INSTITUTO SUPERIOR TÉCNICO
Universidade Técnica de Lisboa

Production and Purification of the $p16^{INK4a}$:

A Naked Mole-Rat's Resistance Protein to Cancer

João Alberto Pacheco Marques de Vasconcelos e Sá

Thesis to obtain the Master of Science Degree in

Biological Engineering

Examination Committee

Chairperson: Professor Duarte Miguel de França Teixeira dos Prazeres (Departamento de Bioengenharia (DBE), Instituto Superior Técnico)

Supervisors : Professora Maria Ângela Cabral Garcia Taipa Meneses de Oliveira (Departamento de Bioengenharia (DBE), Instituto Superior Técnico)

Professor Alexander Dikiy (Norwegian University of Science and Technology)

Member of the Committee: Professor Arsénio do Carmo Sales Mendes Fialho (Departamento de Bioengenharia (DBE), Instituto Superior Técnico)

November 2012

Preface

The thesis was performed at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway, under the Erasmus Student Program. The idea for the project was concocted by the head of the Laboratory of Structural Aspects of Biomolecules Activity (SABA), and the laboratory work was performed in laboratories at the Department of Biotechnology. I would like to thank the warming reception of my supervisors in Norway: Prof. Alexander Dikiy and Dr. Elena Shumilina, and express them my gratitude for giving me the chance to do my thesis abroad.

I am also thankful to Prof. Ângela Taipa, my supervisor in my home institution, for her support and care, and to Prof. Pedro Matias of ITQB for answering some questions about crystallography.

This thesis is dedicated to my parents, who allowed me to have an education.

Abstract

The aim of the present work was the purification of $p16^{INK4a}$ protein from naked mole-rat (nmr) with the objective of performing NMR experiments. These experiments would serve to obtain the necessary information for determination of its tridimensional structure.

Naked mole-rat (*Heterocephalus glaber*) is a remarkable mammal in which spontaneous cancer has never been observed. Experiments showed that the protein $p16^{INK4a}$, encoded in the INK4a/ARF locus, plays an important role in tumor suppression by supporting early contact inhibition, a mechanism absent in human cells. Moreover, this locus uniquely changed in nmr, generates shorter forms of both protein products synthesized from this locus. Thus, analysis of the structure and function of $p16^{INK4a}$ may provide an explanation for the roles of these proteins in cancer. Due to the recent completion of the nmr's genome project, the proposed work is original and innovative.

In a first phase, the plasmid pET28B in *E. coli* BL21 (DE3) which produced p16 with a histidine tag was used, and in a second phase the plasmid pGEX-4T1 in *E. coli* BL21 (DE3) which produced p16 with a GST tag was used. In both expression systems pure concentrated p16 was not achieved, due to precipitation of the protein which was not overcome, despite several conditions have been tried.

Thus, it seems that the precipitation is due intrinsically to p16 and possible alternatives are to use more soluble fusion partners, co-expression with chaperones or even perform NMR or x-ray crystallography experiments with the protein still fused to GST.

Keywords: p16INK4a, naked mole-rat, purification, production, NMR, cancer

Index

List of Tables, Figures and Abbreviations

List of Tables

List of Figures

List of Abbreviations

ITC – isothermal titration calorimetry
MALDI-TOF – matrix-assisted laser desorption/ionization time-of-flight mass spectrometer
MS – mass spectra
NMR – nuclear magnetic resonance
nmr – naked mole rat
SDS-PAGE – sodium dodecyl sulfate polyacrylamide gel electrophoresis

1 Introduction

1.1 Interest in the study of Naked Mole-Rat

The naked mole-rat (*Heterocephalus glaber*) (nmr) is a strictly subterranean mammal which caught scientists' attention because of its extraordinary life expectancy: around 30 years, making him the longest-living rodent (Gladyshev, Wang, Guojie, Park, & Sunyaev, 2011). And this is surprising because it goes against the classic Oxidative Stress theory, which correlates an organism mass with its life expectancy. For the naked mole-rat mass its life expectancy is far higher than it is predicted by that theory (Buffenstein, 2005). And this striking characteristic of this species is not explained neither by the low oxygen environments where they live (Buffenstein, 2008) nor by superior antioxidant defenses (Buffenstein, 2006).

1.2 Link between Longevity and Cancer Resistance

One hypothesis to explain the longevity of naked mole-rat is the strong connection between longevity and cancer, as "cancer is a disease that generally affects most mammalian species and is usually believed to be an unavoidable accompaniment of aging" (Horsnby et al, 2010). It happens that in naked mole-rat, death because of cancer has never been observed, and necropsies have not revealed incidental tumours (Hornsby et al 2010). Two distinguishing tumor suppressing mechanisms have been found on these animals which are unique in the mammals kingdom, from what science knows at the moment. One of them consists on the safety mechanism that makes tumorous cells lose their anchorage and therefore fail to grow in new tissue. In addition, transplanting naked mole-rat's fibroblats expressing oncogenes into mice is not sufficient to cause tumours, as it is with humans and mice. The phenomenon was attributed to the cells' transplantation and therefore loss of anchorage, to which human telomerase reverse transcriptase (hTERT) may confer resistance (Hornsby et al, 2010).

1.3 Early Contact Inhibition

Another tumor suppressing mechanism uniquely found in nmr until now is the early contact inhibition mediated by $p16^{INK4a}$. This mechanism is, it seems, absent in human and mice. Contact inhibition is a key anticancer mechanism that arrests cell division when cells reach a high density. But what is particular of naked mole-rat is that, besides having the regular contact inhibition existent in human and mice with the induction of the protein $p27^{KIP1}$, it has an early contact inhibition where the protein $p16^{INK4a}$ plays a major role. These two mechanisms are temporally separated, and p27 is induced as a backup mechanism when the p16-mediated protection fails (Figure 1) (Seluanov et al, 2009).

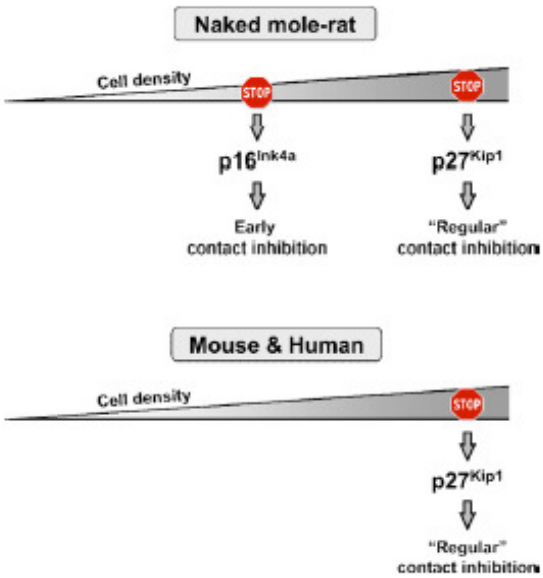

Figure 1 – A model comparing the contact inhibition in naked mole-rat and in mouse and human. Adapted from (Seluanov et al, 2009).

This means that naked mole-rat fibroblasts begin to proliferate but reach confluence upon the formation of the first cell-cell contacts while mouse fibroblasts reach a much higher cell concentration before confluence - Figure 2.

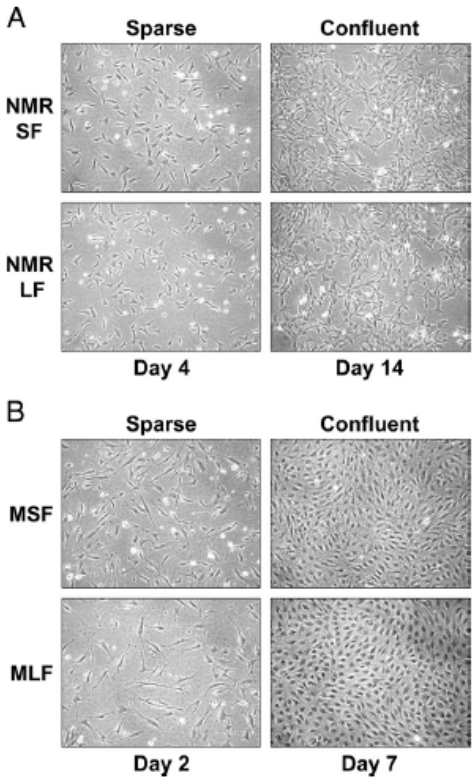

Figure 2 – Growing and confluent naked mole-rat (A) and mouse (B) fibroblasts. SF – skin fibroblasts; LF – lung fibroblasts. From (Seluanov et al, 2009).

Analysis of expression levels of $p16^{INK4a}$ and $p27^{KIP1}$ in naked mole-rat and in mouse and human showed without any doubt a dramatic difference of the role of those proteins in growth arrest (Figure 3). These Western blots show very slight induction of p27 in confluent nmr cells in contrast with the heavy induction in human and mouse fibroblats. As for p16 induction, confluent nmr cells show a strong induction and apparently non-existent in human and mouse fibroblasts.

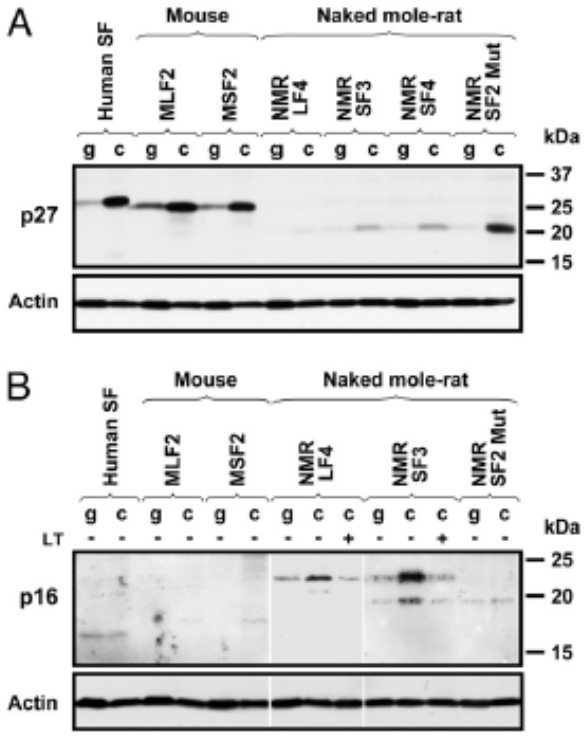

Figure 3 – Western blots analysis for p27 (A) and for p16 (B) in growing and confluent human, naked mole-rat and mouse fibroblasts. g – growing cells; c – confluent cells; LF – lung fibroblasts; SF – skin fibroblasts; NMR Mut – mutated naked mole-rat cells that lost early contact inhibition. Adapted from (Seluanov et al, 2009).

Although p16 is not induced in humans as much as in nmr, Table 1 shows there is a strong correlation between loss of p16 due to loss (point mutation, genetic deletions or promoter methylation) in its locus, and cancer. Generally, about one third of cancers in USA show inactivation of p16 locus, making its loss one of the most frequent lesions of human malignancy.

Table 1 – $p16^{INK4a}$ innactivation in most common cancers in the United States of America. From (Sharpless, 2005).

Tumor type	Annual incidence	Frequency p16 loss[a]	Total cases p16 loss
Prostate	162.0	0.04	6.5
Breast	134.0	0.21	28.1
NSCLCa	106.6	0.64	68.2
Colorectal	66.4	0.27	17.9
Bladder	48.3	0.59	28.5
NHL	37.7	0.33	12.4
SCLC	35.7	0.03	1.1
Head and Neck	31.7	0.68	21.6
Melanoma	31.1	0.65	20.2
Kidney	24.4	0.05	1.2
Endometrium	23.9	0.18	4.3
Leukemias	23.3	0.61	14.2
Pancreas	21.9	0.85	18.6
Ovary	16.2	0.25	4.1
Stomach	16.1	0.44	7.1
Thyroid	13.5	0.03	0.4
Esophagus	11.0	0.70	7.7
Multiple Myeloma	10.1	0.60	6.1
Other	158.5	0.20	31.7
Total	972.4	0.31	299.9

Therefore, the study of p16 is an important issue for unraveling the mechanism of cancer and longevity.

1.4 p16 Genetics and Metabolic Functions

The protein p16 is expressed when the INK4a open reading frame is transcribed hence protein $p16^{INK4a}$. This tumor suppressor $p16^{INK4a}$ binds to cyclin-dependent kinases 4 and 6 and shares the same gene locus with tumor suppressor $p19^{ARF}$ (Figure 4), which regulate pRb and p53 pathways, respectively, which are non-overlapping. These proteins are encoded in alternative reading frames and therefore are not isoforms and have no aminoacid sequence similarity. This genomic structure is practically unique in the mammalian genome (Sharpless, 2005). Both proteins possess significant *in vivo* tumor suppressor activity. In (Seluanov et al, 2009), it is shown that dual inactivation of both proteins is a required condition to obtain cancerigenous nmr's cells. Despite appearing a design flaw, the structure of the locus probably encompasses functional benefits, as a multi-functional tumor repressor locus perfected through evolution to sense a variety of somewhat generic, cancer-associated stimuli, and respond to these by halting the growth of would-be cancer cells.

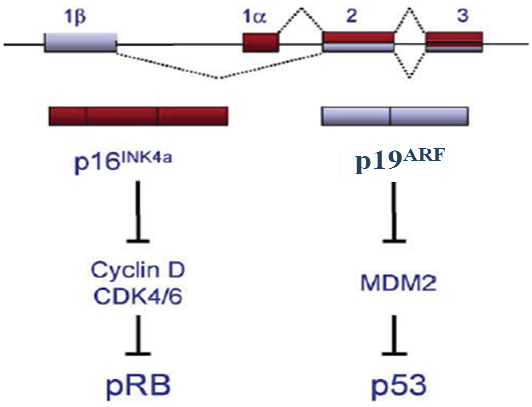

Figure 4 – The INK4a/ARF locus: the open reading frame for $p19^{ARF}$ (in blue, in naked mole-rat, $p14^{ARF}$ for human) and for $p16^{INK4a}$ (in red) are shown. Each has a unique first exon that then splices to a common second exon, but in alternate reading frames. Adapted from (Sharpless, 2005).

Figure 5 shows that $p16^{INK4a}$ and $p19^{ARF}$, despite having the same locus, are not redundant, and reside at the core of a signaling network that principally governs cell-cycle entry and exit (Lowe and Sherr, 2003). In this network, $p16^{INK4a}$ abrogates the kinase activity of cyclin-dependent kinases (CDK) 4 (Serrano et al, 1993) and 6 (Russo et al, 1998), which otherwise would phosphorylate the retinoblastoma protein (pRb) and therefore promote S-phase entry (Sharpless, 2005).

