β-trefoil fold (CNL), cyanovirin-N like fold with LysM domain from *Magnaporthe oryzae*, yeast adhesin fold from *Candida glabrata*, and β-propeller fold of AFL1 [7].

1.1.5 Applications of lectins in biotechnology

During the last decades, fungal lectins have been intensively used in research laboratories for glycomics studies, including the production of several microarrays which could represent emergingtools for high-throughput analysis of glycans and glycoproteins.

In addition, lectins have been proposed as one of the most promising agents to be used against insect pests, being engineered into various crops, such as rice, tobacco, potatoes, and wheat. Since the implementation of transgenic herbicide-tolerant plants does not display noteworthy side

effects, this might be the correct solution to reduce the volume of toxic compounds released into the environment.

Moreover, certain types of lectins, such as *Arisaema jacquemontii*lectin, have demonstrated anti-insect activity, increasing the mortality or delaying the development of insects. The latter may be due to the control of the enzymatic activity of larvae, leading to an increase in esterase activity and to a decrease in acid phosphatase activity [6].

Although the number of purified and characterized lectins is considerably high, only few of them have been shown to possess antifungal properties, which is the case of *Gastrodia elata*lectin that can be activated by the fungus *Thricoderma viride*, spreading in the vascular cells of both the roots and the stems, helping to fight infections [8].

However, as previously described in paragraph 1.1.1, the most interesting feature of lectins is their ability to bind, specifically and in a reversible way, certain carbohydrate epitopes exposed on human cell surfaces. In particular, some epitopes are overexpressed in pathological situations, such as inflammation, early stage cancer, and metastasis. Recently, a lectin isolated from *A. aegerita*, specific for GlcNAc, has been shown to possess antitumor properties towards hepatoma cells, being potentially useful for cancer diagnosis.

Moreover, due to their natural source origin as well as their high specificity, the activity of lectins has already been tested both *in vivo*, *in vitro* and in several human case studies, demonstrating the correlation between their binding patterns and their biological effect in various tumours.

Specifically, the agglutination process, known as one of the main events responsible in arresting tumoral progression, is primarily done by binding to the glycoprotein receptors on cell membranes, causing the interruption of any further migration. Thanks to this, lectins can affect

not only cancer cells, but also the immune system, being able to alter the production of various protein kinases, interleukins and other proteins, as described in paragraph 1.3.

Lately, lectins have been shown to affect both the autophagy and the apoptosis processes, by modulating representative signalling pathways involved in the Bcl-2 family, caspase family, p53, PI3K/Akt, ERK, BNIP3, Ras-Raf and ATG families [9].

1.1.6 Fungal lectins and cancer

As described by Tannock I.F. in 2001, solid tumours display a hypoxic inner core in which cells are in their plateau phase, either non-cycling or slowly progressing through their cycle. Unfortunately, since tumour cells are also quite resistant to cell cycle-specific chemotherapy, the latter is more efficient towards cells in exponential growth, rather than those in their lag phase.

In order to make the treatment more efficient, the physiology of the tumour should be carefully considered, as well as its components and cell types; classification of antineoplastic agents by their selective toxicities toward oxygenated and hypoxic tumour cells is therefore of paramount importance [10].

In the light of the above considerations, it might be worth exploring different ways to treat cancer progression and, in this regard, the distinguishing features of lectins could be of great relevance. As previously described in paragraph 1.1.2, malignant cells display an altered glycosilation on their surface, either caused by neo-synthesis or incomplete synthesis, and glycan modification is responsible for promoting cell mobility, cell adhesion and contributing to the invasive phenotype. The presence of shorter glycans, such as the Thomsen-Friedenreich epitope, also known as T antigen (Galβ1-3GalNAcα1-Ser/Thr), has been observed not only in 90 % of human carcinomas,

but it has also been shown to be related to the early stages of cancer. Although the mechanism of the production of short glycans is not clearly understood, the most corroborated theory highlights the mutation of a chaperone which is required for the galactosyl-transferase activity that modifies O-linked glycans [11].

Interestingly, in normal epithelium, the T antigen is concealed by the addition of sialic acid residues, as well as for sulphates or other sugar chains to form complex O-glycans. In this context, several fungal lectins have shown to interact with the T antigen and to exhibit antiproliferative activity on various tumour cell lines [12-14].

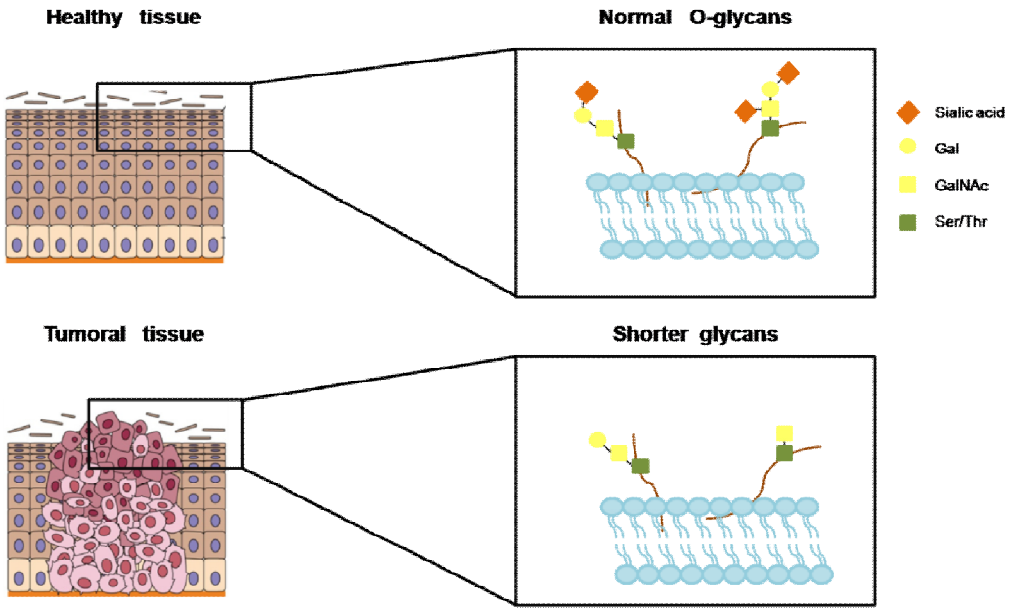


Figure 3: Cartoon representing the sugars exposed on malignant cells surfaces, compared to those on healthy cells.

Therefore, the use of lectins in fighting cancer progression might have limitless potential, ranging from the possibility to exploit their own antiproliferative activity, to the possibility of creating complex systems, such as nanoparticles.

Antiproliferative/Antitumor Activity			
Source of Lectin	**IC_{50}**	**Cell Type/Target**	**Ref.**
Agaricus bisporus (ABL)	50 μg/mL *	HT-29	[50]
Agrocybe aegerita (AAL galectin)	NA	S-180, Hela, SW480,SGC 7901, MGC80-3, BGC-823, HL-60	[56]
Amanita phalloides (*A. phalloides* lectin)	1.7 μg/mL *	L1210	[66]
Armillaria luteovirens (ALL)	2.5 μM 5 μM 10 μM	MBL2 HeLa L1210	[69]
Boletopsis leucomelaena (BLL)	15 μg/mL	U937	[182]
Clitocybe nebularis (CNL)	NA	Mo-T, Jurkat	[79]
Cordyceps militaris (CML)	0.5–0.6 mg/mL *	HepG2	[183]
Flammulina velutipes (*F. velutipes* lectin)	13 μM	L1210	[86]
Ganoderma capense (*G. capense* lectin)	8 μM 16.5 μM 12.5 μM	L1210 HepG2 M1	[88]
Grifola frondosa (GFL)	25 μg/mL *	HeLa	[184]
Hericium erinaceus (HEA)	56.1 μM 76.5 μM	HepG2 MCF7	[94]
Inocybe umbrinella (*I. umbrinella* lectin)	3.5 μM 7.4 μM	HepG2 MCF7	[99]
Lactarius flavidulus (LFL)	8.90 μM 6.81 μM 7.4 μM	HepG2 L1210 MCF7	[106]
Paecilomyces japonica (PJA)	NA	SNU-1, AsPc-1, MDAMB-231	[130]
Paxillus involutus (*P. involutus* lectin)	NA	A-549, HCT-8	[185]
Pholiota adipose (PAL)	2.1 μM 3.2 μM	HepG2 MCF7	[133]
Pleurotus citrinopileatus (*P. citrinopileatus* lectin)	5 mg/kg of body weight/day ¥	S-180 in ICR mice	[137]
Pleurotus eous (PEL)	2 μg/mL 50 μg/mL	MCF-7, K562, HEP-2 SK-N-MC	[139]
Pleurotus ostreatus (POL)	1.5 mg/kg bodyweight/day ¥	S-180, H-22	[32]
Polyporus adustus (*P. adustus* lectin)	NA	M1, Herto, S180	[144]
Psathyrella asperospora (PAL)	0.48 μM	HT29	[186]
Russula delica	0.88 μM 0.52 μM	HepG2 MCF 7	[154]
Schizophyllum commune (SCL)	30 μg/mL	KB	[156]
Tricholoma mongolicum (TML1 & TML2)	NA	P815, PU5-1.8	[187]
Volvariella volvacea (VVL)	17.5 mg/kg body weight ¥	S-180	[188]
Xerocomus chrysenteron (XCL)	NA	Hela, NIH-3T3	[189]
Xylaria hypoxylon (*X. hypoxylon* lectin)	1.24 μM NA	M1 HepG2	[162]

Figure 4: Table representing the antiproliferative/antitumor and fitogen activity of some mushroom lectins [6].

1.1.7 Examples of lectins with anticancer activity

During the last decades, a large number of lectins with anti-proliferative effect has been discovered, among which a very interesting one is that isolated from *Agaricus bisporus*, the most common edible species in Western countries. The *Agaricus bisporus* lectin (ABL) is a 64 KDa homo-tetramer, structurally characterized in the Biocrystallography Laboratory of Verona. Besides its high affinity for the T antigen, ABL has also been shown to possess antiproliferative activity towards human epithelial cancer cells. Its monomer is a single domain composed of six β-strands that arrange themselves to form a β-sandwich. There are two sugar binding sites per monomer with different specificity, interacting 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).

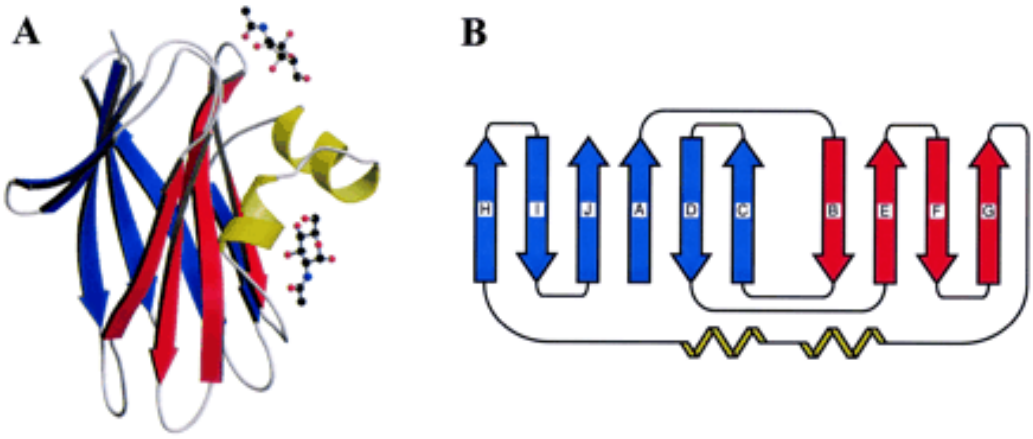


Figure 5: Overall structure and folding of ABL. A) Ribbon representation of the monomer and ball-and-stick representation of the binding sugar N-acetylgalactosamine. B) Topological diagram of the monomer [15].

Another lectin displaying comparable antitumoral properties was purified from *Boletus edulis*, and it was named BEL-β-trefoil because of its characteristic fold. Overall, the protein is a dimer and the

fold of each monomer is that of the β-trefoil family, presenting a β-barrel with a pseudo-three-fold symmetry axis. Each monomer contains three subdomains, called α, β, and γ, respectively, having three potential binding sites (Figure 6). Isothermal titration calorimetry assays, alongside with the three-dimensional structure elucidation have shown in detail the interaction of BEL with various sugars, such as lactose, galactose, N-acetylgalactosamine, and the T antigen.

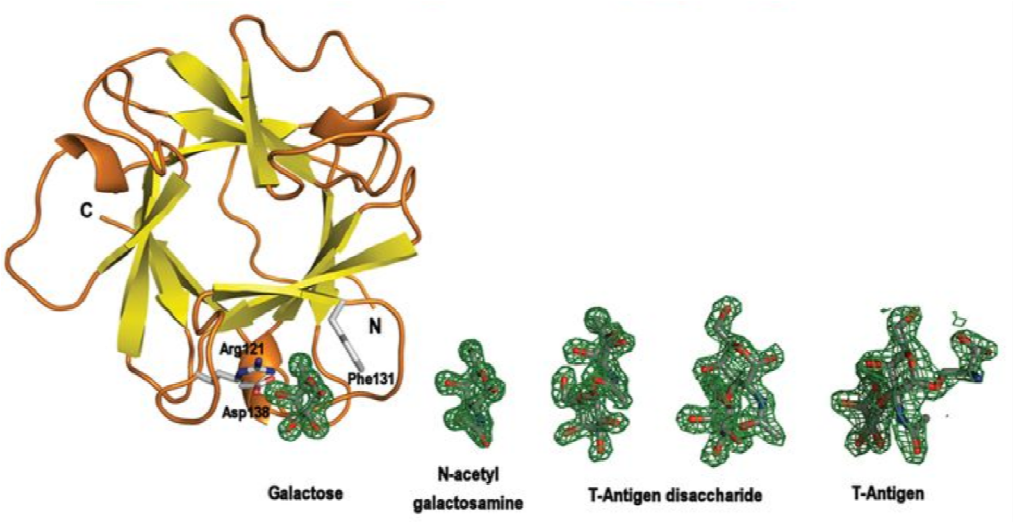


Figure 6: Ribbon representation of a BEL-β-trefoil protomer with electron density of the different saccharides that can occupy the binding site [13].

More recently, a 373 amino acid lectin was isolated from the fruiting bodies of *Pleurotus ostreatus* (POL), also commonly known as edible oyster mushroom; the lectin has the ability to inhibit the growth of human neoplastic cells. The dimeric protein is composed of 22 anti-parallel β-strands, ten of them forming the N-terminus domain, while the remaining strands form the C-terminus domain. The two domains are connected to each other via a 21 aminoacid loop. POL was tested on both human pancreatic cancer cells (MiaPaCa-2) and human hepatocyte carcinoma (HepG2), displaying an evident therapeutic effect [Laura De Stefanis - unpublished data].

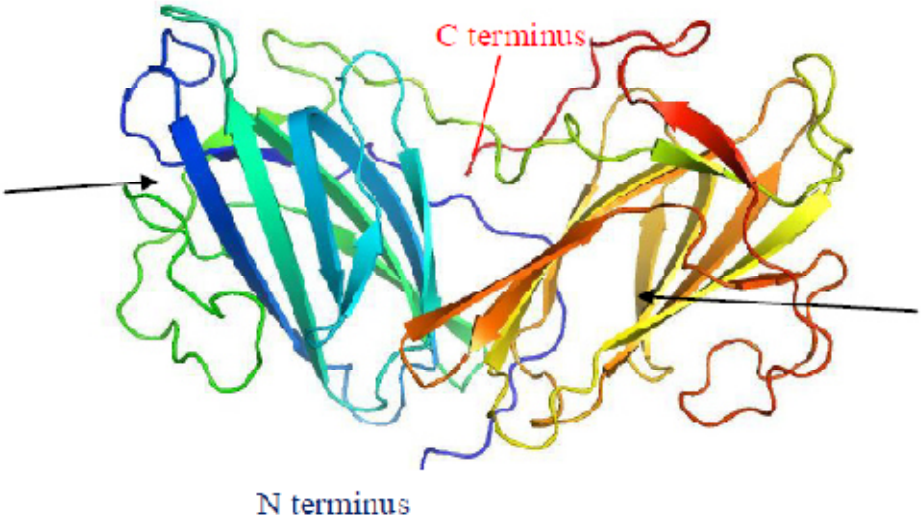


Figure 7: Ribbon representation of POL. The black arrows indicate the cylindrical cavities of the two domains [Laura De Stefanis – unpublished data].

1.2 CBMs and lectins

According to the literature [16], carbohydrate-binding module (CBM) is just one among the many families of proteins involved in the synthesis, modification, recognition, and binding of carbohydrates. As already mentioned, sugar recognition is achieved by lectins, a family of proteins that, thanks to their diversity and selectivity for sugar recognition, is able to perfectly cope with the complexity of carbohydrates.

Interestingly, CBMs and lectins share not only the ability to recognize sugars, but also their presence in all the kingdoms of life, the mechanisms of binding and, most important, fold similarity.

However, the two families are separated by a very fine line, sometimes making it difficult to distinguish between them. Since they share amino acid sequence similarity, several lectins are

classified as CBMs, which is the case of ricin toxin B from *Ricinus communis* and Wheat Germ Agglutinin that reside in family CBM13 and CBM18. In addition, the human lectin malectin, which plays a role in binding and processing the N-linked glycan in endoplasmatic reticulum [17], is part of the family of CBM57.

The most prominent feature that distinguishes CBMs from lectins is the ability of the first to participate in the synthesis and modification of carbohydrates, in addition to their recognition. Whereas the most commonly known property of lectins is their ability to agglutinate cells containing carbohydrates or glycoconjugates on their surfaces, CBM25 and 26 from *Bacillus halodurans* have been shown to provoke a strong agglutination effect on amylopectin.

Overall, the line between the two families still remains not completely defined and subject of study, therefore further consideration will be needed to make it clearer.

Since lectins are proteins that bind carbohydrates it can be stated that all lectins contain at least one carbohydrate binding domain, while CBMs are domains that can also be present in enzymes, as we will see in paragraph 1.3.

1.3 Carbohydrate-protein interactions: a general outlook

The surface of mammalian cells is dominated by the presence of the glycocalyx, also known as pericellular matrix, a glycoprotein layer that surrounds the cell membranes. Among glycan features, the most important is probably being responsible for the interactions between cells and their environment, influencing them at different levels: from molecular to subcellular functions.

Looking deeply, protein glycosylation takes part in several stages of cell growth, is fundamental for lymphocyte activation, and its machinery alteration is related to the development of tumoral phenotypes. Besides that, carbohydrates also mediate cell-adhesion, coordinate certain

hormoneactivities, and orchestrate the immune system, producing fundamental signals for T-cell growth and arrest.

Glycans can also be potent antigens, with the human immune system distinguishing between them at a very fine level of detail, as it is for the non-human Neu5Gc that, differently from the Neu5Ac, can be recognized by the presence of the extra “O”, causing a massive immune response.

In addition, small carbohydrate moieties decorate the surface of erythrocytes, giving important features to the different blood types, such as the compatibility for transfusions and donations.

To provide more context, pathogens are able to exploit carbohydrates as a binding epitope for the infection of human cells, being able to discriminate between differences in their localization and changes in their structure. Considering that it requires different forms of sialic acid to decide which animal species will be infected, influenza virus is probably the most suitable example of how pathogens can take advantage of the sugars exposed on the cell surface.

Since not enough is known on carbohydrate-binding proteins, scientists have struggled with their three-dimensional structure determination, being one of the challenges of the last decades [18]. However, understanding the 3D structures of a molecule is one of the fundamental prerequisites to fathom the relationship between proteins, elucidating their functions, and comprehending the biochemistry that regulates their recognition processes.

Unfortunately, the flexible nature of carbohydrates, both in their free form or as a part of a glycoconjugate, makes them difficult to crystallize, therefore X-ray experiments are still sparse and sporadic. Spectroscopic methods, which can be used keeping the protein in solution, have been relatively more successful and, together with molecular modelling, are considered powerful tools for structural studies of these compounds. Nevertheless, some X-ray data of highly resolved protein-carbohydrate complexes have been collected from synchrotron sources, giving the

opportunity to compare the experimental structure with that predicted by computational methods.

Up to now, carbohydrate-binding proteins have been categorized in terms of their functions, giving rise to six different families that are summarized in figure8.

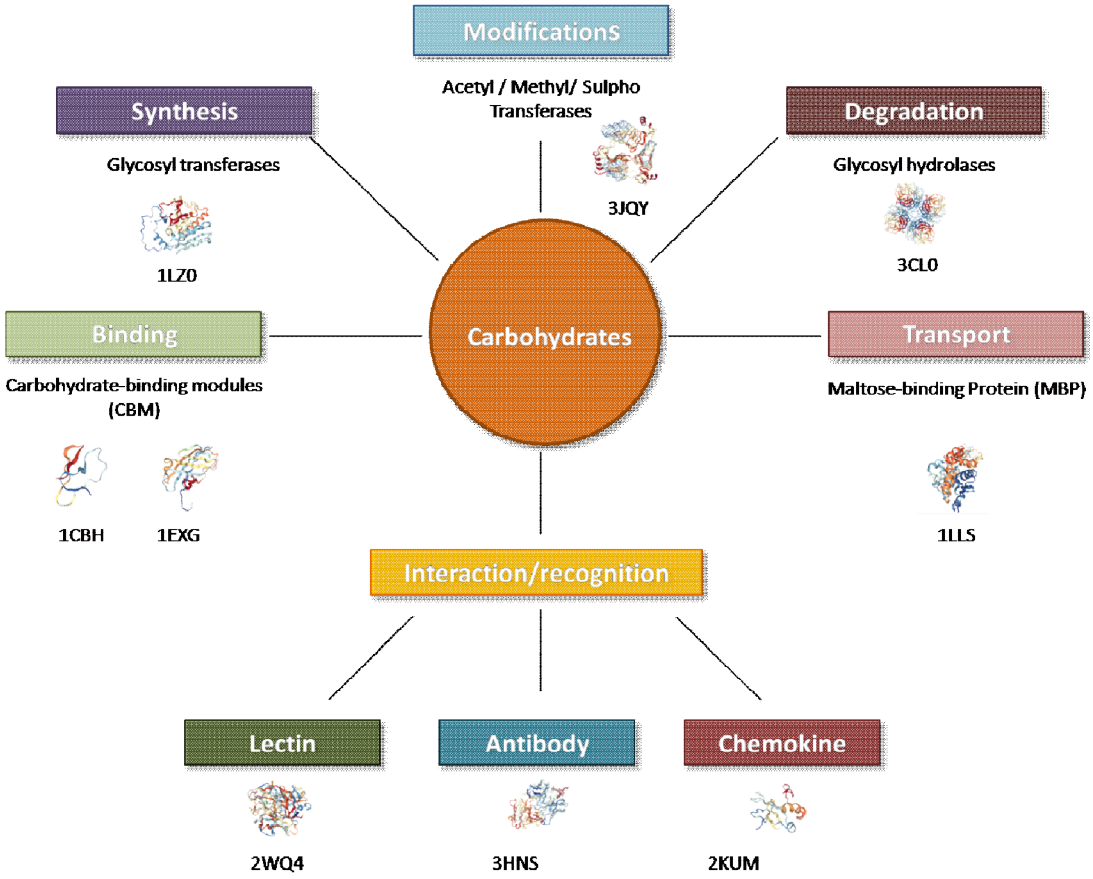


Figure 8: Graphical representation of carbohydrate-binding proteins categorization in terms of functions.

Glycosyltransferases (GTs) handle the synthesis of carbohydrates for both glycoproteins and glycolipids, by catalysing the transfer of a monosaccharide from a sugar donor to an acceptor, and are considered to be one of the most important enzymes in the human body [19]. However, transferases take also care of transferring sulphonyl, methyl or acetyl groups to acceptor molecules, generating sole ligands from carbohydrate substrates, and therefore playing a fundamental role in cell adhesion and signalling.

In addition, some carbohydrate-binding proteins are able to degrade glycosylic bonds of complex sugars, such as cellulose, hemicelluloses, and starch, representing, together with glycosyltransferases, one of the major catalytic machineries of the cell [20].

Interaction and recognition of carbohydrates are typical of three families of proteins, which we will refer to as: lectins, antibodies and chemokines.As previously described in paragraph 1.1, lectins are ubiquitous proteins of non-immune origin, being found in all the living organisms, especially in plants, animals and microorganisms, which specifically and reversibly recognize glycans and induce cell precipitation [7]. Differently, antibodies belong to the family of immunoglobulins, are produced mainly by plasma cells, and are responsible for neutralising pathogens by exploiting the presence of carbohydrates on their cell surfaces. Lastly, small cytokines called chemokines can interact with carbohydrates, giving rise to induction of directed chemotaxis and to the recruitment of immune system cells during infections.

Glycan transport can be accomplished by maltose-binding proteins, monomeric structures of around 42 kDa molecular weight that take care of an effective uptake and catabolism of both maltose and maltodextrin.

Carbohydrate-binding modules (CBMs) are protein domains present in carbohydrate active enzymes that can increase the hydrolytic rate on insoluble substrates by approaching the target to the catalytic module.

1.4 Insight into carbohydrate-binding modules

The CBMs were first termed as cellulose-binding domains for their ability to bind cellulose. Lately, they have been found to possess affinity for other numerous substrates, giving rise to the need of reclassifying these proteins with a more comprehensive definition.

A widely accepted definition of CBMs is a contiguous amino acid sequence within a carbohydrate-active enzyme with a discrete fold, having carbohydrate binding activity [21].

So far, more than 300 putative sequences in different species have been identified, and the binding domains have been grouped into 43 different families, depending on their amino acid sequence, structure and binding specificity [22]. Among these, there are CBMs able to recognize crystalline cellulose, non-crystalline cellulose, chitin, β-1,3-glucans and β-1,3-1,4-mixed linkage glucans, xylan, mannan, galactan and starch, while others exhibit lectin-like specificity and bind to cell-surface glycans.

The CBMs length usually varies between 30 and 200 amino acids and they exist as a single, double, or triple domain. Their location within the enzyme can be both C- or N-terminal and during the last decades some examples of centrally positioned CBMs have been identified.

Both hydrolytic and non-hydrolytic enzymes have been found to possess CBMs, such as xylanases and cellulases, and their presence was also demonstrated also for several polysaccharide-degrading enzymes. For instance, hemicellulase, endomannanase and acetylxylanesterase isolated from *Trichoderma Reesei* contain at least one CBM, and the same is true for the esterase from *Penicillium funiculosum* and for pectate lyase from *Pseudomonas cellulose*. Interestingly, in 2005, Yoshida and colleagues isolated a carbohydrate-binding cytochrome from the White-Rot Fungus *Phanerochaete chrysosporium* [23].

Expansins, which are believed to play a role in non-hydrolytic cell wall expansion, are proteins of about 26 kDa molecular weight, homologues to CBMs, that possess cellulose binding activity. They were isolated for the first time in cucumber seedlings but subsequently their presence was confirmed also in other plant tissues, whose activity was related to cell growth. In addition, expansins are also involved in root growth, especially in response to drought, a feature that might be exploited for manipulating gene expression in crop plants [24].

1.5 Most common folds and structures of CBMs

1.5.1 Fold relationships among different CBMs

During the last decades, about 22 structures of members of CMB family have been solved, making it possible to understand something more about their fold, their structural similarities and properties of the ligand binding sites.

Fold family	Fold	CBMs families
1	β-sandwich	2, 3, 4, 6, 9, 15, 17, 22, 27, 28, 29, 32, 34, 36
2	β-trefoil	13
3	Cysteine knot	1
4	Unique	5, 12
5	OB fold	10
6	Hevein fold	18
7	Unique with hevein-like fold	14

Table I: Summary of the CBM families, categorized by fold similarities.

So far, all the CBMs characterized have structures based on different β-sheets topologies and have been manually classified into seven different fold families (Figure 9), by exploiting DALI, a software

that provides pairwise alignment algorithm for comparing protein structures. Among them, some of the most important CBM families are described below, following table I [22].

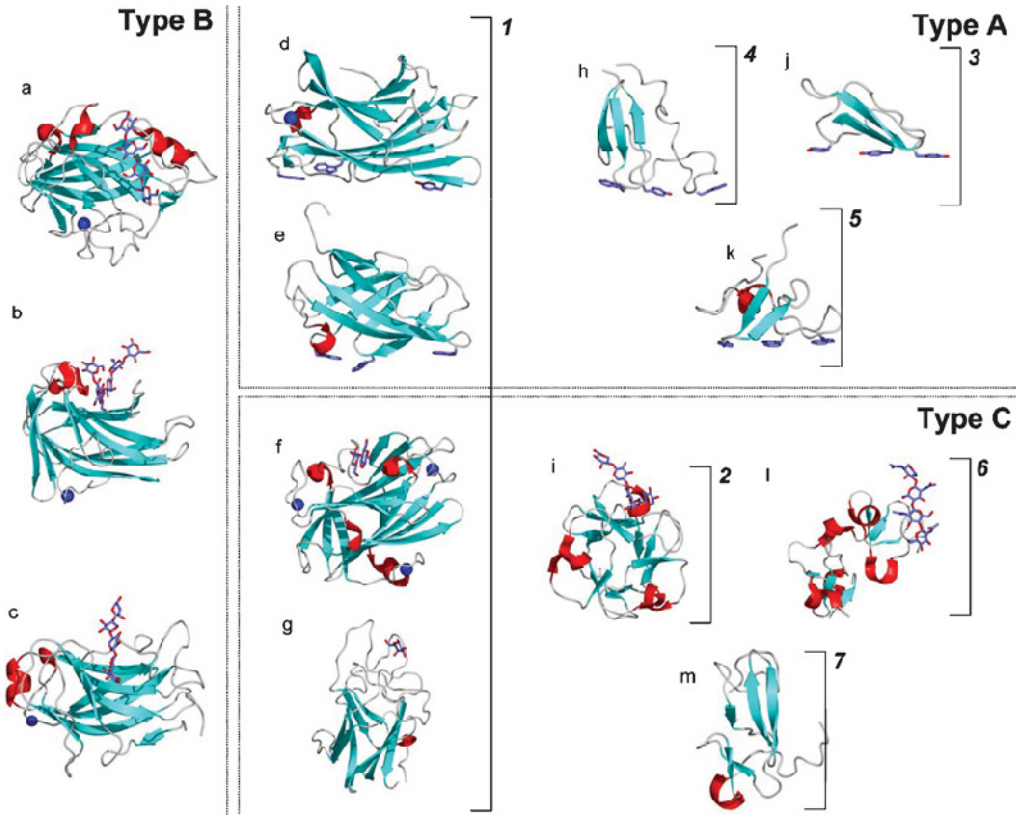


Figure 9 : Dotted lines separate examples of CMBs categorized in terms of different functions (Types A, B and C). Backets with numbers indicate examples of CBMs belonging to the seven fold families (1-7). Starting from family 1 to family 7: β-sandwich, β-trefoil, cysteine knot, unique, OB fold, hevein fold, and unique-containing hevein-like fold. Bound metal ions are represented as a blue spheres, while ligands are depicted in the "liquorice" representation [22].

The β-sandwich fold

It is probably the most abundant fold among CBM family, composed of two β-sheets, made up of three to six antiparallel β strands. Although no sequence similarity has been indicated with other proteins, the CBMs family shares the β-sandwich fold with plant legume lectins, animal galectins, calnexins, and pentraxins. With some exceptions, all the CBMs belonging to this fold have at least

one bound metal ion, which in most cases appears to have a structural function. For some CBMs the presence of a metal ion is required for ligand binding, and this is the case of the *Paenibacillus polymixa* CBM, whose structure in complex with xylotriose and Ca^{2+} has been solved by Jamal-Talabani and colleagues. After solving its 3D structure, the unique calcium-dependent binding was evident, showing the coordination of the central xylose moiety to a hepta-coordinated calcium ion, through the O2 and the O3 hydroxyls. In the native structure, on the other hand this position was replaced by Mg^{2+} [25].

Most of the proteins with a β-sandwich fold show a precise localization of the ligand-recognition site, positioned on the same face of the β-sandwich (Figure 10). However, although families 6 and 32 display the same fold, their ligand-binding site is placed at the edge of the β-sandwich.

The β-trefoil fold

The β-trefoil is the second most frequent fold within the CBM family, and it is commonly associated with the ricin toxin B-chain, a heterodimeric plant lectin whose structure was determined by Rutenber et al. in 1991 [26]. It is characterized by 12 strands of β-sheet, forming six hairpin turns. A β-barrel is formed by six of the strands, together with three hairpin turns, while the other three hairpin turns form the so called “hairpin triplet”. Each subunit of the fold can be seen as a contiguous amino acid sequence with two-hairpin structures and four β-strands, giving rise to the characteristic trefoil shape with a pseudo-3-fold axis. Thanks to the 3-fold symmetry, some of the CBMs are prone to bind more than one carbohydrate molecule and, taking advantage of the β-trefoil multivalence, to interact simultaneously with more than one glycan, as is the case of the ricin toxin B-chain, *Streptomyces lividans* and *Streptomyces olivaceoviridis* xylanases.

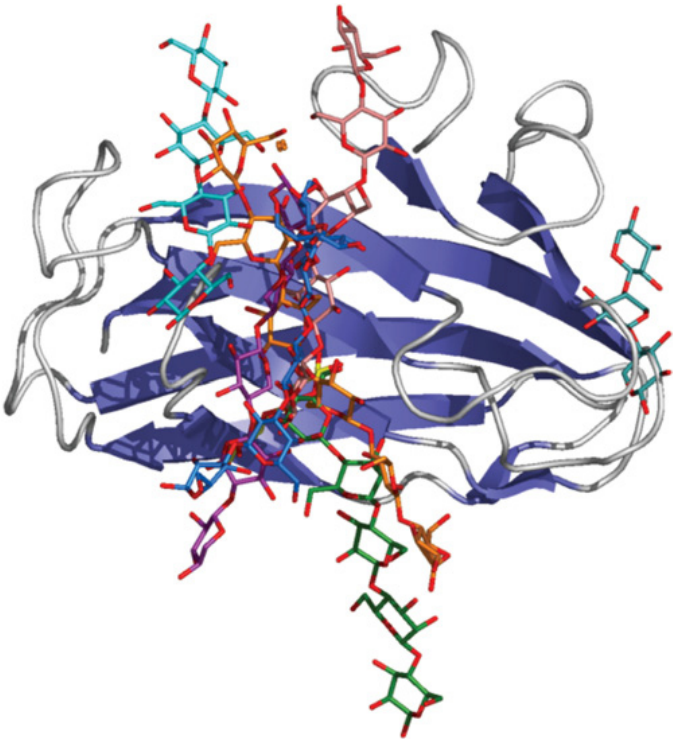


Figure 10: Cartoon representing the location of ligand-binding site in the β- sandwich CBMs (Type B). The picture was produced by overlapping the Cα carbons of all the members of the family whose structure in complex with ligands was available. The ribbon representation of Cel9B from *Cellulomonas fimi* is shown as a representative of the β-sandwich family. Ligands are coloured as follows: laminariohexaose from TmCBM4-2 (blue, PDB code 1GUI), cellohexaose from PeCBM29-2 (orange, PDB code 1GWM), mannopentaose from TmCBM27 (green, PDB code 1OF4), xylotriose from CsCBM6-3 (aqua, PDB code 1NAE), xylopentaose from CjCBM15 (purple, PDB code 1GNY), cellotetraose from CcCBM17 (cyan, PDB code 1J84), cellopentaose from CfCBM4-2 (pink, PDB code 1GU3) [22].

The hevein fold

The CBMs belonging to the family of the hevein domain are small, composed of approximately 40 amino acids, and were identified for the first time as chitin-binding proteins. They are mainly composed of coil, with a slight presence of β-sheets and α-helix, as depicted in figure 9. Despite their dimension, proteins having this fold can bind astonishingly large carbohydrates, as was reported for Wheat Germ Agglutinin (WGA), that is able to bind the chitotetrasaccharide.

The cellulose-binding and OB (oligonucleotide/oligosaccharide binding) folds

The members of the OB and the cysteine knot fold are small polypeptides composed of 30-60 amino acids, whose structure contains only β-sheet and coil. The majority of these CBMs have been shown to possess planar carbohydrate-binding sites with aromatic residues, able to bind preferably cellulose and chitin.

In addition, in different protein structures solved using NMR, no obvious hydrophobic planar surface that could contain a cellulose-binding site has been detected. Furthermore, in CBM10 from *Cellvibrio japonicas* Xyn10A, the xylan binding site is on a different face compared to other proteins belonging to the same family, probably indicating a degree of convergent evolution between them.

1.5.2 Relationships between structure and function of CBMs

Despite the fact that CBMs can be grouped in seven families, according to fold similarities, they display significant differences in terms of aminoacid sequence, binding sites, and functional elements, therefore an additional characterization, based on structural and functional similarities, has been proposed.

Accordingly, CBMs have been divided into three different types, summarized in table II and figure 9: surface-binding CBMs (Type A), glycan-chain-binding CBMs (Type B), and small-sugar-binding CBMs (Type C).

Type	Fold family	CBMs families
A	1, 3, 4, 5	1, 2a, 3, 5, 10
B	1	2b, 4, 6 15, 17, 20, 22, 27, 28, 29, 34. 36
C	1, 2, 6, 7	9, 13, 14, 18, 32

Table II : Summary of the CBM families, categorized by fold similarities.

Surface binding CBMs – Type A

This class of CBMs, to which families 1, 2a, 3, 5, and 10 belong, is characterized by the ability to bind insoluble and highly crystalline carbohydrates, such as chitin and cellulose. Despite the fact that the presence of aromatic amino acids within the binding site is well-conserved among CBMs, the architecture of the binding site platform is not, being flexible and complementary to the ligand. Type A CBMs have been shown to possess limited or no affinity for soluble carbohydrates, but unique features in binding crystalline ligands, whose interaction is driven by positive entropy and thermodynamic forces [27].

Glycan chain binding CBMs – Type B

During the last decades, X-ray crystallography and NMR data have been exploited in order to better understand the nature of the CBMs carbohydrate-binding sites. Data collected have revealed that they are extended in Type B, comprising different sub-domains able to bind individual sugar units and appearing likely to be either grooves or clefts.