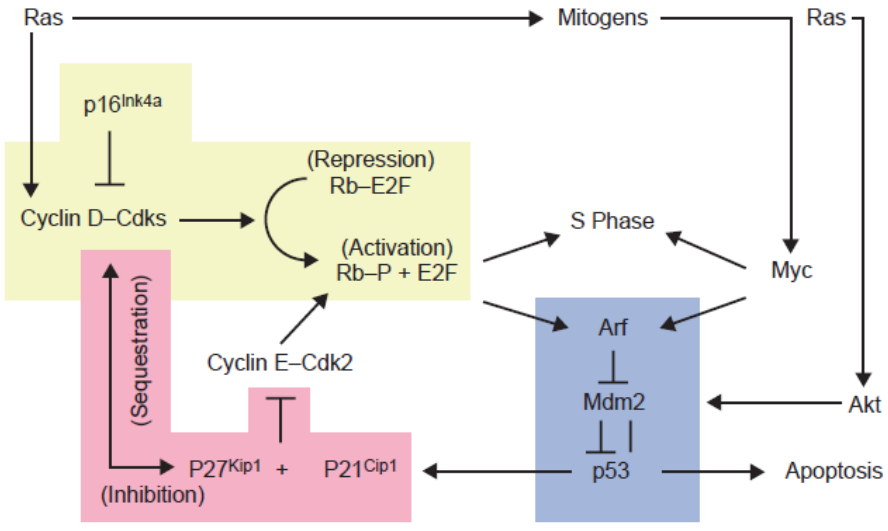

Figure 5 – The INK4a/ARF network. From (Lowe and Sherr, 2003).

However, $p16^{INK4a}$ protein enters in more biological processes, which is supported by the various locations where the protein is present inside the cell - Table 2.

Table 2 – Structural Classification of Proteins (SCOP), Protein Structure Classification (CATH) and Gene Onthology (GO) Terms for p16^{INK4a} with PDB code 1DC2.

Domain Annotation: SCOP Classification (version 1.75) — Hide

Domain Info	Class	Fold	Superfamily	Family	Domain	Species
d1dc2a_	Alpha and beta proteins (a+b)	beta-hairpin-alpha-hairpin repeat	Ankyrin repeat	Ankyrin repeat	Cell cycle inhibitor p16ink4A	Human (Homo sapiens) [TaxId: 9606]

Domain Annotation: CATH Classification (version 3.4.0) — Hide

Domain	Class	Architecture	Topology	Homology
1dc2A00	Mainly Alpha	Alpha Horseshoe	Serine Threonine Protein Phosphatase 5, Tetratricopeptide repeat	

Gene Product Annotation: GO Terms — Hide

Polymer	Molecular Function	Biological Process	Cellular Component
CYCLIN-DEPENDENT KINASE 4 INHIBITOR A (P16INK4A) (1DC2:A)	cyclin-dependent protein kinase inhibitor activity protein binding protein kinase binding NF-kappaB binding	cell cycle checkpoint G1 phase of mitotic cell cycle G1/S transition of mitotic cell cycle mitotic cell cycle negative regulation of cell-matrix adhesion induction of apoptosis cell cycle cell cycle arrest Ras protein signal transduction negative regulation of cell proliferation negative regulation of cell growth negative regulation of NF-kappaB transcription factor activity positive regulation of smooth muscle cell apoptosis senescence-associated heterochromatin focus assembly negative regulation of phosphorylation negative regulation of cyclin-dependent protein kinase activity negative regulation of transcription, DNA-dependent replicative senescence positive regulation of macrophage apoptosis positive regulation of cellular senescence	nucleus cytoplasm cytosol senescence-associated heterochromatin focus

1.5 p16^{INK4a} protein in Naked Mole-Rat

Due to recent genome sequencing of nmr, it was observed that there are two early stop codons in the second exon (Figure 6) and there is low sequence similarity in the last exon (Figure 7) suggesting an enshortment of the protein – resulting in a 14 kDa protein (Gladyshev et al, 2011) – when comparing with mammalian orthologs. It is remarkable that in nmr this protein has been altered in such dramatic way, whereas it is conserved in other organisms.

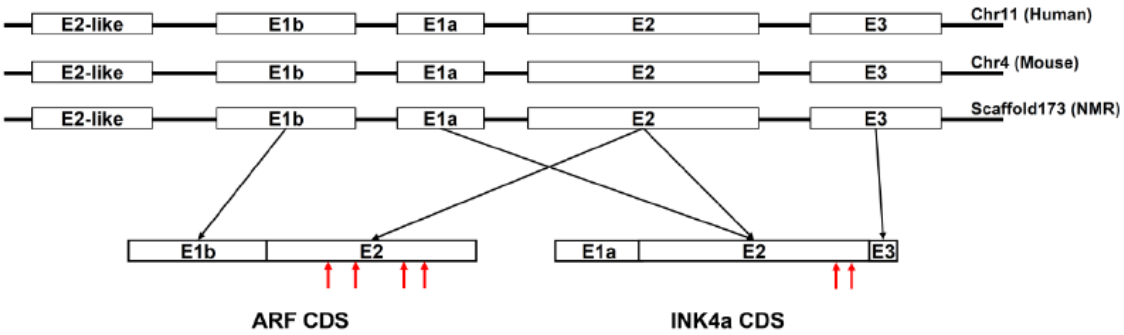

Figure 6 – The INK4a/ARF locus within the genome. E1a, E1b, E2 and E3 are exons. The E2-like region is a sequence with homology to E2. Stop codons detected in naked mole-rat genes are indicated with red arrows. From (Gladyshev et al, 2011).

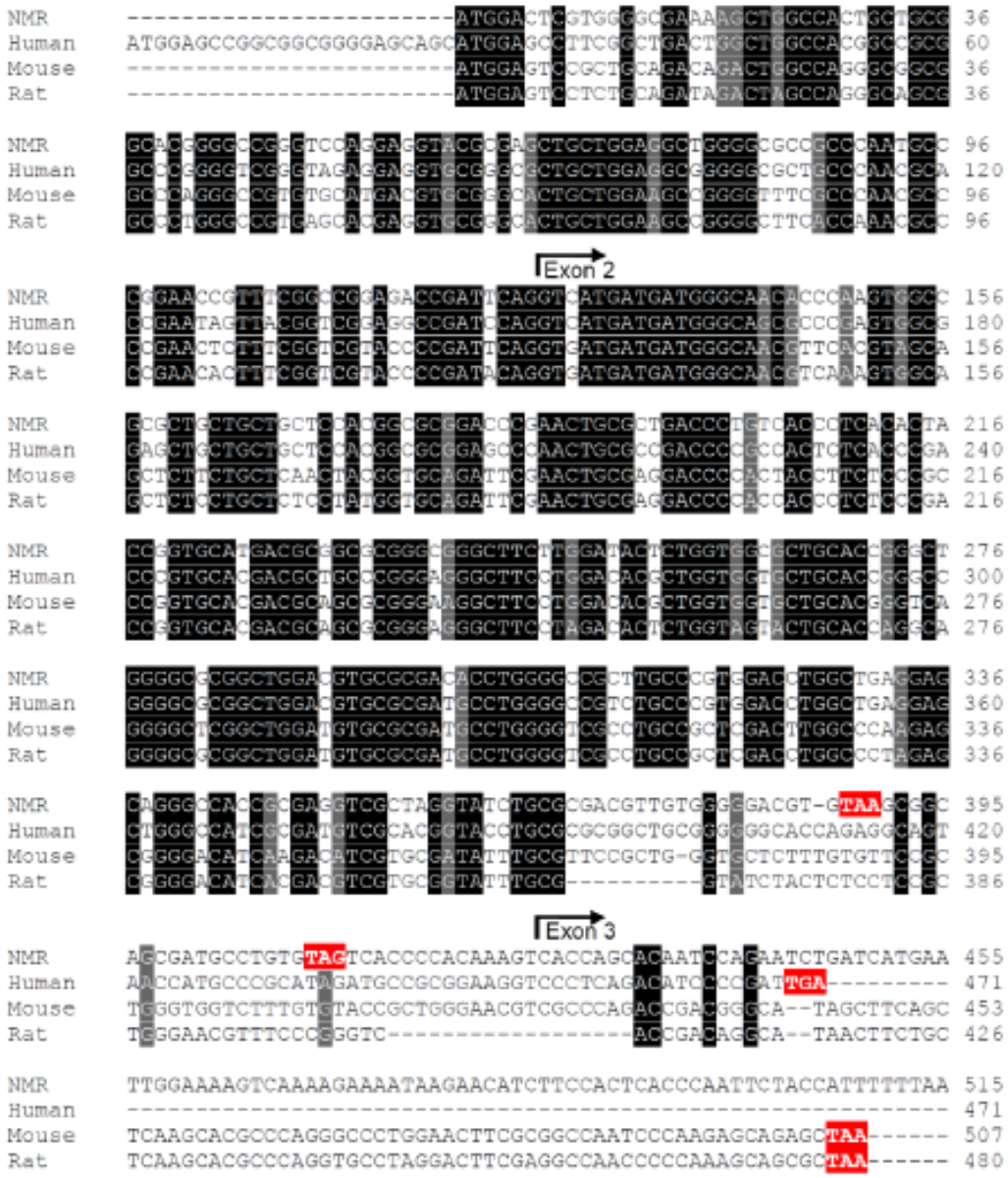

Figure 7 – Alignment of mammalian INK4a coding regions. Stop codons are shown in red. From (Gladyshev et al, 2011).

Despite the possible shorter protein, the four ankyrin repeats were intact and Thr69 (in red), a residue important for CDK6 binding, was conserved, so the function of the protein may be partially preserved (Gladyshev et al, 2011).

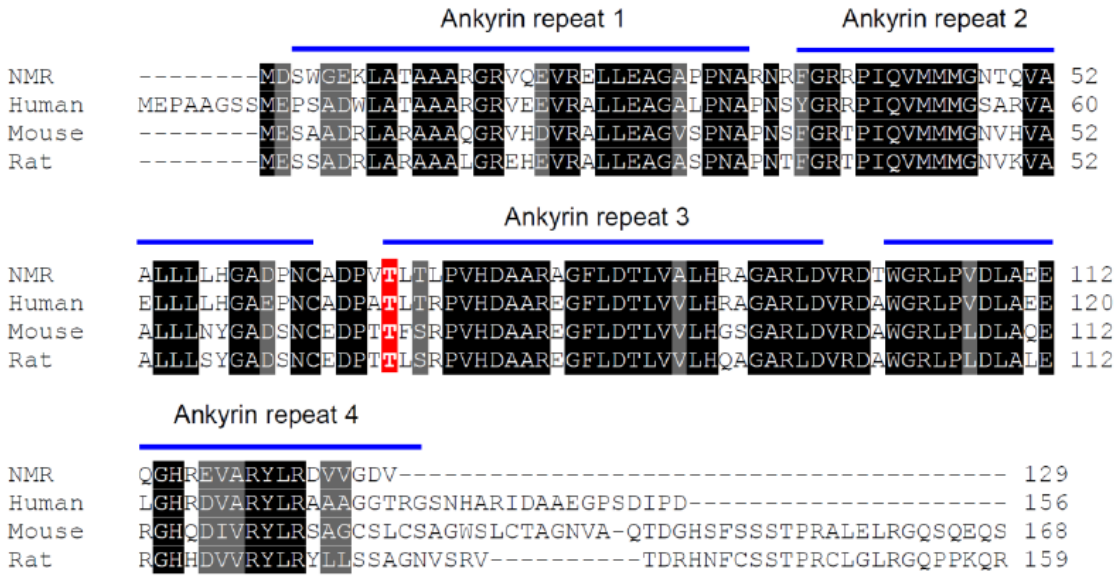

Figure 8 – Alignment of mammalian INK4a protein sequences. Location of four conserved ankyrin repeats is shown by blue lines above the sequence. The conserved threonine residue important for CDK6 binding is shown in red. From (Gladyshev et al, 2011).

In summary, the naked mole-rat's p16 has a molecular weight of 14083.0 Da (with the aminoacid sequence of Figure 9), has a theoretical pI of 6.41 predicted by the bioinformatic tool of Expasy ProtParam, is composed of 129 aminoacids which only one of them is a cysteine residue so there are no intra-disulfide bonds, has the four ankyrins intact and presents great orthology with human p16, with the exception of the last exon that it is absent (79% of the aligned aminoacids are the same).

MDSWGEKLATAAARGRVQEVRELLEAGAPPNARNRFGRRPIQVMMMGNTQVAALLLLH
GADPNCADPVTLTLPVHDAARAGFLDTLVALHRAGARLDVRDTWGRLPVDLAEEQGHRE
VARYLRDVVGDV

Figure 9 – Aminoacid sequence of p16^{INK4a} protein in naked mole-rat, from the N-terminus to the C-terminus.

1.6 p16 Tertiary Structure in the literature

Until the moment, no naked mole-rat's p16 protein structure was yet reported to date but one research group has determined it for human p16 [(Byeon et al, 1998), (Byeon et al, 2000)] by NMR, and another solved the structure of the human p16-CDK6 complex by X-ray crystallography (Russo et al, 1998).

The low quantity of solved structures of p16 up to date is explained on one hand because the protein has a strong tendency to aggregate and therefore it's very difficult to obtain a crystal of free p16, and on the other hand, its conformational flexibility and protein aggregation forces NMR experiments to use rather low concentrations of 0.2-0.4 mM, thus requiring long acquisition times which lead to data problems due to irreversible denaturation (Byeon et al, 1998).

The tertiary structure determination of human p16^{INK4a} (Figure 10) and analysis revealed that the loops connect the four ankyrin motifs and are oriented perpendicularly to those helical axes, show less defined structure and that loops 2 and 3 (Figure 11) are stabilized by His83, which is an aminoacid conserved in both human and naked mole-rat (Byeon et al, 1998).

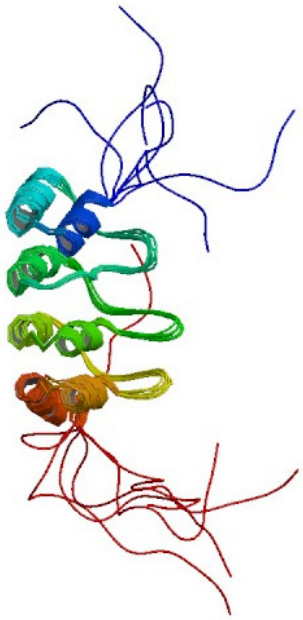

Figure 10 – Human p16^{INK4a} structure. From (Byeon et al, 2000).