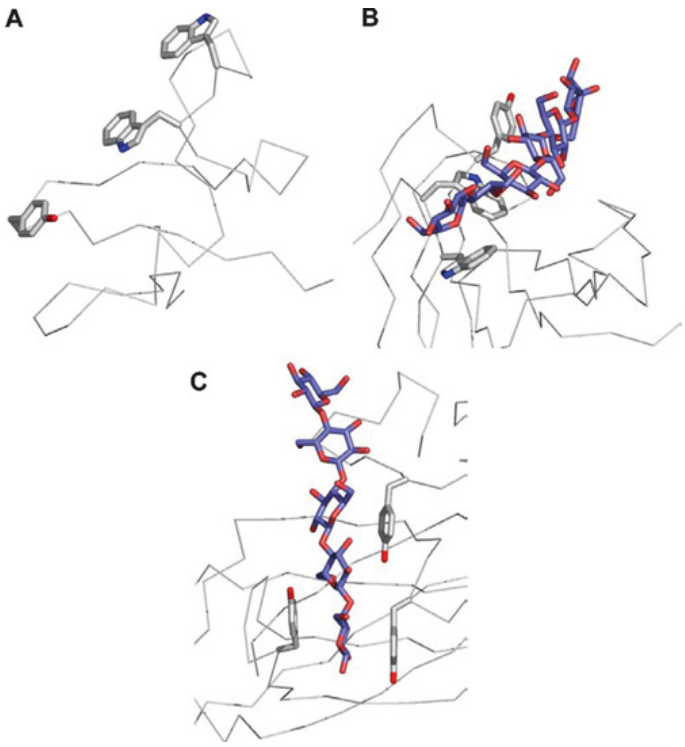


Figure 11 : Different types of binding site depending on aromatic amino acid organization. The C-α backbone is depicted as grey cylinders, while the aromatic amino acids are shown as grey liquorice representations, as well as for the bound oligosaccharide, which is represented in blue. A) Planar platform characteristic of family 10 of Type A CBMs. B) Twisted platform, displayed by family 29 of Type B CBMs. C) Sandwich platform highlighted in family 4 of Type B CBMs [22].

Small sugar binding CBMs – Type C

Over the last years, the use of CBMs in biotechnology has been established and confirmed not only by experimental work, but also by the publication of hundreds of papers and patents, giving rise to the public interest for the topic.

This particular class of CBM possesses the ability to bind mono-, di- and tri-saccharides, displaying lectin-like behaviour and lacking the extended binding-site groove typical of Type B CBMs. Although it may be extremely difficult to discriminate between Type B and C CBMs, their hydrogen-bonding network presents some dissimilarities, with the Type C CBMs having the more extensive net, in agreement with their lectin-like features.

Interestingly, members of families 13 (Ricin-toxin chain-B), 14 and 18 (Wheat Germ Agglutinin), which belong to Type C CBMs, were first categorized as lectins, thanks to their ability to bind small sugars. However, after the discovery of their presence as domains also in glycoside hydrolases, researchers have started to consider them, together with families 9 and 32, as part of type C CBMs.

1.6 *Tremella fuciformis*: a Chinese mushroom with antitumoral properties

Among all the different mushrooms that produce lectins, during the last decades one in particular has drawn attention due to its interesting properties; a fungus we will refer to as *Tremella fuciformis*, also known as white-jelly or silver-ear mushroom. Due to its anticancer [28], anti-inflammatory [29], anti-oxidant [30], and neuroprotective activities, it has been used for a long time in Chinese medicine. *Tremella fuciformis* is rich in content of proteins, dietary fibers, vitamins, chitin, tri-terpenoids and polysaccharides, whose structure consists of a linear backbone of (1→3)α-D-mannan with side chains composed of fucose, xylose and glucuronic acid (9:1:4:3).

1.7 Aim of this thesis

Since some *Tremella fuciformis* proteins have been investigated for their potential therapeutic properties and, according to the literature, have been shown to possess anticancer activity, one of the aims of this thesis was to assess its crude extract looking for the presence of one or more lectins. After identifying a reliable and efficient purification method, particular attention was paid

to the determination of the three-dimensional structure of the lectin isolated using X-ray diffraction of single crystals. In addition, the lectin ability to bind sugars was studied by using isothermal titration calorimetry and, a fine characterization of its organization and a study of its behaviour in different media was undertaken using differential scanning calorimetry and fluorescence spectroscopy. *Tremella fuciformis* lectin (TFL) was also biologically characterized by testing its effect on various tumoral cell lines.

In addition, the expression of recombinant *Tremella fuciformis* lectin was set up in order to improve the yield and to open the way to protein engineering.

After the development of the purification protocol and the complete biological characterization, the sample turned out to be a good candidate to be used for *in vivo*analyses, specifically for tumoral targeting. During the last decades, in fact, many studies on this subject were conducted, in which lectins were used to direct nanoparticles for targeted drug delivery [31]; more specifically the use of a PLGA or chitosan support is of interest. PLGA is a copolymer widely used in medicine and its employment is approved by the Food and Drug Administration, mainly because of its biocompatibility, biodegradability and ability to cross the blood brain barrier (BBB). On the other hand, chitosan is a polysaccharide prepared by treating crustacean shells, it is also biodegradable and biocompatible, able to cross the BBB [32] but has not been approved by the FDA yet.

2. Materials and Methods

2.1 From protein purification to structure determination

Understanding the function of different proteins, especially if they have important biological implications, is one of the key features of structural biology, a branch of molecular biology that has gained remarkable attention during the last decades.

Nowadays, among all the methods available for elucidating protein structure, such as nuclear magnetic resonance (NMR), small angle X-ray Scattering (SAXS) and cryo-electron microscope, X-ray crystallography is probably the most reliable. However, X-ray crystallography involves a series of delicate steps that lead the researcher to face and deal with a considerable number of issues: selection of the target molecule, purification, cloning, expression and crystallization are only few of them (Figure 12).

Since high protein concentration is required to prepare diffraction-quality crystals, purification is often one of the earliest problems to cope with. Moreover, the presence of flexible tails and interdomain regions, as well as the solubility of the sample, contribute to making crystallization one of the most critical bottlenecks in the entire workflow. In addition, even in the presence of particularly suitable samples, crystallization still remains a complex and multiparametric procedure that requires meticulous optimization.

2.1.1 Purification

Ideally, after identifying the protein to work with, a purification protocol should be set up by choosing one or more biochemistry techniques among those available.

According to the literature [33], every protein can be purified using its charge, solubility, size and binding affinity.

By increasing the salt concentration, a method referred to as “salting out”, protein solubility can be extremely affected, causing selective protein precipitation and separation from the rest of the mixture. Salts can then be removed from the samples by using dialyses membranes. Gel-filtration chromatography can also be used, leading to the separation of proteins based on their size. In this method molecules flow at different velocities inside the column gel and the largest come out first, as a smaller volume is accessible to them. Usually, a gel-filtration column - also known as size-exclusion column – consists of porous beads of dextran, agarose or polyacrylamide, which are insoluble but highly hydrated polymers. In addition, a net charge at a certain pH, either positive or negative, can be exploited to perform ion-exchange chromatography. Specifically, positively charged proteins can be separated by using a negatively charged resin, such as carboxymethyl-cellulose (CM). Oppositely, a positively charged resin, such as diethylaminoethyl-cellulose (DEAE), is often used to isolate proteins having positive net charge.

Lastly, specific binding properties can also be particularly useful in purifying proteins, allowing to perform the so called “affinity chromatography” and retaining only molecules that bind tightly and covalently to the specific resin.

2.1.2 Crystallization

Once isolated, the pure protein fraction can be concentrated and used for performing crystallization trials, a step in which a large number of conditions are screened in order to find the most promising that can lead to crystal growth.

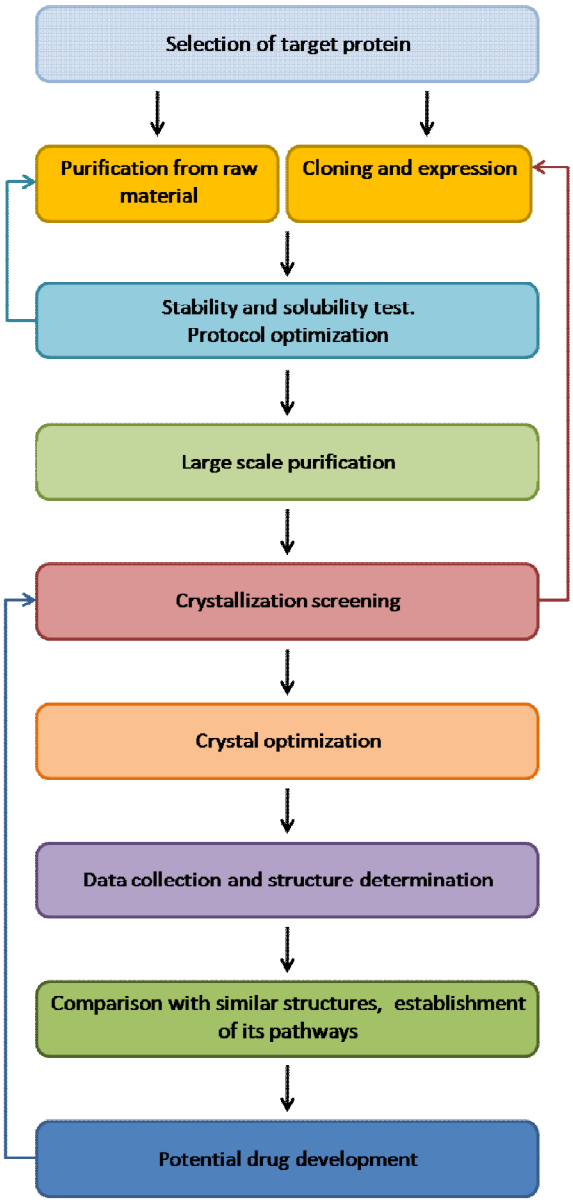


Figure 12: Protein purification and crystallization workflow. Two main bottlenecks are highlighted by arrows: protein purification and crystallization screening (Adapted from Chayen N.E. and Saridakis E., 2008).

Since crystallization involves many delicate stages, it is considered one of the most complicated processes within the structural biology field. The phases common to all molecules that undergo a crystallization process are: nucleation, crystal growth, and cessation of growth. Nucleation is the most difficult to achieve, both theoretically and experimentally, because it involves the transition between a wholly disordered state and an ordered one.

In addition, macromolecular crystallization requires a supersaturated environment, that is nothing more than a non-equilibrium condition, in which some molecules, that exceed the solubility limit, are present in solution. Assumedly, this occurs through the formation of paracrystalline or partially organized intermediates, leading to the yield of small and ordered assemblies, which are referred to as “crystal nuclei”. The new-formed nuclei offer the ideal surface for crystal growth, which is a process better understood than nucleation. In order to obtain a supersaturated environment, the relationship between molecules and solvent in the crystallization media is strongly perturbed, promoting solid state formation. This may be achieved by applying several strategies, such as: the alteration of protein properties (pH or ionization state) and the modification of interactions between molecules (addition of ions). Some of them are listed in the table below.

1	Direct mixing – Batch method
2	Temperature alteration
3	Salt concentration alteration – Salting out
4	pH alteration
5	Water removal
6	Macromolecule concentration
7	Solubilizing agent removal

Table III: Different methods for obtaining a supersaturated environment [34].

As soon as a low protein concentration in solution is reached, crystals stop immediately to grow. A summary of the crystallization process is presented in figure 13, which represents a phase diagram.

Although a supersaturated environment should be easily achieved, it might not be a sufficient condition for crystals to grow, therefore it is advisable to try as many strategies as possible. Ideally, in order to screen the highest number of parameters as possible, a matrix of different conditions should be carried out. Currently, combining the variety of solutions offered by the companies, together with the speed, the accuracy and the high precision of a liquid handler, a large number of different combinations can be investigated, thus increasing the possibility to obtain crystals.

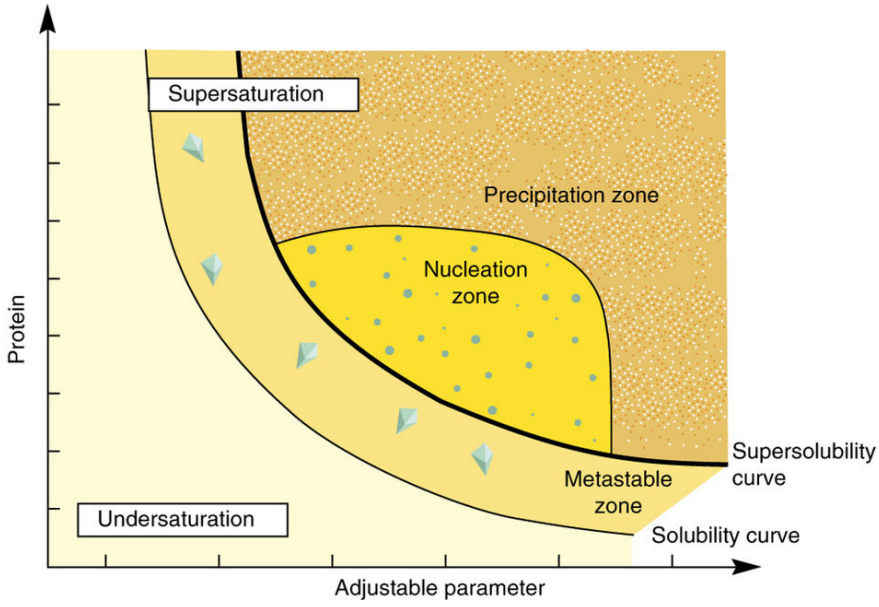


Figure 13: The phase diagram for macromolecule crystallization can be sharply divided in three different zones: the undersaturated zone, the supersaturated zone and the precipitation zone. In turn, the supersaturation region can be split into metastable region, in which the nuclei will develop to form crystals, but no nucleation will occur and the labile one, which both processes are expected to occur. The precipitation zone, is a zone of very high supersaturation in which precipitates might be expected to form. Crystals can only be grown from a supersaturated solution, and creating such a solution supersaturated in the protein of interest is the immediate objective in growing protein crystals [35].

In order to grow protein crystals, a physical device is required, to allow to easily alter the solubility parameters. Hanging drop and sitting drop procedures for vapour diffusion represent the most popular and recommended methods for the majority of investigations. In both experiments, a droplet of protein is placed into a sealed container, either hanging or sitting and equipped with a reservoir, which contains the same buffer but at a higher salt concentration than the protein solution. At the beginning, the precipitants concentration in the reservoir is higher than in the protein droplet, causing a situation of under-saturation. As water evaporates from the sample drop into the reservoir, an increase in protein and precipitant concentration into the drop arises, which causes a shift toward a supersaturated system. Although these two methods seem essentially very similar, they differ in terms of protein droplet orientation. Specifically, in the hanging drop method the protein is placed right above the reservoir, while in the sitting drop, it is positioned horizontally, in a separate micro-well, as illustrated in figure 14.

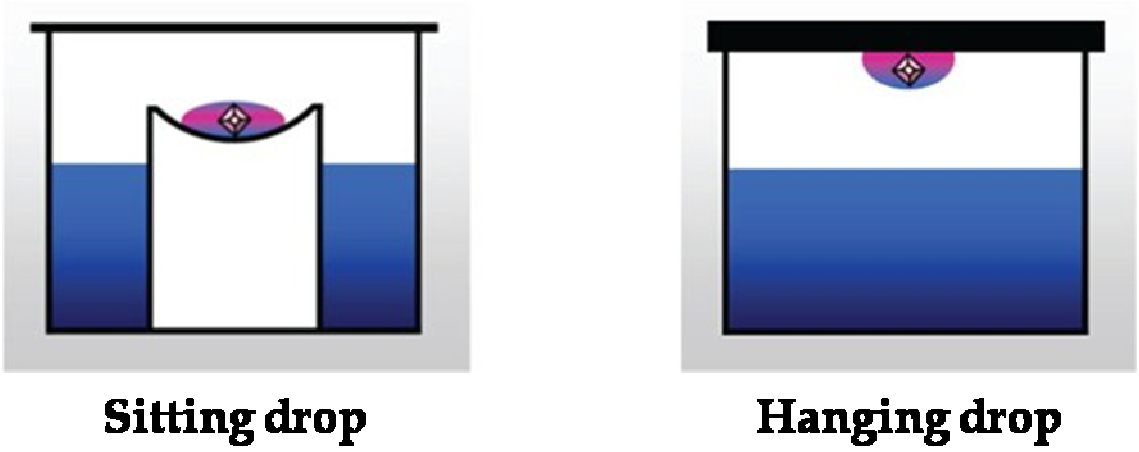


Fig 14: Sitting drop and hanging drop techniques.

Although protein crystallization is considered the mostdifficult goalto reach, structure solution can also be quite complex. Specifically, the other stumbling block is phase determination, a key step during the process of protein structure solution.

2.1.3 Solving the phase problem

As briefly discussed in paragraph 2.1.2, one of the main bottlenecks in protein structure elucidation is represented by the phase problem. Briefly, after recording the diffraction pattern, only intensities and therefore amplitudes can be obtained, while any information regarding the phases is lost. Unfortunately, to calculate the electron density, both amplitude F_{hkl} and phases ϕ_{hkl} of the structure factors are required, as can be seen from the following equation:

$$\rho(xyz) = \frac{1}{V} \sum_{h=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} |F_{hkl}| e^{2\pi i \phi_{hkl}} e^{-2\pi i(hx+ky+lz)}$$

where V stands for the volume, while $|F_{hkl}|$ for the structure factor amplitude and ϕ_{hkl} for its phase.

To overcome this problem, four methods are available:

1) Patterson synthesis;
2) Direct methods;
3) Heavy atom isomorphous replacement;
4) Anomalous scattering.

The first is a Fourier summation based only on the experimentally observed intensities $F(hkl)^2$, which is particularly suitable for small molecules. Briefly, the outcome is a vector map that can usually result in a rough structural approximation, which is further improved by therefinement process.

Structure elucidation by means of direct methods is, on the other hand, based on mathematical relationships between the reflections, without the need of intuitive or subjective steps.

Differently from the previously described methods, both the isomorphous replacement and the use of anomalous scattering rely on the presence of heavy atoms in the crystal, as markers that provide information about phases. Although these methods share acommon ground, the theory beyond them is slightly different in terms of signal source and data available.

The isomorphous replacement method requires the introduction of at least one heavy atom in the sample, which can be accomplished via the so called “soaking process”. The sample thus prepared, differs from the nativeonly by the presence of an electron density peak corresponding to the heavy atom site, thus leading to observable changes in the X- ray diffraction pattern. The intensity difference between the native crystal and derivative, rising from datasets collected at the same wavelength, can be further exploited in order to solve the phase problem, by applying the following formula:

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

where $F_{PH}$is the structure factor of the heavy atom derivative and $F_P$that of the native protein. Consequently, F_{PH} can be derived by the vector addition of F_P to F_H, which is the contribution of the heavy atoms to the structure factor of the derivative.

Depending on how many derivatives are used, single isomorphous replacement (SIR) or multiple isomorphous replacement (MIR) can be performed [36,37].

However, a heavy atom can also act as a source of anomalous signal, causing an imbalance between Friedel’s pairs, from which the calculation of the heavy atom substructure can be performed. Briefly, after the heavy atom introduction, the atomic scattering factor can no longer

be considered as a real number, thus leading to the breakdown of Friedel's law, as can be seen from the following equation:

$$F_{PH}(+) \neq F_{PH}(-)$$

$$F_{PH}(+) = F_H(+) + iF''_H(+) + F_P(+)$$

$$F_{PH}(-) = F_H(-) + iF''_H(-) + F_P(-)$$

Depending on how many different wavelength datasets are collected, single anomalous diffraction (SAD) or multiple anomalous diffraction (MAD) can be performed.

Beside the above-mentioned methods, the "phase problem" can also be overcame by combining the isomorphous replacement together with the anomalous signal from the same crystal (SIRAS) or from multiple derivatives (MIRAS).

Depending on how many different wavelength datasets are collected, single anomalous diffraction (SAD) or multiple anomalous diffraction (MAD) can be performed.

Beside the above-mentioned methods, the phase problem can also be solved by combining the isomorphous replacement together with the anomalous signal from the same crystal (SIRAS) or from multiple derivatives (MIRAS).

2.1.4 Heavy atom derivative preparation

As briefly discussed in paragraph 2.1.3, in order to perform phasing, either the introduction of at least one heavy atom in the crystal, or the presence of native scatterers (i.e. sulphur), is required. However, since each protein crystal exists as two phases – a solid phase of molecules packed in an

open lattice and a liquid phase of solvent occupying the channels – the addition of heavy atoms might represent an issue in terms of crystal packing and protein structure perturbation. Fortunately, during the last decades, a broad number of strategies have been developed in order to carry out the experiment by displacement of the solvent only, without disturbing the crystal structure. The most common and rational approach is the replacement of an already existing amino acid residue with a synthetic analogue containing a heavy atom, or the further addition of heavy compounds in order to promote their reactions with surface amino acids.

Following the first strategy, the heavy atom analogues must resemble the replaced residue, both in term of charge and shape. The most widely diffused and successful example is selenomethionine, introduced in the position of methionines in proteins that can be expressed recombinantly. Two common examples of heavy metals that bind to proteins are mercury, that may be substituted onto aromatic rings, and iodine, which may form good but unstable aromatic derivatives [38,39].

The direct binding of the heavy atom salts to amino acid side chains, is often a non-specific reaction, involving their incorporation by either co-crystallization or soaking. Since the first method involves the incubation of the protein with the heavy compound before crystallization, it is considered a quite risky procedure, that can lead to protein conformational changes and non-isomorphous crystal formation. Even if soakingof the crystal with heavy atom salts seems to be the most reliable method, it might still give rise to difficulties, such as crystal decay or loss of isomorphism.

Beside the previously described methods, another derivatization strategy, which can be extremely useful in the case of delicate crystals, is the quick soak with heavy halides or heavy alkali solutions [40]. Unlike the previously described method, the quick soak allows to reach higher concentration

(up to 1 M) in a very short period of time (usually less than 1 minute), which in turn helps to cryoprotect the crystals and to reduce the risk of damages.

Although the perfect method for preparing heavy atom derivatives does not exist, but it is rather a "trial and error" process, some general guidelines can be followed in order to maximise the chances to obtain a good diffraction quality crystal.

In 1976, Blundell and Johnson divided the heavy atoms into two categories, based on their charge and ability to form bonds: "hard" and "soft". Following their guidelines, hard ligands bind to their target without covalent modification, while the soft ones tend to covalently bind different groups, such as: thiol, imidazole and sulfhydryl. Whereas platinum, gold and mercury atoms are part of the "soft" ligand category, lanthanides are examples of "hard" ligands, binding their target via carbonyl groups and side chains containing oxygen.

In particular, some heavy atoms are able to specifically recognise and bind certain side chains, which is the case of K_2PtCl_4 that reacts mostly with methionines and of many mercury compounds that tend to bind to cysteines. In particular, K_2PtCl_4, $KAu(CN)_2$, K_2HgI_4, $UO_2(C_2H_3O_2)$, $HgCl_2$, para-chloromercury benzoic acid sulfonate (PCMBS) and $K_3UO_2F_5$ have been indicated as the most powerful compounds for crystal derivatization, by Boggon and Shapiro, being renamed the "magic seven" [41].

However, after choosing the heavy atom compound to use, a large number of other parameters - pH, buffer, concentration, and hydrophobicity of the environment - must be considered in order to perform crystal derivatization [36,37].

2.1.5 Data collection and analysis

Once good crystals are produced, harvested and soaked into a cryo-protectant solution, they are ready to be exposed to X-ray radiation. In order to do that, they are mounted onto cryo-loops and positioned under the X-ray beam by a goniometer, which enables the operator to adjust the orientation of the crystal with respect to the laboratory coordinates. In addition, the sample must be kept at a suitable temperature during the data collection, in order to avoid radiation damage (100 K).

The actual routine data collection, especially within synchrotron facilities, is highly automated, and uses a strategy planner that can make the experiment extremely efficient. After collecting only a few images, crystal symmetry and orientation, unit cell parameters, and crystal resolution limit can be calculated from the orientation matrix, allowing to implement the collection strategy in order to maximize both the resolution and the completeness of the dataset.

However, a certain number of decisions have to be made by the user and all the parameters have to be checked before starting the data collection, alongside with a fine tuning of the beam wavelength in case anomalous scatterers are present.

After collecting the datasets, every frame undergoes integration, a step that enables the position of each reflection and its intensity to be recorded.

Subsequently, the intensities obtained from different frames are combined in one set of structure factors and scaled together [42].

Since the intensities are proportional to the square of the structure factor amplitudes, a complex number which contains information about both the amplitudes and the phases is required for

electron density computation. Unfortunately, as briefly described in paragraph 2.1.3, the phases cannot be directly recorded during the X-ray diffraction experiment, giving rise to the phase problem. As previously discussed, a wide number of methods are available to overcome this bottleneck, which allow to calculate the phases and to calculate the electronic density.

Depending on the quality of the electron density maps, model building can be more or less demanding, an issue that is particularly important especially in the case of the models that are affected by bias, such as thoseresulting from molecular replacement. The map of an experimentally phased electron density, on the other hand, gives the operator the possibility to start to build the entire model from scratch.Once a contiguous backbone is built, the individual residues can be fitted and adjusted into the electron density and real space refinement performed to select the appropriate conformer.

Subsequently, in order to correct various and small errors generated by this step, the model is subjected to the reciprocal space restrained refinement, which involves the global minimization of the differences between the observed and calculated structure factors. The two methods are alternated in multiple rounds, until a final model containing as few errors as possible, is obtained.

In the end, an incomplete model will have empty electron density blobs that can be filled by adding water into only peaks above a certain density level and not closer than 2.3 Å to other atoms. The thus-obtained model undergoes a restrained maximum likelihood refinement, in which the wrongly positioned atoms are moved away from the incorrect position and only those placed in real density are kept in their original position. When no more new atoms need to be added, the model is ready for inspection, the last step of model building, in which accuracy of coordinates and R-values are examined.

2.2 *Tremella fuciformis* lectin purification and structural studies

The starting material for protein purification was the edible mushroom *Tremella fuciformis*, a mushroom that is commercially cultivated only in China. Due to its gelatinous aspect and white basidiocarps, alongside with the fact that it can be found on dead trunks, it is also commonly known as “snow fungus” or “silver ear mushroom”. In addition to its high content in terpenoids, fibers, vitamins and chitin, *Tremella fuciformis* is also rich in proteins and polysaccharides, two of its most important bioactive compounds. Interestingly, its polysaccharide matrix has been shown to have not only a significant hypoglycaemic activity on diabetic mice, but also to act as radio protector against 8-Gy gamma irradiation damages [43].

In addition, the crude extract of *Tremella fuciformis*has been found to possess several beneficial effects on the body; among the most important are the anti-tumoral, the anti-inflammatory, the antioxidant, and the immune-regulating properties [28-30, 44].

In the light of the above considerations and using the scarce information available in the literature, an extraction and purification protocol was set up in order to isolate the lectin we will refer to as TFL (*Tremella fuciformis* lectin).

2.2.1 Purification protocol

Fifty grams of dried mushrooms were ground to a fine powder in a blender and 500 mL of 20 mM Tris with 0.1 M NaCl pH 7.5 were added to the powder and let to stir for 5 hours. The insoluble material was removed by centrifugation (17000 g x 40 minutes) and 250 mL of buffer were added

to the pellet in order to perform a second overnight extraction. After joining the supernatant of both extractions, a precipitation with 80% saturated ammonium sulphate was carried out to separate the desired protein from the supernatant. The pellet, separated by centrifugation (17000 x 60 minutes), was resuspended in 10 mL of 20 mM Tris buffer pH 7.5 and dialysed against the same buffer in order to dissolve the precipitated proteins.

The lectin-containing-extract was loaded onto a hog gastric mucin affinity column which was thenextensively washedwith 20 mM Tris buffer pH 7.5, until the OD at 280 nm became negligible. The lectin was eluted from the column by changing the pH of the elution buffer to 11.4, which was then replaced by a neutral pH buffer in order to allow the concentration process. The solution was concentrated under nitrogen pressure using an Amicon Millipore with 3.5 KDa cut-off ultra-filtering membrane. The next purification step used a Sephacryl S-100 HR gel filtration column (composed of cross-linked copolymer of allyl dextran and N,N'-methylene bisacrylamide). The gel filtration step was performed using an ÄKTA PRIME system, which allowed to set up the detection wavelength (280 nm) and the other chromatographic run parameters (flow: 0.5 mL/min, 1.5 mL fraction).

An alternative second method to purify TFL that did not use the affinity column, was to load the crude extract onto a diethylaminoethyl cellulose column (DEAE) previously equilibrated with the same buffer. The column was washed until reaching a negligible absorbance at 280 nm and then TFL was eluted by applying a linear gradient between the initial buffer and 20 mM Tris with 200 mM NaCl. Following the ion-exchange column, a size-exclusion chromatography was also performed as described above. The presence and the purity of TFL were assessed with SDS-PAGE 15%.

2.2.2 Crystallization trials and heavy atom derivative preparation

Purified TFL was concentrated to 20 mg/mL by using Centricon devices (Millipore). Crystallization trials were set up in eight 24-well plates, using both the hanging-drop and the sitting-drop vapour-diffusion methods, at 20 °C (192 different conditions, from Molecular Dimension Limited, were tested).

After filling up the reservoir with 200 µL of precipitant solution, 1 µL of the precipitant was mixed with the same volume of protein solutionand the dropwas deposited on a previously silanized cover slip, used to seal the reservoir. When the sitting drop technique was used, the reservoir was filled with the same amount of precipitant solution, and 3 to 5 µL of it were mixed together with the same volume of protein solution and placed on the horizontal well, as previously described in paragraph 2.1.2.

After setting up the trials, half of them were stored at 20 °C and the remaining at 4 °C, in order to monitor the effect of the temperature on crystal growth. In addition, crystallization conditions were modified by screening the addition of different percentages of DMSO and ethylene glycol into the mother liquor, starting from 0.5 % up to 2 %.

In order to solve the phase problem, due to the lack of a suitable probe to perform molecular replacement, once crystals grew, the preparation of heavy atom derivatives was necessary for phasing. Thus,crystals were soaked with Platinum, Mercury and Gold compounds and transferred to the cold room (4 °C), to allow the process of molecule derivatization to take place. Briefly, a tiny amount of heavy atom powder was dissolved in different drops containing the crystals and letto react for 15 minutes up to 2 hours, at 4 °C [45].

All the crystals, both before and after derivatization, were then immersed into the cryoprotection solution (90 % mother liquor and 10 % glycerol), mounted onto a loop and flash freezed in liquid nitrogen for X-ray exposure.

For data collection aiming at describing the protein binding sites, the same protocol was applied to sugar soaking: in particular, glucose, lactose, sucrose and N-acetylgalactosamine were tested.

2.2.3 X-ray data collection and structure determination

After soaking in cryoprotectant solution and mounting onto cryo-loops, crystals were positioned under the X-ray beam using the instrumentation available at the ESRF.

Final datasets of native TFL were collected at the beamline BM-30A of the European Synchrotron Radiation Facility (ESRF) in Grenoble, while heavy atom derivative datasets were collected at beamline ID-29, particularly suitable for performing high resolution MAD and SAD.

After collecting a couple of preliminary images, an automated indexing was performed and the orientation matrix determined, allowing to obtain several fundamental data, such as: unit cell parameters, crystal orientation, resolution limit and space group. Using these data, an accurate data collection strategy was designed, with the aim of maximizing both the completeness and the resolution of the resulting data set. In every case, the crystal was rotated recording small angle frames, so as to improve the signal to noise ratio.

After predicting the position of the reflections on the frames, the data sets underwent the integration process, leading to the intensity of each reflection on the image. Combining and

scaling together the integrated values from different images, the structure factor amplitudes were obtained [42].

The TFL data sets were indexed using multiple strategies, with both XIA2 and XDS and MOSFLM, to better control the result [46].

After indexing and integrating, the program AIMLESS was run in order to scale and merge multiple observations into an average intensity.

As previously described in paragraph 2.1.3, the electron density cannot be obtained from intensities only, therefore a further calculation is required, assuming that the square root of the intensity is proportional to the structure factor amplitude ($|F_{hkl}|$). The next step of solving the phase problem was approached in this particular case, with the method of Single Isomorphous Replacement with Anomalous Dispersion (SIRAS). The structure was solved using the program CRANK2 (via Buccaneer) pipeline of JsCoFE suite was used and yielded not only the substructure determination, but it was also used for density modification and model building. Briefly, CRANK2 is the successor of CRANK, a program for automatic X-ray structure solution of proteins of unknown fold, particularly suitable for weak anomalous signal and low resolution data.

The preliminary model obtained by CRANK2 was used as a probe to perform a structural superposition with GESAMT, a tool that enables to find structural homologous by using a pairwise algorithm. After having analysed the output of GESAMT one of the chains was identified (PDB code: 4AVR)to perform Molecular Replacement (MR) using high resolution native data. At this stage, since initial phases needed to be improved, a density modification was performed using the software Parrot and implementing not only solvent flattening, but also histogram matching and non-crystallographic symmetry averaging.

Refinement of the final model obtained by Parrot, was carried out by RefMac 5, with manual revisions using the program Coot. In the end, solvent molecules were added to the model in order to obtain better statistics.

2.3 Molecular biology techniques used to produce recombinant TFL

With the aim of improving the yield and opening the way to further protein engineering, the recombinant form of TFL was produced, by incorporating its sequence (GenBank accession number EF152774) into a synthetic commercial gene.

Briefly, after removing the portion corresponding to the signal peptide and to the stop codon, the remaining fragment was optimized for *E.coli* expression and a synthetic gene was generated through the Invitrogen GenArt Gene Synthesis service.

The pet22b(+) vector was chosen for cloning purposes, a plasmid carrying ampicillin resistance, an N-terminal pelB signal sequence for potential periplasmatic localization, plus an optional C-terminal His-Tag. In addition, by using this vector, protein expression is controlled by the T7 RNA polymerase promoter, which can be induced by the addition of Isopropyl-β-thiogalactoside, also commonly known as IPTG. Since the vector does not present a thrombin cleavage site, this was added during primer design stage.

The sequence coding for TFL was amplified by polymerase chain reaction (PCR), by using the following primers specific for cloning into the pet22b(+) vector, between the NdeI and XhoI restriction sites:

TFL forward:

GGCGGCCATATGGATGACACCATTTATATC (5' → 3')

TFL reverse:

GGCGGCCTCGAGGCTGCCGCGCGGCACCAGATCGAATTCCCAGGTAAC (5' → 3')

Where CATATG is the restriction site for NdeI, CTCGAG is XhoI restriction site and GCTGCCGCGCGGCACCAG is the thrombin cleavage site.

The PCR was set up as following:

Step	Temperature (° C)	Time (sec)	
Inizialization	95	180	
Denaturation	95	30	34 cycles
Annealing	60	30	
Elongation	72	45	
Final elongation	72	300	

The PCR product was checked on a 0.8 % agarose electrophoresis gel, in order to confirm the expected size of the amplicon and subsequently purified by using the GenElute™ Gel Extraction kit, provided by Sigma Aldrich.

After digesting both the insert and the vector (pet22b(+) – figure 15) with the same restriction enzymes (NdeI and XhoI), a ligation reaction was carried out.

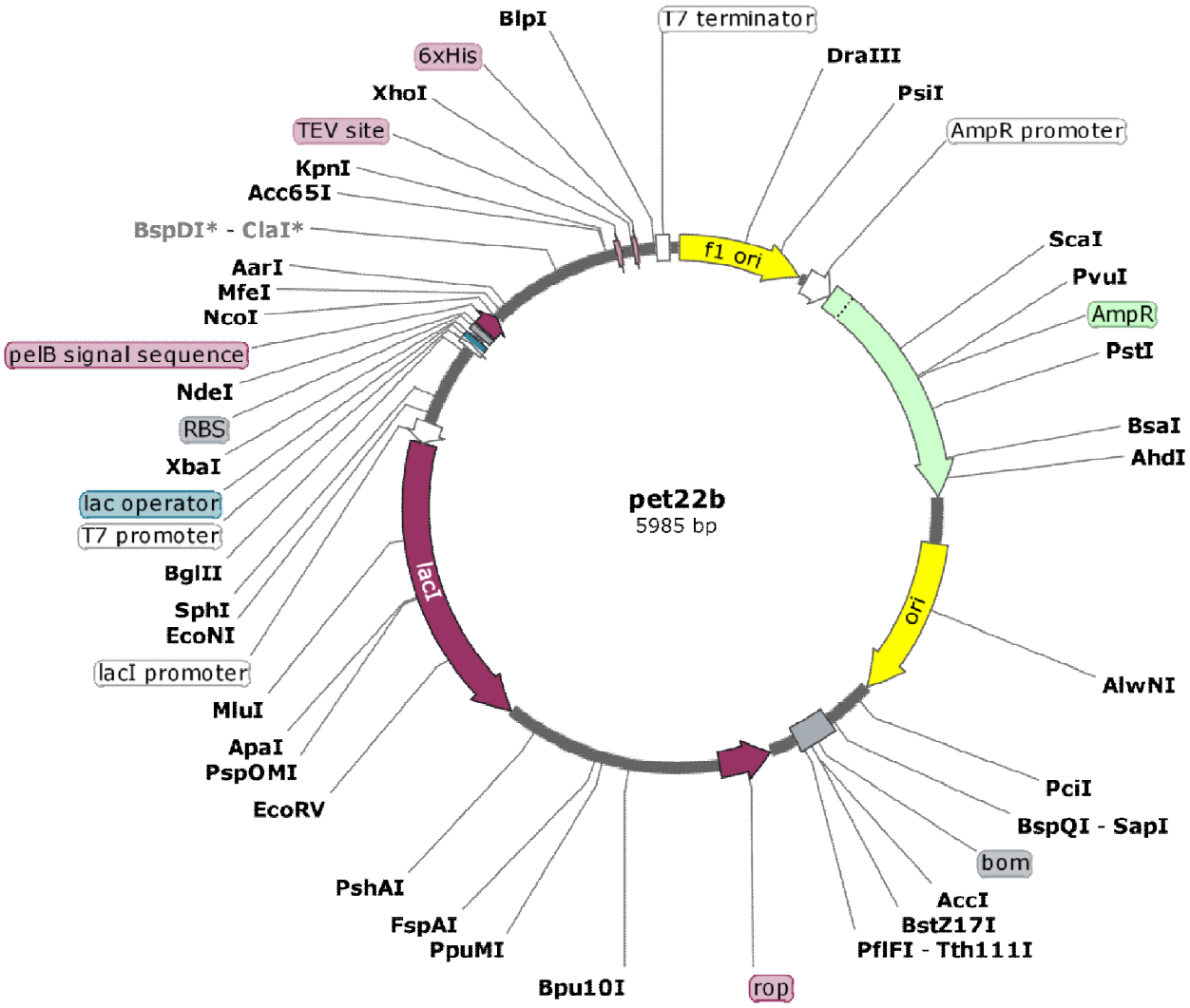


Figure 15: pet22b(+) vector map.

The fragment coding for the TFL sequence was inserted into pet22b(+) and used to transform, by heat shock, *E.coli* XL-1 Blue competent cells, a strain particularly suitable for cloning purposes. The cells were plated on Luria Bertani (LB) agar, supplied with ampicillin (100 µg/mL) for positive selection. Then, five to six colonies were selected and checked by colony PCR, in order to verify the

presence of the sequence of interest. Only positive colonies were grown in liquid LB supplied with ampicillin (100 μg/mL), at 37 °C, and purified by the GenElute^{TM} Miniprep Kit provided by Sigma Aldrich, with the aim of isolating the ones containing the fragment. Subsequently, BL21(DE3), Rosetta, Origami and SG *E.coli* expression strains were transformed by heat shock.

In order to monitor and make the protein expression more efficient, each cell strain was induced with 0.5 mM of IPTG (1:1000 dilution from 0.5 M stock) and two temperature were tested (20 °C and 28 °C).

The pellets isolated from each experiment were resuspended in appropriate buffer (20 mM Tris-HCl pH 7.5, 0.5 M NaCl).The lysing process was carried out by the addition of 1 mg/mL lysozyme, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 1 mM Triton X-100, followed by sonication cycles. Soluble and insoluble fractions were isolated by centrifugation at 8000 g for 15 minutes at 4°C and loaded onto a 15% SDS-PAGE electrophoresis gel. Protein expression was confirmed by Western Blot, using an antibody specific for His-Tag recognition.

After identifying the best conditions leading to the obtainment of soluble TFL, a 1 L culture was prepared and protein purification was carried out. A single colony of Origami *E.coli* was selected and grown in 10 mL of LB broth, supplied with ampicillin (100 μg/mL), at 37 °C. As soon as the cells started to replicate, 1 mL of the cell suspension was transferred into 1 L of the same medium, containing the equivalent antibiotic. After reaching anO.D. = 0.6-0.8, the culture was induced overnight with 0.5 mM of IPTG, at 20 °C.

The next day, after cell lysis, the sample was centrifuged at 8000 g for 15 minutes at 4 °C in order to separate the pellet from the soluble material, which was loaded onto a Nickel-Sepharose column, previously equilibrated with resuspension buffer (20 mM Tris-HCl pH 7.5 0.5 M NaCl). After washing the column until the absorbance (O.D. = 280 nm) was negligible, elution of the

protein was performed by an imidazole gradient, starting from 20 mM and increasing the concentration up to 400 mM, using a FPLC system. The fractions corresponding to the main peak were collected and checked on a 15% SDS-PAGE electrophoresis gel, to assure the presence of the protein of interest, which was later cleaved with thrombin (1 U/µL). Once concentrated, the protein fraction was loaded onto a Sephacryl S-100 HR gel filtration column and eluted with 20 mM Tris-HCl ph 7.5 and 0.15 M NaCl. Finally, the protein was concentrated to 20 mg/mL and crystallization screening was performed, following the same procedure previously described in paragraph 2.2.2.

In addition, recombinant TFL was expressed to obtain the selenomethione substituted protein, in non-methionine auxotrophic *E.Coli*. Since methionine feedback inhibition is promoted by the addition of the amino acid before the induction step, this protocol can be applied to any *E.coli* strain without the need for methionine synthesis deficit strains. A minimal medium, trace elements mix (100x) and vitamins mix (1000x) were prepared following the recipes below. After mixing all the components, the vitamin mix was carefully sterile filtered and kept chilled. Then, the trace elements mix was brought to pH 7.0 with NaOH, sterile filtered and kept chilled. In the end, both the vitamin mix and the trace element mix were added to the minimal medium, following the recipe below.

Vitamin mix (1000x) (0.5 L)

Riboflavin	0.5 g
Niacinamide	0.5 g
Pyridoxine monohydrate	0.5 g
Thiamine	0.05 g

Trace elements mix (100x) (1 L)

EDTA	5 g
$FeCl_3$	0.8 g
$ZnCl_2$	0.08 g
CuCl2	0.01 g
CoCl2	0.01 g
H_3BO_3	0.01 g
$MnCl_2$	1.6 g
$NiSO_4$	tip of spatula
Molybdic acid	tip of spatula

Minimal Medium (1 L)

Na_2HPO_4	6 g
KH_2PO_4	3 g
NaCl	0.5 g
NH_4Cl	1 g
$MgSO_4$ (1M)	2 mL
Glucose (20%)	20 mL
Vitamins mix (1000x)	1 mL
Trace elements mix (100x)	10 mL

Following the transformation of the TFL construct into Origami *E.coli* cells, a single colony was selected and grown in 5 mL of LB media, supplied with ampicillin (100 µg/mL), at 37 °C. As soon as the cells started to replicate, 300 µL of the cell suspension were transferred into 300 mL of minimal medium (1:1000 dilution), containing the equivalent antibiotic, and let grow overnight. The next day, 10 mL of the culture were transferred into 1 L of minimal medium (1:100 dilution). Once the O.D. = 0.6 - 0.8 was reached, 0.5 g of the following amino acids mix were added and mixed thoroughly with a spatula.

Amino acids mix

Lysine	1.0 g
Threonine	1.0 g
Phenylalanine	1.0 g
Leucine	0.5 g
Isoleucine	0.5 g
Valine	0.5 g
L(+) selenomethionine	0.5 g

After 15 minutes, the culture was induced overnight with 0.5 mM of IPTG at 20 °C. The next day, after assessing protein expression, its purification was carried out as previously described for recombinant TFL. Once concentrated up to 20 mg/mL, a crystallization screening was performed, following the same procedure previously described in paragraph 2.2.2.

2.4 *Tremella fuciformis* lectin biophysical characterization

2.4.1 Hemagglutination assay

The hemagglutination assay is a method for titrating molecules based on their ability to bind sugars exposed on red blood cell surfaces. As reported in the literature [47], this technique can be applied to lectins, which are able to hemagglutinate red blood cells, upon sugar binding.

To verify whether TFL possessed this kind of activity, hemagglutination assays were performed. In a U – plate, serial dilutions of TFL (starting from 1 mg/mL down to 0.03125 mg/mL) were mixed with 50 µL of PBS and 50 µL of rabbit red blood cells, and also control wells were included. After incubating for 60 minutes at room temperature, the concentration of active TFL was calculated by looking at the highest dilution producing a positive result. The same protocol was applied to the recombinant form of TFL.

2.4.2 Spectrofluorometric analyses

Fluorescence spectroscopy is a very common and fast method applied to the determination of an analyte concentration in a sample. This technique relies on the ability of the sample to absorb light of a specific wavelength, upon excitation, and to re-emit it at a different wavelength, giving an estimation of its concentration. During the last decades, fluorescence spectroscopy has become a pivotal tool to analyse also protein samples, especially because they possess an intrinsic source of fluorescence: the amino acids.

Among all the amino acids of which proteins are composed, only three are able to produce fluorescence upon excitation: phenylalanine (Phe), tyrosin (Tyr), and tryptophan (Trp). Since thephenylalanine contribution is negligible and tyrosine emission is often quenched, tryptophan is considered as the strongest source of UV absorbance between 280 nm and 360 nm. In addition, it is consistently present in all proteins, and its concentration can vary between 1 mol % and 3 mol % [48].

Considering that the fluorescence properties of tryptophan are strongly affected by the surrounding environment, the spectrofluorometric analysis provides a useful method to follow conformational changes upon denaturation or ligand binding.

All the analyses were performed using a Jasco FP-2800 spectrofluorometer and fluorescence spectra were acquired with excitation and emission 5 nm width slits. The samples were excited at 295 nm and spectra were recorded between 310 nm and 450 nm. In order to monitor the denaturation process, 50 µL of TFL samples (10 µM) were mixed together with 950 µL of 10 mM Tris-HCl at different pH. The same protocol was applied to monitor the unfolding process upon urea addition (0 M – 6 M). The spectra were acquired in triplicates, after 0, 3 and 12 hours from the incubation with different pH buffers and urea concentrations. After normalizing the curves and substracting the background, the peak-shift was analysed and the data were processed by using OriginPro 8. The same protocol was applied to the recombinant form of TFL.

2.4.3 Dynamic Light Scattering and ζ –potential analyses

Due to the fact that it allows to size particles down to 1 nm diameter, Dynamic Light Scattering (DLS) is probably one of the most popular and powerful light scattering techniques. Although it is

particularly suitable for nanoparticles, colloids and micelles, during the last decades it has been extensively applied to proteins, enabling to acquire information not only on their size but also regarding their stability and aggregation. Briefly, DLS correlates the size of the proteins to their Brownian motion, that is their random movement produced after solvent bombardment; the smaller the protein, the faster the Brownian motion will be and vice versa. Then, the size of the protein is calculated with the Stokes-Einstein equation below; the physical quantity thus determined is the hydrodynamic radius.

$$d(H) = \frac{kT}{3\pi\eta D}$$

Where: d(H) is the hydrodynamic diameter, D the translational diffusion coefficient, k is Boltzmann's constant, T the absolute temperature and η the viscosity.

The ζ –potential is a measure of the magnitude of the electrostatic force or the charge repulsion/attraction between particles. Since this physical property is exhibited by any particle in solution, it is one of the most reliable parameters defining the stability of a sample. The technique relies on the Derjaguin-Landau-Verwey-Overbeek (DVLO) theory, which states that the stability of a particle in solution is dependent on its potential energy function (V_T), which in turn is the summation of different contributions.

$$V_T = V_A + V_R + V_S$$

Where V_A and V_R are the van der Waals attractive and the electrical double layer repulsive contributions, while V_S is the potential energy due to the solvent component [49].

TFL samples were diluted to (1 mg/mL) in 20 mM Tris-HCl with 100 mM NaCl and were analysed both in terms of their size and their ζ –potential, at 25 °C, by using the Malvern Zetasizer Nano ZS equipped with a He-Ne laser 633 nm light source.

In order to calculate the error as standard deviation and to obtain a reproducible data set, every measurement was run in triplicate, allowing to increase the precision of the final statistics. Final data were processed by OriginPro 8.The same protocol was applied to the recombinant form of TFL.

2.4.4 Differential Scanning Calorimetry analyses

Thermal analyses are a group of techniques that can be applied to analyse the changes in a physical property of a substance, upon temperature variation. Differential scanning calorimetry is one of them and allows to record the heat of a sample relative to a reference, as a result of a linear temperature ramp.

Applied to protein samples, this technique enables to obtain information on their thermal transition temperatures (melting temperature, T_m), alongside with the enthalpy (ΔH) required for them to reach the unfolded state and the heat capacity variation (ΔC_p) of their denaturation process [50].

Since very few mgs of protein are required to perform this experiment, differential scanning calorimetry has gained much more attention during the last years, being one of the most simple, fast and available thermal analysis methods.

In order to carry out differential scanning calorimetry experiments on TFL samples, a Nano DSC Standard Volume calorimeter, provided by TA Instruments, was used. The reference cell of the instrument was filled with 300 µL of 20 mM Tris, while 300 µL of TFL (120 µM) were loaded into the sample cell. The pressure of the system was set up at 3 atm and a thermal ramp from 10 °C to 100 °C, with a scan rate of 2 °C/min, was selected. Subsequently, the sample was cooled down, conserving the same scan rate and thermal ramp, in order to assess whether the denaturation process was reversible or not.

All the measurements were made three times to allow standard deviation calculation and to assure good quality of the data. The thermograms obtained were analysed by using the Nano Analyze Software provided by TA instruments.

After substracting the background, the melting temperatures corresponding to the edge of each peak were calculated, and where necessary, a peak deconvolution of the thermograms was carried out.

The same protocol was applied to the recombinant form of TFL.

2.4.5 Isothermal Titration Calorimetry analyses

Differential scanning calorimetry or isothermal titration calorimetry belongs also to the group of thermal analyses techniques, being one of the ultimate methods for investigating the variation of the thermodynamic parameters during a binding reaction. The idea behind this technique is the titration of the protein with its substrate, under isothermal conditions. Single peaks corresponding to every injection are obtained, representing the heat released or absorbed upon binding of the two molecules. As soon as the limiting reagent becomes saturated, the heat signal approaches

zero, highlighting the achievement of a steady state in which no more binding sites are available [51].

In order to analyse the ability of TFL to bind sugars, the isothermal titration calorimetry experiments were carried out on a Nano ITC Standard Volume, provided by TA Instruments. The cell of the instrument was filled with 200 µL of TFL (0.185 mM), while the syringe was loaded with 50 µL of different sugars in a range of concentration between 2.5 mM and 8 mM (Glucose, N-acetylgalactosamine, Saccarose and Lactose). Every experiment was performed three times at room temperature (25 °C) and consisted of 25 injections of 2.2 µL each.

The thermograms obtained were analysed by using the Nano Analyze Software provided by TA instruments. After subtracting the buffer background, the binding constant (K_d) and the stoichiometry (n), alongside with the changes in enthalpy (ΔH), the Gibbs free energy (ΔG), and the entropy (ΔS), were calculated.

The same protocol was applied to the recombinant form of TFL.

2.4.6 Circular Dichroism analyses

Circular dichroism (CD) is a rapid and useful technique for a fast and easy determination of protein secondary structure and their folding properties. Briefly, it relies on the difference in absorption between the right-handed and the left-handed circularly polarized light. A CD signal can be only observed when a molecule is optically active (chiral), either because it is linked to a chiral centre, or it is placed in an asymmetric environment, or simply because of its structure. Information about secondary structure can be obtained by performing far UV CD, which records the protein absorption from 240 nm and below, in order to highlight the presence of peptide bonds: $\eta \rightarrow \pi^*$

and π → π* transitions, for instance. On the other hand, by recording the signal between 230 nm and 300 nm, the presence of aromatic amino acids can be assessed, a method known as near UV CD. Since every aromatic amino acid has its own characteristic absorption peak, the acquisition of the near UV spectra is considered as a valuable fingerprint of the tertiary structure of each protein sample [52].

In order to acquire preliminary information on the TFL structural organization, circular dichroism analyses were carried out in collaboration with the laboratory of Professor Daniele Dell'Orco, based in the Biological Chemistry Section of the University of Verona.

Briefly, TFL protein samples were diluted to 20 µM in order to perform Near UV, whereas for acquiring the Far UV spectra a lower concentration was used (10 µM). Since sodium chloride absorbs UV light below 200 nm, all the samples were diluted in phosphate buffer without salt, in order to avoid possible interferences during the measurements. Far UV spectra were recorded in a range between 200 nm and 250 nm, while near UV spectra were recorded from 250 nm to 320 nm. Every experiment was run in triplicate, at 25 °C.

The same protocol was applied to the recombinant form of TFL.

2.4.7 XTT cells proliferation assay

Due to its association with mitochondria, cytoplasm, and endosome/lysosome compartments, tetrazolium salts have become one of the preferred reagents to monitor the metabolic activity of the cells. Although first generation tetrazolium salts (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, MTT) were characterized by a strong and net positive charge, responsible for their uptake via the plasma membrane, the second generation tetrazolium salts

(2,3-Bis-(2-Methoxy-4-Nitro-5-Sulfophenyl)-2H-Tetrazolium-5-Carboxanilide, XTT) present some slight differences. Differently from its ancestor, XTT forms water-soluble compounds called formazans and requires an intermediate electron acceptor for its reduction. In addition, it has a strong negative charge, it is largely cell-impermeable, and its reduction takes place via trans-plasma membrane electron transport [53].

Since 1988, the XTT cell proliferation assay is used as a powerful procedure to measure cell growth and drug sensitivity in tumour cell lines.

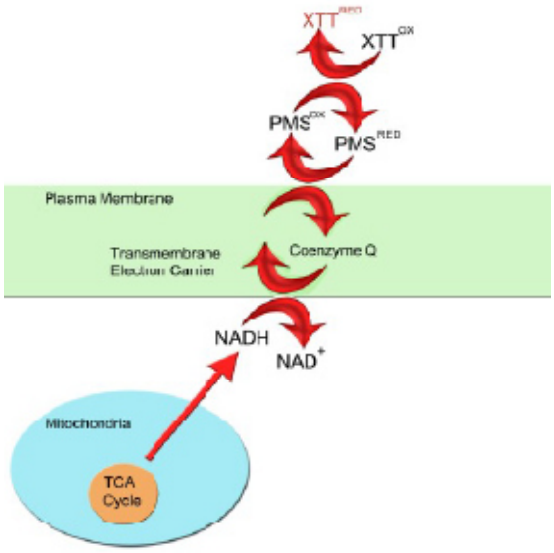


Figure 16: The colorimetric reduction of the XTT molecule by cellular enzymes.

In order to test TFL biocompatibility with the human body, theXTT proliferation assay was performed on fibroblast (control), SK-MEL28 cells (skin malignant melanoma) and A375 cells (skin malignant melanoma), in collaboration with the laboratory of Dr. Maria Teresa Valenti, based on the Internal Medicine Section of the University of Verona. With the aim of evaluating a potential

dose-dependent effect, three increasing concentrations of TFL were tested: 20 mg/mL, 50 mg/mL, and 100 mg/mL.

The XTT proliferation assay was set up in a 96-well plate flat bottom, by seeding around 20000 cells per well, which were counted using the Bürker chamber. Later, the cells were supplied with the appropriate growth medium and incubated in it for 24 hours. After being washed twice with freshly made phosphate buffer (PBS), they were treated with maintenance medium composed of Roswell Park Memorial Institute 1640 (RPMI), 2% of Foetal Bovine Serum (FBS), 1 % of penicillin and streptomycin, and 1% of glutamine. The maintenance medium was also supplied with different ascorbic acid (AsA) concentrations, ranging from 250 mM to 500 mM, and the plate was further incubated for 48 hours. After washing the plates as previously described, the tetrazolium salt provided by Chemicon International was added to the wells, following the protocol described by the company.

Absorbance at 440 nm was recorded after 24 and 48 hours from the treatment. All the measurements were acquired three times to allow standard deviation calculation and to assure good quality and reproducibility of the data. In the end, the data were normalized and the signal of the media alone was subtracted as background.

2.4.8 Cell migration tests via time-lapse video microscopy

Cell migration is an active process, particularly relevant not only for tissue repair, immune system function and embryonic development, but also for tumour invasion. It is an extremely complicated machinery, which involves both chemotactic signals and intrinsic cues [54]. Tumour cells possess the unique ability to adapt themselves to different environments and conditions, being able to

change their morphologies in order to stay motile. According to an *in vivo* study performed by Friedl P. and colleagues [55], tumoral cell migration is usually faster and more aggressive than that of healthy cells, displaying a wider range of motion modes.

In order to assess whether TFL samples could have a positive effect in arresting or reducing tumour cell migration, time-lapse video microscopy was performed following a commercial protocol. Briefly, after growing the cells in a 48-well plate to 80% confluence, the cultures were scratched by using a pipette tip, and thoroughly washed with PBS. Then, different concentrations of TFL (0 µg, 20 µg, 50 µg and 100 µg) were added to the cultures, together with fresh growth media to allow further cell proliferation. Cell migration was analysed after 24 h and 48 h, by using a EVOS Cell imaging system, and the final data was calculated as relative migration distance (RMD), using the equation:

$$RMD(\%) = \frac{(A - B)}{A} * 100$$

Where A and B are the width of the scratches before and after incubation with TFL samples, respectively.

To allow standard deviation calculation and to assure good quality data, every measurement was performed in triplicate.

2.4.9 Gene expression analyses via real-time PCR

Due to the fact that changes in fluorescence can be detected during the reaction, real-time PCR, also known as two-step quantitative RT-PCR, has probably become the method of choice for

analysing gene expression. The reaction consists of four steps: first the RNA has to be isolated from the target tissue, then it has to be reverse-transcribed to cDNA and further amplified. Only during the last step, the amount of the selected transcript will be calculated from the exponential phase of the reaction [56,57].

Using this technique, the effect of TFL on different genes, involved both in cancer progression and arrest, was analysed. Briefly, cell cultures were grown and treated as previously described in paragraph 2.4.7. After 48 hours, they were harvested using a cell scraper, thoroughly washed with PBS and stored at -80 °C. The RNA extraction was carried out by Rneasy Protect Mini Kit, purchased from Qiagen, alongside with the use of QIA shredder columns in order to optimise the entire workflow. Later, First-Strand cDNA Synthesis Kit (GE Healthcare) was used to reverse-transcribe the previously obtained RNA, setting up the following PCR parameters and using the following DNA primers Oligos.

Eno1 gene:

Reverse TCCAGGCCTTCTTTATTCTC

Forward TTACCACAACCTGAAGAATG

SSBP1 gene:

Reverse TCACATATTGATATGCCACG

Forward GGTGATGTCAGTCAAAAGAC

Step	Temperature (° C)	Time (sec)	
Inizialization	95	120	
Denaturation	95	15	} 40 cycles
Annealing	60	60	
Elongation	60	60	
Final elongation	72	300	

In order to obtain the final mRNA quantification, Actin Beta (ACTB) and Glyceraldehyde-3-Phosphate Dehydrogenase (GAPDH) were used as housekeeping genes.

2.4.10 Quantification of cytokine production via ELISA assay

One of the most powerful and reliable methods to study inflammation and its consequences on our body is to monitor its characteristic markers, such as the cytokines: a very broad family of small proteins, remarkably important in cell signalling. However, cytokines are only a small percentage of the massive and extremely complicate network of different cell types that coordinate their efforts to achieve a common goal: to trigger the immune response. In fact, B cells, T cells, macrophages, neutrophils, basophils, eosinophils, and mast cells play also a role during the inflammatory response.

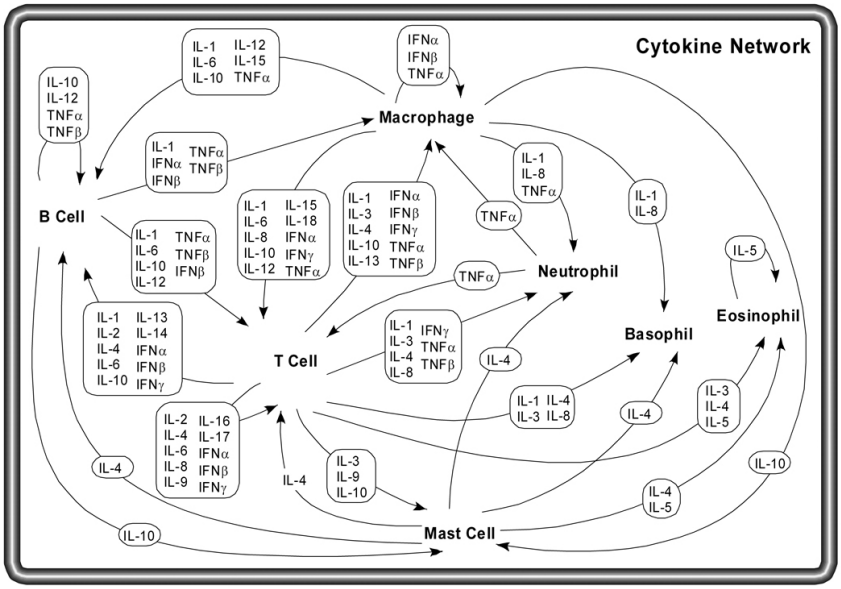


Figure 17: Cytokine network involving macrophages, B cells, T cells, neutrophils, eosinophils, basophils, and mast cells which communicatewith each other using secreted cytokines.

All the previously described cell types have something in common: they communicate with each other using secreted cytokines, therefore their analysis has become a cornerstone during the last decades.

The measurement of cytokine concentrations was carried out by an enzyme-linked immunosorbent assay (ELISA), in collaboration with the laboratory of Professor Stefano Dusi, based in the Medicine Section of the University of Verona. The blood samples were obtained from healthy volunteers, upon their written informed consent and the approval of the ethical committee, thanks to the Blood Transfusion Centre of the University of Verona. After separating the buffy coat from the venous blood, monocytes were isolated by Ficoll-Hypaque and Percoll density gradients (GE Healthcare Life Science) and further purified through the human monocyte isolation kit II (MiltenyiBiotec), following what stated in Zenaro E. et al. [58]. The thus obtained monocyte population was assessed to be almost pure (99%) by FACS analysis. In order to produce immature dendritic cells (iDCs), monocytes were supplied with RPMI 1640, heat-inactivated 10% low endotoxin FBS, 2 mM L-glutamine, 50 ng/mL GM-CSF, and 20 ng/mL IL-4 and further incubated

in 5% CO_2, at 37 °C, for 5-6 days. As described before, the purity of the population obtained was assessed via FACS analyses, and was primarily composed of CD1a+ cells (98%).

After treating the cells with 10 μg/mL and 50 μg/mL of native and recombinant TFL, the Ready-Set-Go Elisa Kit, purchased from Bioscience (San Diego, CA), was used to assess the concentration of the following cytokines in the cell culture supernatant: IL-6 (range 2-200 pg/ml), IL-12 (range 4-500 pg/ml), and TNF-α (range 4-500 pg/ml).

The same procedure was followed to treat TFL samples after encapsulation into a PLGA nanoparticle.

2.5 PLGA nanoparticle synthesis and characterization

During the last decades, scientific studies on targeted drug delivery have significantly grown, in particular those devoted to limit the side effects on healthy cells. Ideally, the perfect delivery system would be that able to selectively and efficiently transport the chemotherapeutic agent to the target tissue, causing the least amount of damages to the healthy cells. Recently, nanoparticles have been extensively studied as a powerful tool not only to reduce the interaction with the immune system, but also to allow the use of a lower concentration of drugs, thus leading to the reduction of therapeutic costs. Among the reagents used to prepare nanoparticles, Poly Lactic-co-Glycolic Acid (PLGA) seems to be one of the best choices, since the particles are biocompatible and biodegradable and they have been recently approved by the Food and Drug Administration (FDA) [59,60].

Therefore, in order to test the immune system response to TFL-loaded nanoparticle administration, nanoparticles were prepared following the modified solvent displacement (MSD) method [61]. Briefly, 10 mg of PLGA powder (75:25) were dissolved in 1 mL of polyethylene glycol (PEG) 400, to allow diffusing phase formation. After sonication, polyvinyl alcohol 1 % (PVA) was added to the pre-formed phase, under magnetic stirring (600 rpm) for 6 hours. The thus formed nanoparticles were thoroughly washed with PBS by centrifugation at 12000 rpm for 20 minutes at 4 °C and were ready to be used for further studies.

3. Results and Discussion

3.1 Purification of TFL from *Tremella fuciformis* fruiting bodies

Each batch of protein was prepared following the protocol previously described in paragraph 2.2.1. After being ground to a fine powder, 45 g of dry fruiting bodies of *Tremella fuciformis* were homogenised and centrifuged. A first and rough protein separation was carried out by adding ammonium sulphate (80%) to the supernatant, and the pure fraction was subsequently separated first by ion-exchange and then by size-exclusion chromatography. Alternatively, TFL was first eluted from a hog gastric mucin column with 20 mM Tris buffer at pH 11.4 and further isolated by size-exclusion chromatography. Both of the above described methods have led to obtaining a pure protein fraction. A representative elution profile of the final gel filtration step used by both methods is shown in figure 18.

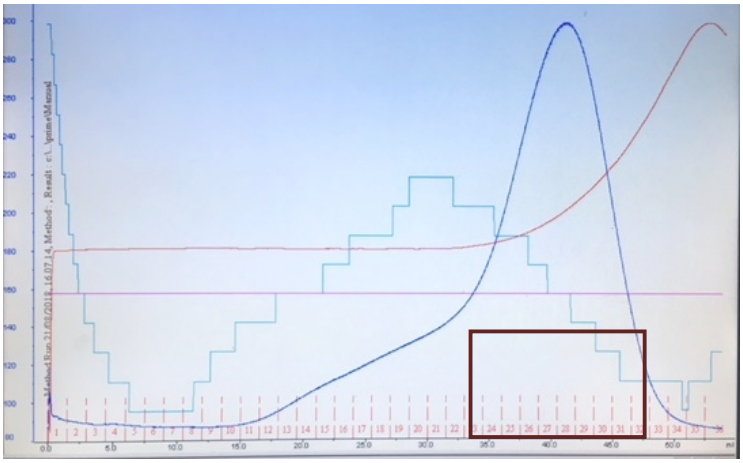


Figure 18: TFL elution chromatogram from a Sephacryl S-100 HR gel filtration column. The blue line corresponds to the UV absorbance, while the red indicates the conductivity.

Fractions corresponding to the highest peak, underlined dark red in figure 18, were analysed by SDS-PAGE gel and those containing pure TFL were pooled and concentrated for further analyses.

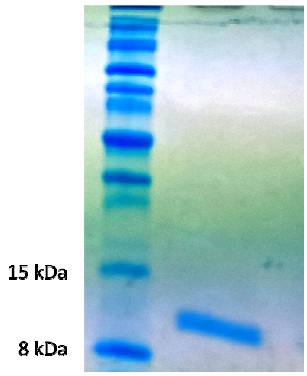


Figure 19: SDS-PAGE of the pure TFL fraction isolated from dried fruiting bodies.

Although the amount of protein obtained was consistently reproducible for both methods used, the final sample was purer and more homogeneous after the DEAE and SEC columns, therefore it was the product of that protocol that was chosen for the purification of the samples for crystallization trials.

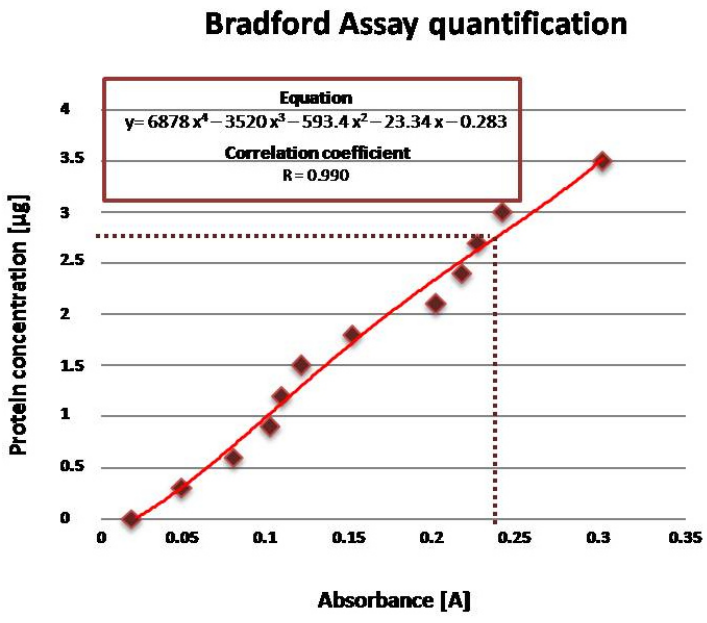


Figure 20: TFL yield after the purification process, assessed by both the Bradford Assay and an SDS-PAGE gel elaboration via ImageJ.

The amount of protein was determined using the Bradford protein assay, and the yield, starting from 45 g of *Tremella fuciformis* dried fruiting bodies, was 29 mg of pure protein. This result was confirmed estimating the protein on an SDS-PAGE gel using ImageJ, an open-source program that allows to evaluate proteins loaded on gels.

3.2 Structural studies of TFL

3.2.1 Crystallization trials and heavy atom derivative preparation

Following the purification process, pure TFL was concentrated up to 20 mg/mL and used to set up different crystallization trials. Both the hanging drop and the sitting drop techniques were used and in total, 192 conditions were tested. After 48 hours, crystals grew in one condition at 20 °C: 0.1 M Tris pH 8.5 and 1.5 M potassium phosphate dibasic (Figure 21).

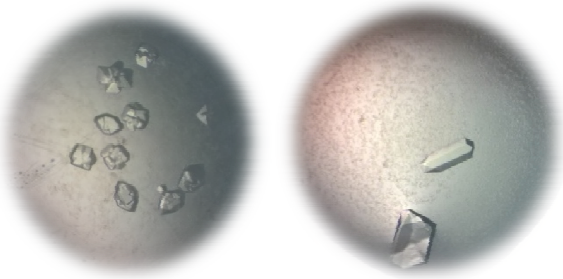


Figure 21: On the left crystals grown in 0.1 M Tris pH 8.5 and 1.5 M potassium phosphate dibasic, at 20 °C. On the right, crystals grown with the addition of 1% DMSO, at 4 °C.

However, since crystals were highly twinned, crystallization conditions were further refined by varying the salt concentration and the pH of the buffer. Moreover, the further addition of 1% DMSO or 1 % polyethylene-glycol was tested in order to facilitate the growthoflargersingle

crystals. An additional change was to incubate the plates at 4 °C in order to slow down the growth process and to allow the formation of more organized nuclei. Due to the temperature drop, crystals appeared after 3 to 4 days (Figure 21).

The condition that led to the best crystals was 0.1 M Tris pH 8.5, 1.5 M potassium phosphate dibasic and 1% DMSO, at 4 °C.

Heavy atom derivatives were prepared by adding the chemical powder directly to the drops cointaining the already formed crystals. The crystals were allowed to soak for a maximum period of 24 hours, at 4 °C and samples were collected at regular intervals, in order to monitor the derivatization reaction. A variety of heavy atoms were screened, including platinum, gold and mercury compounds. The best derivative was obtained upon soaking in mercury (II) chloride ($HgCl_2$) for 2 hours, and its data were processed as illustrated in paragraph 2.2.3.

Subsequently, some of the best crystals were soaked with various carbohydrates, with the aim of finding the carbohydrate binding site. In particular, N-acetylgalactosamine, glucose, lactose, sucrose, and N-acetylglucosamine were tested. For each sugar, 10 mM stock solutions were prepared and added to the already formed crystals, obtaining a final concentrations in the drops between 1 mM and 5 mM. The crystals were allowed to soak for a maximum period of 24 hours, at 4 °C and samples were collected at regular intervals, in order to monitor the derivatization reaction. In addition, considering the high solubility of the above mentioned sugars, also co-crystallization attempts have been made, keeping a 1:3 protein:ligand ratio.

After 48 hours, crystals grew at 20 °C, in the same condition that yielded crystals for the apo-protein: 0.1 M Tris pH 8.5 and 1.5 M potassium phosphate dibasic.

3.2.2 Computational methods for model prediction

Since the structure of a protein can provide important information and additional knowledge in understanding its function and mechanism of action, alongside with the fact that it is considered a great challenge from the scientific point of view, computational methods were used to obtain a model prediction for TFL. For this purpose, the web-based server Phyre2 (Protein Homology/analogy Recognition Engine V 2.0) was used, which allowed to obtain a full three-dimensional model, starting from the sequence of a protein described in the literature and expected to be the same isolated in paragraph 3.1.

The idea beyond Phyre2 is that, during the evolution process, the structure of a protein is usually more conserved than its amino acid sequence, making it possible to accurately model the structure upon sequence alignment. In addition to the final model, Phyre2 gave useful information on secondary structure and disorder prediction, alongside with an in-depth analyses of protein sequence and domains.

The output of Phyre2 (figure 22) was based on the *Tremella fuciformis* immunomodulatory protein sequence (UniProt code – A1YZD0) described in the literature [43]. The model was built using the structure of a peptidoglycan-binding protein from *Bacillus subilitis* (PDBcode – 3D30). Specifically, 97 residues out of 107 (92% of the sequence) were modelled with 100.0 % confidence by the single highest scoring template [62].

As shown in figure 23, secondary structure and disorder prediction revealed the presence of a high percentage of beta strands (41%), along with a small percentage of alpha helixes (13%) and non-ordered region (11%).

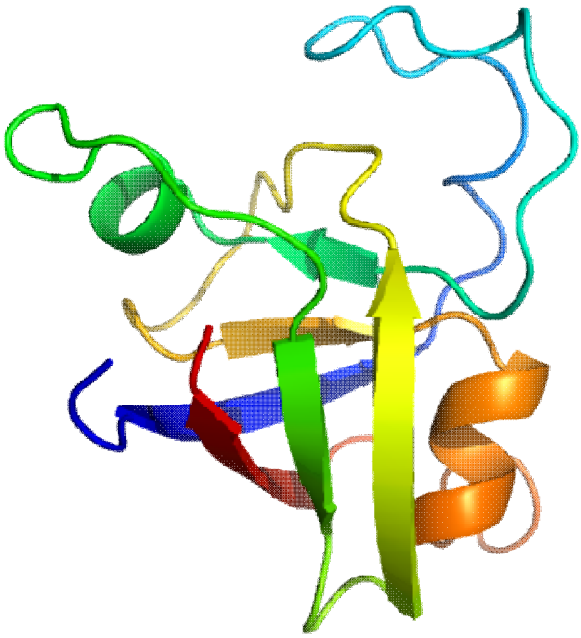


Figure 22: Three-dimensional model prediction obtained by Phyre2 server. Image coloured by rainbow N→ C-terminus. Model dimension (Å): X= 37.009 Y= 36.165 Z= 31.564.

Although in the end this model was rather similar to the experimentally determined structure, it was not adequate as a search probe for molecular replacement.

Following a brief soaking in the cryo-protectant solution (90 % mother liquor and 10 % glycerol), the final datasets of the apo-protein were collected at the beamline BM-30A of the European Synchrotron Radiation Facility (ESRF), in Grenoble, France. The data were indexed and integrated either with XDS, XIA2 or MOSFLM. The crystals grown in the above-mentioned conditions wee all monoclinic (space group $P12_11$) with cell parameters: a= 61.62 Å, b= 61.83 Å, c= 67.84 Å and α=β= 90 ° and γ= 106.87 °. The number of molecules in the asymmetric unit was four and the maximum resolution 1.28 Å.