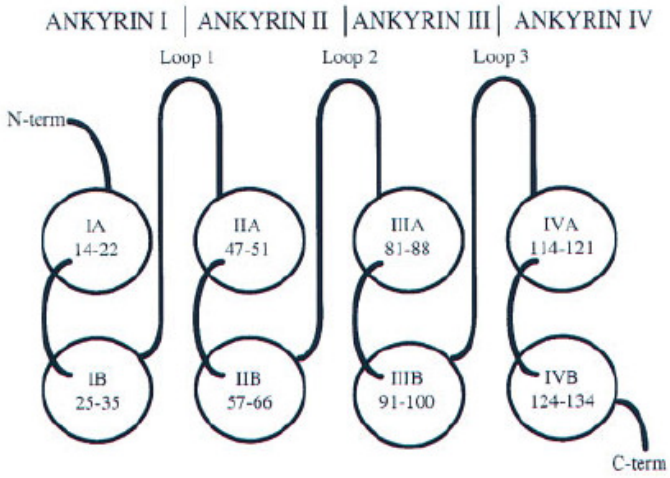

Figure 11 – Topology diagram of the human p16 structure. Helices (in circle) are perpendicular to the plane of the page, and the residue numbers forming the helices are indicated. From (Byeon et al, 1998).

The solvent-accessible surface representation of p16 (Figure 12) shows two possible important features for the binding to CDK4: the presence of clusters of charged groups on the surface, in red, and a pocket located on the right side of the representation (Byeon et al 1998).

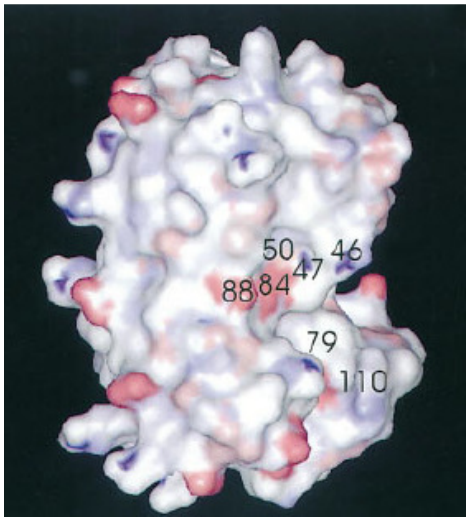

Figure 12 – Solvent-accessible surface representation of human p16. Positive and negative charges are indicated by blue and red, respectively. From (Byeon et al, 1998).

Other characteristics of the protein include its high flexibility on the minutes to hundreds of hours time-scale due to the lack of nonexchangeable NH protons (Byeon et al, 1998), its marginal stability due to denaturation experiments that yielded the low denaturation stability of 1,94 kcal/mol (Byeon et al, 1999) and its very limited conformational flexibility on the pico-to-nanosecond time-scale due to the heteronuclear 1H-15N NOE measurements (Byeon et al, 1999).

1.7 Functional Study of p16 protein in the literature

As for the functional study of p16 protein on its interactions with cyclin-dependent kinases, some initial research has been conducted. In respect to human p16-CDK4 complex, three residues (human H66, D84 and R124) were found to be possible key elements in that interaction by doing p16 mutant activity tests and respective 2D NMR analysis (Byeon et al, 1998). However two of those residues are not conserved in naked mole-rat aminoacid sequence. In the same article, yeast two-hibrid system experiments followed by site-specific mutagenesis allowed to discover the location of binding in CKD4 (Byeon et al, 1998), and predict the 3D interaction between those proteins.

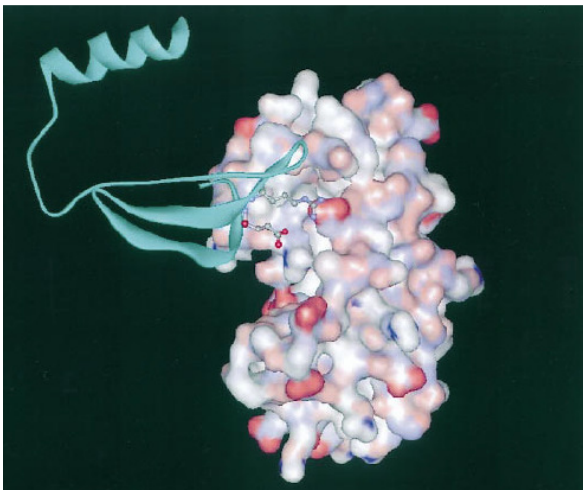

Figure 13 – Working model for the human p16-CDK4 complex. p16 protein is in solvent-accessible surface representation while only part of CDK4 that was found to interact with p16 is represented in blue. This docking experiment was performed with Global Range Molecular Matching (GRAMM). From (Byeon et al, 1998).

The interaction between human p16 and CDK6 was also studied (and the structure of the complex solved) and this study led to the conclusion that the first and second loops appear to be involved in the interaction (Russo et al, 1998).

1.8 Protocols used in the literature to obtain human p16 protein's sample for NMR experiments

Human p16 was expressed in soluble form as a glutathione S-transferase (GST) fusion protein in *E. coli* BL21 (DE3) (Novagen). The cell lysate was purified on a glutathione –agarose column; p16 was cleaved on-column with thrombin and further purified by an S-100 column equilibrated with 4 mM HEPES buffer containing 1 mM DTT and 5 mM EDTA (pH 7.5). Protein concentrations used [for NMR] varied from 0.2-0.4 mM (Byeon et al, 1998). In (Byeon et al, 2000) where $p15^{INK4b}$ (which shares a common ancestor from 350 million years ago with $p16^{INK4a}$ (Sharpless, 2005) and therefore is similar) was studied and the protocol was identical to (Byeon et al, 1998), with additional information about the isotopically labeled protein which was grown in M9 media, about the typical yields of p15 protein which were 2-4 mg/L in rich media and 1-3 mg/L in minimal media. The authors also reveal that the samples were lyophilized and stored at -80 ºC before use and the samples for NMR experiments contained 4 mM HEPES, 1 mM DTT and 5 µM EDTA in 95% H2O / 5% D_2O or 100% at pH = 7.5. For the determination of $p18^{INK4c}$ (Li et al, 1999) (which shares a common ancestor with $p16^{INK4a}$) the exact same protocol of (Byeon et al, 1998) and (Byeon et al, 2000) was followed. Finally, for the other related protein – $p19^{INK4d}$ – 15 mg/mL of that protein (from mouse) were obtained (Holak et al, 1998), which corresponds to 0.9 mM, by following the protocol in (Luh et al, 1997) in which the

differences to the protocols used in (Byeon et al, 1998) and (Byeon et al, 2000) consists in that the protein was affinity-purified on glutathione-Sepharose beads (Pharmacia) and that after thrombin removal by an anti-thrombin resin (Sigma), the exchange buffer used was composed of phosphate buffer (20 mM, pH 7.5) containing 100 mM NaCl, 1 mM EDTA and 5 mM DTT.

1.9 Structural and Functional Analysis of p16 by NMR

1.9.1 Importance of determining nmr's p16 structure

The hypothesis underlying this work is that the tumor suppressor p16 (see section 1.4) has improved or other biological functions in naked mole rat than in human due to its extraordinary cancer resistance (see section 1.1). Thus, testing to this hypothesis implies determining the three-dimensional structure of nmr's p16, in order to compare to the already available human one (see section 1.6). Comparing structures is a prerequisite to compare biological functions, as the biological function of a protein depends on its native conformation (Horton et al, 2006). The biological functions of interest for p16 protein would be the interaction with cyclin-dependent kinases 4 and 6 present in the network that govern the cell entry to the S-phase (Figure 5).

1.9.2 Choosing NMR as a technique for determining nmr's p16 structure

At the moment, the three techniques that independently enable the determination of a protein structure are: x-ray crystallography, NMR spectroscopy and *in silico* prediction. The later is not experimental therefore there isn't the certainty that the obtained structure is equal to the biological on (Li, 2010). Moreover, as seen in section 1.61.7, there is a the very low possibility of obtaining crystals of p16 protein in order to perform x-ray crystallography, due to its strong aggregation properties, and so far there are no academic papers reporting x-ray studies of p16 protein. This problem exists, in spite of p16 belonging to the ankyrin repeat family (Figure 8) whose many of its elements have been solved by x-ray crystallography (30 out of 38 in Protein Data Bank). This lead to choose NMR spectroscopy as the technique to determine nmr's p16 structure (as it has been also chosen for human p16, see section 1.6). NMR spectroscopy is a prodigious technique that has impacted all aspects of protein structure investigations, even X-ray crystallography, whose structural targets may be selected and refined using NMR spectral properties. Moreover, the availability of solution structures of proteins cannot be underestimated, even in cases in which crystal structures exist (Altieri and Byrd, 2004).

1.9.3 NMR experiments for structure determination

For possible interpretation of spectra, nmr's $p16^{INK4a}$ shall be expressed in cells in mineral media enriched with 15N and 13C isotopes (NMR active nuclei). The availability of the 15N, 13C double enriched protein would make it possible to use all available modern NMR pulse

sequences involving triple resonance experiments in order to obtain unambiguous 1H, 15N and 13C NMR spectral assignment. These data would be subsequently used in NOE-based 2D and 3D NMR experiments to determine distance geometrical constraints. The dihedral angle constraints for the protein would be also determined. The obtained structural constraints would be used as an input for the calculation of the three-dimensional structures of the proteins in solution. The calculated final structures would be made available through their coordinate deposition in the PDB data bank, while the data concerning NMR spectral assignment would be deposited to BioMagResBank (BMRB).

1.9.4 Problems & Solutions of structure determination by NMR

Despite the obvious benefits of NMR over x-ray crystallography, the interpretation of NMR spectra is laborious and time-consuming, usually taking several months on average (Li, 2010). Fortunately, there are promising advances in software tools for automation of protein structure determination using NMR (for such steps as resonance assignment, followed by assignment of NOE (nuclear Overhauser enhancement) interactions (now intertwined with structure calculation), assembly of input files for structure calculation, intermediate analyses of incorrect assignments and bad input data, and finally structure validation). At the moment near completely automated structure determination is possible for small proteins, and the prospect for medium and large sized proteins looks bright (Altieri and Byrd, 2004). Several software tools are being developed for different steps of structure determination. Table 3 lists some for determination of protein structure after resonance assignment. Tools for automatic peak picking and error tolerant backbone assignments have also been designed (Li, 2010). Moreover, a recent article claims to have made an important breakthrough on designing an algorithm (FLYA algorithm) that performs backbone and side-chain resonance assignment better than current alternative algorithms, having solely as input peak lists and the aminoacid sequence of the protein (Schmidt and Güntert, 2012). In that article, the algorithm was tested with three proteins with the same size range as naked mole-rat's p16 and the excellent results showed that FLYA assigned correctly 96-99% of the backbone and 90-91% of all resonances that could be assigned manually. Another quick and attractive strategy to determine naked mole-rat' p16 protein would be to resort to a structure based assignment using an AMR system, that uses an available ortholog structure (in this case: human p16 protein) as a reference structure (Li M. , 2010). These are all promising new tools for the goal of automatization, which will certainly shorten the time for spectra interpretation.

Table 3 – Summary of programs for automated structure calculation having resonance assignment as input (adapted from (Altieri and Byrd, 2004)). MD stands for molecular dynamics.

Program	MD engine	Utility
ARIA	CNS XPLOR	Ambiguous NOE restraint generation, spin diffusion correction, iterative structure calculation, analysis
AutoStructure	XPLOR CNS DYANA	NOE, torsion angle and hydrogen bond restraint generation, NOESY assignment, distance matrix calculation
BACUS/CLOUDS		NOESY assignment, distance matrix calculation
CANDID/ATNOS	DYANA	NOESY peak analysis, NOESY peak assignment, restraint generation, iterative structure calculation
NOAH	DIAMOD DIANA	NOESY assignment, NOE restraint generation, torsion angle restraints, iterative structure calculation
SANE	AMBER DYANA	NOESY assignment, restraint generation, structure calculation
PASD	XPLOR-NIH	Probability analysis of NOE restraints and simultaneous structure calculation

1.9.5 Experiments for functional analysis of nmr's p16

The dynamic properties of a protein can often determine its functional peculiarities. Therefore, elucidation of protein mobility on the broad time scale range (picoseconds-seconds) represents an important source of information required for establishing structure-function relationships within the investigated class of proteins. Availability of the NMR spectral assignment data (see previous paragraph) can be exploited to further analyze NMR experiments aimed at determining the longitudinal (R1) and transverse (R2) relaxation rates of backbone 15N nuclei as well as the hetero-nuclear (1H-15N) NOE. These parameters, determined for each residue, constitute a primary source of information to determine the mobility (order parameter, S2) of the protein backbone on the picosecond-nanosecond time scale. Additionally, from the analysis of other NMR experiments (T2 CPMG, T1ρ and H/D-N proton exchange experiments) the protein mobility would also be monitored on the longer (microsecond–second-minutes) time scales. Therefore, the described approach would lead to the complete determination of the $p16^{INK4a}$ mobility features. Then, these data can be correlated with the structural features of $p16^{INK4a}$ and their interacting partners (cyclin-dependent kinases 4 and 6, Figure 4) as well as with protein reactivity.