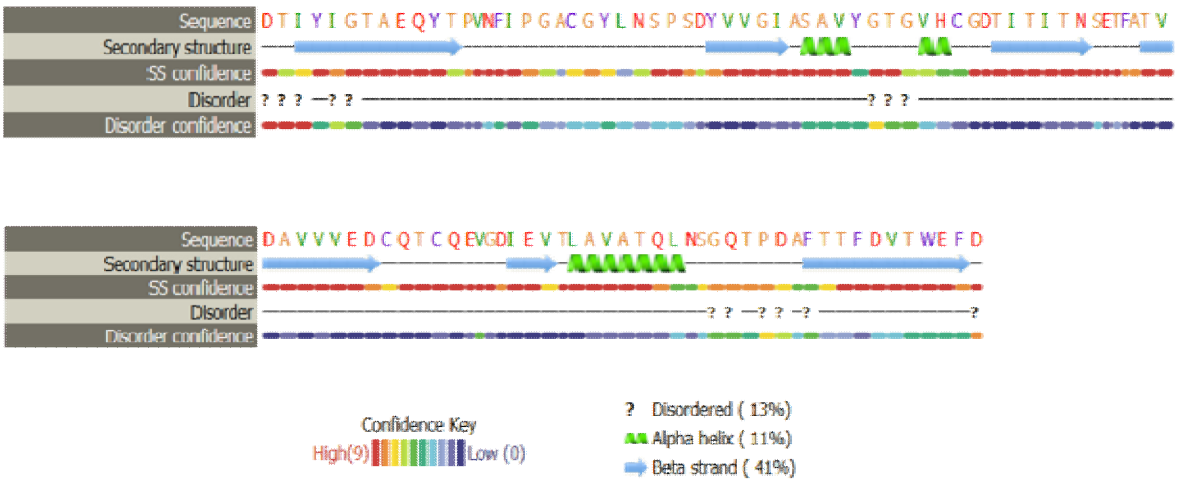


Figure 23: Secondary structure and disorder prediction, assigned by the Phyre2 Protein Fold Recognition Server.

3.2.2 Data collection and structure determination

Considering the need of tuning the wavelength, the heavy atom derivative datasets were collected instead at the beamline ID-29 of the European Synchrotron Radiation Facility (ESRF). The crystals soaked in mercury were also monoclinic (space group $P12_11$) but with slightly different cell parameters: a= 61.62 Å, b= 61.83 Å, c= 67.84 Å and α=β= 90 ° and γ= 106.87 °. In this case the maximum resolution reached was 2.08 Å.

TFL structure has been successfully elucidated following the 1st CCP4 Structure Solution workshop (18-23 February 2018), held at Ben-Gurion University of th Negev, in Beer-Sheba, Israel. Considering the difficulties encountered in performing phasing, mostly because of the weak anomalous signal and the presence of some high-resolution ice rings, a combination of innovative programs was used.

The phases were determined using the SIRAS method, exploiting the anomalous signal coming from the lower resolution map (2.08 Å).

Dataset	TFL apo
Space group	$P12_11$
a (Å)	61.62
b (Å)	61.83
c (Å)	67.84
α (°)	90
β (°)	106.87
γ (°)	90
Resolution range (Å)	64.92-1.28
Observed reflections	301327
Independent reflections	95668
Multiplicity	3.3 (3.4)
Rmerge (%)	5.9 (39.0)
I/σ	28.6 (3.2)
Completeness (%)	99.9
Reflections in refinement	86087
Rcryst (%)	25 (27)
Rfree (%)	31 (33)
r.m.s.d on bond angles (°)	1.20
r.m.s.d on bond length (Å)	0.01
r.m.s.d on planar groups (Å)	0.008
Chiral volume deviation ($Å^3$)	0.085
Average B-factor ($Å^2$)	17.73

Table IV: Data collection and refinement statistics. The values between parentheses refer to the highest-resolution shell i.e. 1.28 Å.

R_{merge} was calculated with the following equation: $\frac{\sum_h \sum_i |I_{ih} - \langle I_h \rangle|}{\sum_h \sum_i \langle I_h \rangle}$, while $\frac{\sum ||F_{obs}| - |F_{calc}||}{\sum |F_{obs}|}$ was used to define R_{cryst}. Rfree was calculated choosing a randomly subset of the diffraction data (5%) not included in the refinement.

Later on, molecular replacement was performed against the high resolution native dataset (1.28 Å) and the same dataset was also used for refining purposes.

The CCP4 web application JsCoFE was used as a platform to run all the jobs that led to the structure solution [147].

SIRAS phasing, including the location of the heavy atom sites, was performed by using the automated structure solution package CRANK2, which has been shown to be particularly powerful in case of weak anomalous signal or low resolution data set [150]. Differently from other phasing tools, its novel multivariate approach, enables to combine in a single process, phase information from the experimentally collected X-ray data, density modification, and model building.

Considering that the three steps are no longer performed separately but rather simultanueosly, the final solution does not rely only on successive approximations of the experimental data.

In addition to the above mentioned features, it is worth noticing that CRANK2 can also adjust the total number of iteration according to the quality of the inpunt dataset, therefore the weaker the signal, the higher the number of interations performed.

Analougously to other pipeline, CRANK2 consists of several steps and can be started or ended by the user at each of the following stages:

1. Substructure determination with SHELXD [152], using the output obtained from SHELXC [152];

2. Substructure improvement with REFMAC coupled with SHELXE [152] phasing;

3. Density modification carried out either with PARROT or SHELXE;

4. Model building with Buccaneer iterated with REFMAC model refinement;

5. Final model refinement stage with REFMAC.

Concerning TFL structure solution, CRANK2 was ran from step 1 to step 5, inserting “Hg” and “4” as substructure atom element symbol and expected number in the asymmetric unit, respectively.

The anomalous data reflection in a merged MTZ file, the anomalous scattering coefficients f’ and f’’ and the wavelength of the data collected were inserted as input as well.

The preliminary model thus obtained was used to perform a structural superposition with GESAMT, a tool that enables to find structural homologous by using a pairwise algorithm [151]. Although the latter is often use for comparing two structures, in this particular case it has successfully been exploited in order to scan the PDB archive for structural hints.

After the evaluation of the results, both in terms of Q-score and RMSD, the best candidate (PDB code: 4AVR) was identified and a single chain isolated from the PDB file was used to perform molecular replacement against the high resolution native dataset. Since TFL sequence was not known yet, a poly-alanine chain was used during molecular replacement stage.

At this stage, the initial phases still needed to be improved, therefore a density modification step was performed using Parrot [153], improving not only solvent flattening, but also histogram matching and non-crystallographic symmetry averaging. The native dataset of reflections, as well as for the set of phases from the molecular replacement model were selected as input.

Considering the high resolution of the native dataset and the level of detail of the electron density map, the replacement of the poly alanine chain with the correct residues was easily carried out.

The model was refined with RefMac 5 [149] with manual revisions using the program Coot [154], as described in paragraph 2.2.3. Lastly, water molecules were added to improve the statistics and the R value.

After refining, the interactive web tool PISA (Protein Interfaces, Surfaces and Assemblies) was used to check whether the presence of four molecules in the asymmetric unit could be indicative of a possible interaction or rather being only a crystallographic artefact.

To validate the model obtained, Polygon and MolProbity from Phenix suite were used [155, 156].

Calculation of the Ramachandran plot revealed that 99.01 % of the residues were in the most favoured region, while the remaining 0.9 % was in the allowed region and no outliers were detected. The average B-factor for the protein atoms was 17.73 for the main chain.

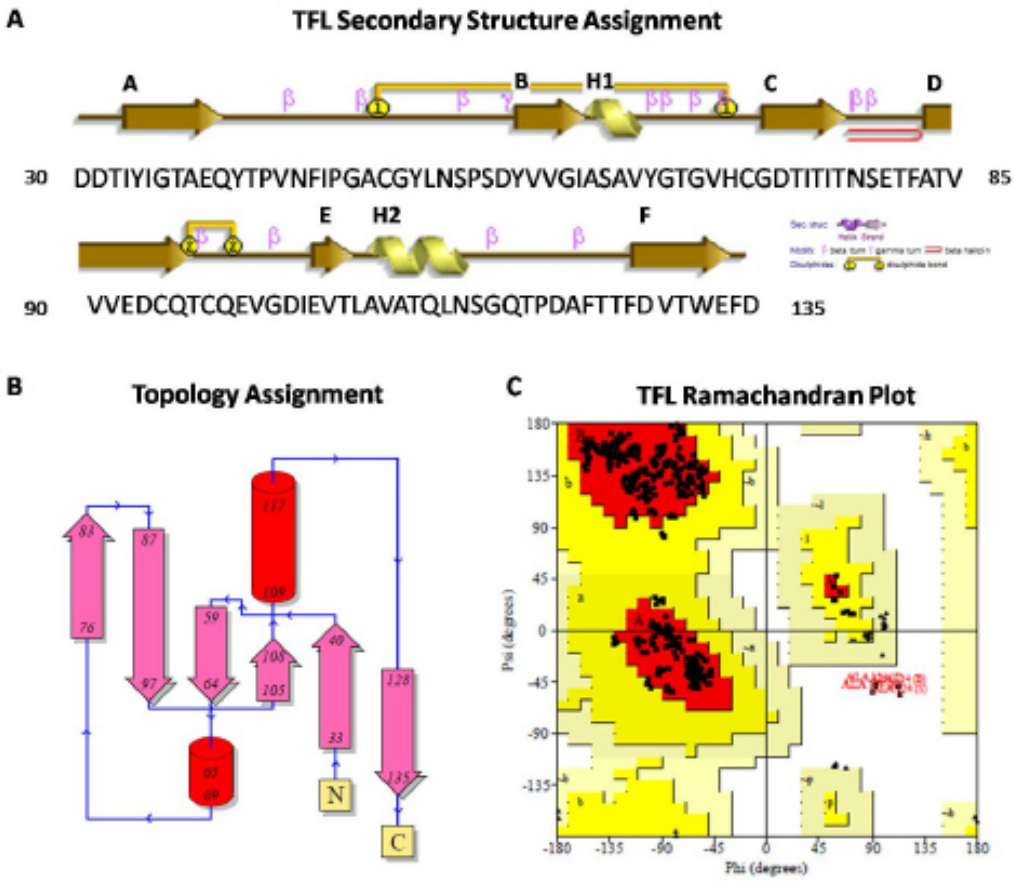


Figure 24: TFL Secondary Structure Assignment (A), Topology (B), and Ramachandran Plot (C) performed by the PDBsum web server. Beta-strands are named A-F, while the two alpha helices are termed H1 and H2.

Confirming the DLS measurements (paragraph 3.3.2), TFL is a monomer in the crystals and its peptide chain is 107 amino acids long.

Protein secondary structure was assigned using thePDBsum web server (Figure 24) [63]. The program identified a β-barrel, composed of the following β-strands: A (residues 33-40), B (residues 59-64), C (residues 76-83), D (residues 87-97), E (residues 105-108), F (residues 128-135), separated by turns and coils (Figure 24).

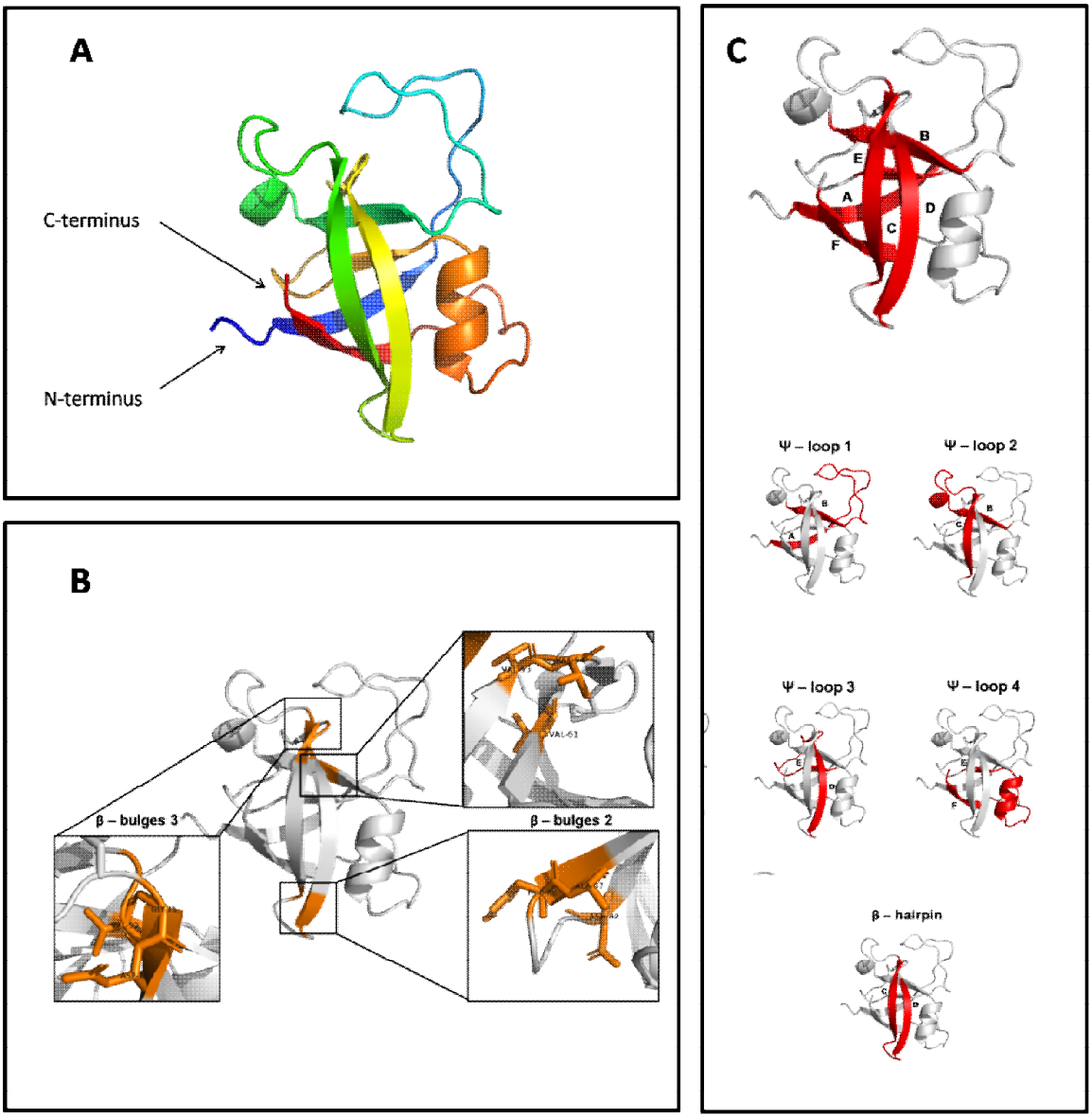


Figure 25: Panel showing the three-dimensional structure of apo TFL (A), composed mainly of β-structure (B and C).

The connection between β – strands B and C gives rise to the formation of a β – hairpin, while four Ψ – loops are formed by the β – strand pairs A-B, B-C, D-E, and E-F, respectively.

In addition, two α-helices were recognizedand named H_1 (residues 65-68) and H_2 (residues 110-116), respectively. According to table V, 13 β-turns, 3 β-bulges and one γ-turn were identified.

Motif	Residue
β – strand A	Tyr33 – Gln39
β – strand B	Val60 – Ala64
β – strand C	Thr77 – Asn82
β – strand D	Thr88 – Cys97
β – strand E	Ile106 – Val 108
β strand F	Phe128 Phe 134
β bulges 1	Val 61 Val93 Val94
β – bulges 2	Asn82 – Phe86 – Ala87
β – bulges 3	Val93 – Gly75 – Asp76
α – helix H1	Ser65 – Tyr68
α – helix H2	Leu110 – Leu116
γ turn	Asp58 – Val60
β – turn 1	Val43 – Ile46
β turn 2	Gly48 Gly51
β turn 3	Ser55 Asp58
β – turn 4	Tyr68 – Gly71
β – turn 5	Gly69 – Val72
β – turn 6	Gly71 – Cys74
β – turn 7	His73 – Asp76
β – turn 8	Asn82 – Thr85
β – turn 9	Ser83 – Phe86
β turn 10	Cys97 Cys100
β – turn 11	Glu102 – Asp105
β – turn 12	Asn117 – Gln120
β – turn 13	Asp123 – Thr126
Disulphide bond 1	Cys50 – Cys74
Disulphide bond 2	Cys97 – Cys100

Ψ – loop 1	β sheet
Ψ – loop 2	
β hairpin	
Ψ – loop 3	
Ψ – loop 4	

Table V: TFL Secondary structure assignment.

The electron density clearly showed the presence of two disulphide bridges, connecting cysteine 50 to 74 and cysteine 97 to 100 (Figure 26). It is worth mentioning that, the residues before cysteines 50 and 97, could be readily modelled into the electronic density.

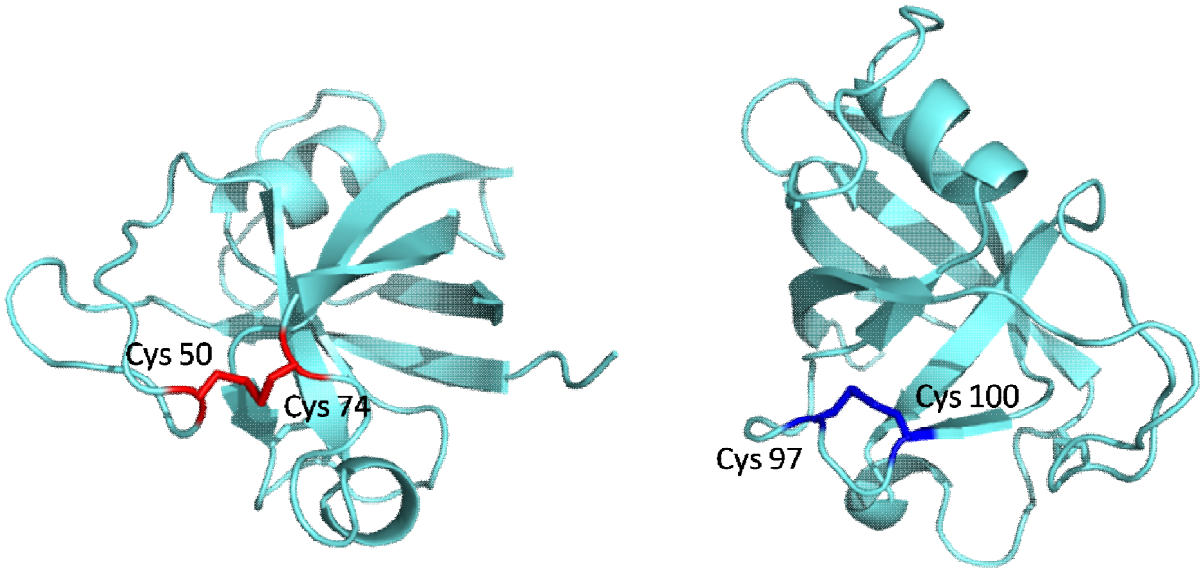


Figure 26: Stick representation of the disulphide bridges of TFL, coloured red (Cys 50 – Cys 74) and blue (Cys 97 – Cys 100).

No electron density ascribable to sugars was detected in the maps, confirming the factthat TFL does not possess glycosilation sites. In addition, no electron density attributable to metal ions (Ca^{2+} or Mg^{2+}) was identified, andtherefore it is assumed that TFL activity is accomplished without the need of cofactors.

In order to look for fold homologues among previously discovered proteins, a fast secondary-structure-based method, implemented in the program CATHEDRAL, was used. One of the main features of this program is that it exploits the double-dynamic programming algorithm, which is used to align members of the target fold groups against the query protein structure to identify the closest relative and assign domain boundaries [64].

The output of CATHEDRAL revealed a high structural homology with the carbohydrate binding module of the apo *Clavibacter michiganensis* expansin (PDB code 4JJO), also confirmed by superposition (RMSD = 2.0 Å). In particular, the SSAP score, a value that measures structural similarity residue by residue, was found to be around 90%.

Not surprisingly, this protein has been included in the family of lectins or of sugar binding proteins, which suggests the existence of a probable common ancestor.

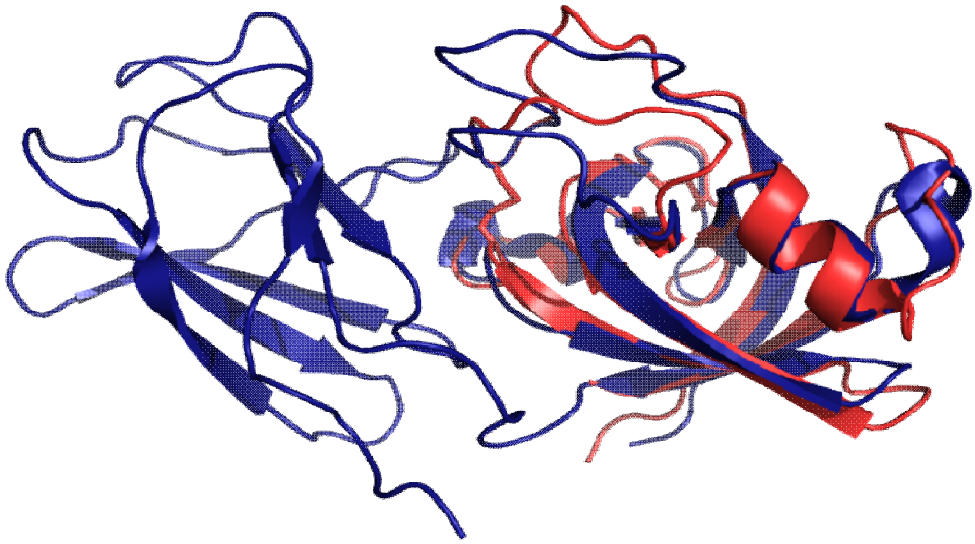


Figure 27: Superposition of TFL and apo *Clavibacter michiganensin* expansin (PDB code 4JJO), which belongs to the family of the cellulose-binding proteins.

Even if crystals soaked with N-acetylgalactosamine diffracted to a resolution too low (4-4.5 Å) for ligand fitting in the density maps, collected crystal data after different soaking times gave a variation in the unit cell parameters. For instance, in the crystals frozen after 15 minutes of interaction with the ligand, a and b measured 62.50 Å and 62.70 Å respectively, while c measured 69.01 Å, being indicative of the fact that N-acetylgalactosamine interacts with TFL, forcing it to rearrange its conformation and unit cell parameters.

3.3 Biophysical characterization of TFL

3.3.1 Hemagglutination assay

A hemoagglutination assay was performed on TFL samples, in order to check for this activity.

The pure protein fraction was concentrated to 1 mg/mL, and subsequently serially diluted down to 0.03125 mg/mL and tested against all the blood groups.

The assay revealed no apparent activity for TFL, placing the protein in the subfamily of merolectins, confirming what previously stated in paragraph 1.1.1.

3.3.2 Dynamic Light Scattering and ζ – potential analyses

According to the model presented in paragraph 3.2.3, DLS analyses confirmed the monomeric state of TFL samples. Analysing the size distribution plot of the scattered signal acquired, the average hydrodynamic diameter of the lectin was found to be 2.373 nm ± 0.248, displaying a good homogeneity and monodispersion (Polydispersion index = 0.4 ± 0.1), and revealing the absence of aggregates.

The ζ – potential of TFL in PBS buffer at pH 7 was 15 mV, confirming a good stability of the protein. Interestingly, the ζ – potential increased with decreasing pH, indicating a loss of stability of TFL and its tendency to aggregate, as shown in figure 26. In addition, a MonoQ anion exchange column separated different TFL isoforms in a fast, sensitive and high-resolution way. According to the elution profile, the higher is the elution volume of each fraction, the lower is the ζ – potential associated, as described in figure 28.

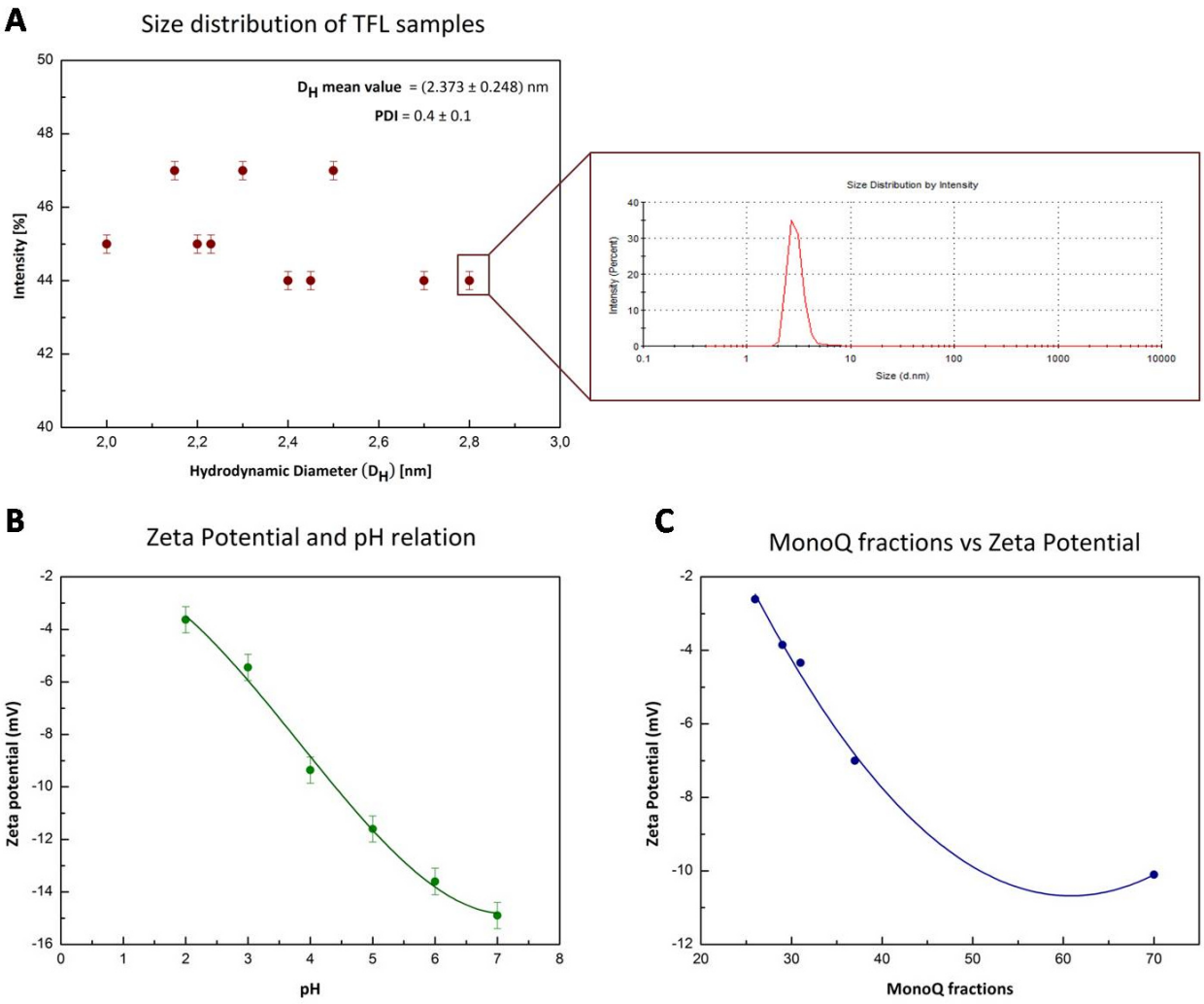
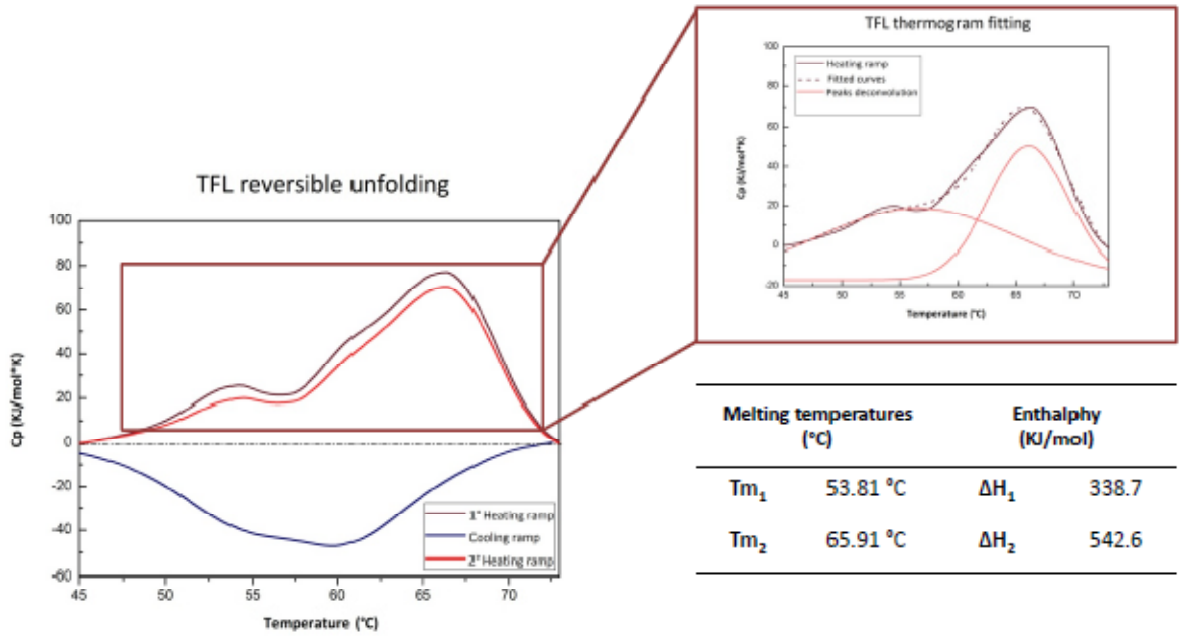


Figure 28: Average diameter of TFL calculated from a series of ten samples. The inset shows the size distribution plot related to a single measurement (2.8 nm).

3.3.3 Differential Scanning Calorimetry analyses

In order to assess TFL thermal stability, DSC measurements in scanning mode were performed, as shown in figure 29. According to the thermogram, two different endothermic peaks were observed, suggesting that TFL probably does not undergo a two-state unfolding process. The melting temperatures associated with the peaks were: 53.81 °C and 65.91 °C, respectively. A second heating ramp was performed in order to investigate the reversibility of the thermal

unfolding. Interestingly, although most of the protein unfolds irreversibly, TFL denaturation was confirmed to be reversible. The enthalpy variation calculated for the second heating ramp was around 90 % of the first, confirming the reversibility of the transition.



Melting temperatures (°C)		Enthalphy (KJ/mol)	
Tm_1	53.81 °C	ΔH_1	338.7
Tm_2	65.91 °C	ΔH_2	542.6

Figure 29: DSC thermogram highlighting the reversibility of TFL denaturation process. The inset shows the fitting of the curve and the deconvolution of the two peaks.

The small change in the heat capacity between the two peaks is ascribable to aggregate formation, which in turn results in less unfolded protein available to refold during the cooling ramp. Moreover, the cooling ramp (blue coloured) indicates that an exothermic transition occurred, which can be associated to the refolding process of TFL.

The fitting of the curve was consistent with a non-two-state model, i.e. characterized by the presence of two components, having ΔH of 338.7 KJ/mol and 542.6 KJ/mol, respectively. Although multidomain or dimeric proteins display the presence of two peaks in the DSC thermogram, TFL is

monomeric and possesses only one domain. However, the previously described behaviour could be due to the early unfolding of the α-helical portion of the molecule (first peak) compared to the β-barrel (second peak), giving rise to the development of two different intermediates, or microstates [65]. According to Molero and colleagues, the assumption that a protein can only exist in a given number of well-defined and distinct states is not necessarily true. In fact, the transition between the folded and the unfolded state of a protein cannot be considered as a chemical reaction, because it does not involve any breakage or formation of new and strong covalent bonds, but rather a re-organization of several weak interactions. In addition, the assumption of the existence of several microstates implies that the barriers between them must have a free-energy lower than the thermal energy RT.

Therefore, for certain kinds of analyses, it is preferable to consider the folding/unfolding transition of a protein as a continuous process involving an ensemble of microstates, rather than imposing well-definedmacrostates (native and unfolded) *a priori*. In addition, every microstate can be defined in terms of a given set of different protein/solvent conformations, implying the presence of a very high population for each intermediate degree of unfolding.

3.3.4 Isothermal titration Calorimetry analyses

Calorimetric assays have revealed the ability of TFL to selectively bind sugars, confirming that the protein belongs to the family of the lectins. In particular, isothermal titration calorimetry allowed to further characterize the interaction of TFL with different sugars from the thermodynamic point of view, resulting in the evaluation of the K_d and all the other parameters associated to the binding. Among the different sugars tested, the one giving the best result, in term of K_d, was the N-acetylgalactosamine, also well known as the monosaccharide building block of Tn antigen.

Although it was not possible to get more information about the ligand binding site of TFL, by means of X-ray crystallography, some of the datasets collected after soaking with N-acetylgalactosamine, highlighted variations in the cell parameters, being indicative of a possible structural rearrangement.

Titration with 8 mM N-acetylgalactosamine was performed on a volume of 0.25 mL of 0.185 mM of TFL, setting up 25 injections of 2.2 µL each.

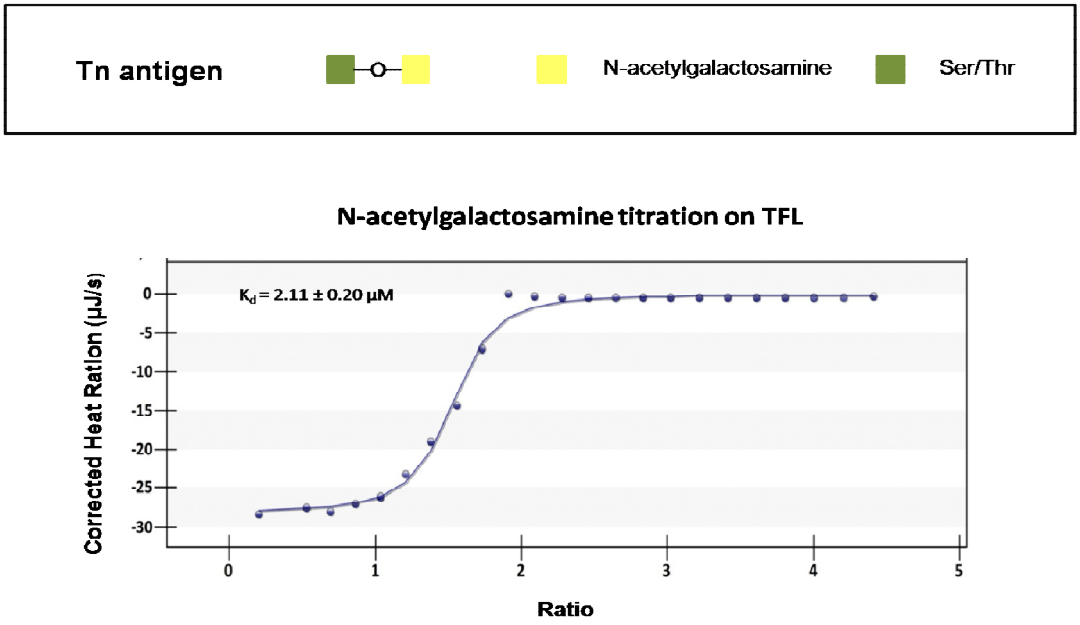


Figure 30: ITC thermogram representing the titration of N-acetylgalactosamine, the building block of Tn antigen, on TFL.

Fitting of the data, based on the multiple-site model, revealed the presence of one binding site with a dissociation constant of 2.11 ± 0.20 µM. Besides confirming the interactions with the sugar, the ITC plot also provides important information on the nature of the forces responsible for the binding. In particular, polar interactions tend to contribute favourably to the enthalpic component, while entropically favoured interactions imply a higher hydrophobicity. According to table VII, the binding of TFL to N-acetylgalactosamine involves hydrogen bonding and hydrophobic

interactions, as indicated by the negative or favourable binding enthalpy (ΔH) and entropy factor (-TΔS). Therefore, since the enthalpy associated to the binding was found to be - 28.04 KJ/mol, while the Gibbs free energy was - 32.09 KJ/mol, resulting in a negative entropy (- 4.34 KJ/mol), the interaction can be considered entropy-driven (Table VI). The main feature of the latter is to display a more hydrophobic character compared to an enthalpy-driven, which tends to be led by hydrogen bonding and van-der-Waals forces.

Ligand	TFL [mM]	n	K_d [µM]	K_a [µM]	ΔG [KJ/mol]	ΔH [KJ/mol]	-TΔS [KJ/mol]
GalNAc [2.5 mM]	0.185	2	2.11 ± 0.20	0.47 ± 0.10	-32.09	-28.04	-4.34

Table VI: Kinetic and themodynamic parameters of the interaction of TFL with N-acetylgalactosamine.

3.3.5 Spectrofluorometric analyses

As described in paragraph 2.4.2, one of the most valuable features of intrinsic protein fluorescence is the high sensitivity of tryptophan to its local environment, whose emission spectra can be analysed in order to look for conformational transitions, subunit association, substrate binding, and denaturation. Althoughtyrosine and phenylalanine emit also fluorescence in response to environmental changes, tryptophan is very sensitive to collisional quenching, therefore its signal was used to monitor TFL behaviour in different pH and urea concentrations.

The emission of tryptophan was acquired from 305 nm to 445 nm and the peak shift was monitored and further plotted against different pH conditions and urea concentrations. Analysing the spectra, at neutral pH and non-significant urea concentration, the maximum Trp emission was

found to be around 335-340 nm, indicating that a Trp is partially buried into the TFL structure [Ghisaidoobe, Chung - 2014]. There is only a single Trp in the TFL sequence, it is amino acid number 132 and it is positioned towards the C-terminus of the chain. As expected from the X-ray model, it is partially surrounded on one side by the β-barrel.

In addition, no evident peak shift was detected between 0 M and 2 M urea concentrations at pH 2 and 8, highlighting the stability of TFL in a wide range of conditions (Figure 31).

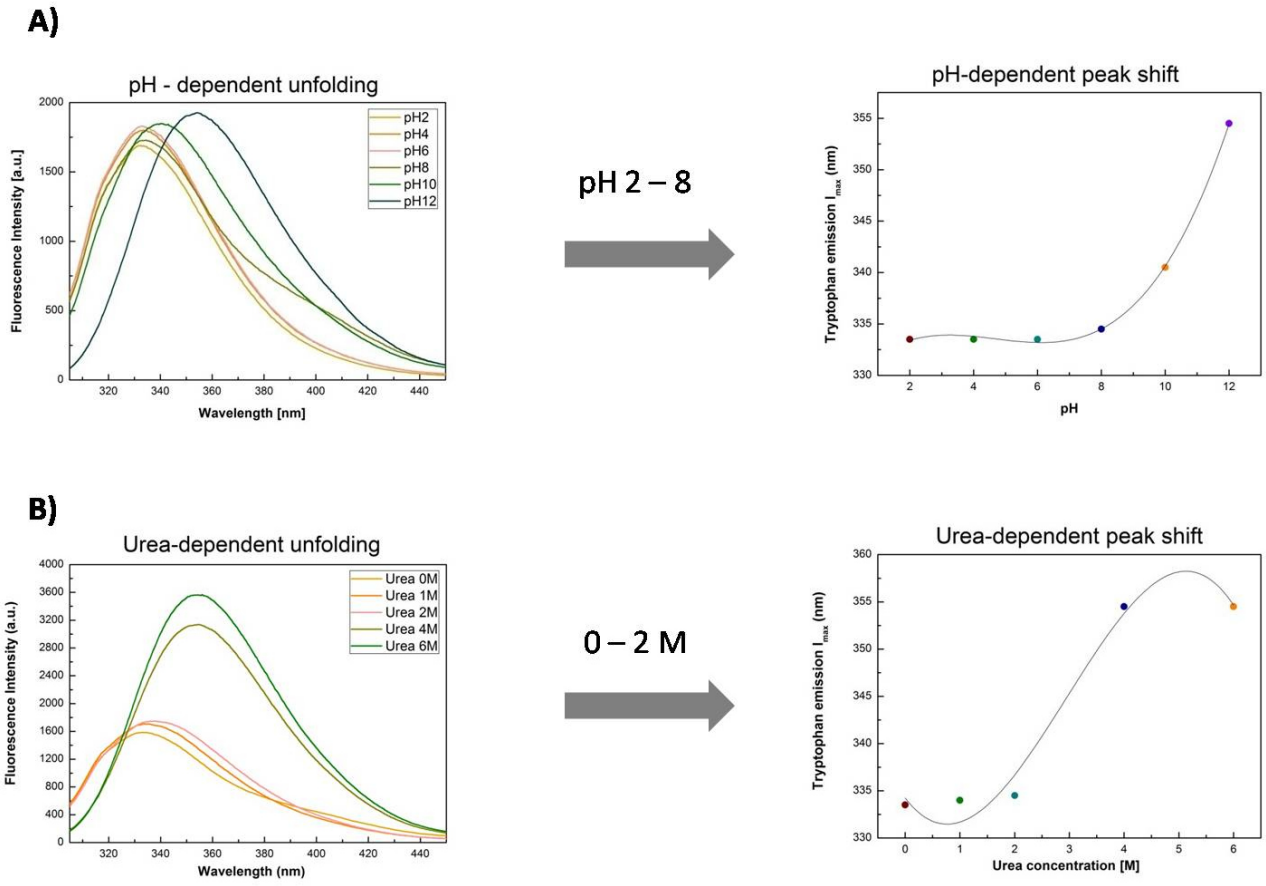


Figure 31: Fluorescence spectra of Trp emission related to TFL stability in different pH conditions (A) and urea concentration (B).

Only the data acquired after 24h are shown as representative of the general behaviour of the protein.

Although several trials to use fluorescence spectroscopy to determine the K_d of sugar binding were carried out, no acceptable data were obtained, most probably because the cavity dedicated to

host carbohydrates is placed far from the Trp residue. Analogously, it was not possible to monitor the titration of TFL with different metal ion cofactors, such as calcium and magnesium, therefore X-ray diffraction remained the only technique that could give this kind of information.

3.3.6 Circular dichroism analyses

As previously described in paragraph 2.4.6, circular dichroism is a simple tool for the rapid determination of the secondary structure and folding properties of proteins, either obtained from an expression system or purified from natural sources. Circular dichroism is commonly known for its ability to determine the folding state of a protein and, if it the case, to detect if mutations can affect its conformation or stability.

In the light of the above consideration, in collaboration with the laboratory of Professor Dell'Orco, based in the Biological Chemistry Section of the University of Verona, both far-UV and near-UV experiments were performed, before the X-ray structure of the protein had been determined.

As shown in figure 32, the far UV spectrum has revealed a strong negative peak around 210 nm, consistent with a high percentage of β-sheets, while the weak positive peak around 235 nm can be ascribed to the presence of non-repetitive secondary structure [52]. Overall, these data are in accordance with the three-dimensional structure and confirm the highly organized character of TFL, being mainly composed of β-sheets, with only a low percentage of α-helix and non-organized regions.

The presence of aromatic amino acids was confirmed through the analysis of the near UV spectrum, in which one main negative peak between 280 nm and 290 nm can be recognized.

According to the sequence, about one tenth of the total number of amino acids include an aromatic ring.

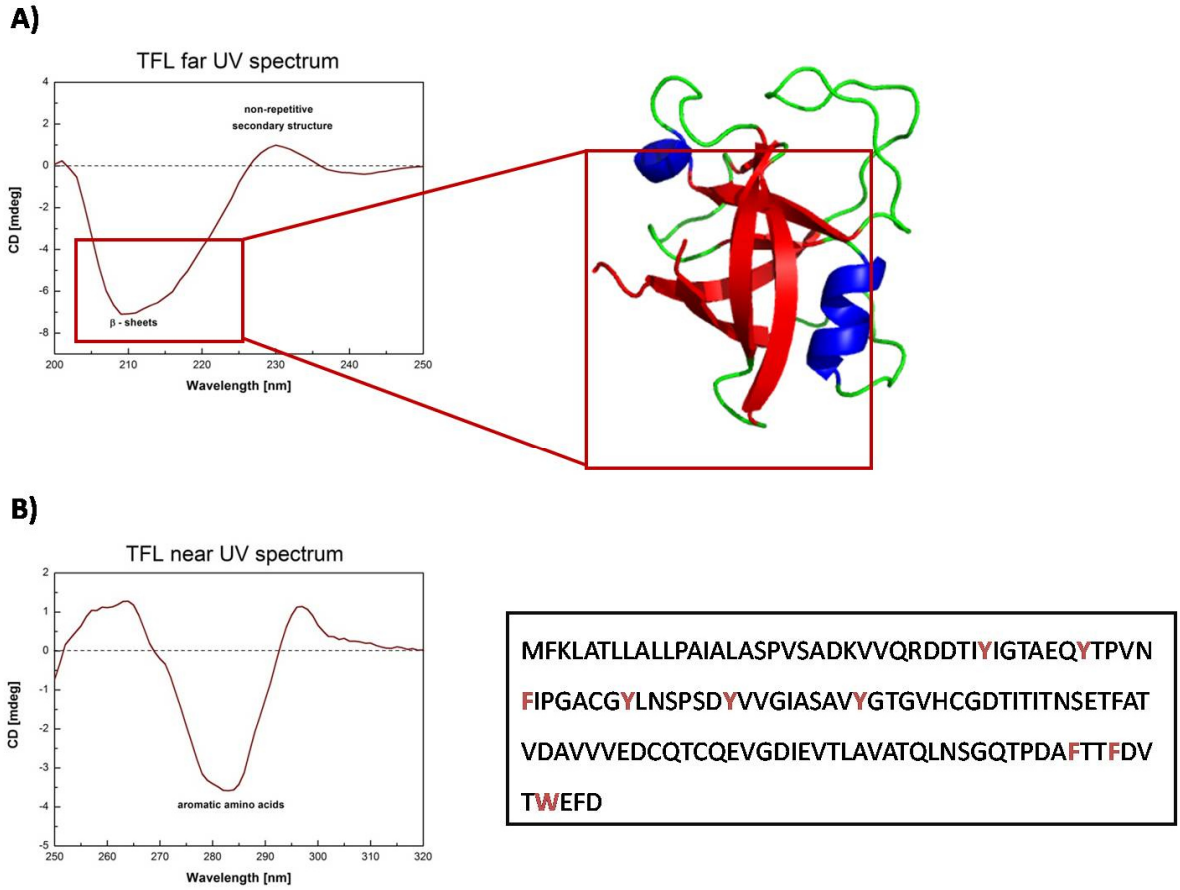


Figure 32: A) Far UV spectrum of TFL, showing the presence of two peaks, ascribable to the β-sheets and to non-repetitive structure in the molecule. B) Near UV (B) spectrum of TFL, indicating the presence of several aromatic amino acids. Analyses carried out in collaboration with the laboratory of Prof. Daniele Dell'Orco.

3.3.7 XTT cell proliferation assay

During the last decades, a variety of tetrazolium compounds have been exploited to detect cell viability, among which XTT is the most commonly used. According to the literature [66], only viable cells with active metabolism are able to convert XTT into a product with absorbance of 570 nm,

called formazan, therefore it is particularly suitable to assess the effect of different compounds on various cell lines.

The assay was used to assess the effect of TFL on both malignant and healthy cellsin collaboration with the laboratory of Dr. Maria Teresa Valenti, based in the Internal Medicine Section of the University of Verona. First, theTFL biocompatibility was examined. In particular, three different TFL concentrations (20 mg/mL, 50 mg/mL and 100 mg/mL) were tested on two tumour cell lines (A375 and SK-MEL 28) and on ahealthy cell line (fibroblasts), and cell growth was monitored for 24 and 48 hours. According to figure 33, upon TFL treatment, either after one day or two days, about the 80 % of the cells were still alive, confirming the biocompatibility of the lectin and thus paving the way for further investigations.

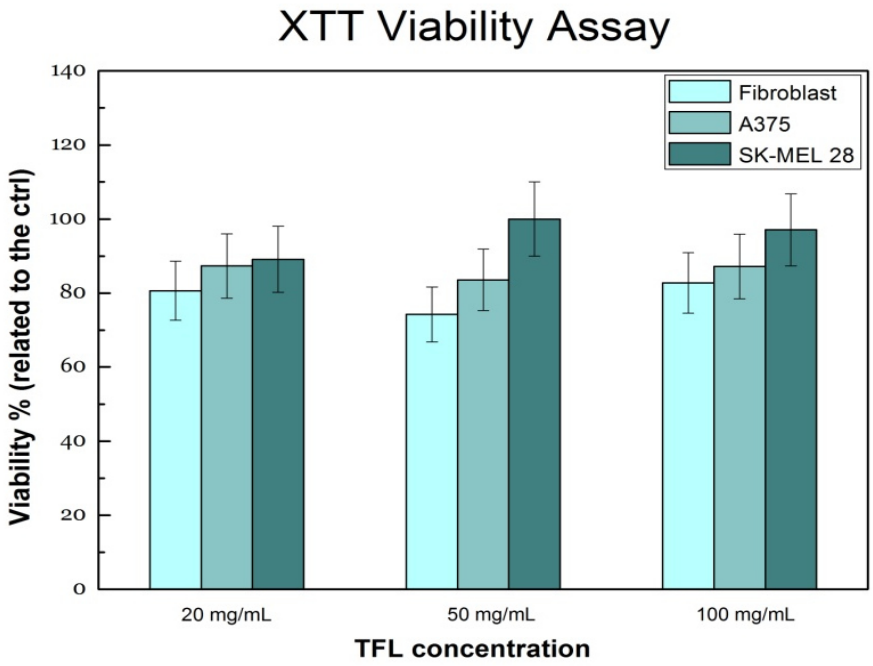


Figure 33: TFL biocompatibility tested on three different cell lines (Fibroblast, A375 and SK-MEL 28), by using the XTT viability assay. Analyses carried out in collaboration with the laboratory of Dr. Maria Teresa Valenti.

3.3.8 Cells migration tests via time-lapse video microscopy

According to the literature [67], besides being a key property of live cells, migration is a critical step during disease processes, such as inflammation and cancer. The idea to analyse the speed of wound closure and cell migration after making a scratch into a cell culture, allows to understand how the cell migratory behaviour was affected by different TFL doses. With the aim of reducing the variability of the results, all the tests were performed on the same cell lines previously used in paragraph 3.3.7.

Cell migration was recorded by a time-lapse microscopy system, taking snapshot pictures at regular intervals and then calculating the relative migration distance for each of them.

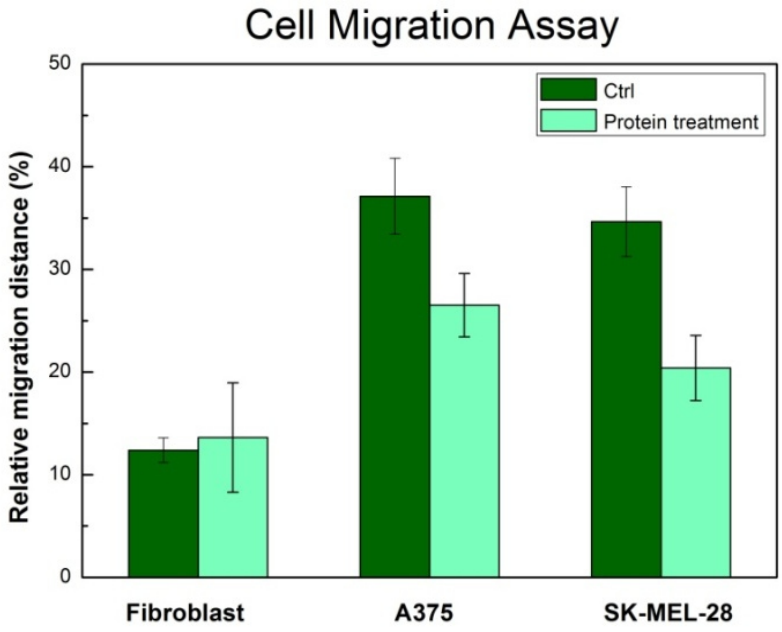


Figure 34: Analysis of the speed of wound closure, upon TFL treatment, monitored by real time microscopy. Analyses carried out in collaboration with the laboratory of Dr. Maria Teresa Valenti.

As shown in figure 34, upon TFL treatment, the speed of scratch closure in tumour cell lines seemed to be reduced, if compared to the healthy control. Interestingly, this was shown to be a rather new behaviour for a lectin, which are best known for their ability to inhibit cell proliferation and to cause apoptosis [68]. This result encouraged us to undertake more detailed studies.

3.3.9 Gene expression analyses via real time PCR

During the last decades, reverse transcription followed by real-time PCR (RT-qPCR) has been considered as the gold standard for quantifying gene expression [69]. Considering that certain genes are strictly related either to the proliferation or the arrest of malignant cells progression, the possibility to monitor their expression might be extremely useful to better understand the effect of antitumor proteins. Therefore, this technique was used to analyse the effect of TFL on the expression of two genes, highly involved not only in early stagesofcancer but also in metastatic stages: they are the SBBP1 and ENO1 genes.

As shown in figure 35, upon TFL treatment, the expression of SBBP1 in both malignant cell lines is highly upregulated, compared to the healthy line. On the contrary, the expression level of ENO1 was found to be downregulated, if compared to the control. Interestingly, while the latter is involved in cell transformation, promotes invasion and metastasis formation [Song Y., 2014], SBBP1 is associated with acetyltransferase p300 and enhances p53 acetylation, being essential for an effective action of tumour suppression [70].

In the light of the above consideration, the ability of TFL to affect the expression of ENO1 and SBBP1 genes might be extremely promising, helping the human body to face and fight appropriately the tumour invasion.

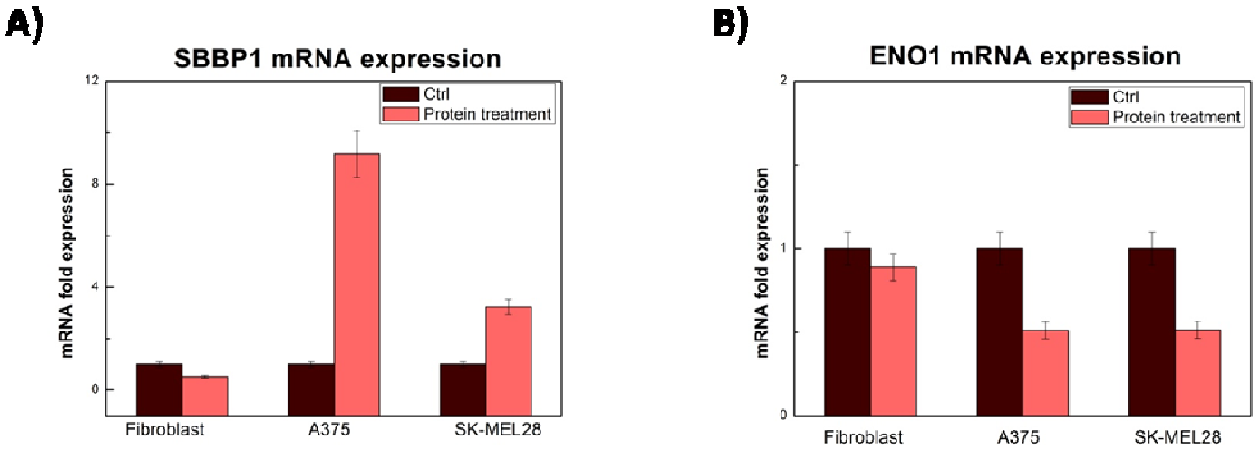


Figure 35: Expression quantification of two main genes involved in cancer progression and arresting (SBBP1 and ENO1), by means of RT-PCR. Analyses carried out in collaboration with the laboratory of Dr. Maria Teresa Valenti.

In addition, it might be probably worth analysing other genes related to cancer progression, such as NPM1, PTGES3, and PARK7, involved in telomere, chromosome and mitochondrial morphology maintenance, respectively.

3.3.10 Quantification of cytokine production via ELISA assay

The *in vivo* immunomodulatory effects of several lectins have been reported during the last few years, paving the way for a new possible use of this property [71-73]. Thanks to the fact that manylectins survive the gastrointestinal tract and enter intact into the blood circulation [74], they are able to stimulate the innate and the adaptive immune response. In particular, dietary lectins have shown the ability to interact with several components of the immune system, causing the development of a proinflammatory state, as well as the stimulation of T cell proliferation and the production of various cytokines (IL and TNF).

Sheu and colleagues studied the ability of a protein that turned out to be TFL to stimulate the immune systemthrough cytokine production which was quantified by means of ELISA assays. In collaboration with the laboratory of Professor Stefano Dusi, based in the Internal Medicine Section of the University of Verona, the TFL was tested in order to confirm or disprove this property. In particular, the quantification of IL-6 and TNF gave the best results, confirming that TFL has a stimulating effect on the immune system, but not as high as that elicited by LPS addition.

In addition, and as stated in Sheu et al., the ability of TFL to stimulate the production of TNF could promote the polarization of macrophages, leading to their activation and toa higher production of IL-12. Together with IL-6, the latter might be able to enhance Th1 cell development, known to be involved in type I inflammation, possibly resulting in the killing of intracellular microorganisms and tumour resistance [75]. However, since the activation of a violent immune system reaction can also cause several side effects, such as autoimmunity and the development of severe allergic reactions, one of the main concerns was to find a way to tune the immune response.

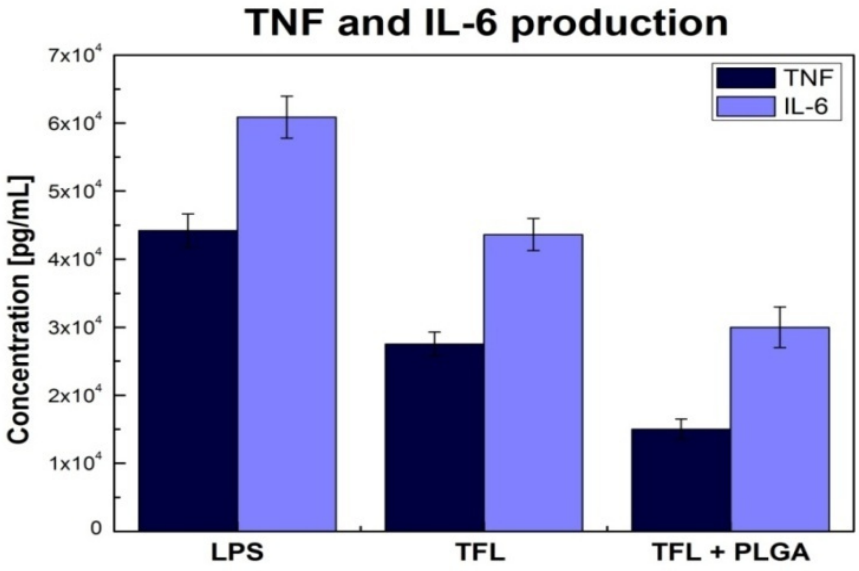


Figure 36: Quantification of cytokine production during the inflammatory process, upon TFL treatment. Analyses carried out in collaboration with the laboratory of Prof. Stefano Dusi.

To do this, TFL was either embedded on a core of PLGA nanoparticlesor linked on their surface, exploiting the main features related to nanoparticle production, such as the size, the possibility to easily tailor their surface, and their biocompatibility.

As shown in figure 36, after encapsulating TFL inside PLGA nanoparticles, a reduction of cytokine production was observed. In addition, by linking potential antitumor moleculesto the nanoparticle surface, a minimization of the side effects against healthy cells could be obtained, making the treatment more efficient and highly directed only to the inflammation site. Moreover, the encapsulation of chemotherapeutic agents into nanoparticles also allows to reduce the dose to be used, minimizing its dispersion into the human organism.

3.4 Recombinant TFL

The expression of recombinant TFL in *E.coli* was undertaken both to open the way to protein engineering and to improve the yield. Its sequence was incorporated into a synthetic gene (accession number EF152774) and the expression and purification steps were carried out as described in paragraph 2.3.

Thanks to the presence of a His-Tag portion at the C-terminus, the purification was carried out by using a Nickel-sepharose affinity column, and the retained protein was then eluted performing an imidazole gradient from 20 mM to 400 mM.

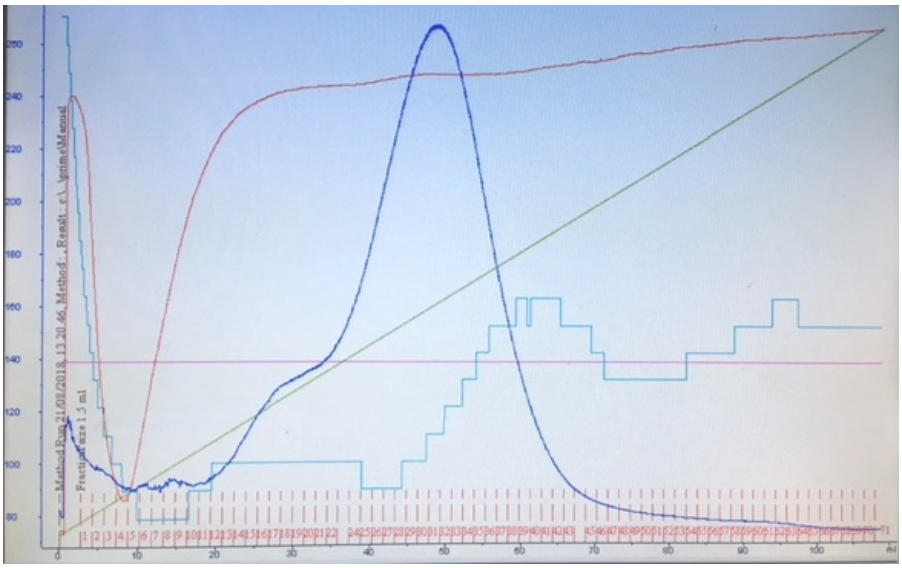


Figure 37: Elution profile of TFL after Ni-Sepharose affinity column.

After elution from the IMAC column, the fractions under the main peak were collected and checked by means of anSDS-PAGE electrophoresis gel. The fractions containing the pure protein were pooled and dialyzed against Tris-HCl buffer (pH 8) to perform thrombin digestion. After a further purification step involving a Sephacryl S-100 HR gel filtration column (Figure 38), purity was assessed by gel electrophoresis and the protein was concentrated to set up the crystallisation trials.

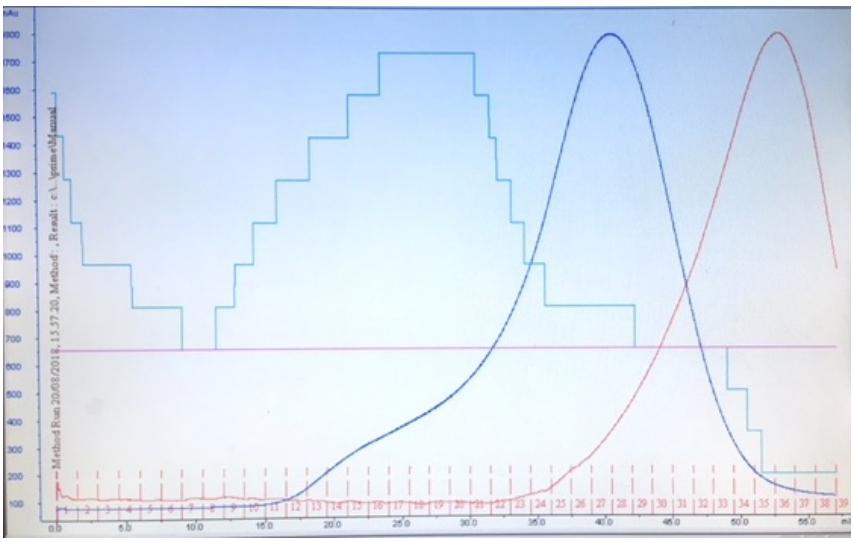


Figure 38 : Elution profile of TFL after Sephacryl-100 HR size exclusion column.

After one week crystals of TFL were obtained in two conditions:

- 0.1 M Sodium chloride, 0.1 M Magnesium chloride hexahydrate, 0.1 M Sodium citrate pH 5.5, 30% PEG 400;

- 1.5 Sodium phosphate monobasic monohydrate, 0.1 Sodium citrate pH 5.5;

The best crystals were soaked into cryo-protectant solution (90 % mother liquor and 10 % glycerol), and the final dataset were collected for the two crystal forms at beamline ID-29 and ID-30A-3 of the European Synchrotron Radiation Facility (ESRF), in Grenoble, France. After being indexed and integrated either with XDS, XIA2 or MOSFLM, the crystals grown in the above-mentioned conditions were found to be both monoclinic (space group $P12_11$) with cell parameters: a= 63.01 Å, b= 64.02 Å, c= 66.15 Å and α=β= 90 ° and γ= 107.45°. The number of molecules in the asymmetric unit is four and the maximum resolution obtained was 2.0 Å.

The TFL available model previously built as described in paragraph 3.2.3, was used as a search probe to perform molecular replacement, using the program Phaser-MR. The models were subjected to the refinement process using Refmac5 and further manual revision by Coot.

Interestingly, the thus obtained three-dimensional modelsturned out to have better final statistics when compared with the previously solved apo TFL structure. In particular, the R-work was 19 %, while the R-free was 24 %. This was probably due to the absence of ice rings in the images collected, which has led to a better data.

Recombinant TFL was correctly folded as shown by the superposition of its structure with that of the purified protein (Figure 39).

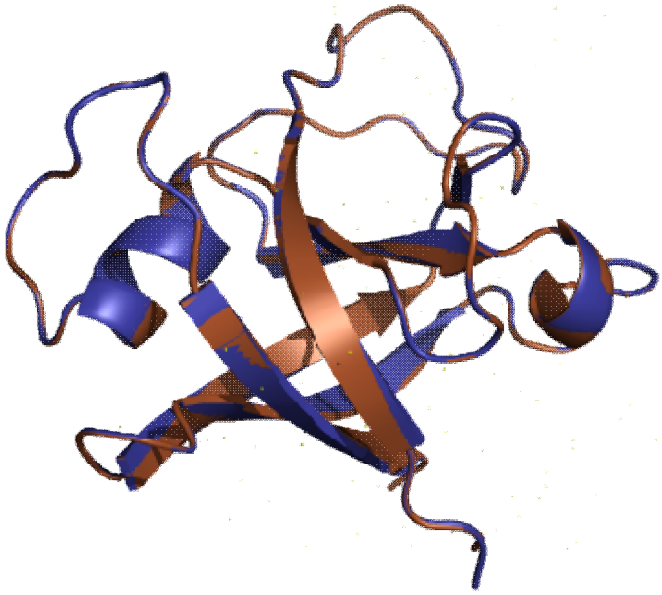


Figure 39: Superposition of the purified TFL (blue model) and the recombinant form (brown model).

Overall, the successful expression of abundant TFL in *E.coli*, (15 mg) confers to the lectin interesting therapeutic potentials, both as an antineoplastic agent and a tumour marker, making it an extremely useful candidate for further applications.

4. Conclusions and Future Perspectives

A new lectin, named TFL, was isolated from the Chinese edible mushroom *Tremella fuciformis*using an optimized purification protocol that allowed to obtain a good yield (29 mg of pure protein starting from 45 grams of dry mushrooms) of its apo form. Following its purification, the three-dimensional structure of the lectin was determined by means of single crystal X-ray diffraction, revealing that it belongs to a well known structural family with a CBM, called β-barrell. In addition, TFL has shown to possess micromolar affinity for N-acetilgalactosamine, the building block of Tn antigen, an exceedingly important epitope which is exposed only on malignant cells surface.

In addition, TFL was tested for its ability to effectively inhibit tumour proliferation on different human cancer cell lines, revealing a powerful effect in reducing human skin malignant cells migration. Moreover, the expression analyses of certain genes involved in tumour progression and arresting, highlighted the attitude of the lectin to help in facing and fighting cancer, without any apparent side effect on healthy cells. This effect is probably mediated by the interaction of TFL with the Tn Antigen, the main carbohydrate expressed on tumour cell surfaces. However, the mechanism of inhibition is not known yet, even if, according to the literature, lectins have not been shown to induce the apoptotic process.

Interestingly, the ability of TFL to stimulate the production of TNF and IL-6 has been demonstrated, thus leading to the activation of an immune system response, a new feature of the lectin family which has never been described before. In addition, the possibility to tune this effect has been explored either by encapsulating TFL into PLGA nanoparticles or by linking it on their surface.

However, one of the questions which remains still unaddressed concerns the correct amount of TFL to be used in order not to provoke a violent immune system reaction, therefore further

studies *in vivo* will be needed to investigate this behaviour. One of the strategies would be to mutate TFL in order to make it as similar as possible to a human counterpart so far not yet identified, and not recognizable as foreign molecule by the immune system.

In addition, the obtainment of more information about the three-dimensional structure of the ligand binding site would be extremely useful in order to understand in depth TFL binding properties and its activity. Furthermore, once known the residues involved in carbohydrates binding, the possibility to use TFL as therapeutic agent could be explored. To accomplish the objective, a combination of computational methods and X-ray crystallography would be a perfect tool to further investigate the properties of the protein-carbohydrate complex.

5. Appendix

In parallel to my main PhD project, I have also worked on another topic, that I have carried out while I was spending my period abroad at King's College in London.

During my 9-month stay in the laboratory of Professor Roberto A. Steiner, I have addressed the question of the role of kinesin-1 isoform in neurological diseases, focusing mainly on the cargo recognition process. In order to do that I have used biophysical approaches, with the aim of elucidating the ability of kinesin-1 in selecting adaptors which rely on the "Y-motif", like the protein JIP_1.

5.1 Introduction

5.1.1 Intracellular transport: a general overview.

Intracellular transport is one of the most important and puzzling processes in the eukaryotic cells. Like an extremely crowded metropolis, the cell needs to find the best way to cope with its large number of macromolecules, not only by establishing a good transport and communication network, but also by taking care of its organelle organization.

In all eukaryotes, intracellular transport is accomplished through the cytoskeleton, a highly complex and dynamic structure that determines the shape of the cell and its spatial organization. Although the cytoskeleton function relies on three polymers: intermediate filaments, actin filaments, and microtubules (MTs), only the latter two are involved in motor-driven transport. Overall, their arrangement gives rise to the so called "road network" for molecular motors, which enables active transport and drags intracellular material, commonly known as "cargo", along the cytoskeletal filaments, to various vesicles and organelles.

The movement of motor proteins on their “tracks” is made possible by the hydrolysis of adenosine triphosphate (ATP), leading to MT conformational changes resulting in a step along the filament itself. According to the literature [75], motor proteins have been grouped into three sub-families: actin-dependent myosin, dynein, and kinesin. In particular, actin-myosin is involved in muscle contraction, non-muscle cell movements, and cell crawling. On the other hand, the sorting of intracellular components and their long-distance transport is accomplished by kinesin and dynein, that are responsible for plus-directed and minus-directed movement, respectively.

Differently from actin and myosin, kinesin and dynein possess the ability to move along the microtubules (MTs) and to specifically bind their cargoes. Unfortunately, even slight alterations in this mechanism, which often occur in a pathological context, are responsible for the development of serious neurodegenerative diseases, such as Alzheimer’s, Huntington’s, and Parkinson’s [76-78].Moreover, kinesin and dynein are often hijacked by viruses and bacteria to fulfil their replication processes, helping them to spread and colonise their hosts [79-81].

In addition, apart from its involvement in transportation pathways, cytoskeletal transport is also essential during cell morphogenesis, functioning and survival, as will be discussed later.

5.1.1 Intracellular transport: historical remarks

In the late 18th century, the first example of intracellular movement was observed by B. Corti in *Charaglobularis* (Fragile stonewort), an alga belonging to the family of *Characeae*. Subsequently, enormous progresses were made in contrast-microscopy, which allowed to identify, for the first time, both the existence of organelles and of protein complexes [82,83]. When analysing the squid giant axon, the presence of sub-microscopic vesicles (30-50 nm) was detected, highlighting the

continuous nature of axonal transport. During the same decade, intra-flagellar transport (IFT) was reported for the green alga *Chlamydomonas* and described as the fast and bidirectional movement of granulo-like vesicles along the flagella, and that was one of the earliest descriptions of motions unrelated to dynein.

In the late 19th century, thanks to the development of even more sophisticated microscopy techniques, alongside with the use of the green fluorescent protein for tagging organelles, a clear picture of the intracellular movements was provided.

In particular, in 1984, W.E. Balch and colleagues described the mechanism of cell-free transport by exploiting a mixture of purified Golgi compartments isolated both from donor and acceptor cells. Briefly, after infecting donor cells - lacking the N-acetylglucosamine transferase enzyme - with vesicular stomatitis virus (VSV), the associated G proteinwas produced [84]. Then, by exploiting the binding between N-acetylglycosamine and VSV G protein, the transportation path within the Golgi compartments was tracked [85].

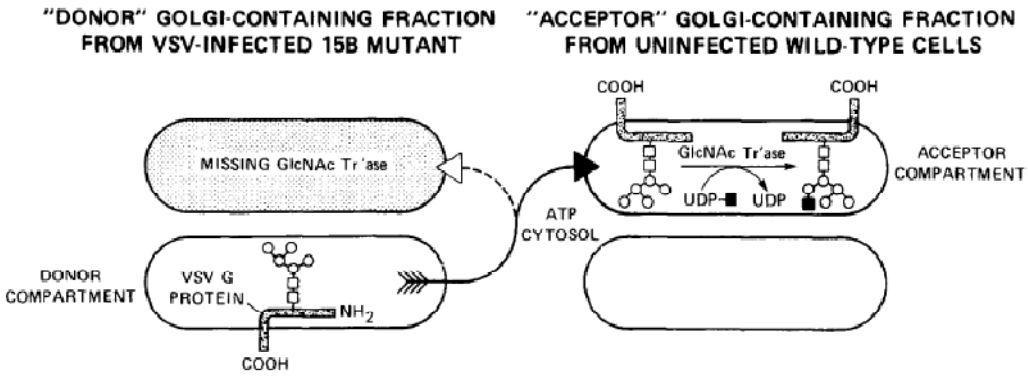


Figure 40: Representation of the cell-free transport, described by W.E. Balch and colleagues by exploiting a mixture of Golgi compartments belonging to both acceptor and donor cells.

After 1985, a series of new discoveries concerning cytoplasmatic transport were published, highlighting, among other facts, the bidirectional nature of microtubule transport [86]. Moreover,

the presence of a motor transport called “kinesin”, both in squid and mammalian brain was assessed, confirming its ability to power active transport, upon ATP binding and hydrolysis, and its unique plus-end directed activity [87].

However, although astonishing progresses have been made in understanding the mechanisms beyond intracellular transport and motor protein functions, these are still not completely understood and a more in-depth analysis is well justified.

5.1.2 Cytoskeleton arrangement

Among the filaments forming the cytoskeleton, microtubules are probably the most rigid and complex structure, being responsible for highly dynamic assembly and disassembly mechanisms. Their ability to cover incredibly long distances - approximately the same order as the diameter of the cell - alongside with their high stability, underline their importance in intracellular transport.

However, as previously described in paragraph 5.1, only kinesin and dynein carry out long-distances transport, while myosin is involved in force generation and cell locomotion.

According to the literature [75], MTs must comply with some *a priori* contradictory constrains, being extremely resistant to thermal fluctuations and mechanical stress, but also prompt to respond to rapid changes. MTs are organized in polymers, whose fundamental unit is composed of a tubulin heterodimer of 8 nm inlength, made up of an α- and a β-tubulin monomer [88].