The full information regarding the molecular determinants for the interaction between two proteins must involve knowledge of both thermodynamic parameters and molecular details of

the interaction. Therefore, a double approach should be followed in the case of the interaction between $p16^{INK4a}$ and the interacting proteins of interest (cyclin-dependent kinases 4 and 6).

a) It is in general of great interest to determine the nature of forces that stabilize interaction between two proteins. The thermodynamics of such association are characterized by the stoichiometry of the interaction (n), the association constant (Ka), the free energy (ΔGb), enthalpy (ΔHb), entropy (ΔSb), and heat capacity of binding (ΔCp). In combination with structural information, the energetics of binding can provide a complete dissection of the interaction and aid in identifying the most important regions of the interface and the energetic contributions. Isothermal titration calorimetry (ITC) represents the most quantitative means available for measuring the thermodynamic properties of a protein-protein interaction. ITC can be used to directly measure the equilibrium binding by determining the heat evolved on the association of $p16^{INK4a}$ with its interactive partners. In a single experiment, the values of the binding constant (Ka), the stoichiometry (n), and the enthalpy of binding (ΔHb) can be determined. The free energy and entropy of binding will be determined from the association constant. The temperature dependence of the ΔHb parameter will be measured by performing the titration at varying temperatures, allowing us to derive the ΔCp term. The thermodynamic parameters for both nmr and human $p16^{INK4a}$ with respective interactive partners will be compared. The information obtained from such studies can aid in the elucidation of the factors that determine the specificity of the interactions.

b) NMR spectroscopy is a technique that allows us to monitor the interaction between two or more molecules. This feature is broadly used for the development of new drugs, to monitor protein-protein or protein-DNA interactions, and to study interactions between enzymes and their substrates/inhibitors. NMR spectroscopy could be applied for this purpose because the technique allows the detection of very weak interactions between molecules in solution. The nuclei chemical shift values are extremely sensitive indicators of the nucleus's chemical environment. If some of the molecule's residues are directly involved in the interaction with another molecule, the chemical shift and relaxation rate values of the relevant nuclei will be changed with respect to the corresponding values in the native molecule. The stronger the interaction is between molecules, the larger alterations of the indicated parameters from the involved nuclei are observed. Therefore, NMR allows to qualitatively monitor the interaction by providing information on the identity of amino acid residues involved in such interaction and, in addition, it provides with the information on the relative strength of the single local interaction. Thus, data concerning on the interaction of $p16^{INK4a}$ and its most important functional partners (cyclin-dependent kinases 4) can be generated. First, 15N-enriched nmr proteins would be titrated by the non-isotopically enriched selected interactive proteins, closed and prepared as described above for $p16^{INK4a}$. The presence of 15N nuclei in $p16^{INK4a}$ proteins allows monitoring this interaction by the analysis of spectral changes occurring within the 1H-15N HSQC spectrum upon addition of the interactive protein to $p16^{INK4a}$, while the 1H-15N HSQC NMR spectrum of non-enriched interactive protein would be silent and would not interfere. Titration studies (a number of HSQC spectra of $p16^{INK4a}$ protein upon increase of concentration of interactive

proteins) would provide with the information on the interaction stoichiometry and protein-protein binding constant, while directly indicating which $p16^{INK4a}$ residues are involved in the protein-protein interaction. The same set of NMR experiments would be performed for comparison purposes with human $p16^{INK4a}$ and its partners (cyclin-dependent kinases 4 and 6). These data would represent an extremely important source of experimental information since they can be subsequently used for the calculation of a structural model of their protein interactions (see below). Finally, they may lead to understanding of the differences in functional properties between nmr and human ortholog, which in turn may have important biomedical implications.

Determination of structures of protein-protein complexes is a very important tool in biochemistry because it can reveal the structural basis for the interactions between two bio-macromolecules and can provide hints about the functional roles of investigated proteins. Thus, the three-dimensional structures of the most relevant protein interactions, in which $p16^{INK4a}$ are involved, should be determined. The experimentally determined protein binding constant (see above) would allow the determination of the protein concentration required to form the complete complex. The strategy described above would be repeated in order to perform NMR spectral assignment of the complex. Additionally, taking into consideration the high molecular weight of the complex, TROSY (transverse relaxation optimization spectroscopy) based experiments would be used to decrease the signal linewidths and thus improve spectral resolution. Furthermore, semi-experimental biocomputing approach would be used to calculate the intermolecular structures of proteins encoded in nmr and human genomes. Again, the observed adaptations may indicate fundamental biomedical differences in the studied organisms.

Finally but only if necessary, site-directed mutagenesis would be used to produce protein variants to examine protein structure and function of nmr's p16 as it has been already done to some extent for human p16 (see section 1.7).

Summing up, all these experiments and studies might unveil important structural and functional differences, between nmr's p16 and the human ortholog, and thus understanding the role of p16 in conferring cancer resistance to nmr (Figure 14).

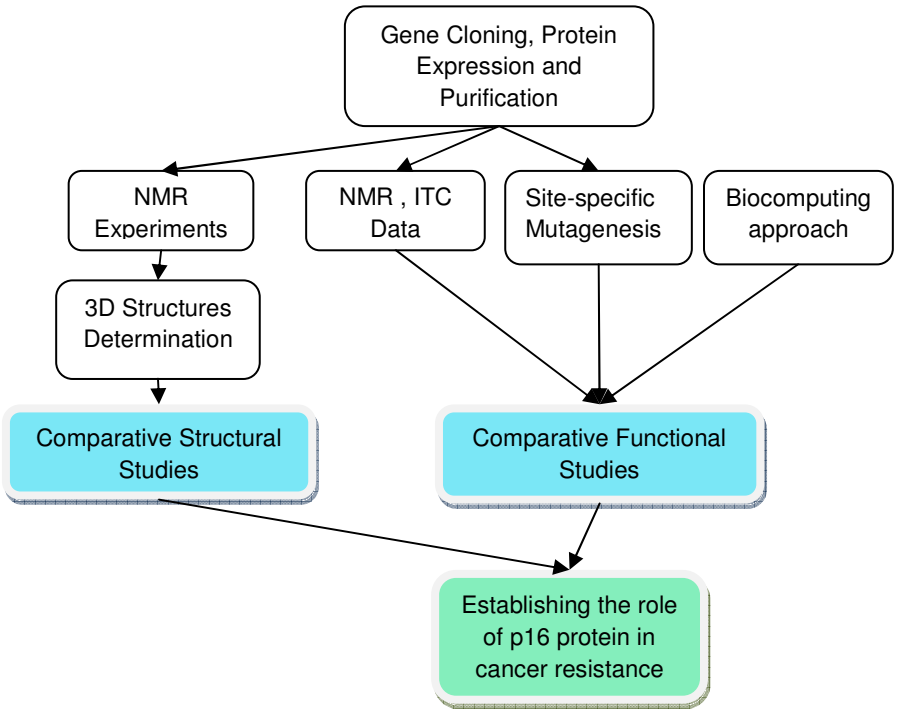

Figure 14 - Schematic representation of the experiments and analysis in a holistic study for understanding the role of p16 protein in cancer resistance mechanisms.

2 Objectives

The aim of the work was to develop the purification protocol for p16 protein, in a way that the protein solution would be suitable for NMR experiments. For that purpose, at least 95% of purity is required and the protein should be dissolved in a low salt buffer with no sensitive hydrogens. Moreover, the aimed concentration of the protein in NMR sample should be of at least 0.2 mM. Due to the highly expensive enriched media to be used in NMR experiments, the concentration goal had to be achieved in a minimum volume of 0.4 mL characteristic of a NMR sample by using 1 L of cell culture, meaning a minimum p16 quantity of 8×10^{-5} moles.

The long-term objective would be to obtain naked-mole rat p16 protein structure by NMR, which could be compared to the already known human p16 structure in order to see if there are 1) structural differences; 2) different binding affinities / specificities to cyclin-dependent kinases between the proteins. These results could reveal if naked-mole rat p16 is a key factor for conferring resistance to cancer in naked-mole rat through natural selection or if the differences in biological processes are just a consequence of the different anti-cancer mechanisms in that rodent.

3 Methods

3.1 Methods

Throughout the months of experiments, the protocol has been refined due to changes in the means to get the purified protein and to drawbacks that were corrected. Therefore, in this section, the methods described refer to the last version of the protocol. In addition, some experiments used different plasmids for expression, different affinity chromatography columns, different concentration procedures and different dialysis procedures which will be extensively listed.

3.1.1 Cell Culture and Expression of p16 fused to tag protein

There were two expression systems used. In the first one tried, p16 was expressed with a fused six histidine tag on its N-terminus, in *Escherichia coli* BL21 (DE3) cells transformed with pET28b plasmid. The his-tag used has a size of 2171.3 Da and a 20 aminoacids sequence of MGSSHHHHHHSSGCVPRGSH (which means his-p16 is composed of 149 aminoacids), where after the six histidines there is a cleavage site for thrombin (VPR↓GS). Expasy ProtParam bioinformatic tool predicts, for the joined sequence of his-tag and p16, a theoretical pI = 7.20 (for a total molecular weight of 16236.3 Da). When histidine tail is cut by thrombin, the pI of cut p16 is 6.57 and has a molecular weight of 14364.3 Da.

The second one expressed p16 fused with a GST tag on its N-terminus, in *Escherichia coli* BL21 (DE3) cells transformed with pGEX-4T1 plasmid. This sequence of 361 aminoacids - Figure 15 – has a molecular weight of 41059.4 Da and a pI = 6.23 according to the bioinformatic tool of Expasy ProtParam. When GST is cut by thrombin, the pI of cut p16 is 6.40 and has a molecular weight of 14853.9 Da.

N terminus MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYYIDGDVKLTQSMA
IIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTH
PDFMLYDALDVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPRGSPE
FGRPMDSWGEKLATAAARGRVQEVRELLEAGAPPNARNRFGRRPIQVMMMGNTQVAALLLLHGADPNCADPVTLTL
PVHDAARAGFLDTLVALHRAGARLDVRDTWGRLPVDLAEEQGHREVARYLRDVVGDV **C terminus**

Figure 15 – GST-p16 aminoacid sequence in the pGEX-4T1 plasmid that was used in experiments. p16 aminoacid sequence is colored green and the thrombin cleavage site is colored in blue.

The plasmids with the recombinant protein also conferred resistance against an antibiotic for selective pressure in culture medium. Plasmid pET28b made *E.coli* cells resistant to kanamycin and plasmid pGEX-4T1 to ampicillin. Thus, cell cultures were grown with certain concentrations of these antibiotics (see 8.1.4.1).

The pre-culture had usually a volume of around 1% of the final cell culture, was inoculated from cell colonies from agar plates and was incubated at 37 ℃ and at 220 rpm usually overnight. After reaching a high cell concentration, some volume of pre-culture was transferred to cell culture solution as inoculum, and incubated at 37 ℃ and at 220 rpm. Once reached an optical density of 0.6 at 600 nm, IPTG 1 M was added to solution to reach final concentration of 1 mM IPTG. Then, the cell culture was usually incubated for 2 hours at 25 ℃ and for 6 hours at 16 ℃.

3.1.2 Centrifugation and Sonication

The cell culture was transferred to pre-weighted centrifugation flasks adapted to SLA-1500 rotor, and centrifuged at 6000 rev/min for 10 minutes in Sorvall RC 5B Plus centrifuge. The supernatant was discarded, the flasks with cell were weighted again to determine cells weight and frozen at -20 ℃.

To sonicate the cells, those were unfrozen and added 5 mL of Cell Suspension Buffer per 1 g of cells in the flask. The flasks with buffer were stirred for 30 minutes at 4 ℃. Afterwards, the cell suspension was transferred to a sonication flask and sonicated at 30% Cycle, 3.5 of Power and for 5 minutes per 10 mL of cell suspension in Branson Sonifier 250. The lysate was centrifuged in flasks adapted to SS-34 rotor at 19 000 rpm for 40 minutes, and centrifuged again with the same conditions in new flasks after adding Protamine Sulfate to a final concentration of 0.4%. Finally the lysate was filtered through a 0.45 µm filter, and debris and protamine sulfate precipitate in the flasks were frozen at -20 ℃.

3.1.3 Capture Purification of p16

The capture purification step was done in native conditions to preserve its natural folding and presumably enhancing its solubility, by adding β-mercaptoethanol (a reducing agent) to buffers and adjust buffers to pH around 7.5.

3.1.3.1 For his-p16

A BIO-RAD chromatography column was filled with 0.8 mL of immobilized nickel by nitrilotriacetic acid (Ni-NTA) 50% EtOH resin per L of cell culture was used for immobilized metal-ion affinity chromatography (IMAC). Its ethanol was removed of the resin by centrifuging it at 700 rpm for 2 minutes. This procedure was repeated two times with Binding Buffer to be used later in the purification. Then, the lysate was added to the resin and left to gentle agitation for 50 minutes at 4 ℃. The mixture was centrifuged in Eppendorf 5403 at 500 rpm for 4 minutes. After removal of supernatant, Binding Buffer was added to the resin, left to gentle agitation at 4 ℃ for 5 minutes and then centrifuged in the same conditions before. This addition of Binding Buffer was repeated until 50 mL were in contact with the resin. The resin was then transferred to a BIO-RAD column, which was cleaned by successive washings with 60 mL of Wash 10 mM Imidazole solution per L of cell culture, of 30 mL of Wash 13 mM Imidazole solution per L of cell

culture and of 20 mL of Wash 15 mM Imidazole solution per L of cell culture. The elution of the protein attached to the resin was performed with 10 mL of 250 mM Imidazole solution per L of cell culture.

3.1.3.2 For GST-p16

A GSTrap HP column was used which was inserted in a ÄKTA FPLC chromatography system. The chromatography system was controlled using the UNICORN software, the solutions were passed through a 0.45 µm filter before use and pressure tests were done before placing the column in order to determine the maximum pressure of the system that the chromatography column could support, as explained in page 10 of its system manual. Finally, after chromatography, the column was cleaned with distilled water and then stored in 20% (v/v) ethanol.

Cleaning of a used GSTrap HP column was performed with 2 column volumes of guanidine hydrochloride, immediately followed by 5 column volumes of Binding Buffer or PBS, and/or with 4 column volumes of 70% ethanol immediately followed by 5 column volumes of Binding Buffer or PBS, as said in GSTrap HP instructions. The flows used for loading the lysate, washing the column and eluting the target protein were, unless the pressure prevent it, the ones recommended by GSTrap instructions. The buffers used for loading the protein were standard Binding buffer or PBS as recommended by the manufacturer. After use, the column was cleaned with 20% ethanol and stored at 4 ℃.

3.1.3.2.1 Eluting GST-p16

The elution buffers used were various but contained a glutathione concentration of at least 30 mM, three times above what is recommended in GSTrap HP instructions.

3.1.3.2.2 Eluting cut p16

Cut p16 was eluted with various buffers containing biotinylated thrombin. Afterwards, a buffer containing a glutathione concentration of at least of 30 mM was used to remove GST tag or even uncut GST-p16 off the column.

3.1.4 Polishing Purification

A column of Sephacryl S-100 HiPrep 16/60 was used to further purify p16, equilibrated with HEPES buffer. After use, the column was cleaned and stored in 20% ethanol. Loading of samples onto the column and their preparation were done following manufacturer's recommendations.

3.1.5 Thrombin cleavage and capture

Biotinylated thrombin was added to target protein, either bound to the column or eluted, in a quantity that assured complete cut based on a reported activity of 1 U/mg of target protein in

Novagen Thrombin Kit instructions, for 8-16 h at room temperature. Occasionally, thrombin cleavage buffer from Novagen was added to improve cleavage conditions. When the cleavage was on-column, the column was sealed with the top cap and the stop plug.

After protein cleavage, biotinylated thrombin was captured by adding Streptavidin Agarose in a ration of 16 µL per unit of biotinylated thrombin. The resin was then removed by placing the mixture (sample and resin) into a centrifugal filter with a microporous membrane and centrifugating it at 500 xg for 5 minutes.

3.1.6 Dialysis of protein samples

For doing the NMR experiment, the protein should be dissolved in low salt buffer. Thus, the exchange buffer was NMR Buffer (see 8.1.3.6) 0.45 µm filtered and either the dialysis was done in a Spectra/Por® Dialysis Membrane MWCO: 6-8000 Da membrane bag or in a Micro Dialysis Capsule QuixSep capsules. In the former, the protein sample was inserted in a dialysis membrane bag with 2 cm per mL of protein sample and was then dialysed in 100x protein sample volume NMR Buffer for 1 hour and finally in 200x overnight. In the later, the dialysis time was 30 min per 0.1 mL of sample.

3.1.7 Concentration of protein samples

When using the centrifugal filters of Amicon® Ultra MWCO 10,000 Da, the filters were firstly washed two times (one for each membranes placed perpendicularly to the radial direction), and with distilled water by centrifuging in Eppendorf 5403 centrifuge at 4760 rpm for 4 minutes in order to remove the trace amounts of glycerine. After discarding the permeate, the protein sample is applied and centrifuged once at 4000 g for an amount of time depending on the wanted final volume, determined by consulting the tables in the product manual.

3.1.8 Protein Concentration Tests

The protein concentration was analyzed by using UV spectroscopy set to a wavelength of 280 nm, and sometimes also 260 nm to check for the presence of nucleic acids impurities by calculating the absorbance ratio at those wavelengths.

The blank was always, unless said, the buffer solution where the protein was dissolved. In addition, the stability of the blank was checked sometimes by waiting some seconds when using Ultrospec 2000 or by performing various measurements when using Nanodrop (see section 8.1.1).

Similarly, the most important protein samples were measured several times. And sometimes, their value of absorbance was immediately converted to mass concentration by using the Lambert-Beer Law (Equation 1 and Table 4).

$$A = \varepsilon \,.\, l \,.\, \frac{C_M}{MM}$$ **Equation 1**

Table 4 – Lambert-Beer Factors values for different conditions. The values in the table for extinction coefficient assume that all cysteines are reduced. The values in parenthesis for extinction coefficient and molecular weight are for p16 protein after cleavage of its tag. Values taken from Expasy ProtParam bioinformatic tool.

Lambert-Beer Factors		
	Ultrospec 2000	Nanodrop
$\boldsymbol{l}$	1 cm	1 mm
	His-p16 (of pET-28b)	GST-p16 (of pGET-4T1)
$\boldsymbol{\varepsilon}$ ($M^{-1}cm^{-1}$)	12 490 (12 490)	55 350 (12 490)
$\boldsymbol{MM}$ ($g.mol^{-1}$)	16 236.3 (14 364.3)	41 059.4 (14 853.9)

By using Adobe Photoshop CS4 version 11.0 software, the signals of protein bands in a SDS-PAGE gel were quantified. The parameter for transforming the signal intensity into protein concentration was determined by a protein test sample that was both measured in UV spectrometer and visualized in SDS-PAGE.

3.1.9 Protein Purity Tests

In order to test if the protein sample had the required purity for performing NMR experiments, SDS-PAGE and MALDI-TOF tests were performed. MALDI-TOF was used as an expedite test that determines the molecular weights of the protein sample. It was performed on a Bruker Autoflex III smartbeam instrument. The sample was previously desalted by a filter resin. The resin was washed with 50% acetonitrile (ACN) in water and equilibrated with 0.1% trifluoracetic (TFA) in water it. 10 µL of sample were loaded into the resin and the latter was again washed with 0.1% TFA in 5% ACN in water. Protein sample was eluted with 5 µL of a 0.1% TFA in 50-95% ACN in water and was mixed with the matrix solution (20 mg/ml sinapinic acid in acetonitrile: water: TFA (50:50:0.1)) and applied onto the sinapinic acid in ethanol matrix. The laser was fired 2000 times at the matrix which vaporises the samples. Analysis of Mass spectra was acquired using Flex control and Flex analysis software and a graph was plotted against intensity and m/z ratio.

SDS-PAGE was used for assessing the purity and concentration evolution of the protein sample through the purification procedures. Samples for loading in polyacrylamide gels followed the ratio recommended by the manufacturer of 2.5 µL LDS and 1 µL Reducing Agent per 10 µL of sample, and were then heated at 70 ℃ for 10 minutes. The samples were subjected to electrophoresis at 200 V for 40 minutes. Subsequently, gels were Coomassie-stained following manufacturer's instructions.

3.1.10 Liofilization

Before liofilization, the protein sample was frozen with liquid nitrogen. After some hours of liofilization, the tube containing the protein powder was kept at -20 °C.

3.1.11 NMR experiments

Proton NMR experiments were performed with 10% D_2O on a Bruker 600 MHz equipped with a cryoprobe, in order to evaluate if the protein was folded.

4 Results

4.1 *E.coli* BL21 (DE3) transformed with pET28b plasmid: p16 tagged with histidine tail

4.1.1 Expression of his-p16

The expression systems used for his-p16 – *E.coli* BL21 (DE3) transformed with pET28b plasmid – showed to be stably transformed. Moreover, the lysates confirmed the presence of his-p16, as supported by the comparison between non-induced cells and induced cells in Figure 16.

However, the production of his-p16 in lysates was very low: $3.8 \pm 0{,}5 \times 10^{-5}$ mmol per liter of culture. This is below the minimum threshold of protein quantity to satisfy the requirements for sufficient NMR signal (chapter 2).

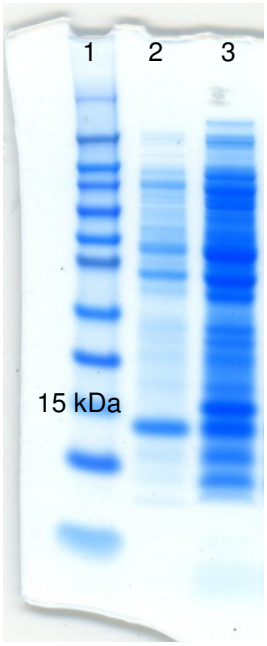

Figure 16 – SDS-PAGE of fractions of different steps during an experiment for purification of his-p16. Lane 1 – MW standards. Lane 2 – Cells not induced by IPTG. Lane 3 – Cells induced by IPTG.

4.1.2 Methods for obtaining p16

To obtain p16 without his tag, there are two options: eluting his-p16 with imidazole and then cut the histidine tail with thrombin, or cleaving his-p16 on-column with thrombin and elute p16. However, after elution there was some variance in the order and type of procedures taken to prepare his-p16/p16 to NMR experiments which is illustrated in Figure 18.

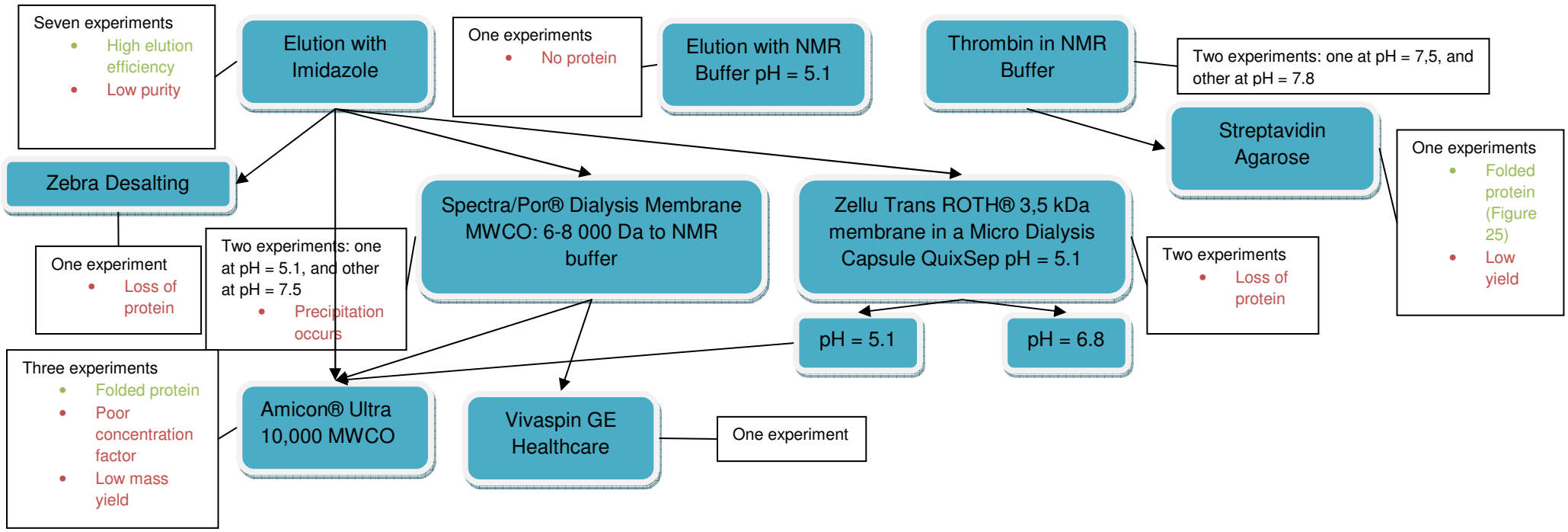

Figure 17 – Schematic representation of the all experiments performed concerning purification and concentration of the protein prior to NMR experiment, for p16 tagged with an histidine tail.

4.1.3 Elution of his-p16 by IMAC

The followed procedure for capture purification failed in providing pure protein, as Figure 18 shows clearly in lanes 5 and 6, where his-p16 was eluted jointly with unselected impurities, usually accounting for approximately 40% of total protein mass. The cause for the impurity was not inefficient washing of the resin, as the respective lanes (lanes 7,8 and 9 in Figure 18) show barely any impurities.

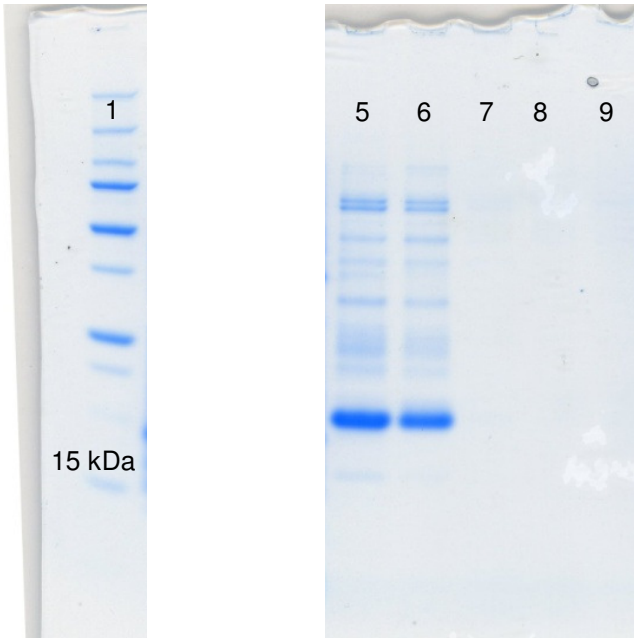

Figure 18 – SDS-PAGE of fractions of different steps during an experiment for purification of his-p16. Eluted his-p16 is contaminated with many impurities. Lane 1 – MW standards. Lanes 5 and 6 – Successive eluted samples by the imidazole solution. Lanes 7, 8 and 9 – Fractions of the first, second and third washings of the resin, respectively.

Thus, instead of eluting with imidazole, it has been attempted to elute his-p16 through an acid NMR buffer (pH = 5.1). This procedure failed to elute his-p16, because of negative results from electrophoresis and UV spectroscopy.

In addition to the observed associated impurities, on average, only 15% of his-p16 produced in cell culture (whose quantity was determined through the lysate) was obtained after elution with imidazole, representing a total quantity of his-p16 in the eluted fractions of 0.18 ± 0.06 mg, almost ten times lesser than needed (chapter 2). The affinity resin Ni-NTA did not casue the low mass yield of his-p16 elution, as its maximum binding capacity was never reached – 10 mg/mL.

4.1.4 Dialysis and Concentration

The dialysis of protein samples caused protein loss, although the purity remained sensibly the same. The dialysis performed transferred the protein solution from Binding buffer to NMR buffer, resulting that less than 60% of the protein was recovered, though with its purity

unaffected (Figure 19). Moreover, in Figure 20 two eluted fractions of the same column were analogously dialyzed (in Micro Dialysis Capsule QuixSep) except with different capsules sizes: the first with 5 mL and the other one with 1 mL. The loss of the protein in the bigger capsule was much greater than the smaller one.

Figure 19 – SDS-PAGE of fractions of different steps during an experiment for purification of his-p16. Lane 1 – MW standards. Lane 5 – His-p16 eluted by imidazole solution. Lane 6 – Eluted fraction after dialysis with NMR buffer at pH = 5.1.

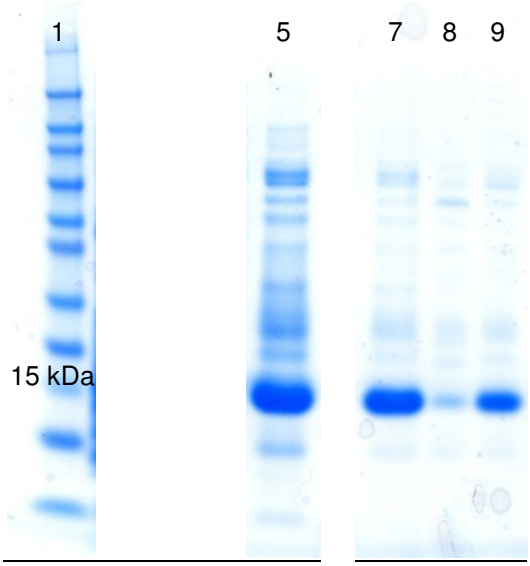

Figure 20 – SDS-PAGE of fractions of different steps during an experiment for purification of his-p16. Lane 1 – MW standards. Lanes 5 and 7 – Two fractions of his-p16 eluted with imidazole. Lanes 8 and 9 – Dialysis of the eluted fractions of lanes 5 and 7, respectively.

Target protein was not concentrated proportionally to volume reduction. For example, a protein sample had its volume reduced 12 times but the protein concentration did not even doubled – the final concentration was 0.42 ± 0.02 mg/mL which is 7 times lower than the

minimum required for NMR experiments. In addition to this, gels (Figure 21) show that there was his-p16 protein loss. This may be due to the fact that the concentration device used is incompatible with imidazole, which has the same chemical group as the histidines in his-p16.