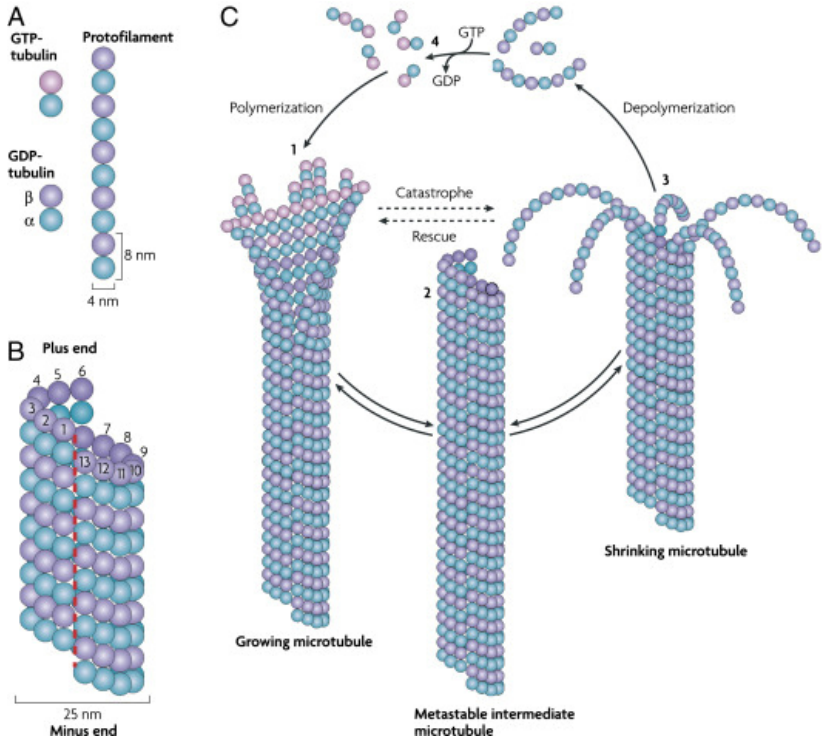


Figure 41: Schematic representation of microtubule arrangement and dynamic instability processes. The α- and β- tubulin dimers associate to form protofilaments of 8 nm each (a), which wrap into a helical structure (b). After reaching a stable and well-organized structure, microtubules can undergo either the depolymerisation process, also known as *catastrophe*, or the inverse transition, called *rescue* (c).

The assembly of tubulin dimers along their long axis leads to the formation of protofilaments, which wrap together to form the microtubule, a helical structure of approximately 25 nm in diameter (Figure 41).

Due to the peculiar asymmetry of the α- and β- subunits, MTs are not only locally polarized but they also have two chemically different ends, named “plus” and “minus”, a feature that is exceedingly important for intracellular transport. In fact, filament polarity is fundamental in determining the walking direction of molecular motors and, conventionally, the term “plus-end” is referred to the portion of the filament that carries the β-tubulin monomer. Interestingly, MTs display the so called “dynamic instability” (Figure 41), resulting in the ability of switching between

phases of growth and shrinkage [89], alternating *catastrophe* (from growth to shrinkage) and *rescue* (from shrinkage to growth) states.

The movement along MTs and their arrangement is coordinated by the self-association and interaction with a broad number of proteins, which have been classified under the name of microtubule-associated proteins (MAPs), and have displayed different structures depending on their function. However, they all share the ability of interacting with the COOH-terminal tubulin domain, alongside with the capacity of stabilizing the microtubules and linking them with other cytoskeletal polymers. MAPs, together with dimethyl sulfoxide, Mg^{2+}, and GTP are essential for *in vivo* assembly of tubulin, taking care of maintaining its stoichiometric ratio after theassembly/disassembly cycles.

During the past years, a large number of MAPs has been identified, both in neuronal and non-neuronal cells, including either high-molecular-mass components (MAP-1A, MAP-1B, MAP-1C) or smaller proteins (MAP-2C) [90]. For the purpose of this study, we will focus on thec-jun N-terminal kinases (JNK) pathway only, which is one of the major signalling cassettes among the MAPK proteins.

However, despite the efforts made in trying to understand thestructure and function of MAPs and their involvement in intracellular transport, many aspects still remain elusive, and require the work of researchers to go deep and increase the knowledge in the field.

5.1.3 Insights into plus-end directed transport

As previously mentioned in paragraph 5.3, plus-end directed transport, also known as “anterograde”, is carried out by the kinesin superfamily motor proteins (KIFs), which are

responsible not only for the secretory vesicle migration to the plasma membrane, but also for the radial movement of the endoplasmic reticulum (ER) membranes and pigments.

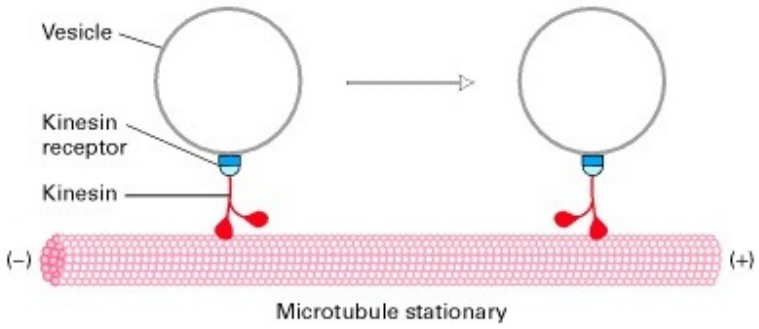


Figure 42: Representation of the plus-end directed transport accomplished by kynesin, highlighting the movement of the vesicletowards the (+) end of the microtubule. (Adapted from [87]).

Crystallographic studies have revealed that KIFs can be described as composed of a head, which contains the ATP-driven motor domain, connected to a tail through a coiled-coil stalk, that mediates protein dimerization. Following a phylogenetic analysis, KIFs seems to be regulated by more than 45 genes and the family contains at least 15 members, named from 1 to 14B, depending on their features and functions. Analysing the arrangement of the motor domain, these families can be further divided into three groups: N-type, positioned at the N-terminus, C-type if located in the carboxy-terminal region, or M-type if in the middle. Interestingly, both N- and M-type kinesins are (+)-end directed motors, while the C-type performs (-)-end directed transport [91].

Although some kinesins are monomeric, which is the case of KIF1A (C351) that plays a fundamental role in neurons, most of them display dimeric organization, being composed of two

heavy chains containing the ATP-driven motor domain and two light chains, appointed to cargo recognition [92].

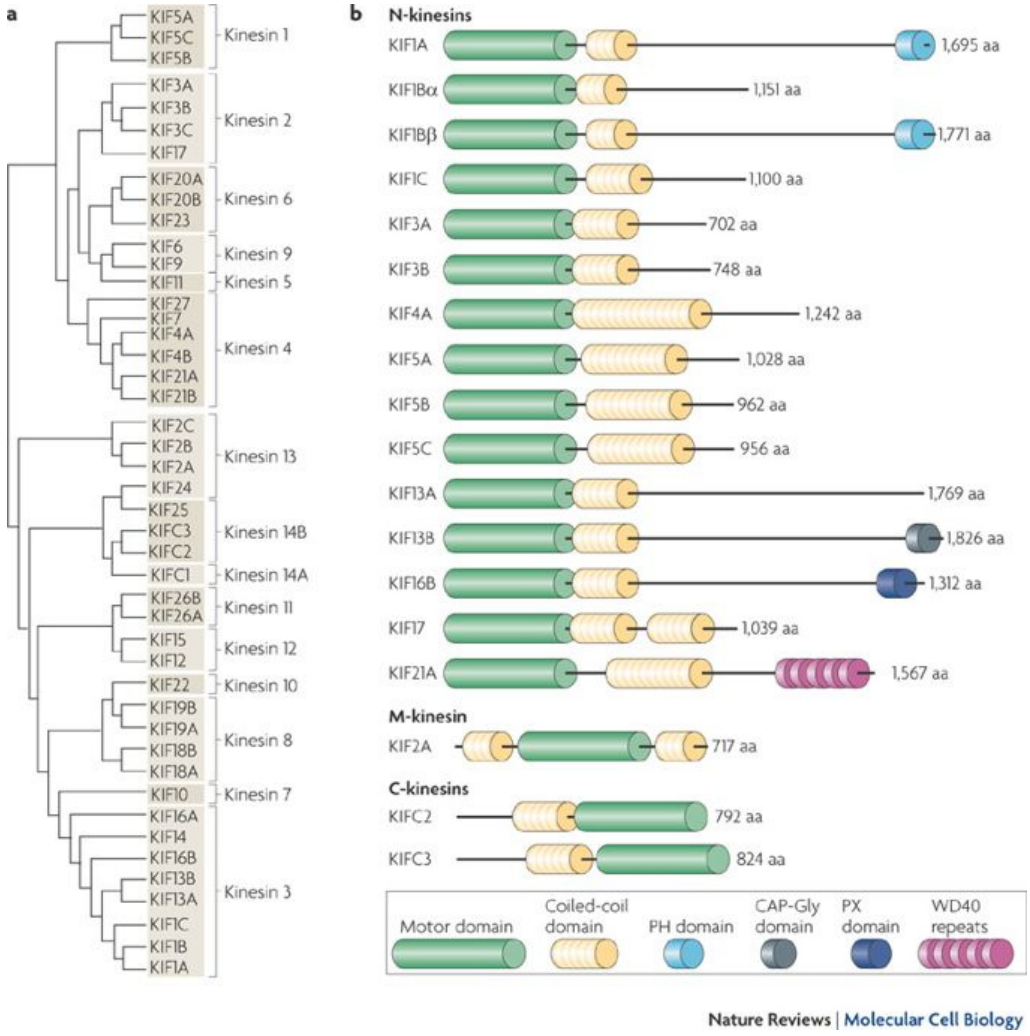


Figure 43: a) The 45 genes that regulate kinesin superfamily motor proteins (KIFs). b) Representation of the domain structure of the most important kinesins, divided into three groups: N-type, M-type, and C-type, according to the position of their motor domain [91].

In vivo, the kinesin heavy chain is associated to the light, a 51 kDa protein that contains a highly conserved region composed of six tandem tetratricopeptide repeats (TPRs), plus various phosphorylation sites and a coiled-coil domain. According to Gauger and Goldstein, the latter is

positioned at the N-terminal region of the light chain and it is essential for heavy chain binding [93].

Kinesin motor domain

In 1996, the structure of the catalytic core of human kinesin was elucidated by F.J. Kull and colleagues, by means of X-ray crystallography, revealing its unique arrangement and features [94]. The kinesin motor domain is composed of an eight-stranded β-sheet, flanked on each side by three α-helixes, plus a small lobe containing three stranded β-sheets.

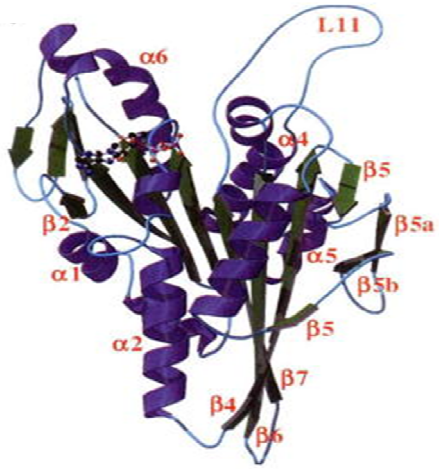


Figure 44: Structure of the motor domain of human kinesin, in complex with MgADP, solved by X-ray crystallography using various isomorphous derivatives [95].

The central β-sheet of the motor domain accommodates a nucleotide-binding cleft which displays a highly conserved motif (GQTGSGKT) responsible for the formation of the phosphate binding loop, also known as "P-loop". The microtubule-binding site is instead located on the opposite side of the core, interacting with tubulin via L11-α4-L12-α5 structural elements, upon ATP hydrolysis.

In addition, the presence of highly conserved switch regions has to be highlighted since they are believed to be responsible for kinesin movement. Specifically, switch I (SSRH) and switch II (DLAGSE), are able to change their conformation during the ATP hydrolysis cycle, displaying a similar behaviour and structural homology to G-proteins. According to Parke and colleagues, switch I and II seem to be involved in closing the nucleotide-binding cleft, via salt-bridge formation, offering the perfect environment to hydrolyse ATP molecules [96]. Lang J. and colleagues, used an optical approach to identify two more structural elements essential for accomplishing the motility process, which are the N- and the C-terminus of the head domain, named cover strand (CS) and neck linker (NL), respectively [97].

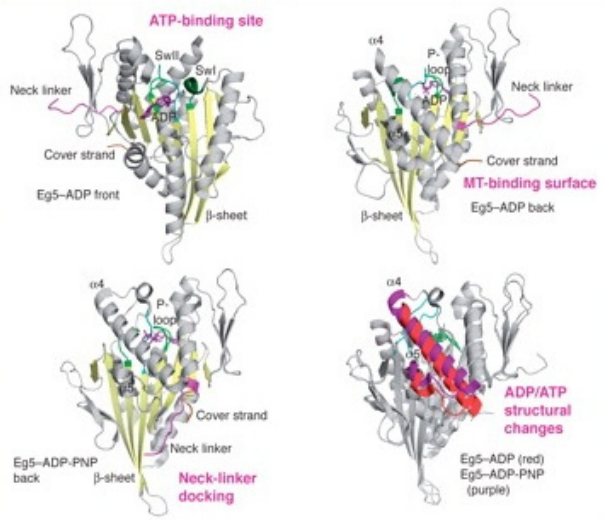


Figure 45: Representation of kinesin catalytic core. Adapted from [98].

Although the catalytic core is extremely well conserved among kinesins belonging to different families, cellular function and motility can largely vary, since they are carried out by regions outside the motor domain. Specifically, the region adjacent to the catalytic core is responsible for kinesin directionality, presumably by interacting with the head in order to allow the correct positioning of the stalk [99,100].

Some kinesins lack either the neck linker, which is the case of kinesin 14, having the head directly connected to the end of the stalk, or both the neck linker and the coiled-coil stalk, like kinesin 13.

Kinesin motility

By performing in vitro experiments, two main features of kinesins have been revealed: their step-size and force. Specifically, a two-headed kinesin molecule was found to move in 8-nm steps [101] and to exert a force of 6 pN [102], suggesting its ability to bind only to an α- or β-tubulin monomer.

Recently, the motility cycle of kinesin-1, one of the most extensively studied family members, has been described by Wonmuk Hwang and colleagues [103]. As previously mentioned in paragraph 5.4.1, in order to walk on the MTs, kinesin forms a dimer, and uses one molecule of ATP per step. After ATP hydrolysis, when only ADP is present in the environment, it reaches the "nucleotide-free APO state", in which it unbinds from the MTs, that displays a lower higher affinity for it.

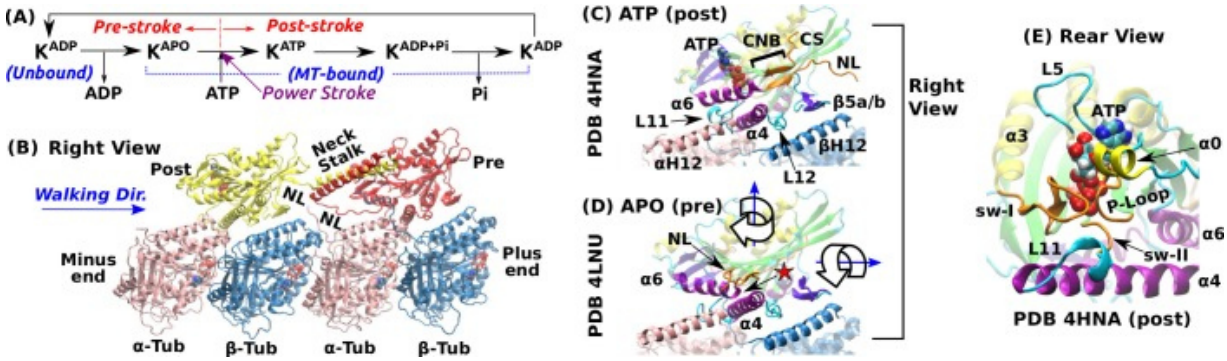


Figure 46: Representation of the structure and the motility cycle of kinesin. A) ATPase cycle of the motor head during pre- and post-stroke states. B) Kinesin dimer bound to the MT, showing the post- and pre-stroke states of the rear and the front heads, respectively. C-D) Comparison between the orientation of α-4 and α-6 during pre- and post-stroke states. E) ATP-binding pocket alone [103].

Following ATP binding, the cover strand and the neck linker undergoes the formation of the cover-neck bundle (CNB), generating a forward force called “power stroke”. Interestingly, the motor head orientation differs during the pre- and post-stroke state, being tilted rightwards to the (+)-end direction in the first case, and rotated in the opposite way in the second.

Since the ATP hydrolysis rate is 10-fold higher in MT-bound kinesin rather than in the isolated molecule, the nucleotide pocket must be allosterically controlled by the MT interface [94]. Interestingly, the MT-contacts can be divided into three different groups: the rear (L11 and α-4), the front (L8/β5), and the middle (L12 and α5). While the rear and the middle interact with α- and β-tubulin and are present in all the states, the front one lacks non-polar contacts, being less robust in pre-stroke states. The rear contactundergoes important conformational changes, being located after switch II, only when the motor head is tightly bound to the MT.

Although the members of the kinesin superfamily share some common features, they display important differences in terms of motility, underlying the evolutionary divergence and the diverse cellular functions.

Kinesin cellular functions

As previously described (paragraph 1.4), there are approximately 15 subfamilies of kinesins and, according to the phylogenetic analysis, some of them are able to perform similar cellular functions in different cells or organisms [104]. Vesicles transport to the presynaptic region of the axon is accomplished by kinesin-1, being responsible not only for mediating the signal transmission but also for preventing the development of severe neurological disorders. Moreover, proteins belonging to the kinesin-1 subfamily are able to easily bind either lysosomes or mitochondria, as well as receptors and adaptor proteins, fulfilling the role of a cargo transporter and regulating

signal transduction [105]. Due to its interaction with disease-related proteins, kinesin-1 is believed to be involved in many neurological disorders, such as Alzheimer's, Huntington's, and Parkinson's diseases.

Differently, kinesin-2 proteins regulate flagella and cilia growth and development, being essentially the main motor that powers intra-flagellar transport [106].

Other types of kinesins carry out their functions during mitosis and meiosis, having essential roles not only in separating centrosomes and attaching chromosomes to spindles, but also in driving their assembly. According to recent studies [107], kinesins seem to play a role during kinetochore disassembly, producing tension on its fibres and depolymerising the microtubules to allow chromosome movement. Kinesin-2 mutations and further dysfunctions are often associated to cell cycle arrest, thus leading not only to cell death but also to female sterility [104].

In support of this theory, many other experiments have been performed during the last decades, highlighting, once more, the kinesin fundamental role during meiosis and mitosis. To mention a few examples, kinesin-5 and 6 are responsible for generating the sliding force in mitotic spindle, while kinesin-14 is essential during anastral oocyte spindle assembly in *Drosophila*. Moreover, kinesin-7 seems to allow a good communication between the spindle and the chromosome, whereas kinesin 4 and 13 are mainly kinetochore motors that take care of either assembly or disassembly of microtubules [108].

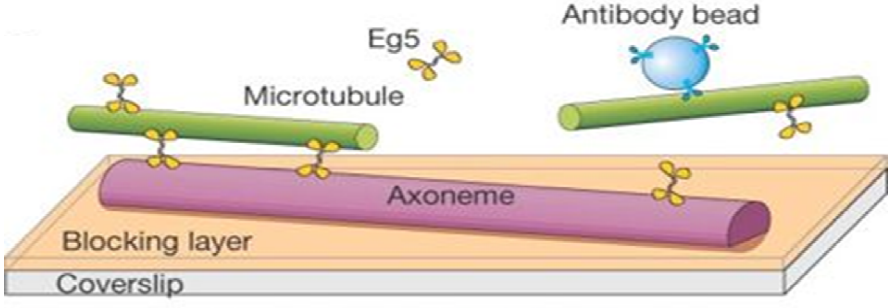


Figure 47: *In vitro* kinesin-5 (Eg5) operational arrangement during microtubule sliding. Immobilized axonemes (magenta) onto a coverslip surface are attached by MTs (green), via Eg5.

5.1.4 Kinesin and its binding partners

Although the knowledge of kinesin transporters is still limited, the need for additional binding partners to fulfil their role, has been clearly demonstrated. By performing a two-hybrid screen, a molecular biology technique used to test physical interactions between proteins, some of them have already been identified, leading to a better understanding of their involvement in intracellular transport.

According to the literature, both heavy and light chains of kinesin seem to play a role during cargo recognition, allowing their movement along the microtubules upon protein interactions. However, due to the lack of the light chains, in certain species of fungi, such as *Neurospora crassa*, the C-terminal portion of the heavy chain is necessary and sufficient for binding cargoes (Seiler et al. 2000). Interestingly, in higher eukaryotes, the loss of KLC functions can lead to severe neuronal defects, sometimes leading to death [109,110].

In 1998, Verhey and colleagues established the role of KLCs in promoting the inhibition of KHC functions when cargoes are not bound to it, a state which we will refer to as “inactive state” that will be further discussed in paragraph 5.3 [105].

However, according to the literature, the two most important regions for binding cargo appear to be the TPR one, alongside with the heavy chain stalk region. Specifically, the KLC TPRs are able to recognize and specifically bind various numbers of proteins belonging to the JNK signalling pathway family, called JIP (JNK-interacting proteins) [111]. So far, JIP proteins have been identified

both in *Drosophila, C.elegans* and mammals, and have been shown to be encoded by four different genesnamed JIP_1, JIP_2, JIP_3, and JIP_4.

Interestingly, all of them have been proposed to act as scaffolding proteins for the JNK- kinase signalling pathway and to take part as linkers between the motor proteins and their cargoes.

According to Yasuda et al., JIP_1 and JIP_2 share 50 % sequence homology, similar domain structures, and can either form homodimers or heterodimers. On the other hand, JIP3 and JIP4 don't share any sequence relationship with both JIP_1 and JIP_2 but can still form oligomers with the latter and among themselves. According to the literature [112,113], JIP proteins bind simultaneously to three components of the JNK signalling pathway, which are the following: MAP kinase kinase (DLK and MLK3), MAP kinase kinase (MKK4 and MKK7), and MAP kinase JNK.

Moreover, in 1999, Meyer and colleagues described the ability of JIP_1 and JIP_2 to interact with rhoGEF and ApoER2, which are the RhoGTPase nucleotide exchange factor and the Apolipoprotein E receptor 2 respectively.

JIP proteins

As previously described in paragraph 5.5, JIP proteins act as signalling scaffolds for JNK kinases, playing a role in linking cytoplasmatic signalling molecules to ion channels, maximizing the signal propagation and maintaining pathway specificity.

Concerning their localization, JIP proteins have beenshown to be extremely abundant at synaptic junctions during brain development and to be highly polarized at the tips of neuronal cells, corresponding to the regions where plus end MTs are located.

As mentioned before, in mammals there are four genes encoding the JIP family members but none of them displays sequence similarity to yeast scaffold proteins [114]. According to the literature, JIP_1 and JIP_2, also known as Islet Brain 1 and 2 (IB-1 and IB-2) share structural homology, containing at the C-terminus a Srchomogy 3 (SH_3) domain plus a phosphotyrosine-binding domain (PTB) [115].

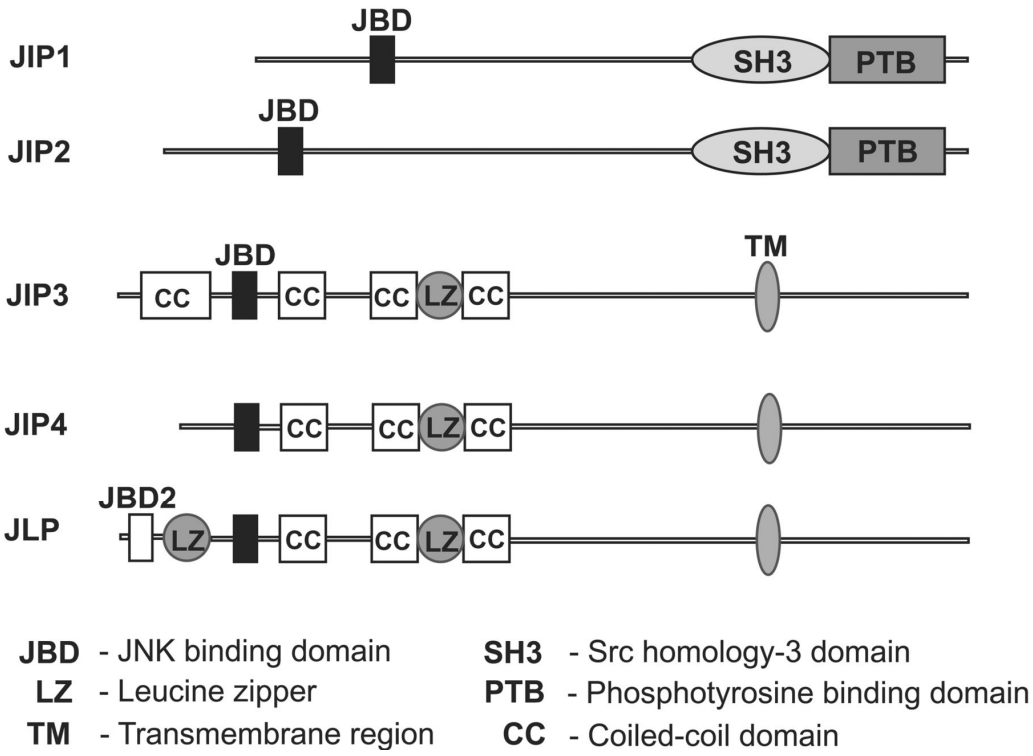


Figure 48: Mammalian JIP family members can be divided into two different groups, according to their structure. As can be seen from the schematic representation, JIP_1 and JIP_2 share structural homology, likewise JIP_3 and JIP4. Differently from JIP_4, its variant JLP, contains an additional leucine zipper domain (LZ) and a second region able to bind JNK (JBD2) [116].

With the only exception of the JBD domain, JIP_3 and JIP_4 have completely different sequences, possessing a transmembrane region at their C-terminus, plus a coiled-coil region and a leucine zipper at the N-terminus [117].

Although all the mammalian JIP proteins are mainly expressed in brain, a high content of JIP_1 and JIP_2 has also been found in pancreatic β-cells and endocrine tissues, whereas JIP_3 and JIP_4 are present in testis, heart, kidney, and liver.

5.1.5 Kinesin-1 isoform

Kinesin-1 is one of the members of the kinesin superfamily of microtubule motor proteins. According to the literature[75], kinesin-1 is involved in a number of roles, such as the movement of vesicles in squid giant axons and the dynamics of the Golgi apparatus, mitochondria and the endoplasmic reticulum.

Like the other members of the kinesin superfamily, kinesin-1 is also composed of two motor subunits and two light chains, named KHCs and KLCs respectively. According to Gyoeva and colleagues [118], the molecular mass of the subunits can vary in different species, being in a range from 110 to 130 kDa for KHCs, and from 51 to 76 KDa for KLCs.

In addition, the kinesin-1 heavy chain is able to hydrolyse ATP and to interact with MTs, either at their globular or amino-terminal portions [119,120]. Concerning the light chain, its carboxy-terminal tail region plays a role in binding membranes and it is responsible for cargo recognition and attachment [121]. On the other hand, the TPR domain of KLC displays a highly conserved sequence which is composed of a leucine-phenylalanine-proline (LFP) triplet, flanked by negatively charged amino acids, such as aspartic and glutamic acid. Recently, the structure of KLC_2^{TPR} with its $SKIP^{WD}$ cargo has been elucidated by Pernigo and colleagues [122], revealing the presence of six TPR (TPR1-6) repeats arranged to form a right-handed super-helical conformation. Interestingly, a non-TPR helix is positioned between TPR5 and TPR6. A coiled-coil region links together the two KHC subunits, which are responsible for MTs binding and ATP recognition and hydrolysis. As

shown in figure 10, a short neck region connects the head to the stalk of KHC, which is in turn intersected by two hinge domains, responsible for kinesin-1 flexibility and folding.

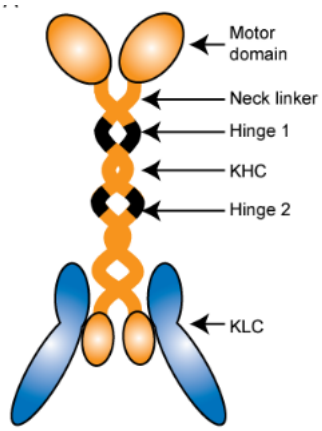


Figure 49: Representation of the structure and the major subunits of kinesin-1.

Depending on the ionic concentration, kinesin-1 can be found *in vitro* in two different conformations: a folded state, which is supposed to be the "inactive" form of the motor transport, and an extended molecule, that is thought to be the "active" conformation. The folded state seems to promote the interaction of the KHC C-terminus with the N-terminal portion of itself, a mechanism that will be analysed in depth later (paragraph 5.6.1).

As previously described, the KHC and KLC subunits can interact via the heptad repeat region and both of them are able to bind cargoes and seem to be essential for kinesin-1 activity [93]. However, during the last decades, many examples of transport in which KLC binding is not required, have been discovered. In addition, in lower organisms, the light chain subunits of kinesin are not expressed, which is the case of the fungus *Neurospora crassa*[123]. According to Kanai and colleagues [124], due to deletion of thekif5 gene, three different isoforms of kinesin heavy chain are expressed, named KIF5A, KIF5B, and KIF5C respectively. Whereas theA and C isoforms are mainly expressed in neurons, KIF5B shows a glial cell distribution pattern. During the last years,

the presence of mRNA belonging both to KIF5A and KIF5C has been reported in human epithelial cells, thus raising the hypothesis that the three isoforms could be equally expressed at least in some cell types [125].

In a similar way to that of KHC, the regulation of KLC expression relies on four different genes, leading to the production of isoforms from 1 to 4 [126]. Interestingly, while KLC3 can only be found in spermatids, and KLC1 is abundant in neuronal tissue, KLC1 is ubiquitously expressed. Furthermore, at least nineteen KLC1 variants have been reported, named KLC1A-S, probably produced by alternative mRNA splicing.

Overall, despite the majority of the interactions between cargos and KLC relies on the presence of the tetratricopeptide repeat region, most of the mechanisms involved remain still elusive.

Kinesin is autoinhibited when not involved in active transport

Although the mechanism of kinesin-1 inhibition is still not yet understood, some hypotheses have been proposed during the last decades. Specifically, it seems that the auto-inhibition process relies on the binding between theKHC tail and the motor domain of kinesin-1 itself, preventing the futile waste of ATP when it is not bound to cargoes. Furthermore, by adopting the inactive conformation, kinesin-1 can be accumulated at the plus end of MTs [127].

In 2010, Rice and colleagues demonstrated the fundamental contribution of the heptad repeat-TPR linker region in destabilizing the interaction between the MTs and the heavy-chain carboxyl-tail [128].In addition, they showed the effect of the TPR domain in reducing the affinity of the KHC head for its C-terminal region, a mechanism involving both steric and electrostatic factors.

More recently, Kaan and colleagues solved the 2.2 Å structure of a kinesin motor domain dimer in complex with one of its tails containing the IAK region, showing the main differences with the apo-dimer [129]. The conformation of the two catalytic cores are similar, varying only in terms of angle

of attachment of the coiled-coil region. According to their results, the IAK region seems to be necessary and sufficient for binding.

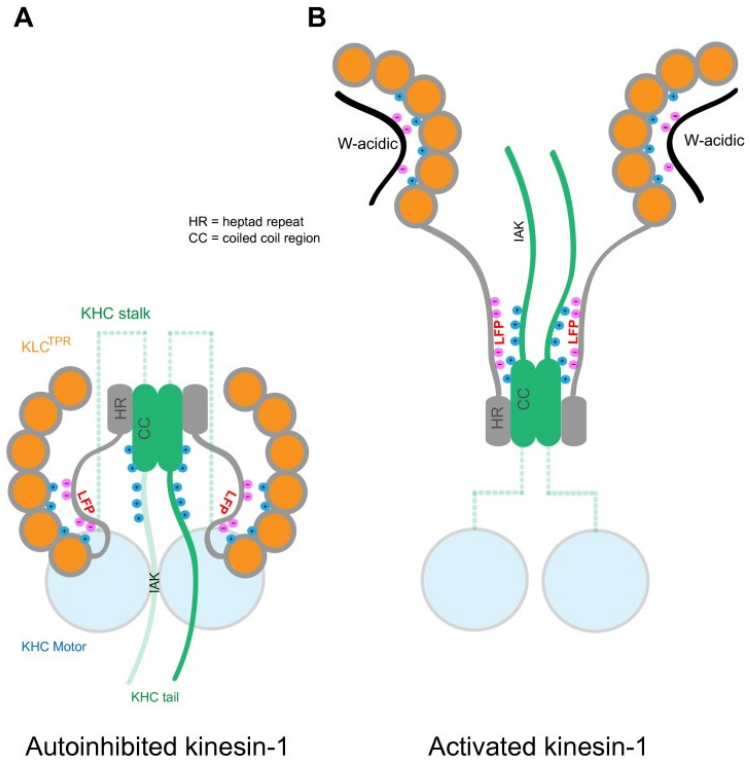


Figure 50: Representation of KHCs coloured green and KLCs coloured orange/grey, in their auto-inhibited state (A) and in their cargo-bound states (B).

Kinesin-1 cargo binding

Recently, Yip and colleagues suggested a possible model for kinesin-1 inhibition, involving the LFP portion of the KLC subunits [130]. Briefly, they have shown how a mutational disruption of the LFP motif can enhance the ATPase activity of purified kinesin-1. In the light of the above considerations, they proposed the interaction of the heavy-chain tail as a result of the displacement of the linker from the TPR, mediated by the W-acidic motif. In addition, a similar behaviour was found for both CSTN1 and KLC1, suggesting that the same principle could also be applied to other light-chain isoforms.

As previously discussed in paragraph 5.5, to accomplish active transport, a number of adaptor proteins are required, which provide a link between kinesin-1 and its cargoes [131]. However, while vesicular cargoes can bind both to KHCs and KLCs, the recognition process is often mediated by the interaction of the KLC^{TPR} domain with specific residues on the adaptor sequences. Specifically, a tryptophan flanked by acidic residues (WD/E or WD/Q) - that can either be aspartic acid or glutamic acid – represents the most common recognition peptide within adaptors.

Interestingly, recent works have demonstrated the presence of "W-acidic" regions in the SifA-kinesin interacting protein (SKIP), in the neuronal protein calsyntenin-1 (CSTN-1), in dynein intermediate chain (DIC), nesprin-2, gadkin, and cayman ataxia protein (BNIP-H) [79, 132-139]. In particular, the structure of $KLC2^{TPR}$ with its $SKIP^{WD}$ cargo has recently been solved by Dodding M.P. and colleagues, revealing structural differences between the apo- and the cargo-bound form of $KLC2^{TPR}$. Upon cargo recognition, the rigid movement of the N-terminus of both TPR2 and 3, leads to the formation of an optimal pocket for $SKIP^{WD}$, promoting its binding. As shown in figure 12, all the residues of the $SKIP^{WD}$ cargo are involved in the complex formation, stabilizing not only TPR2 and 3, but also the internal helix of TPR4, leading to the development of a concave groove carrying apositive electrostatic charge.

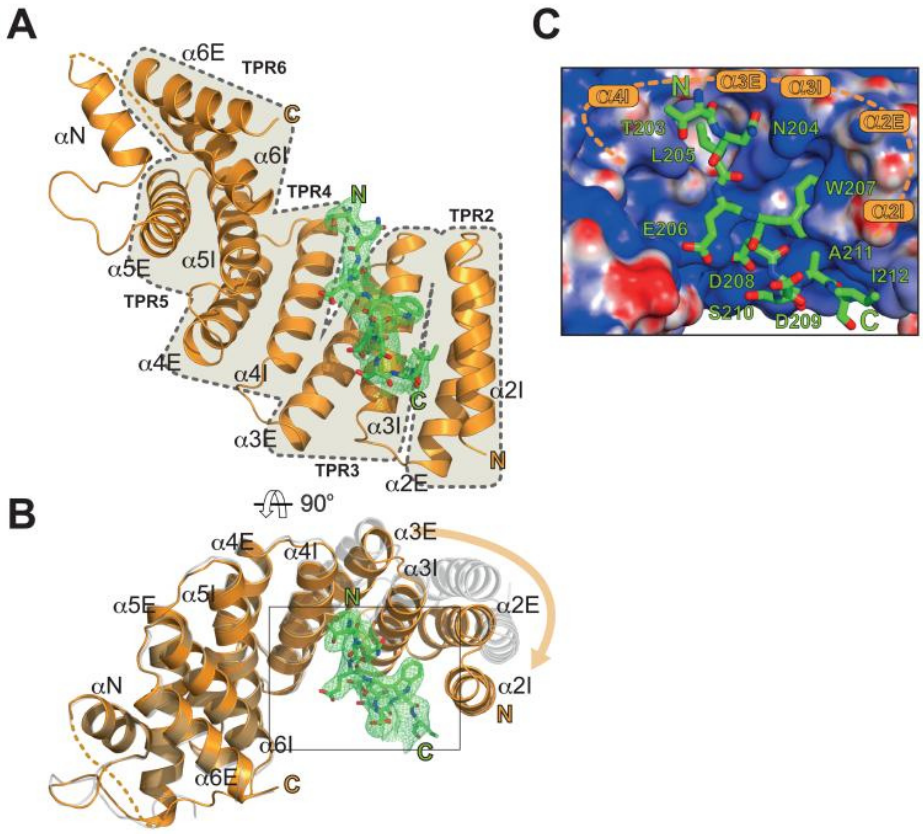


Figure 51: Structure of the KLC2^{TPR} and SKIP^{WD} complex, representingthe SKIP^{WD} motif in two orthogonal orientations.

However, the role of every “W-acidic” motif can vary depending on the protein groupexamined. For instance, in the case of SKIP, the first tryptophan motif (WD) is more effective in promoting virus transport than the second, whereas for calsyntenin, bothare necessary to perform intracellular enveloped virus transport.

Overall, according to Dodding and colleagues, the presence of at least one “W-acidic” motif seems to be necessary and sufficient to mediate viral transport, suggesting that aprotein with a single tryptophan may recognise and specifically attach kinesin-1 [79]. Nevertheless, this process is subordinate not only to protein binding affinity, but also to sequences outside the motif, and to the stoichiometric relationship between KLC - binding motif and individual molecules.