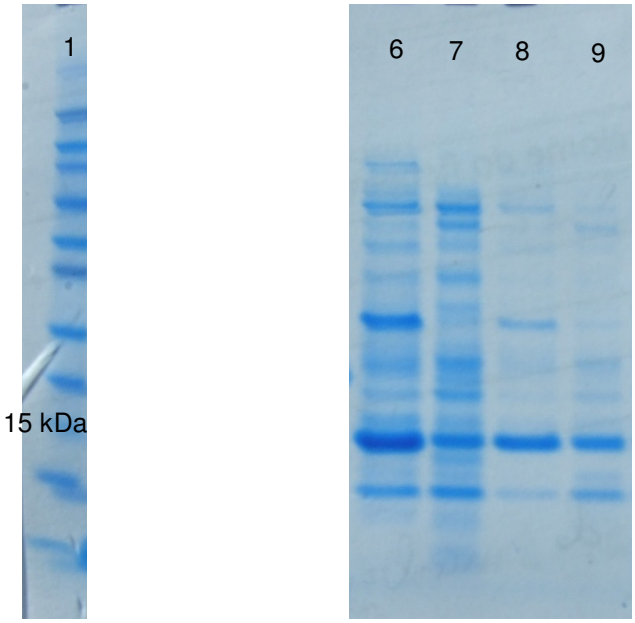

Figure 21 – SDS-PAGE of fractions of different steps during an experiment for purification of his-p16. Lane 1 – MW standards. Lanes 6 and 7 – First eluted sample by the imidazole solution before and after concentration. Lanes 8 and 9 – Second eluted sample by the imidazole solution before and after concentration.

4.1.5 Cleavage of eluted his-p16 by thrombin

Cutting the tail of histidines off the target protein p16 was proven to be possible, by MALDI-TOF (Figure 22) and by SDS-PAGE (Figure 23). Figure 22 shows the MS results for cut p16 eluted by NMR buffer after incubation with thrombin while lane 14 in Figure 23 shows simultaneously uncut and cut protein, by eluting the resin with 250 mM Imidazole. The eluted was incubated with thrombin for some hours, but at suboptimal conditions which explain the incomplete cut of his-p16 (approximately only 50% was cut). Thus, the lane presents three bands: one of p16 protein, other of uncut his-p16 and a last one of histidines tail.

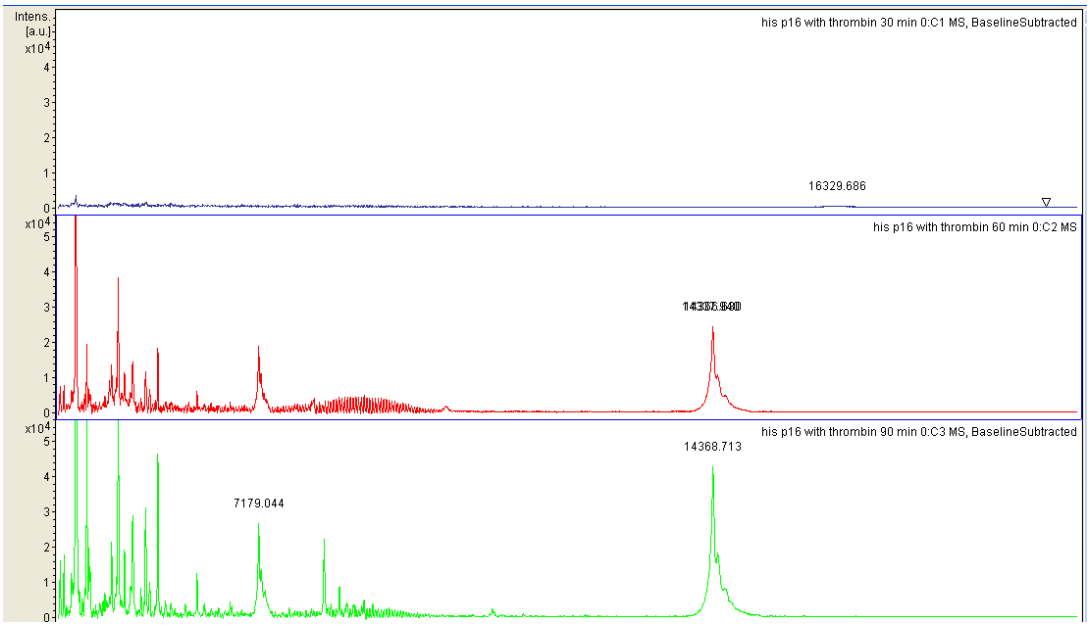

Figure 22 - MS spectras after 30, 60 and 90 minutes of on-column cleavage by incubation with thrombin in NMR buffer.

Figure 23 - SDS-PAGE of fractions of different steps during an experiment for purification of his-p16. Lane 1 – MW standards. Lane 14 – Fraction eluted by the imidazole solution after addition of thrombin.

4.1.6 Folding assessment of his-p16 and cut p16 by 1H NMR

Despite the low concentration and therefore the impossibility to perform proper NMR experiments, it has been decided to verify if his-p16 and p16 (after being cut from the histidine tail) were folded. Figure 24 and Figure 25 are 1H NMR spectra suggesting that his-p16 and p16 are folded in NMR buffer at pH = 5.1 and pH = 7.5, respectively. The signals are spread from -1

or 0 ppm til 9.5/9 ppm meaning that the protein is folded. Otherwise all signals would be dispersed from 7 to 8 ppm and up to approximately 1 ppm.

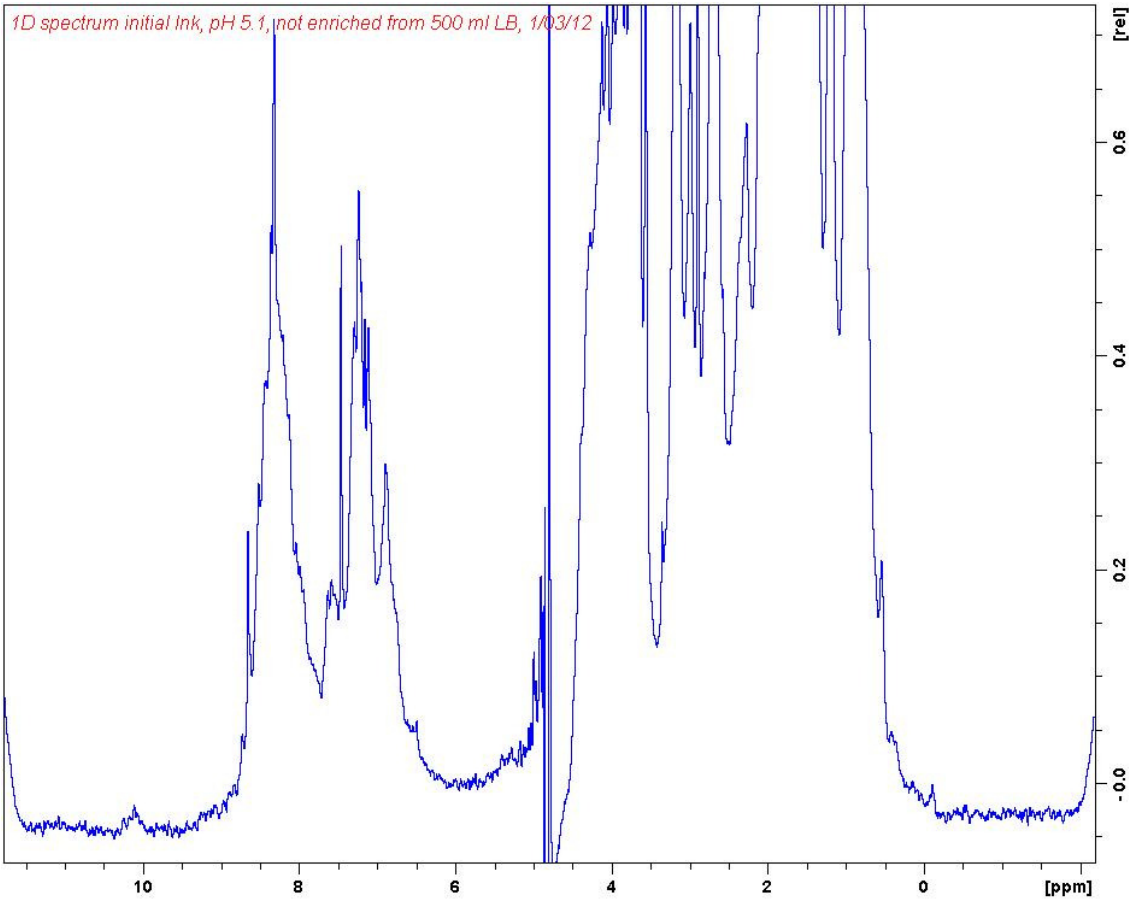

Figure 24 – Proton NMR spectra for his-p16 in NMR Buffer at pH 5.1.

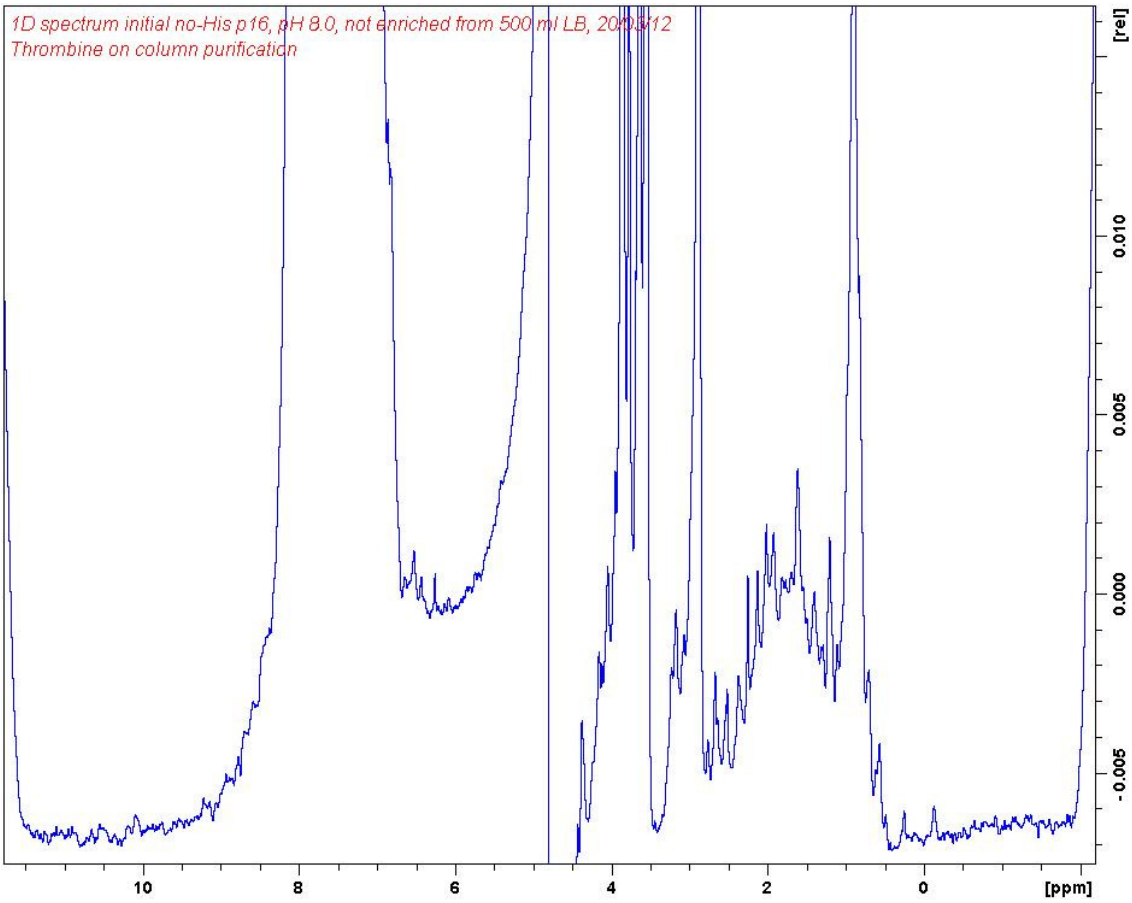

Figure 25 – Proton NMR spectra of p16 protein in NMR Buffer at pH 7.5 with contamination of a 250 mM Imidazole solution.

4.2 *E.coli* BL21 (DE3) transformed with pGEX-4T1 plasmid: p16 tagged with GST protein

4.2.1 Expression of GST-p16

The production of his-p16 was insufficient to satisfy the requirements exposed in chapter 0, so another expression system was attempted: *E.coli* BL21 (DE3) transformed with pGEX-4T1 plasmid where p16 is tagged with GST protein. The expression systems showed to be stably transformed.

The production of GST-p16 in lysates was around 1.4 ± 0.6 x 10^{-4} mmol per liter of culture which means that a production of one liter surpasses the minimum threshold of protein quantity to satisfy the requirements for sufficient NMR signal (chapter 2).

4.2.2 Affinity chromatography of GST-p16

In the first place, an increase of protein purity was observed when compared with the previous purification procedure for his-p16. For this purpose, it is sufficient to compare Figure 21 with Figure 32.

Secondly, using a FPLC system to perform protein purification gave a much better control on real time than the chromatography procedure used with his-p16, as a UV spectrometer installed in it enables measurement of eluted protein concentration online (Figure 26).

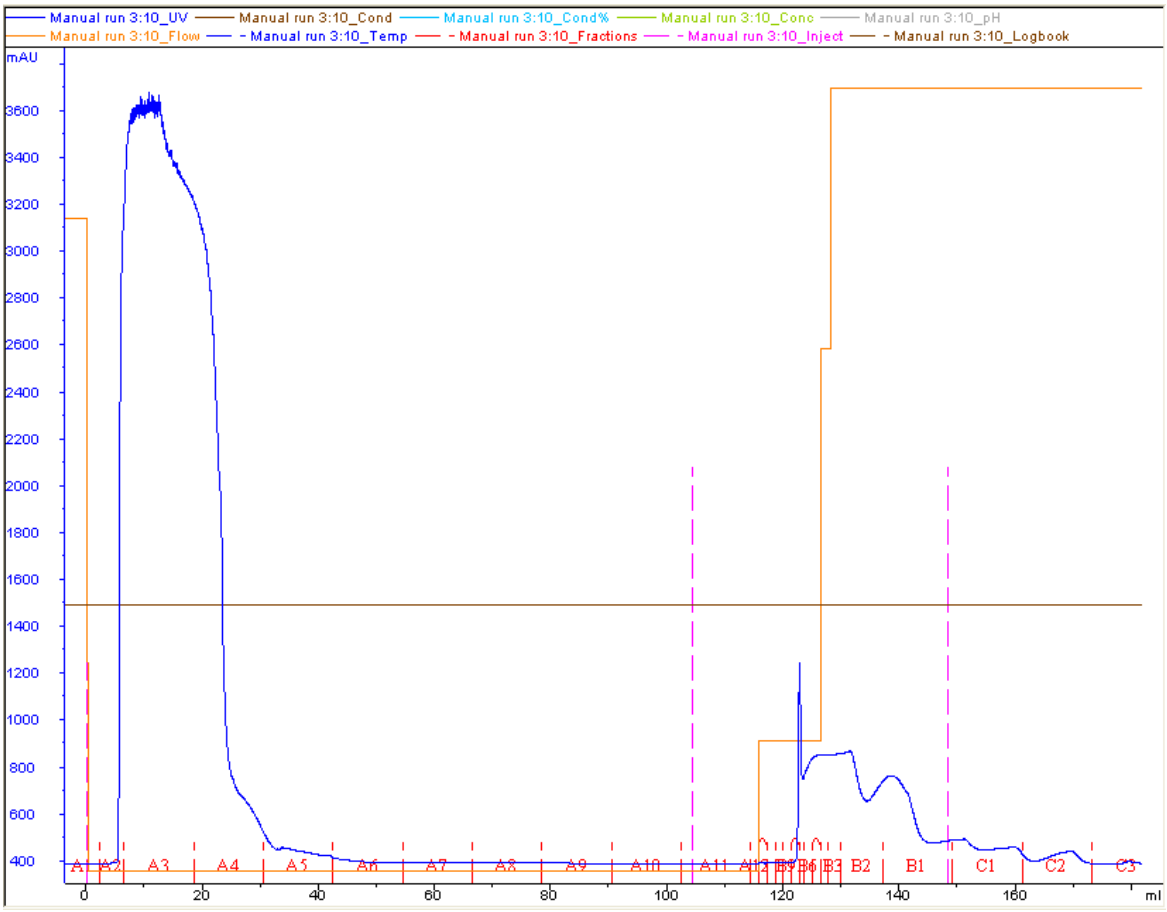

Figure 26 – A chromatogram depicting the purification profile of GST-p16 protein. A FPLC system has the advantage of allowing the user to watch on real time the different phases of the purification: flow-through of unbound proteins in lysate (fractions A2- A5), following return of absorbance baseline (280 nm) (fractions A6-A10) and the eluted GST-p16 by the elution solution (fractions A11-B11).

To obtain p16 without GST tag, two options were used: eluting GST-p16 with glutathione (to compete for the column) and then cut GST tag with thrombin, or cleaving GST-p16 on-column by thrombin and elute p16. For each of those two elution possibilities, there was some variance in choosing the binding buffers, as an attempt to maximize the binding and elution of the target protein in the GSTrap column (Table 5).

The binding efficiency of GST-p16 to column in PBS buffer was on average 58.9 ± 0.7 %, as measured by bands of non-binding fraction in SDS-PAGE (Figure 27 and Figure 28).

Table 5 - Summary of the experiments of purification of p16 concerning the use of buffers for Binding and Elution in GSTrap (marked by "x"). The first three columns are for elution buffer that elute GST-p16 while the others elute p16.

Elution Buffer / Binding Buffer	50 mM Glutathione in Binding Buffer	30 mM Glutathione in Binding Buffer	30 mM Glutathione in HEPES Buffer, pH = 7.5	0,8% Biotinylated Thrombin in HEPES Buffer, pH = 7.5	0,8% Biotinylated Thrombin in PBS Buffer, pH = 7.5	0,4% Biotinylated Thrombin in Binding Buffer	0,15% Biotinylated Thrombin in NMR Buffer
300 mM NaCl, 20 mM Na_2HPO_4, pH = 7.3	x						
NMR Buffer, pH = 7.3						x	
300 mM NaCl, 50 mM Na_2HPO_4, 2 mM Beta-Mercaptoethanol, 2 mM EDTA, pH = 7.3		X					x
PBS Buffer, pH = 7.3			2 x	x	x		

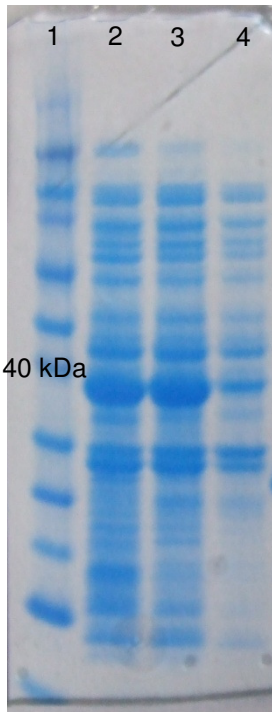

Figure 27 - SDS-PAGE of a purification of GST-p16 on-column cleavage by thrombin and using HEPES as elution buffer. Lane 1 – MW standards. Lane 2 – Lysate. Lane 3 – Lysate after nucleic acid removal. Lane 4 – Non-binding fraction.

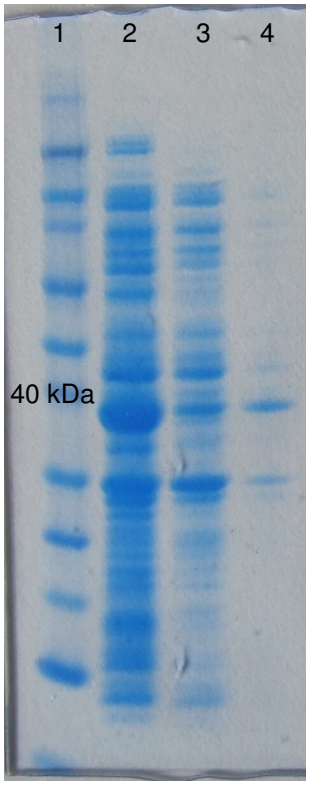

Figure 28 – SDS-PAGE of a purification of GST-p16 on-column cleavage by thrombin and using PBS as elution buffer. Lane 1 – MW standards. Lane 2 – Cells induced by IPTG. Lanes 3 and 4 – Non-binding fraction.

4.2.2.1 *Cleavage of eluted GST-p16 by thrombin*

The option of eluting GST-p16 failed quantitatively, despite giving a Gaussian-type elution profile - Figure 29. With respect to the GST-p16 in the lysate, on average only 6% of molar equivalent was eluted. Consequently, the mass of p16 eluted was far below lower (less than 20% of molar quantity) what is required for NMR experiments (chapter 2). The glutathione concentration used was of 30 mM, three times more than the suggested in GSTrap HP instruction, therefore glutathione is not to the cause for low elution yield. The ionic strength of the buffer used for elution does not seem to be an important variable, as a HEPES buffer and Binding buffer have been used with no striking result differences. It should be noted that HEPES buffer has been previously used for NMR experiments of human p16, so in principle should be adequate to nmr p16. Moreover, HEPES buffer has an ionic strength within the same range of the standard elution buffer recommended in GSTrap HP instructions (10 mM Tris-HCl, pH = 8.0).

Elution of GST-p16 also resulted in very low purity, around 50% - Figure 30. This is inadequate for NMR experiments.

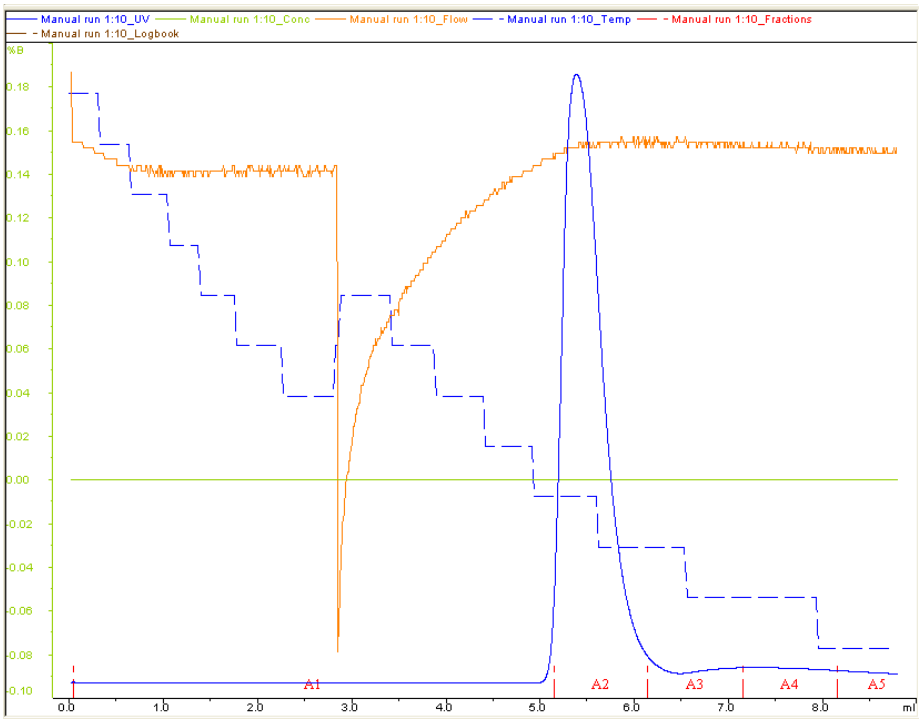

Figure 29 – Chromatogram zoomed on the peak of protein eluted with glutathione.

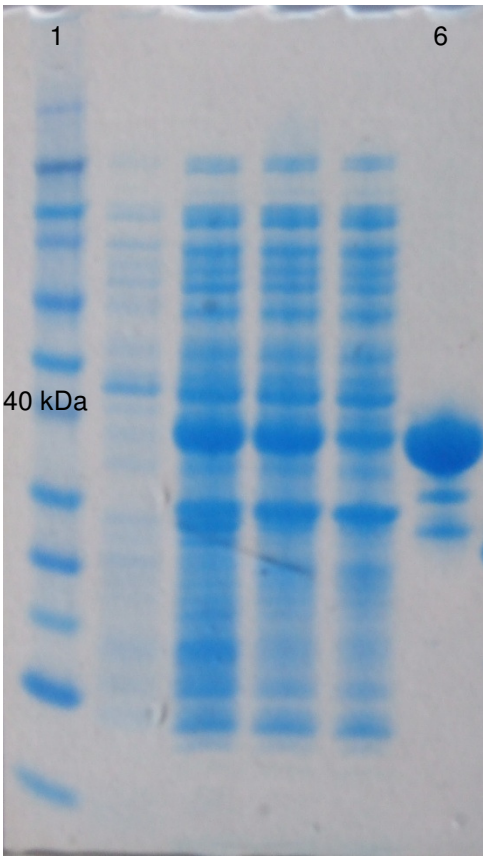

Figure 30 - SDS-PAGE showing the impurities eluted along with GST-p16 eluted by the imidazole solution. Lane 1 – MW standards. Lane 6 – Eluted fraction of GST-p16 by the imidazole solution.

After elution, thrombin was added to the protein eluted sample. The cleavage of GST took place at room temperature as the results show clearly that: 1) the intensity of the MS peak of target protein (15 kDa) increases consistently with incubation time, as it happens for GST peak (Figure 31), and 2) a SDS-PAGE (Figure 32) that shows two bands at 26 and 15 kDa and no band at 41 kDa. As matter of fact, in all experiments where SDS-PAGE was performed, analysis on the bands confers a cleavage efficiency above 97%.

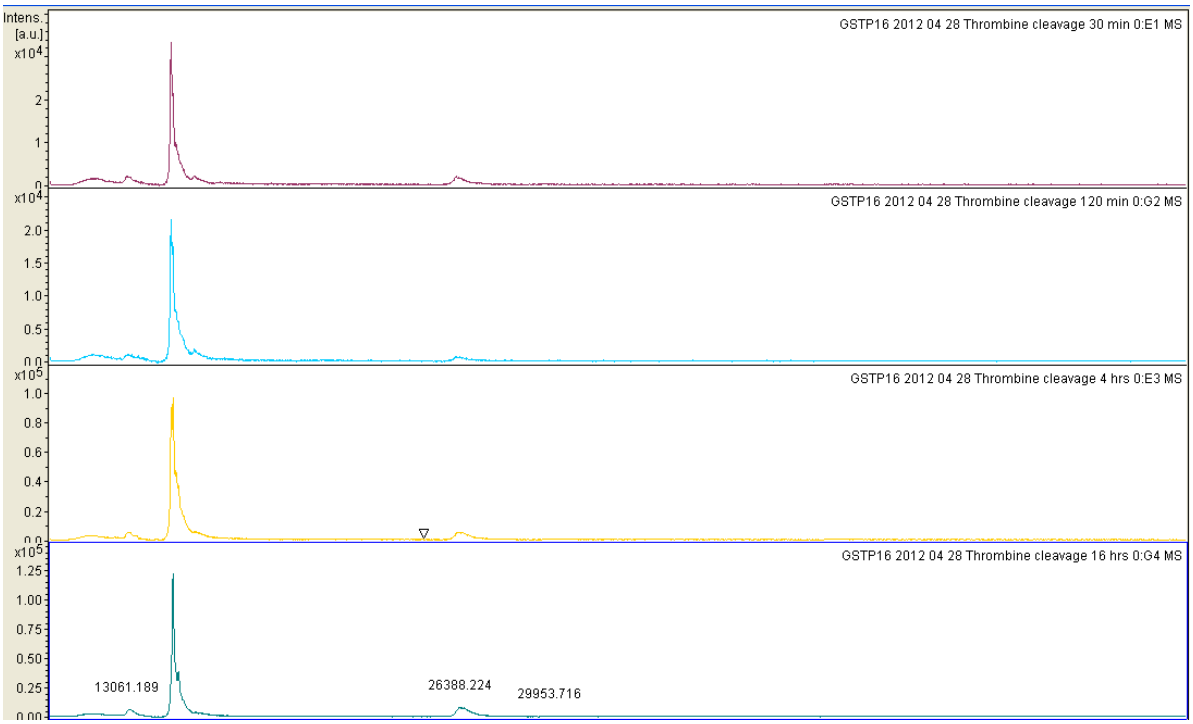

Figure 31 - MALDI-TOF of several samples of GST-p16 with thrombin along incubation time. The signal for cut p16 increases during incubation time.