In the light of the above considerations, it is clear that KLC binding to its cargo is dependent not only on the protein involved, but also on the surrounding environment.

Recently, some proteins containing the WD/E motif have been shown to possess the ability to distinguish between KLC_1 and KLC_2, being able to selectively bind only one of them. On the other hand, according to immunofluorescence analysis, W-acidic proteins like SKIP and Calsyntenin, are capable of recruiting both of the isoforms, highlighting their ability of interacting with them outside the context of infection.

Besides the strong interaction with W-acidic motifs, kinesin-1 has been shown to be able to interact with other peptides, such as JIP_1, an adaptor that is involved in the transport of the amyloid precursor containing vesicles. Specifically, kinesin-1 is able to bind the C-terminal of JIP_1, named JIP^{c11}, whose amino acid sequence is characterized by the presence of a tyrosine flanked by acidic amino acids (YTCPTEDIYLE), also known as “Y motifs”. Differently from “W-acidic” sequences, the latter has a clear preference for the KLC_1 isoform rather than KLC_2, being necessary and sufficient for binding [140]. Recently, an example of Y-motif dependent binding has been clearly illustrated by Steiner R.A. and colleagues, by solving the 2.3 Å structure of the complex between KLC_1^{TPR} and the C-terminus of TorsinA, also known as “dystonia-related protein”. As shown in figure 13, the recognition of theY-acidic and W-acidic motifs is accomplished by partly overlapping binding sites which are slightly displaced along the channel generated by the TPR super-helical arrangement. While the W-acidic motif binds to the N-terminal end of it, between TPR2 and TPR3, the Y-acidic motif is more shifted toward the C-terminal portion, between TPR3 and TPR4.

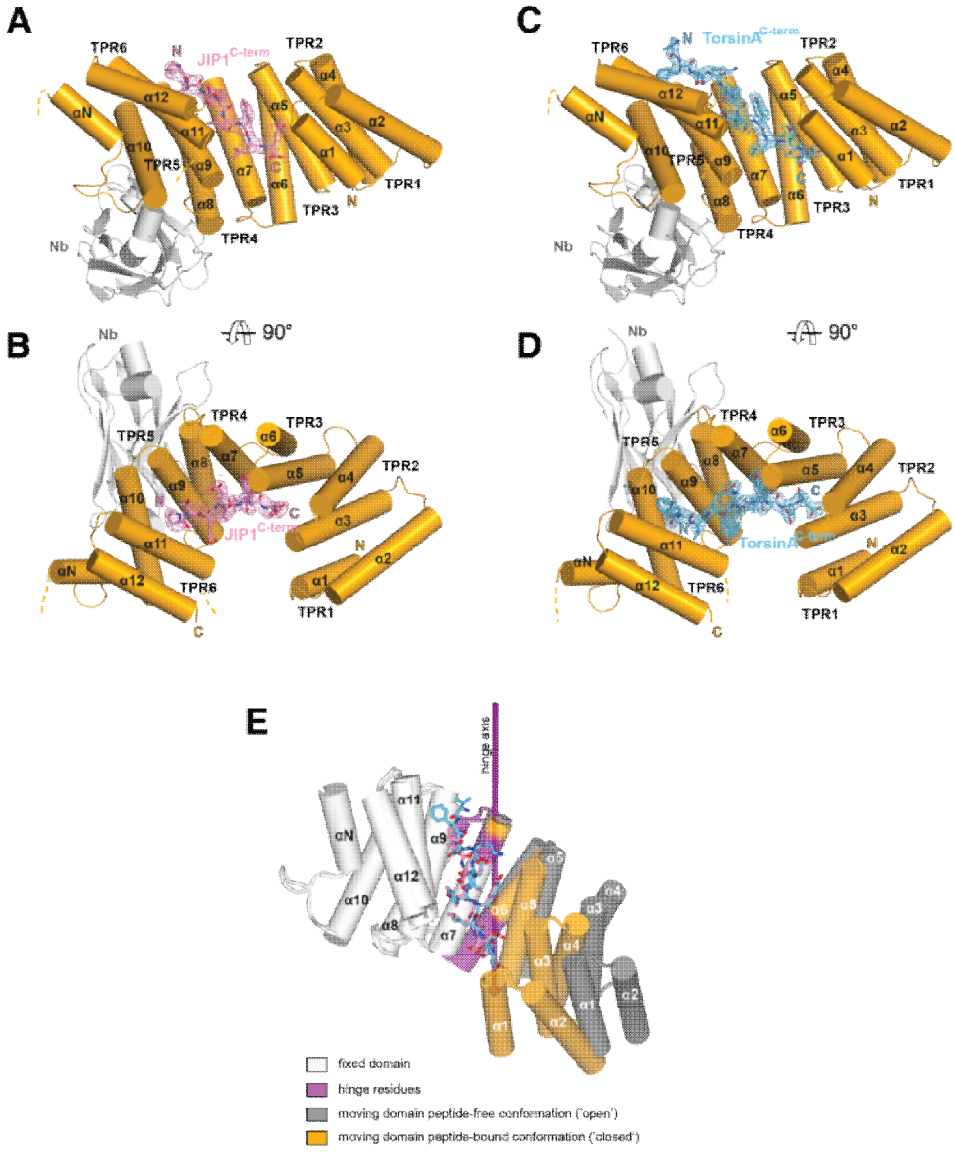


Figure 52: Structure of KLC1^{TPR} and JIP^{C11} complex, representing JIP^{11} motif in two orthogonal orientations (A-B). Structure of KLC1^{TPR} and TorA complex, representing TorA motif in two orthogonal orientations (C-D). Representation of the binding site shift along the inner channel generated by the TPR right-handed super-helical arrangement.

However, despite the above consideration, JIP1^{C11} is insufficient to promote transport activation, requiring further interactions with the heavy-chain tail binding protein Fez1 and the coiled-coil region of the heavy-chain itself [134, 141, 142].

5.2 Aim of this work

In the light of the above consideration, the present work focuses on the use of both a structural and a biophysical approach to elucidate the ability of kinesin-1 in selecting adaptors which rely on the "Y-motif", like the previously described JIP1. A combination of techniques for the analyses of protein-protein interactions will be exploited, mainly involving ITC and FP assays to assess the binding affinity between KLCTPR and the C-terminus of "Y-motif" adaptors.

Moreover, considering the difficulties encountered in collecting X-ray data for the JIP_1^{Cterm} - KLC_{1T}^{PR} complex, nanobodies will be exploited as crystallization chaperones, with the aim of improving the diffraction resolution. In particular, among a pool of 29 tested, three of them will be chosen to test their specificity towards KLC_1^{TPR}, rather than KLC_2^{TPR}, by means of isothermal titration calorimetry.

5.3 Materials and Methods

5.3.1 Constructs preparation

Molecular biology techniques for cloning

Codon-optimized DNA sequences encoding the TPR domain of mouse KLC_1 (residues 205-496, Uniprot Q5UE59, KLC_1^{TPR}) were purchased from Genscript and subcloned between the NdeI/XhoI sites of a pET28 vector (Novagen). The latter was further fused either to the C-terminal eleven amino acids of human JIP_1 (residues 701-711, Uniprot Q9UQF2, $JIP^{C\text{-}term}$), or to the C-terminal twelve amino acids of mouse TorsinA (residues 322-333, Uniprot Q9ER39, $TorsinA^{C\text{-}term}$), via a (Thr-GLy-Ser)$_5$ or (Thr-GLy-Ser)$_{10}$ flexible connector.

Nanobodies R2KL27, R2KL29, and R2KL76 were generated following the work of Pardon and colleagues. Briefly, llamas were immunized with purified KLC_1^{TPR} – $JIP^{C\text{-}term}$, by the VIB Nanobody Service Facility of Vrije Universiteit Brussel (VUB), leading to the obtainment of a panel of 36 antigen-specific nanobodies, supplied in a pMECS vector. Among them, only a subset having the most divergent amino acidic sequences were cloned between PstI/BstEII sites of a pHEN6c vector, resulting in the expression of C-terminally His_6-tagged nanobodies in the periplasmatic space of *E.coli*.

The above mentioned work has been performed by Dr. Stefano Pernigo (Professor Roberto A. Steiner Laboratory) before my arrival.

Chimeric protein expression protocol

All the chimeric proteins used in the present study were expressed in the *E.coli* BL21(DE3) cell strain. Briefly, single colonies were picked from a replica plate and grown in LB media supplemented with Kanamycin, under constant shaking (220 rpm), at 37 °C, for 3 to 4 hours. Then, the small-scale culture was used to inoculate 4x1 L cultures that were further incubated at 37 °C, until reaching an OD of 0.4-0.5.

Since in pET28 protein expression is controlled by the T7 RNA polymerase promoter, protein synthesis was induced by the addition of 0.4 mM Isopropyl-β-thiogalactoside (IPTG) and carried out at 18 °C. After 14 to 16 hours, cells were harvested by centrifugation at 8000 rpm for 20 minutes and the pellet was either used straightaway or stored at -80 °C.

Chimeric protein purification protocol

Cell pellets harvested followingthe mentionedexpression protocol were resuspended in buffer A (8-10 mL per each g of pellet) supplemented with protease inhibitor cocktail and 5U/mL of

Benzonase endonuclease. Cell lysis was carried out by several sonication steps and the insoluble material was further spun down by centrifugation at 20000 g for 40 minutes, at 4 °C. Once collected, the supernatant was filtered using a 0.22 µm pore size microfilter and loaded onto a Ni-HisTrap HP column for immobilized metal affinity chromatography (IMAC), previously equilibrated with buffer A.

After eluting with an imidazole gradient, all the fractions containing the target protein were collected, loaded onto an SDS-PAGE electrophoresis gel, and checked for their purity. The pure fraction was further dialyzed against imidazole-free buffer A, at 4 °C, overnight.

Removal of the purification tag was carried out by incubating the samples with thrombin protease covalently bound to agarose, for approximately 6 hours.

A filtration step over a gravitational column was required in order to get rid of the beads. The eluate was further filtered using a 0.22 µm pore size microfilter and loaded onto a Ni-HisTrap HP column previously equilibrated with buffer A. The flow-through containing the untagged material was collected, concentrated and the last purification step was carried out by size exclusion chromatography (SEC) on a 16/60 HiLoadSuperdex 75 column, previously equilibrated with low salt buffer A.

Nanobody expression protocol

The three nanobodies (RKL27, RKL29, and RKL76) used in the present study were expressed in *E.coli* WK6 cells.

Briefly, single colonies were picked from a replica plate and grown in TB media supplemented with Ampicillin, under constant shaking (150 rpm), at 37 °C, overnight. Then, the small-scale culture was used to inoculate 1 L culture, supplemented with 0.1 % glucose and further incubated at 37

°C, until reaching an OD of 0.6-0.8. Subsequently, the temperature was lowered to 28 °C and protein synthesis was induced by the addition of 1 mM of IPTG. After 14 to 16 hours, cells were harvested by centrifugation at 6283 rpm for 30 minutes and the pellet was either used straightaway or stored at -80 °C.

Both the nanobodies and the chimeric protein expressions were carried out with the help of Dr. Magda Chegkazi (Professor Roberto A. Steiner Laboratory).

Nanobody purification protocol

Cell pellets harvested with the above described expression protocolwere resuspended in TES buffer (12 mL for 1 L culture) and vigorously shaken for 1 hour, at 4 °C. Subsequently, 18 mL of TES/4 were added to the mixture and further incubated for an additional hour, as described above. After centrifuging for 30 minutes at 4 °C, the periplasmatic fraction was purified first by IMAC and then by SEC using a HiLoadSuperdex 75 column.

Both the nanobodies and the chimeric protein purifications were carried out with the help of Dr. Magda Chegkazi (Professor Roberto A. Steiner Laboratory).

Buffer A	**Low salt buffer A**	**TES buffer**
50 mM HEPES buffer at pH 7.5	50 mM HEPES buffer at pH 7.5	0.2 M Tris pH8.0
500 mM NaCl	150 mM NaCl	0.5 mM EDTA
5 mM β-mercaptoethanol	5 mM β-mercaptoethanol	0.5 M sucrose

5.3.2 Biophysical characterization

Isothermal titration calorimetry

All the samples for ITC measurements were prepared following the same protocol, in order to reduce the variability of the results. Briefly, after being extensively dialyzed in low salt buffer, the samples were centrifuged at 10000 rpm for 2 minutes at 4 °C, to get rid of possible aggregates.

The experiments were carried out in a MicroCal ITC instrument, setting a temperature of 20 °C, and the two KLC^{TPR} isoforms were tested for their ability to bind nanobodies 27, 29, and 76.

After fitting the data, the binding constants and the other thermodynamic parameters were calculated with the Origin software package, assuming a singlebinding site.

All the ITC measurements concerning KLC_1^{TPR} were performed by Conor Treacy and Dr. Stefano Pernigo (Professor Roberto A. Steiner Laboratory), while the measurements regarding KLC_2^{TPR} were carried out with the help of Dr. Magda Chegkazi (Professor Roberto A. Steiner Laboratory). In particular, I took care of preparing the samples and acquiring some of the thermograms regarding nanobodies binding to KLC_2^{TPR}.

Fluorescence polarization

All the peptides used for the FP experiments, both TAMRA-conjugated and non-conjugated, were purchased from Bio-Synthesis (JIP_1^{C11}, YTCPTEDIYLE; $SKIP^{WD}$, STNLEWDDSAI). The experiments were carried out on a BMG Labtech PolarStar Omega plate reader, setting up a temperature of 20 °C.

Briefly, 300 nM TAMRA-labeled peptides were incubated with the indicated protein in low salt buffer A, at increasing concentrations. The estimation of the K_D was obtained assuming a one-site specific-binding model.

For competition experiments, a mixture of TAMRA-JIP_1^{C11} and KLC_1^{TPR} at 300 nM and 2 µM, respectively, were incubated with increasing concentrations of unlabeled $SKIP^{WD}$, resuspended in buffer supplemented with 5 % DMSO. The FP signal was further analyzed and fitted by using the Prism package, in order to derive the IC50. K_I estimation was performed using the online tool (http://sw16.im.med.umich.edu/software/calc_ki/online tool) based on Nikolovska-Coleska et al., 2004 and all the data points were the mean of three replicates [143].

All the FP experiments were carried out with the help of Dr. Stefano Pernigo and Dr. Magda Chegkazi (Professor Roberto A. Steiner Laboratory). In particular, I participated in preparing the samples and collecting part of the data.

Immunoprecipitation

After sample preparation performed with the help of Dr. Magda Chegkazi, following what described in paragraphs 2.1.2 and 2.1.3, immunoprecipitation experiments were carried out by Dr. Yan Yip (Professor Mark Dodding Laboratory). Hela cells were grown under a 5 % CO_2 atmosphere, in DMEM supplemented with 10 % FBS, L-glutamine, and penicillin/streptomycin. $1x10^6$ cells were seeded on a 6-well dish and transfected with 0.4 µg of DNA (HA-KLC_1 isoforms A-F or HA-KLC_2). After 24 hours, transfected cells were lysed with lysis buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 0.1 &Nonidet P-40, 0.1 % Triton X-100) for 10 minutes. All the samples were blotted with 1 µg/mL of the three nanobodies respectively, in 5 % milk TBST or anti-HA antibody overnight. The anti-HIS-

HRP was used as secondary antibody for the nanobody detection and the un-transfected cell lysate as a control (Ctrl).

5.4 Results and Discussion

5.4.1 KLC_1^{TPR} and KLC_2^{TPR} purification

The isoforms 1 and 2 of the kinesin-1 tetramer, named KLC_1^{TPR} and KLC_2^{TPR} respectively, were expressed and purified according to the protocol described in paragraph 5.3.1. Both of the above mentioned proteins were firstly purified by affinity chromatography, exploiting their His-tag and further loaded onto a size exclusion chromatography column, which led to the isolation of their pure fractions.

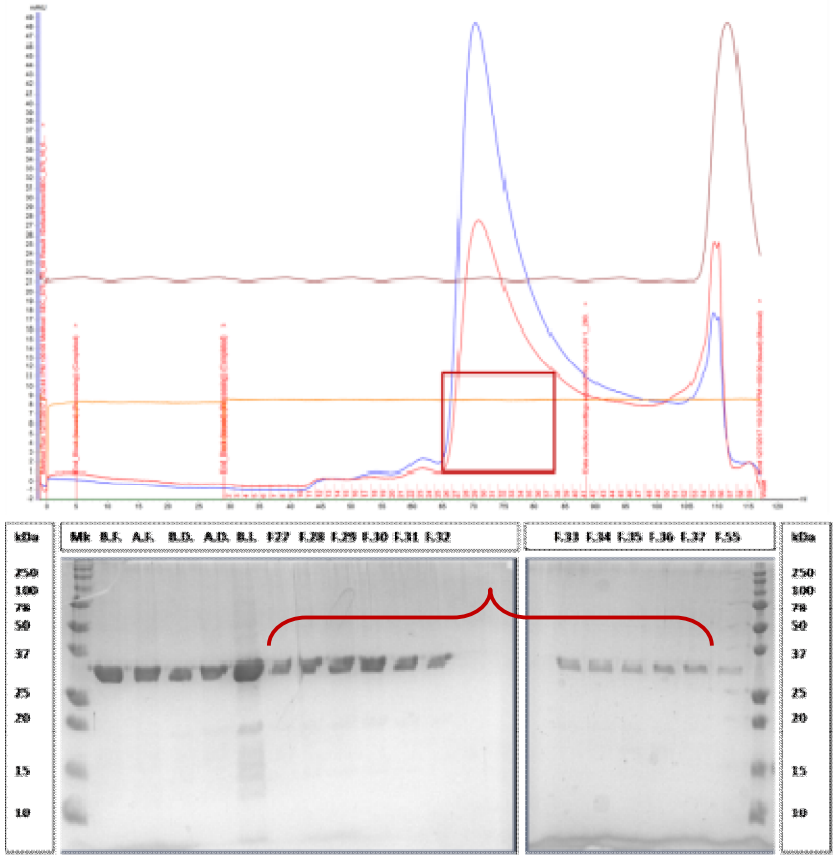


Figure 53: SEC chromatogram of KLC_2^{TPR} and SDS-PAGE gel of the fractions pooled (coloured red).

According to their sequences, KLC_1^{TPR} and KLC_2^{TPR} have molecular weights of about 36 kDa, a value confirmed also by the SEC elution profile and by SDS-PAGE electrophoresis (Figure 53).

According to figure 53, the chromatogram ofthe KLC_2^{TPR} elution profileshows the presence of a main peak and a big shoulder on its right-hand side. Although KLC_2^{TPR} was equally present in all the fractions under the two curves, as highlighted in the SDS-PAGE gel, only fractions from 27 to 37 were pooled, in order to guarantee homogeneity and purity of the protein.

A similar behaviour was observed after the analyses of the size exclusion profile of the KLC_1^{TPR} purification (figure 54). In spite of the presence of one main peak and a left-hand side shoulder, as previously described for KLC_2^{TPR}, all the fractions contained in this case the pure protein fraction, according to the SDS-PAGE electrophoresis gel. However, only fractions from 17 to 25 were pooled and concentrated for further experiments.

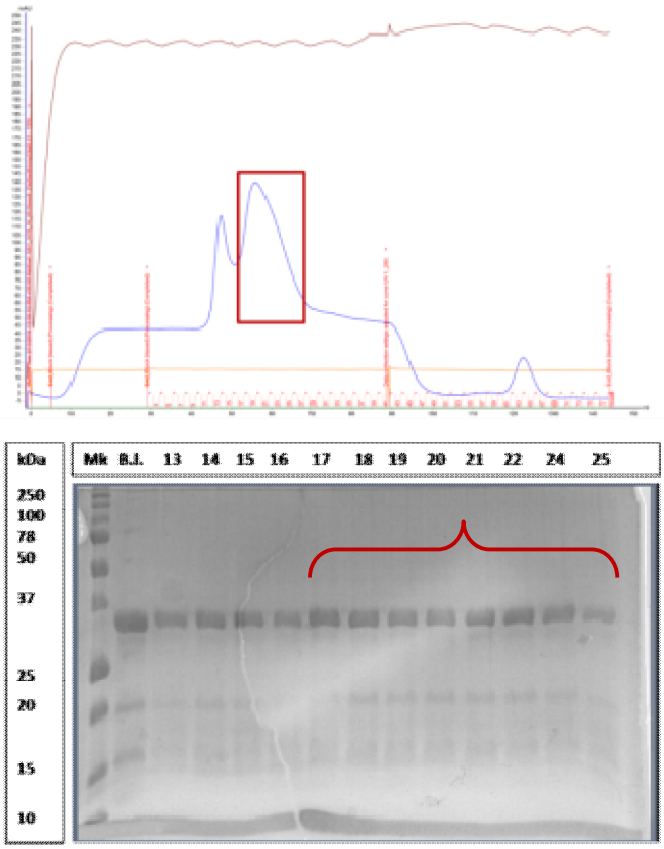


Figure 54: SEC chromatogram of KLC_1^{TPR} and SDS-PAGE gel of the fractions pooled (coloured red).

5.4.2 Nanobody purification

As described in paragraph 5.3.1, nanobodies 26, 29, and 76 were produced following the procedure described in Pardon et al. According to that protocol, nanobodies were initially purified by IMAC, taking advantage of the His_6-tag fused to their C-terminal. Then, a size exclusion chromatography step was required in order to get rid of the contaminants and to obtain the pure protein fraction. As shown in figure 55, the chromatogram indicates the presence of a well-shaped single peak, after a 100 mL void volume. According to the calibration of the 16/60 HiLoadSuperdex 75 column, alongside with the result of the SDS-PAGE electrophoresis, the peak corresponds to a molecular weight of about 13 kDa, a value confirmed by their sequences (table VII).

Nanobodies	Molecular weight (kDa)	Extinction coefficient (ε)
27	13407	2.04
29	13379	2.12
76	13214	2.04

Table VII: molecular weight and extinction coefficient of the three nanobodies used for the experiment.

All the fractions under the curve were pooled, checked by SDS-PAGE electrophoresis, and concentrated for further experiments. Since all the nanobodies were expressed and purified using the same protocol, a representative SEC elution profile, alongside with an SDS-PAGE gel are shown in figure 55.

Overall, the expression and purification of all the nanobodies, allowed to obtain pure and abundant products to be used for further ITC experiments.

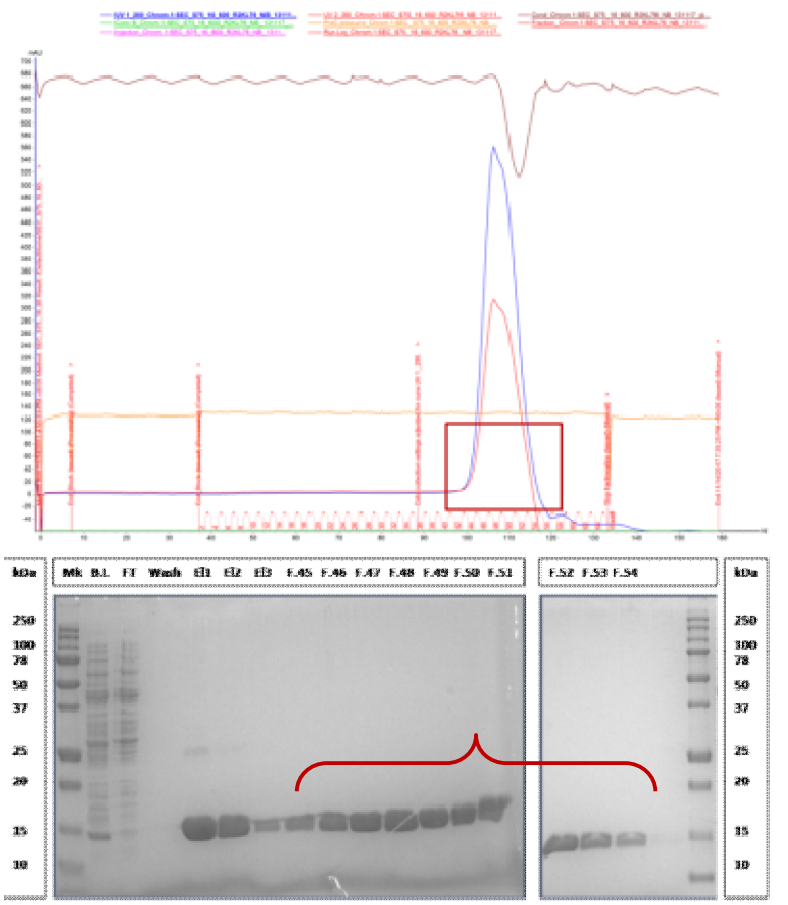


Figure 55: SEC chromatogram of R2KL76 and SDS-PAGE gel of the fractions pooled coloured red.

5.4.3 Isothermal titration calorimetry

During the last decades, nanobodies have been successfully used as crystallization chaperons, thanks to their ability to intimately interact with membrane proteins and complexes. In addition, they can penetrate into small cavities, such as membrane-embedded substrate binding sites, a

feature due to both by their small size, and the relatively long complementary determining region (CDR).

In the light of the above consideration, a pool of 29 antigen-specific nanobodies were produced after the immunization of a llama with KLC_1^{TPR} – $JIP^{C\text{-}term}$, following thementioned protocol (paragraph 2.1.1). Nanobodies number 76, 29, and 27 were selected for further experiments aimed at improving the quality of the$KLC1^{TPR}$ – $JIP^{C\text{-}term}$ crystals (unpublished data from Pernigo et al., 2018).

In order to test the ability of these three nanobodies to discriminate between KLC_1^{TPR} and KLC_2^{TPR}, isothermal titration calorimetry was performed by Dr. Magda Chegkazi and Conor Treacy. In addition, the use ofITC allowed to further characterize the interaction from the thermodynamic point of view, resulting in the determination of the K_D and the other parameters associated to the binding.

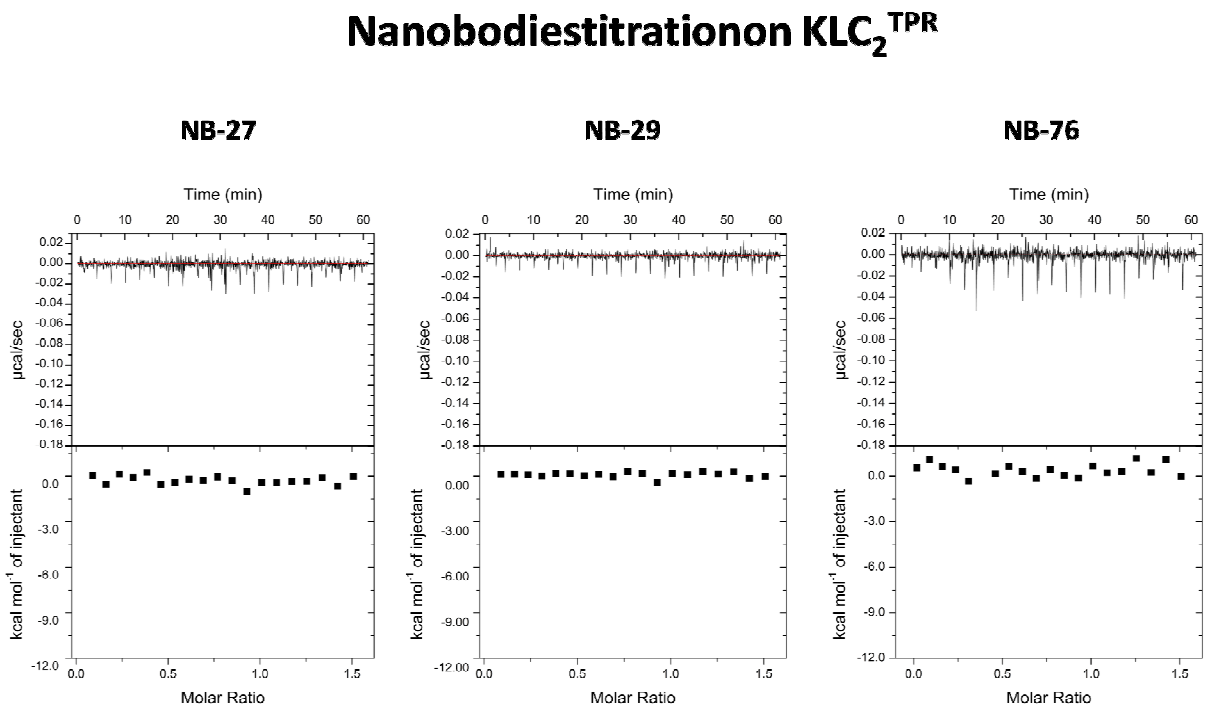


Figure 56: Thermograms of the nanobodies titration on KLC_2^{TPR}. Fitting of the data provided by Dr. Magda Chegkazi.

According to ITC analyses performed before my arrival, all the nanobodies were able to specifically bind KLC_1^{TPR} with nanomolar affinity. According to figure 56, fitting of the data revealed the absence of evident binding between KLC_2^{TPR} and the nanobodies, confirming, once again, their specificity only for KLC_1^{TPR}.

5.4.4 Fluorescence polarization

The three-dimensional structure of KLC_2^{TPR} in complex with the acidic peptide $SKIP^{WD}$, previously solved by Pernigo and colleagues has provided the structural basis for the recognition of Y-acidic and W-acidic motifs, revealing that they are partly recognized by overlapping binding sites.

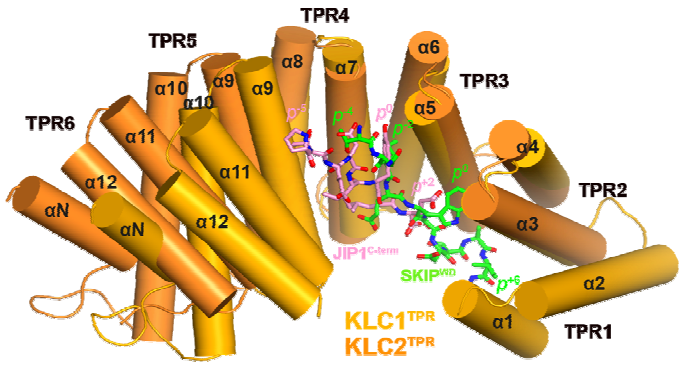


Figure 57: Superposition of the KLC1TPR-$JIP_1^{C\text{-}term}$ and KLC_2^{TPR}-$SKIP^{WD}$ complexes (KLC_1 and $KLC_2$are coloured in gold and orange respectively, while $SKIP^{WD}$ and $JIP_1^{C\text{-}term}$ are represented as pink and green sticks) highlighting the shift of Y-acidic and W-acidic peptides along the inner cavity of the helical KLC^{TPR}.

In particular, they are shifted with respect to each other along the inner channel generated by the TPR right-handed super-helical arrangement, according to figure 57. Whereas the Y-acidic motif binds towards the C-terminal end of the channel, the W-acidic is positioned closer to the N-

terminus. In addition, the binding with the latter requires the presence of the key residue tyrosine and it is further stabilized by the interface between TPR3 and TPR4, whereas a tryptophan sandwiched between TPR2 and TPR3 plays an important role in the Y-acidic motif binding.

In the light of the above consideration, fluorescence polarization was used in order to monitor the K_D of the binding of KLC_1^{TPR} to the Y-acidic $JIP1^{C\text{-}term}$ peptide and to the W-acidic $SKIP^{WD}$ peptide. In addition, the competition between the two acidic motifs was characterized by performing a competitive FP assay (work carried out together with Dr. Stefano Pernigo, Dr. Magda Chegkazi).

A fluorescence polarization assay (figure 58) has shown the ability of unlabelled $SKIP^{WD}$ to specifically compete with TAMRA-$JIP_1^{C\text{-}term}$, in a concentration dependent manner. This result agrees with previous findings, which showed that despite the differences between the two acidic peptides, their binding sites partly overlap, giving rise to the competitive behaviour just mentioned.

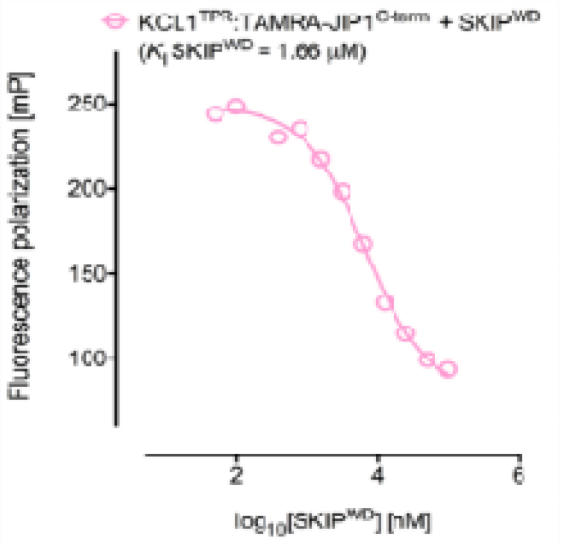


Figure 58: Titration of increasing amounts of non-labelled $SKIP^{WD}$ on KLC_1^{TPR}-$JIP^{C\text{-}term}$ complex.

However, since a single KLC_1^{TPR} binding site is not expected to accommodate both types of peptides at the same time, the partial separation of the two acidic residues could have significant functional consequences.

In order to look deeper, the affinity of TAMRA-$SKIP^{WD}$ for KLC_1^{extTPR} and KLC_1^{TPR} were analysed, with the aim of understanding the effect of the intramolecular LFP-acidic region on the binding. As previously shown for KLC_2^{TPR} [144], the presence of the LFP motif strongly inhibits the binding, revealing an increase in the K_D value of one order of magnitude: from 0.64 ± 0.02 µM to 8.20 ± 0.37 µM.

Interestingly, TAMRA-$JIP^{C\text{-}term}$ was shown to equally bind to KLC_1^{extTPR} (K_D= 2.21 ± 0.06 µM) and to KLC_1^{TPR} (K_D= 0.95 ± 0.09 µM), by Yip and colleagues.

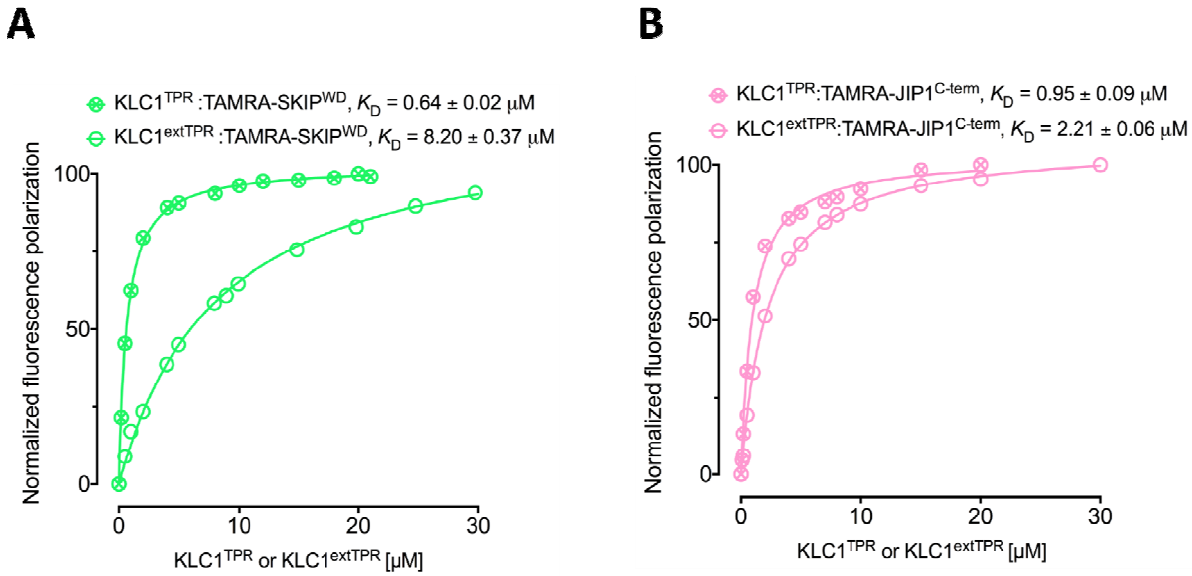


Figure 59: Titration of increasing amounts of TAMRA-$SKIP^{WD}$ (A) and $JIP^{1C\text{-}term}$ (B) on KLC_1^{extTPR}, which includes the LFP motif.

Overall, considering the information acquired in the previous studies [122, 130], the partial separation of the Y-acidic and the W-acidic binding sites on KLC^{TPR} could have significant functional consequences. In particular, while the LFP portion of KLC^{TPR} does not affect the binding of $JIP^{C\text{-}term}$, that area seems not to be compatible with $SKIP^{WD}$ binding, according to the FP data (figure 59).

5.4.5 Immunoprecipitation

In order to validate and to rationalize the high preference of the nanobodies for $KLC_1{}^{TPR}$ over its isoform $KLC_2{}^{TPR}$, previously observed in FP measurements, an immunoprecipitation assay in HeLa cells was carried out by Dr. Yan Yip. Following cell co-transfection with constructs expressing hemagglutinins (HA)-$KLC_1{}^{TPR}$ and (HA)-$KLC_2{}^{TPR}$, a Western Blot assay was performed as described in paragraph 5.3.2.

As can be seen from figure 60, the result of the immunoprecipitation analyses has revealed a high specificity of the three nanobodies for $KLC_1{}^{TPR}$ and its isoforms (1A-F), when compared to $KLC_2{}^{TPR}$,confirming what previously found by ITC measurements. However, it is worth mentioning that since all the signals come from the same exposure, nanobody 27, 29, and 76 have different sensitivities towards $KLC_1{}^{TPR}$, as demonstrated also by the different K_D obtained from ITC data. In particular, the strongest signal was obtained after cell blotting with nanobody 29, while nanobody 27 has shown to be the least specific for KLC1TPR, revealing the weakest intensity band among the three.

Overall, the immunoprecipitation data are consistent with what was previously revealed by ITC experiments, highlighting the specificity of the nanobodies only for isoform 1 of KLC^{TPR}.

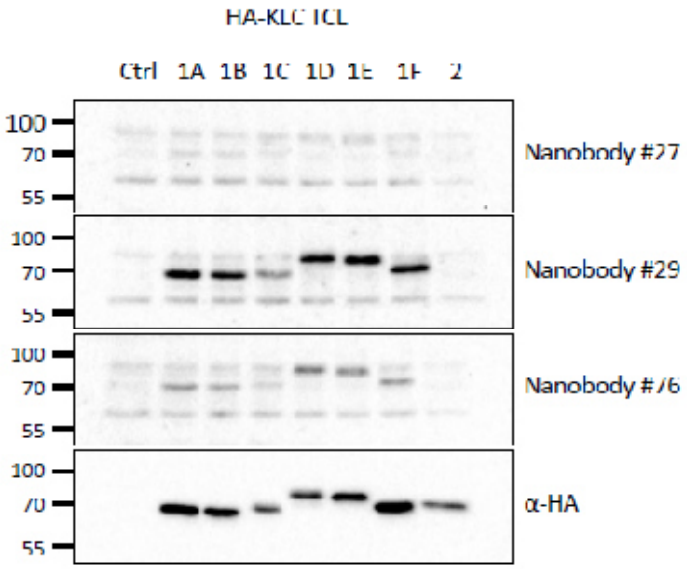


Figure 60: Western Blot analysis of immunoprecipitation showing the specificity of the nanobodies towards KLC_1TPR rather than KLC_2TPR. Image provided by Dr. Yan Yip.

5.5 Conclusions

The kinesin-1 motor protein, responsible for the (+)-end-directed transport of different cargoes along the microtubules, is composed of two motor and two regulatory subunits, named KHCs and KLCs respectively. However, the presence of several adaptors or scaffolding proteins is a fundamental requirement to accomplish the kinesin-driven intracellular transport. Particular attention has to be paid to the TPR domain of KLC, an important regulatory region often engaged by the adaptors, being found in a broad range of proteins and representing the platform for protein-protein interactions.

Since the crystallographic analysis of the TPR region has highlighted its ability to recognize and bind both Y-acidic and W-acidic motifs, one of the main concerns of the present study was to clarify the nature of the binding.

Clearly, upon W-acidic motifs binding, the intramolecular interaction between KLC^{TPR} and the so called “LFP-motif” undergoes important modifications, leading to global changes in its conformation. Consequently, kinesin-1 activity is promoted and the exposition of hidden binding sites on KHC-tail takes place.

To test these considerations, the LFP motif was inserted into both the isoforms of KLC^{TPR}, resulting in a dramatic reduction of their affinity for W-acidic peptides. Interestingly, a different behavior was observed for the Y-acidic $JIP_1^{C\text{-}term}$ peptides, whose affinity for KLC_1^{TPR} was weakly affected.

Overall, these findings suggest a possible coexistence of the LFP motif and the Y-acidic motif, an evidence that combined with previous results allows to propose two different hypotheses for thekinesin-1 activation pathway: either the LFP motif displacement might be responsible for triggering the reaction, or,in alternative,the binding of the JIP_3 adaptor.

However, in order to confirm or reject the above hypothesis, further structural insight into KLC1-JIP1-JIP3 complexes will be required to address the question.

6. References

[1] Nascimento K. S., Cunha A. I., Nascimento K. S., Cavada B. S., Azevedo A. M. and Aires-Barros M. R. An overview of lectins purification strategies. *J. Mol. Recognit*. 25(11):527-41. 2012.

[2] Sharon N. and Lis H. History of lectins: from hemagglutinins to biological recognition molecules. *Glycobiology*. 14(11):53R-62R. 2004.

[3] Guillot J. and Konska G. Lectins in higher fungi. *Biochem. Syst. and Ecol*. 25(3): 203-230. 1997.

[4] Wang H., Ng T. B., Ooi V. Lectins from mushrooms. *Mycol. Res.* 102(8): 897-906. 1998.

[5] Khan F., Khan M.I. Fungal lectins: current molecular and biochemical perspectives. *Int. J. Biol. Chem.* 5: 1-20. 2011.

[6] Singh R.S., Bhari R., Kaur H.P. Mushroom lectins: current status and future perspectives. *Crit Rev. Biotechnol.* 30: 99-126.2010.

[7] Varrot A., Basheer S. M., Imberty A. Fungal lectins: structure, function and potential applications. *Curr. Op. in Struct. Biol.* 23: 678-685. 2013.

[8] Sá R.A., Santos N.D., Silva C.S., Napoleão T.H., Gomes F.S., Cavada B.S., Coelho L.C., Navarro D.M., Bieber L.W., Paiva P.M. Larvicidal activity of lectins from *Myracrodruon urundeuva* on *Aedes aegypti*. *Comp Biochem. Physiol. C Toxicol. Pharmacol.* 149: 300–306. 2009.

[9] Jiang Q.L., Zhang S., Tian M., Zhang S.Y., Xie T., Chen D.Y., Chen Y.J., He J., Liu J., Ouyang L., Jiang X. Plant lectins, from ancient sugar-binding proteins to emerging anti-cancer drugs in apoptosis and autophagy. *Cell Prolif.* 48(1): 17-28. 2015.

[10] Teicher B., Lazo J.S., Sartorelli A.C. Classification of anticancer agent by their selective toxicities toward oxygenated and hypoxic tumor cells. *Cancer Res*. 41: 73-81. 1981.

[11] Häuselmann I., Borsig L. Altered tumor-cell glycosylation promotes metastasis. *Front Oncol*. 4(28): 1-12. 2014.

[12] Capaldi S., Faggion B., Carrizo M.E., Destefanis L., Gonzalez M.C., Perduca M., Bovi M., Galliano M., Monaco H.L. Three-dimensional structure and ligand-binding site of carp fishelectin (FEL). *Acta Crystallogr. D Biol. Crystallogr.* 71(Pt 5): 1123-35. 2015.

[13] Bovi M., Cenci L., Perduca M., Capaldi S., Carrizo M.E., Civiero L., Chiarelli L.R., Galliano M., Monaco H.L. BEL β-trefoil: a novel lectin with antineoplastic properties in king bolete (Boletus edulis) mushrooms. *Glycobiology*. 23(5): 578-92. 2013.

[14] Bovi M., Carrizo M.E., Capaldi S., Perduca M., Chiarelli L.R., Galliano M., Monaco H.L. Structure of a lectin with antitumoral properties in king bolete (Boletus edulis) mushrooms. *Glycobiology*. 21(8):1000-9. 2011

[15] Carrizo M.E., Capaldi S., Perduca M., Irazoqui F.J., Nores G.A., Monaco H.L. 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(11): 10614-23. 2005.

[16] Oded S., Ziv S., Ilan L. Carbohydrate Binding Modules: Biochemical Properties and Novel Applications. *Microbiol. Mol. Biol. Rev.* 70(2): 283–295. 2006

[17] Galli C., Bernasconi R., Soldà T., Calanca V., Molinari M. Malectin participates in a backup glycoprotein quality control pathway in the mammalian ER. *PLoS One*. 6(1): e16304. 2011

[18] Woods R.J. and Tessier M.B. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Curr. Opin. Struct. Biol. 2*0(5): 575-83. 2010.

[19] Qasba P.K. and Ramakrishnan B. X-ray crystal structures of glycosyltransferas. In: Kamerling J.P., Boons G.J., Lee Y.C., Suzuki A., Taniguchi N., Voragen A.G.J., editors. Comprehensive Glycosciences: From Chemistry to Systems Biology. Burlington, M.A.: Elsevier Ltd, 251–281. 2007

[20] Bowman K.G., Bertozzi C.R. Carbohydrate sulfotransferases: mediators of extracellular communication. *Chem. Biol*. 6(1): R9-R22. 1999.

[21] Shoseyov O., Shani Z., Levy I. Carbohydrate binding modules: biochemical properties and novel applications. *Microbiol. Mol. Biol.R*. 70: 283. 2006.

[22] Boraston A.B., Bolam D.N., Gilbert H.J., Davies G.J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.*15(382): 769-81. 2004.

[23] Yoshida M., Igarashi K., Wada M., Kaneko S., Suzuki N., Matsumura H., Nakamura N., Ohno H., Samejima M. Characterization of carbohydrate-binding cytochrome b562 from the white-rot fungus *Phanerochaete chrysosporium*. *Appl. Environ. Microbiol.*71(8): 4548-55. 2005.

[24] Cosgrove D.J. Loosening of plant cell walls by expansins. *Nature.* 407: 321–326. 2000.

[25] Jamal-Talabani S., Boraston A.B., Turkenburg J.P., Tarbouriech N., Ducros V.M., Davies G.J. Ab initio structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules. *Structure.*12(7): 1177-87. 2004.

[26] Rutenber E., Katzin B.J., Ernst S., Collins E.J., Mlsna D., Ready M.P., Robertus J.D. Crystallographic refinement of ricin to 2.5 A. *Proteins.*10(3): 240-50. 1991.

[27] Creagh A.L., Ong E., Jervis E. Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven. *Proc. Natl. Acad. Sci. U.S.A.* 93: 12229 – 12234. 1996.

[28] Ukai, S., K. Hirose, T. Kiho, C. Hara and T. kikura. Antitumor activity on sarcoma 180 of the polysaccharides from *Tremella fuciformis* Berk. *Chem. Pharm. Bull.*20: 2293-2294. 1972.

[29] Ukai S., Kiho T., Hara C., Kuruma L. and Tanaka Y. Polysaccharides in fungi. XIV. Anti-inflammatory effect of the polysaccharides from the fruit bodies of several fungi. *J.* Phnrmacobiodyn.6: 983-990. 1983.

[30] Wu Y., Shan L., Yang S. and Ma A. Identification and antioxidant activify of melanin isolated from *Hypoxylon archeri*, a companion fungus of *Tremella fuciformis*. *J. Basic Microbiol*. 48: 211-221. 2008.

[31] Yin Y., Chen D., Qiao M., Lu Z., Hu H. Preparation and evaluation of lectin-conjugated PLGA nanoparticles for oral delivery of thymopentin. *J Control Release.* 116(3): 337-45. 2006

[32] Malatesta M, Galimberti V, Cisterna B, Costanzo M, Biggiogera M, Zancanaro C. Chitosan nanoparticles are efficient carriers for delivering biodegradable drugs to neuronal cells. *Histochem. Cell. Biol.* 141: 551–558. 2014

[33] Lodish H., Berk A., Zipursky S.L. Molecular Cell Biology. 4th edition. New York: W. H. Freeman; 2000.

[34] McPherson A. Introduction to Macromolecular Crystallography. J. *Chem. Educ.*80(9): 1007. 2003.

[35] McPherson A., Gavira J.A. Introduction to protein crystallization. *Acta Crystallogr. F Struct. Biol. Commun.*70(Pt 1): 2-20. 2014.

[36] Blundell T. and Johnson L. Protein Crystallography. *Academic Press*. 1976.

[37] Bernhard R., Biomolecular Crystallography: Principles, Practice, and Application to Structural Biology. Garland Science. 2010.

[38] Hendrickson W.A., Horton J.R., LeMaster D.M. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. *EMBO J.*9(5): 1665-72. 1990.

[39] Leahy, D. J., Erickson, H., Aukhil, I., Joshi, P. & Hendrickson, W. Crystallization of a fragment of human fibronectin: introduction of methionine by site-directed mutagenesis to allow phasing via selenomethionine. *Proteins.* 19: 48–54. 1994

[40] Methods in Molecular Biology, volume 364: Macromolecular Crystallography Protocols: Volume 2: Structure Determination. Edited by S. Doublié, Humana Press, Inc – Dauter M. And Dauter Z. 2007.

[41] Boggon TJ1, Shapiro L. Screening for phasing atoms in protein crystallography. Structure. 15: 8(7): R143-9. 2010.

[42] Powell H. Integration and scaling. Lecture for the Medical Research Council.

[43] Han C.K., Chiang H.C., Lin C.Y., Tang C.H., Lee H., Huang D.D., Zeng Y.R., Chuang T.N., Huang Y.L. Comparison of Immunomodulatory and Anticancer Activities in Different Strains of *Tremella* fuciformis Berk. Am. J. Chi.n Med. 43(8): 1637-55. 2015.

[44] Lin Z.B., Qin Z.L., Xia H.L., Guan H.C. and Jiao K. Effects of *Tremella* polysaccharides on immunological stahrs and content of cytochrome P450 in mouse liver homogenates. Zhongguo Yao Lì Xue Bao 6: 2O1-20. 1985.

[45] Pike A.C. W., Garman E.F., Krojer T., von Delft F. and Carpentera E.P. An overview of heavy-atom derivatization of protein crystals. *Acta Crystallogr. D Struct. Biol.* 72(Pt 3): 303–318. *2*016.

[46] Battye T.G., Kontogiannis L., Johnson O., Powell H.R., Leslie A.G. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr*. 67(Pt 4): 271-81. 2011.

[47] Sano K., Ogawa H. Hemagglutination (inhibition) assay. *Methods Mol Biol*. 1200:47-52. 2014.

[48] Ghisaidoobe A.B., Chung S.J. Intrinsic tryptophan fluorescence in the detection and analysis of proteins: a focus on Förster resonance energy transfer techniques. *Int. J. Mol. Sci.* 5;15(12): 22518-38. 2014.

[49] Bhattacharjee S. DLS and zeta potential - What they are and what they are not?*J. Control. Release.* 10; 235: 337-351. 2016.

[50] Durowoju I.B., Bhandal K.S., Hu J., Carpick B. and Kirkitadze M. Differential Scanning Calorimetry — A Method for Assessing the Thermal Stability and Conformation of Protein Antigen. *J. Vis. Exp.* (121): 55262. 2017.

[51] Duff M.R. Jr., Grubbs J. and Howell E.E. Isothermal Titration Calorimetry for Measuring Macromolecule-Ligand Affinity. J. Vis. Exp. (55): 2796. 2011.

[52] Kelly S.M., Jess T.J., Price N.C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta.* 10; 1751(2): 119-39. 2005.

[53] Berridge M.V., Herst P.M., Tan A.S. Tetrazolium dyes as tools in cell biology: new insights into their cellular reduction. *Biotechnol. Annu. Rev.*11: 127-52. 2005.

[54] Jain P., Worthylake R.A., Alahari S.K. Quantitative analysis of random migration of cells using time-lapse video microscopy. *J. Vis. Exp.* 13; (63): e3585. 2012.

[55] Alexander S., Friedl P. Cancer invasion and resistance: interconnected processes of disease progression and therapy failure. *Trends Mol. Med*. 18(1): 13-26. 2012.

[56] Wagner E.M. Monitoring gene expression: quantitative real-time rt-PCR. Methods Mol Biol.1027: 19-45. 2013

[57] Kozera B., Rapacz M.J. Reference genes in real-time PCR. Appl. Genet. 54(4): 391-406. 2013.

[58] Zenaro E., Donini M., Dusi S. Induction of Th1/Th17 immune response by Mycobacterium tuberculosis: role of dectin-1, Mannose Receptor, and DC-SIGN. *Leukoc. Biol.*86(6):1393-401. 2009.

[59] Cho K., Wang X., Nie S., Chen Z.G., Shin D.M. Therapeutic nanoparticles for drug delivery in cancer. *Clin Cancer Res*. 1;14(5):1310-6. 2008.

[60] Soppimath K.S., Aminabhavi T.M., Kulkarni A.R., Rudzinski W.E. Biodegradable polymeric nanoparticles as drug delivery devices. *J Control Release.* 29;70(1-2):1-20. 2001.

[61] Ali M.E., Lamprecht A. Polyethylene glycol as an alternative polymer solvent for nanoparticle preparation. *Int. J. Pharm.* 1;456(1):135-42. 2013.

[62] Kelley L.A., Mezulis S., Yates C.M., Wass M.N., Sternberg M.J. The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10(6): 845-58. 2015.

[63] Laskowski R.A. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*. 1; 29(1): 221-2. 2001.

[64] Redfern, O. C., Harrison, A., Dallman, T., Pearl, F. M., Orengo, C. A. CATHEDRAL: a fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures, *PLoS Comp. Biol.* 3(11): e232., 1553-734. 2007.

[65] Halskau O.Jr., Perez-Jimenez R., Ibarra-Molero B., Underhaug J., Muñoz V., Martinez A., Sanchez-Ruiz J.M. Large-scale modulation of thermodynamic protein folding barriers linked to electrostatics. *Proc. Natl. Acad. Sci. U S A.* 24; 105(25): 8625-30. 2008.

[66] Riss T.L., Moravec R.A., Niles A.L., Duellman S., Benink H.A., Worzella T.J. and Minor L.. Cell Viability Assays. Sittampalam G.S., Coussens N.P., Brimacombe K. et al., editors.Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences. 2004.

[67]Justus C.R., Leffler N., Ruiz-Echevarria M., Yang L.V. In vitro cell migration and invasion assays. *J. Vis. Exp.* 1; (88). 2014.

[68]Faheina-Martins G.V., da Silveira A.L., Ramos M.V., Marques-Santos L.F., Araujo D.A. Influence of fetal bovine serum on cytotoxic and genotoxic effects of lectins in MCF-7 cells. *J. Biochem. Mol. Toxicol*. 25(5): 290-6. 2011.

[69]Costantini S., Di Bernardo G., Cammarota M., Castello G., Colonna G. Gene expression signature of human HepG2 cell line. *Gene*. 15; 518(2): 335-45. 2013.

[70]Xu J., Haigis K.M., Firestone A.J., McNerney M.E., Li Q, Davis E., Chen S.C., Nakitandwe J., Downing J., Jacks T., Le Beau M.M., Shannon K. Dominant role of oncogene dosage and absence of tumor suppressor activity in Nras-driven hematopoietic transformation. *Cancer Discov.* 3(9):993-1001. 2013.

[71]Ditamo Y., Rupil L.L., Sendra V.G., Nores G.A., Roth G.A., Irazoqui F.J. In vivo immunomodulatory effect of the lectin from edible mushroom Agaricus bisporus.*Food Funct.* 7(1):262-9. 2016.

[72] Cao ., Zhou M., Wang C., Hou L., Li Y. and Chen L. *Musca domestica* Pupae Lectin Improves the Immunomodulatory Activity of Macrophages by Activating Nuclear Factor-κB. *J. Med.Food.* 15(2): 145–151. 2012.

[73] Ivory C.P., Keller K., Chadee K. CpG-oligodeoxynucleotide is a potent adjuvant with an Entamoeba histolytica Gal-inhibitable lectin vaccine against amoebic liver abscess in gerbils. Infect. Immun. 74(1):528-36. 2006.

[74] Pizzorno J.E. and Murray M.T. Textbook of Natural Medicine, 4th Edition. 2012.

[75]Appert-RollandaC., Ebbinghausabc M. and Santen L. Intracellular transport driven by cytoskeletal motors: General mechanisms and defects. *Physics Reports*. 593(2): 1-59. 2015.

[76] Barghorn S., Mandelkow E. Toward a unified scheme for the aggregation of tau into Alzheimer paired helical filaments. *Biochemistry.*17: 41(50):14885-96. 2002.

[77] Moreno H., Morfini G., Buitrago L., Ujlaki G., Choi S., Yu E., Moreira J.E., Avila J., Brady S.T., Pant H., Sugimori M., Llinás R.R. Tau pathology-mediated presynaptic dysfunction. *Neuroscience.*14: 325:30-8. 2016.

[78] Gan J.K., Morihara T. and Silverman M.A. Atlas stumbled: Kinesin light chain-1 variant E triggers a vicious cycle of axonal transport disruption and amyloid-β generation in Alzheimer's disease. *Bioessays*. 37(2): 131-41. 2015.

[79] Dodding M.P. and Way M. Coupling viruses to dynein and kinesin-1. *EMBO J.* 31: 30(17): 3527-39. 2011.

[80] Dodding M.P., Mitter R., Humphries A.C., Way M. A kinesin-1 binding motif in vaccinia virus that is widespread throughout the human genome. *EMBO J.*16: 30(22):4523-38. 2011.

[81] Boucrot E., Henry T., Borg J.P., Gorvel J.P., Méresse S. The intracellular fate of Salmonella depends on the recruitment of kinesin. *Science.* 20: 308(5725): 1174-8. 2005.

[82] Allen R.D., Metuzals J., Tasaki I., Brady S.T., Gilbert S.P. Fast axonal transport in squid giant axon. *Science.* 10: 218(4577): 1127-9.1982.

[83] Kozminski K.G., Johnson K.A., Forscher P., Rosenbaum J.L. A motility in the eukaryotic flagellum unrelated to flagellar beating.*Proc. Natl. Acad. Sci. U S A.* 15: 90(12): 5519-23.1993.

[84] Balch W.E., Dunphy W.G., Braell W.A., Rothman J.E. Reconstitution of the transport of protein between successive compartments of the Golgi measured by the coupled incorporation of N-acetylglucosamine. Cell. 39(2 Pt 1): 405-16. 1984.

[85] Faini M., Prinz S., Beck R., Schorb M., Riches J.D., Bacia K., Brügger B., Wieland F.T., Briggs J.A. The structures of COPI-coated vesicles reveal alternate coatomer conformations and interactions. *Science.* 15: 336(6087): 1451-4. 2012.

[86]Schnapp B., Vail R., Sheetz M., Reese T.S. Directed movements of organelles along filaments dissociated from giant squid axon.

[87] Vale R.D., Reese T.S. and Sheetz M.P. Identification of a Novel Force-Generating Protein, Kinesin, Involved in Microtubule-Based Motility. *Cell.* 42(1): 39–50. 1985.

[88] Alberts B., Johnson A., Lewis J., Raff M., Roberts K. and Walter P. Molecular biology of the cell, 5th edition. John Boyle. 2008.

[89] Mitchison T., Kirschner M. Dynamic instability of microtubule growth. *Nature.*15-21: 312(5991): 237-42. 1984.

[90] Maccioni R.B., Cambiazo V. Role of microtubule-associated proteins in the control of microtubule assembly. *Physiol Rev.*75(4): 835-64. 1995.

[91] Hirokawa N., Noda Y., Tanaka Y., Niwa S. Kinesin superfamily motor proteins and intracellular transport. *Nat. Rev. Mol. Cell. Biol.* 10(10): 682-96. 2009.

[92] Kikkawa M., Okada Y., Hirokawa N. 15 A resolution model of the monomeric kinesin motor, KIF1A. *Cell.* 21: 100(2): 241-52. 2000.

[93] Gauger A.K., Goldstein L.S. The Drosophila kinesin light chain. Primary structure and interaction with kinesin heavy chain. *J. Biol. Chem.* 25: 268(18): 13657-66. 1993.

[94] Kull F.J., Sablin E.P., Lau R., Fletterick R.J., Vale R.D. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature.*11: 380(6574): 550-5. 1996.

[95] Kull F.J. Motor proteins of the kinesin superfamily: structure and mechanism. *Essays Biochem.*35: 61-73. 2000.

[96] Park J., Aragona B.J., Kile B.M., Carelli R.M., Wightman R.M. In vivo voltammetric monitoring of catecholamine release in subterritories of the nucleus accumbens shell. Neuroscience. 11: 169(1): 132-42. 2010.

[97] Hwang W., Lang M.J., Karplus M. Force generation in kinesin hinges on cover-neck bundle formation. *Structure.* 16(1): 62-71. 2008.

[98] Endow S.A., Kull F.J. and Liu H. Kinesins at a glance*. J. Cell Science* 123: 4000. 2010.

[99]Endow S.A. Determinants of molecular motor directionality*. Nat Cell Biol.* 1(6): E163-7. 1999.

[100]Kozielski F., Sack S., Marx A., Thormählen M., Schönbrunn E., Biou V., Thompson A., Mandelkow E.M., Mandelkow E. The crystal structure of dimeric kinesin and implications for microtubule-dependent motility. *Cell.* 26: 91(7): 985-94. 1997.

[101] Svoboda K., Schmidt C.F., Schnapp J.B. and Block S.M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature.3*65: 721–727. 1993.

[102] Svoboda K., Block S.M. Biological applications of optical forces. *Annu. Rev. Biophys. Biomol. Struct.* 23:247-85. 1994.

[103] Hwang W., Lang M.J. and Karplus M. Kinesin motility is driven by subdomain dynamics. *eLife.* 6: e28948. 2017.

[104] Kim A.J., Endow S.A. A kinesin family tree. *J. Cell Sci.*113 Pt 21:3681-2. 2000.

[105] Verhey K.J., Meyer D., Deehan R., Blenis J., Schnapp B.J., Rapoport T.A., Margolisc B. Cargo of Kinesin Identified as Jip Scaffolding Proteins and Associated Signaling Molecules*J. Cell Biol.* 5: 152(5): 959–970. 2001.

[106] Rosenbaum J.L., Witman G.B. Intraflagellar transport. *Nat. Rev. Mol. Cell Biol.* 3(11): 813-25. 2002.

[107] Gardner M.K., Devoss J.J., Friedman R.S., Wong D.J., Tan Y.X., Zhou X., Johannes K.P., Su M.A., Chang H.Y., Krummel M.F., Anderson M.S. The microtubule-based motor Kar3 and plus end-binding protein Bim1 provide structural support for the anaphase spindle. *J Cell Biol* 180(1): 91-100. 2008.

[108] Kapitein L.C., Peterman E.J., Kwok B.H., Kim J.H., Kapoor T.M., Schmidt C.F. The bipolar mitotic kinesin Eg5 moves on both microtubules that it crosslinks. *Nature.* 5: 435(7038): 114-8. 2005.

[109] Gindhart J.G. Jr., Desai C.J., Beushausen S., Zinn K., Goldstein L.S. Kinesin light chains are essential for axonal transport in Drosophila.*J. Cell Biol.* 20: 141(2): 443-54. 1998.

[110] Rahman A., Kamal A., Roberts E.A., Goldstein L.S. Defective kinesin heavy chain behavior in mouse kinesin light chain mutants. *J. Cell Biol.* 20: 146(6): 1277-88. 1999.

[111] Widmann C., Gibson S., Jarpe M.B., Johnson G.L. Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human. *Physiol. Rev.* 79(1): 143-80. 1999.

[112] Davis R.J. Signal transduction by the JNK group of MAP kinases. *Cell.* 13: 103(2): 239-52. 2000.

[113] Mielke K., Herdegen T. JNK and p38 stresskinases--degenerative effectors of signal-transduction-cascades in the nervous system. *Prog. Neurobiol*. 61(1): 45-60.2000.

[114] Horiuchi D., Barkus R.V., Pilling A.D., Gassman A. and Saxton W.M. APLIP1, a Kinesin Binding JIP-1/JNK Scaffold Protein, Influences the Axonal Transport of Both Vesicles and Mitochondria in *Drosophila*. *Curr. Biol.*15(23): 2137–2141. 2005.

[115] Yasuda J., Whitmarsh A.J., Cavanagh J., Sharma M., Davis R.J. The JIP group of mitogen-activated protein kinase scaffold proteins. *Mol. Cell Biol*. 19(10):7245-54. 1999.

[116] Whitmarsh A.J. The JIP family of MAPK scaffold proteins. *Biochem. Soc. Trans*. 34(Pt 5): 828-32. 2006.

[117] Kelkar N., Gupta S., Dickens M., Davis R.J. Interaction of a mitogen-activated protein kinase signaling module with the neuronal protein JIP3. *Mol. Cell Biol.*20(3): 1030-43.2000.

[118] Gyoeva F.K., Bybikova E.M., Minin A.A. An isoform of kinesin light chain specific for the Golgi complex. *J Cell Sci*. 113(Pt 11): 2047-54. 2000.

[119] Scholey J.M., Heuser J., Yang J.T., Goldstein L.S. Identification of globular mechanochemical heads of kinesin. *Nature.* 23: 338(6213): 355-7. 1989.

[120] Hirokawa N., Pfister K.K., Yorifuji H., Wagner M.C., Brady S.T., Bloom G.S. Submolecular domains of bovine brain kinesin identified by electron microscopy and monoclonal antibody decoration. *Cell.* 10: 56(5): 867-78. 1989.

[121] Skoufias D.A., Cole D.G., Wedaman K.P., Scholey J.M. The carboxyl-terminal domain of kinesin heavy chain is important for membrane binding. *J Biol Chem.*14: 269(2):1477-85. 1994.

[122] Pernigo S., Lamprecht A., Steiner R.A., Dodding M.P. Structural basis for kinesin-1:cargo recognition. *Science.* 19: 340(6130): 356-9. 2013.

[123] Steinberg G., Schliwa M. Characterization of the biophysical and motility properties of kinesin from the fungus *Neurospora crassa. J. Biol. Chem.* 29: 271(13): 7516-21. 1996.

[124] Kanai Y., Okada Y., Tanaka Y., Harada A., Terada S., Hirokawa N. KIF5C, a novel neuronal kinesin enriched in motor neurons. *J. Neurosci*. 1: 20(17): 6374-84. 2000.

[125] Jaulin F., Xue X., Rodriguez-Boulan E., Kreitzer G. Polarization-dependent selective transport to the apical membrane by KIF5B in MDCK cells. *Dev Cell.* 13(4): 511-22. 2007.

[126] Hisanaga S., Murofushi H., Okuhara K., Sato R., Masuda Y., Sakai H., Hirokawa N. The molecular structure of adrenal medulla kinesin. *Cell Motil. Cytoskeleton.* 12(4): 264-72. 1989.

[127] Hackney D.D., Levitt J.D., Suhan J. Kinesin undergoes a 9 S to 6 S conformational transition.J *Biol Chem.* 25: 267(12): 8696-701. 1992.

[128] Seeger M.A., Rice S.E. Microtubule-associated protein-like binding of the kinesin-1 tail to microtubules. *J. Biol. Chem*. 12: 285(11): 8155-62. 2010.

[129] Kaan H.Y., Hackney D.D., Kozielski F. The structure of the kinesin-1 motor-tail complex reveals the mechanism of autoinhibition. *Science.*12: 333(6044):883-5. 2011.

[130] Yip Y.Y., Pernigo S., Sanger A., Xu M., Parsons M., Steiner R.A. and Dodding M.P. The light chains of kinesin-1 are autoinhibited. *PNAS*. 113(9):2418-2423. 2016.

[131] Fu M.M., Holzbaur E.L. Integrated regulation of motor-driven organelle transport by scaffolding proteins. *Trends Cell Biol.*24(10):564-74. 2014.

[132] Aoyama T., Hata S., Nakao T., Tanigawa Y., Oka C., Kawaichi M. Cayman ataxia protein caytaxin is transported by kinesin along neurites through binding to kinesin light chains. *J. Cell Sci.*15: 122(Pt 22):4177-85. 2009.

[133]Araki Y et al. The novel cargo Alcadein induces vesicle association of kinesin-1 motor components and activates axonal transport. *EMBO J.*21: 26(6): 1475–1486. 2007.

[134] Kawano T., Araseki M., Araki Y., Kinjo M., Yamamoto T., Suzuki T. A small peptide sequence is sufficient for initiating kinesin-1 activation through part of TPR region of KLC1. *Traffic.*13(6):834-48. 2012.

[135] Konecna A. et al., Calsyntenin-1 Docks Vesicular Cargo to Kinesin-1. *Mol. Biol. Cell.* 17(8): 3651–3663. 2006.

[136] Ligon L.A., Tokito M., Finklestein J.M., Grossman F.E., Holzbaur E.L. A direct interaction between cytoplasmic dynein and kinesin I may coordinate motor activity. *J. Biol. Chem*. 279: 19201-19208. 2004.

[137] McGuire J.R., Rong J., Li S.H., Li X.J. Interaction of Huntingtin-associated protein-1 with kinesin light chain: implications in intracellular trafficking in neurons. *J. Biol. Chem.*10: 281(6):3552-9. 2006.

[138] Schmidt M.R., Maritzen T., Kukhtina V., Higman V.A., Doglio L., Barak N.N., Strauss H., Oschkinat H., Dotti C.G., Haucke V. Regulation of endosomal membrane traffic by a Gadkin/AP-1/kinesin KIF5 complex. Proc. Natl. Acad. Sci. U.S.A.8: 106(36):15344-9. 2009.

[139] Wilson M.H. and Holzbaur E.L.F. Nesprins anchor kinesin-1 motors to the nucleus to drive nuclear distribution in muscle cells. Development. 1: 142(1): 218–228. 2015.

[140] Zhu H., Lee H.Y., Tong Y., Hong B.S., Kim K.P., Shen Y., Lim K.J., Mackenzie F., Tempel W., Park H.W.Crystal structures of the tetratricopeptide repeat domains of kinesin light chains: insight into cargo recognition mechanisms.*PLoS One.* 7(3):e33943. 2012.

[141] Blasius T.L., Cai D., Jih G.T., Toret C.P., Verhey K.J. Two binding partners cooperate to activate the molecular motor Kinesin-1. *J. Cell Biol.* 1: 176(1):11-7. 2007.

[142] Hammond J.W., Cai D., Verhey K.J. Tubulin modifications and their cellular functions. *Curr. Opin. Cell Biol.* 20(1):71-6. 2008.

[143] Nikolovska-Coleska Z., Wang R., Fang X., Pan H, Tomita Y., Li P., Roller P.P., Krajewski K., Saito N.G., Stuckey J.A., Wang S. Development and optimization of a binding assay for the XIAP BIR3 domain using fluorescence polarization. *Anal. Biochem.* 15: 332(2):261-73. 2004.

[144] Pernigo S., Chegkazi M.S., Yip Y.Y., Treacy C., Glorani G., Hansen K., Politis A., Bui S., Dodding M.P., Steiner R.A. Structural basis for isoform-specific kinesin-1 recognition of Y-acidic cargo adaptors.*Elife.*15(7): e38362. 2018.

[145] Leslie A.G.W. The integration of macromolecular diffraction data. *Acta Cryst*. D62, 48-57. 2006.

[146] Murshudov G.N. and Evans P.R. How good are my data and what is the resolution? *Acta Cryst*. D-69, 1204-1214. 2013

[147] Krissinel E., Uski V., Lebedev A., Winn M., Ballard C. Distributed computing for macromolecular crystallography. *Acta Cryst*. D-74, 143-151. 2018.

[148] Thorn A., Parkhurst J., Emsley P., Nicholls R.A., Vollmar M., Evans G., Murshudov G.N. AUSPEX: a graphical tool for X-ray diffraction data analysis. *Acta Cryst*. D-73, 729-737. 2017.

[149] Murshudov G.N., Skubak P., Lebedev A.A., Pannu S.N., Winn Martyn S., Steiner R.A., Nicholls R., Wynn M., Logen F., Vagin A.S., Long F., Vagin A. REFMAC5 for the refinement of molecular crystal structures. *Acta Crystal.* D-67, 355-363. 2011.

[150] Skubák P. and Pannu N. S. Automatic protein structure solution from weak X-ray data. *Nat. Commun.* 4:2777. 2013.

[151] Krissinel E. Enhanced fold recognition using efficient short fragment clustering. *J. Mol. Biochem.* 1(2), 76-85. 2012.

[152] Sheldrick G.M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystal. D*. D-66, 479-485. 2010.

[153] Cowtan K. Recent developments in classical density modification. *Acta Crystal. D*. D-66, 470-478. 2010.

[154] Emsley P. and Cowtan K. *Coot:* model-building tools for molecular graphics. *Acta Crystal. D.* D-60, 2126-2132. 2004.

[155] Krissinel E. and Henrick K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372(3), 774-797. 2007.

[155] Urzhumseva L., Afonine P.V., Adams P.D., Urzhumtsev A. Crystallographic model quality at a glance. *Acta Crystal. D.* D-65, 297-300. 2009.

[156] Williams C.J., Hintze B.J., Headd J.J., Moriarty N.W., Chen V.B., Jain S., Prisant MG Lewis S.M., Videau L.L., Keedy D.A., Deis L.N., Arendall WB I.I.I., Verma V., Snoeyink J.S., Adams P.D., Lovell S.C., Richardson J.S. and Richardson D.C. MolProbity: More and better reference data for improved all-atom structure validation. *Protein Science*. 27, 293-315. 2018.

Publications

Pernigo S., Chegkazi M.S., Yip Y.Y., Treacy C., Glorani G., Hansen K., Politis A., Bui S., Dodding M.P., Steiner R.A. Structural basis for isoform-specific kinesin-1 recognition of Y-acidic cargo adaptors.*Elife.* 15(7): e38362. 2018.

Glorani G., Marin R., Canton P., Pinto M., Conti G., Fracasso G., Riello P. Pegylated silica nanoparticles: cytotoxicity and macrophage uptake.*J. Nano. Res.* 19(8): 1-14. 2017.

Acknowledgments

I would like to thank Professor H.L. Monaco for giving me the opportunity to be part of his team and for the continuous support during my PhD path. Besides, I would like to thank Dr. Stefano Capaldi and Dr. Massimiliano Perduca for teaching me how to deal with difficult situations.

Thanks to Professor Roberto A. Steiner for accepting me in his lab at King's College: it was an amazing opportunity for my personal and professional development and I had the opportunity to meet new and valuable people, such as Soi Bui, Magda Chegzaki, Stefano Pernigo, Gintare Bucaite, Mitla Garcia Maya, and people from Conte Lab.

Thanks to all the people who gave their contribution to this work, including Professor Daniele Dell'Orco, Professor Stefano Dusi, Dr. Marta Donini and Dr. Maria Teresa Valenti, my thesis would not be the same without their help. Thanks also to the organization committee of the 1st CCP4 Structure Solution workshop and to all the amazing tutors who helped with my dataset, without them I would probably have never managed to elucidate the structure of TFL.

My sincere gratitude goes to Dr. Michele Bovi, Dr. Maria Cecilia González and Dr. Laura De Stefanis, exemplary colleagues, faithful friends and greater supporters. I will always think of you guys with a smile, remembering our helpful discussion, our lovely coffee break, and big laughs.

I woud also like to say a huge thank to my family, for their patience, and their constant support, even when I went throught hard times.

I would like to express my gratitude and my affection to Ms. Georgia Sigala, not only my landlady andflatmate during my stay in London, but rather a special friends, who was able to put some sense into my irrational mind, to help me dealing with difficult situations and to show her constant support and appreciation. I would like to dedicate my protein to her, renaming it "Georgia's peptide".

Last, but not least, I would like to say a big thank to Dr. Riccardo Marin for pushing me every day to do my best, for teaching me to never give up and to be one of my supporters.