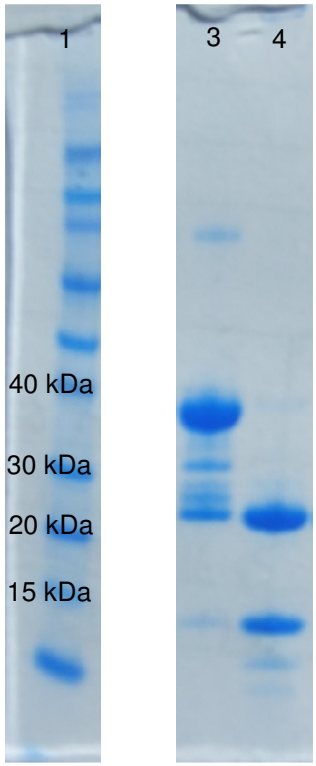

Figure 32 – SDS-PAGE of fractions of different steps during an experiment for purification of p16. Lane 1 – MW standards. Lane 3 – Eluted fraction by glutathione after concentration and desalting. Lane 4 – Eluted protein sample after thrombin cleavage, mixed with Streptavidin Agarose.

However, it seems that cut p16 is not soluble in solution. A trail of precipitate was left in place after thrombin cleavage - Figure 33.

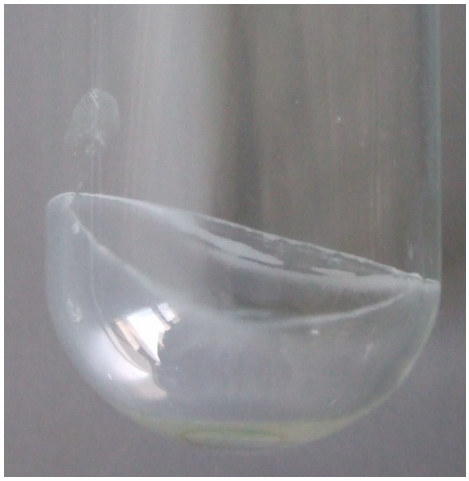

Figure 33 - Sonication tube used for thrombin cleavage of GST-p16, with white precipitate.

4.2.2.2 GST-p16 on-column cleavage by thrombin

GST-p16 bound to column was cut by thrombin in a HEPES, PBS or NMR buffer which resulted in a elution with a Gaussian-like peak (Figure 34). Afterwards, the remaining GST tag and uncut GST-p16 were eluted by glutathione - Figure 35.

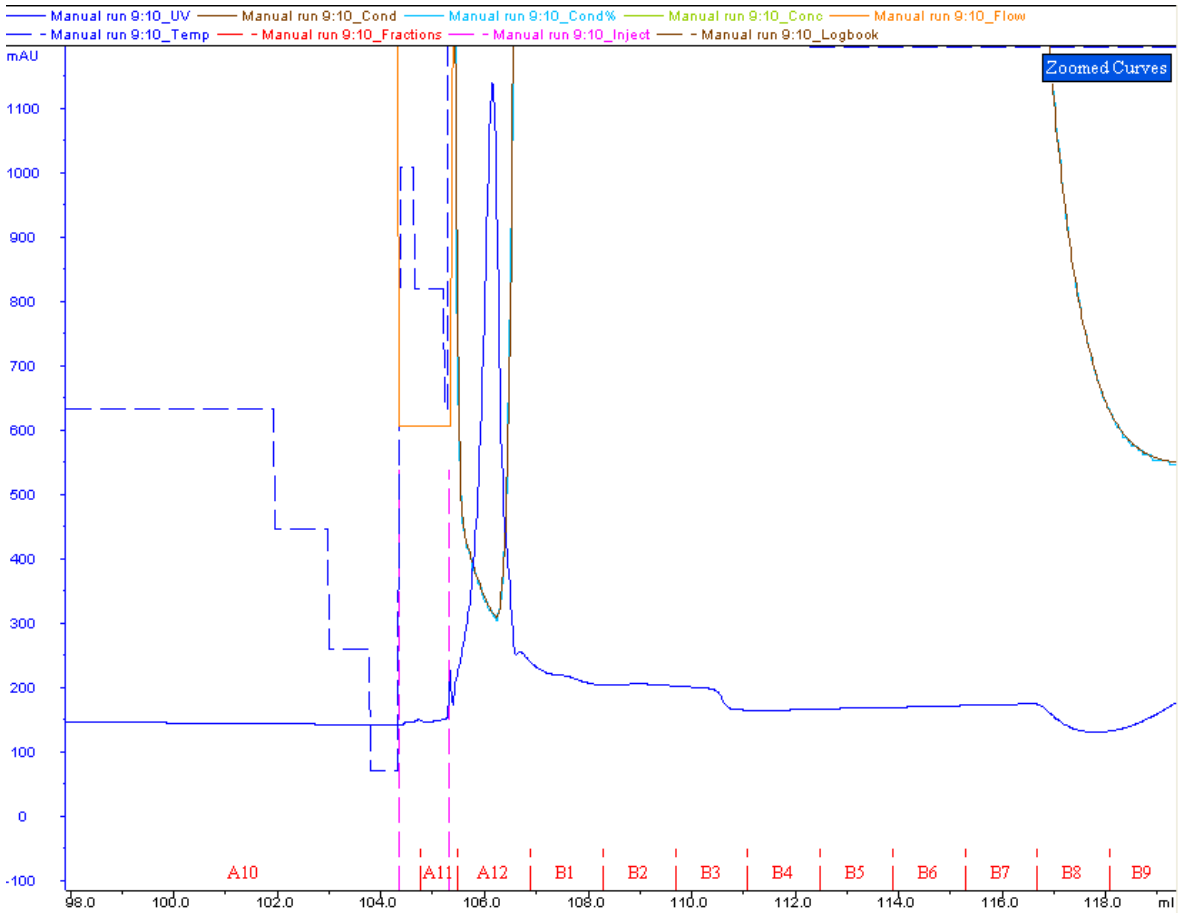

Figure 34 – Chromatogram zoomed in the eluted cut p16 peak. This elution was performed with NMR buffer.

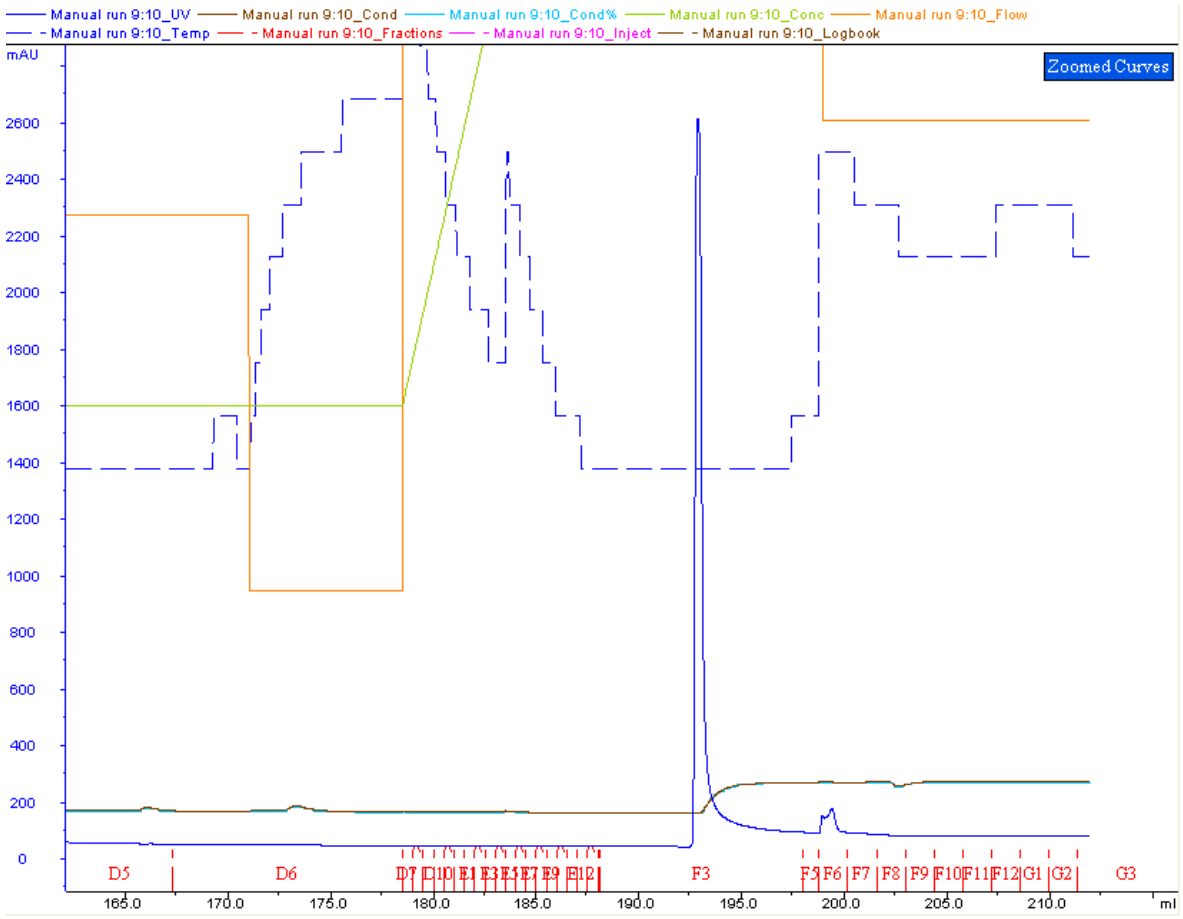

Figure 35 – Chromatogram zoomed in GST/GST-p16 eluted peak, after cut p16 was eluted with NMR buffer.

SDS-PAGE bands of eluted cut p16 indicate that only an average of 5% of molar quantity corresponding GST-p16 in the lysate was eluted. The quantity of p16 eluted was insufficient to perform NMR analysis.

In PBS elution buffer, GSP-p16 (41 kDa) bands are almost completely absent in eluted fraction. As Figure 36 shows, elution with glutathione after elution of cut p16 did only remove GST tag and no GST-p16.

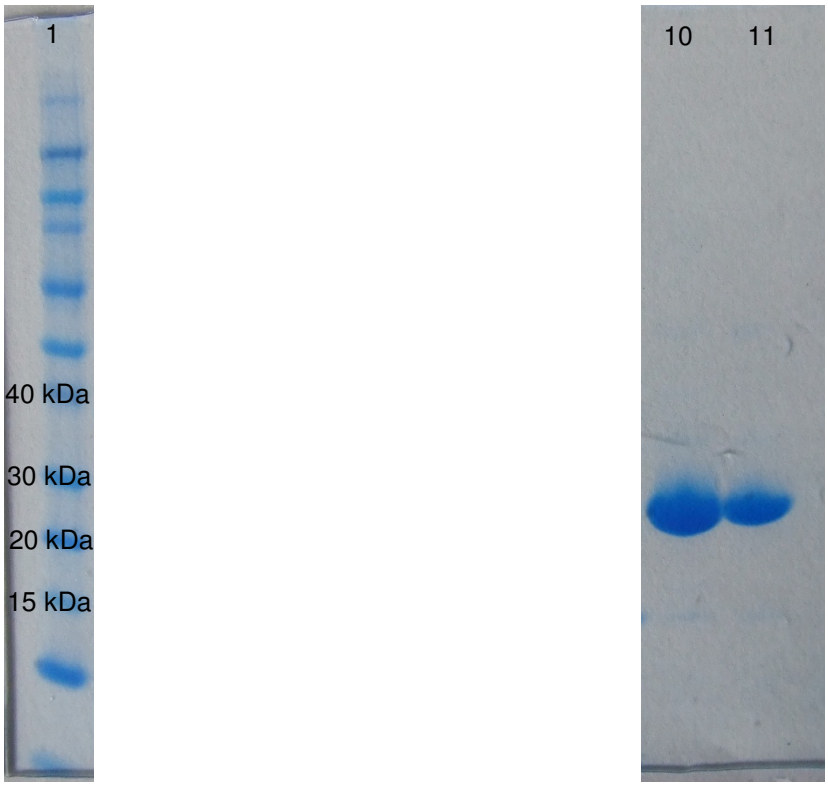

Figure 36 – SDS-PAGE of a purification of GST-p16 on-column cleavage by thrombin and using PBS as elution buffer. Lane 1 – MW standards. Lanes 10 and 11 – Eluted fractions by glutathione after elution of cut p16.

When using solely HEPES as elution buffer, GST tag and uncut GST-p16 were removed off the column - Figure 37. In addition, the chromatogram in Figure 38 shows a broad non Gaussian-like peak reflecting the stringency of HEPES buffer and therefore its inadequacy as elution buffer. The same phenomena happened when using NMR buffer, where cut p16 was eluted along with GST tag - Figure 39 and Figure 40. In contrast, when PBS buffer was used the amount of GST tag in the eluted was much lower - Figure 36.

At the time of removing GST tag (and uncut GST-p16) off the column by using glutathione in HEPES and NMR buffer, cut p16 was present again in the eluted fraction - Figure 37 and Figure 41 – unlike what happened with PBS buffer - Figure 36.

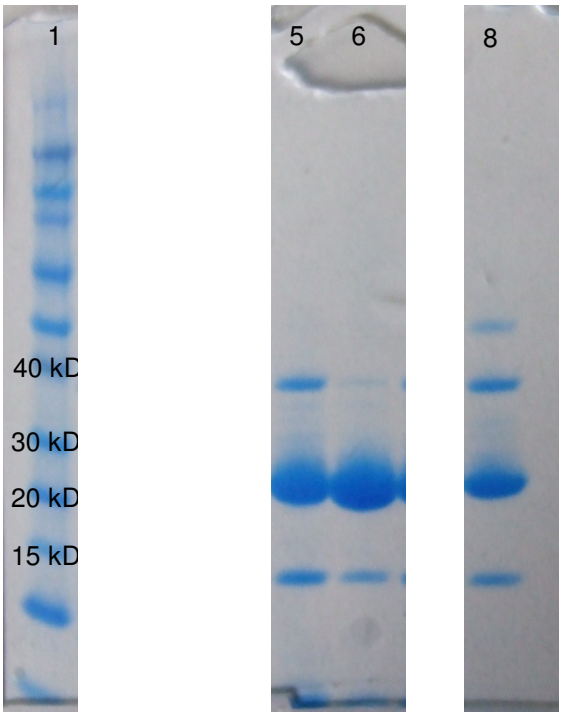

Figure 37 - SDS-PAGE of a purification of GST-p16 on-column cleavage by thrombin and using HEPES as elution buffer. Lane 1 – MW standards. Lanes 5 and 6 – Eluted fractions of cut p16 by HEPES buffer. Lane 8 – Eluted fraction by glutathione after elution of cut p16.

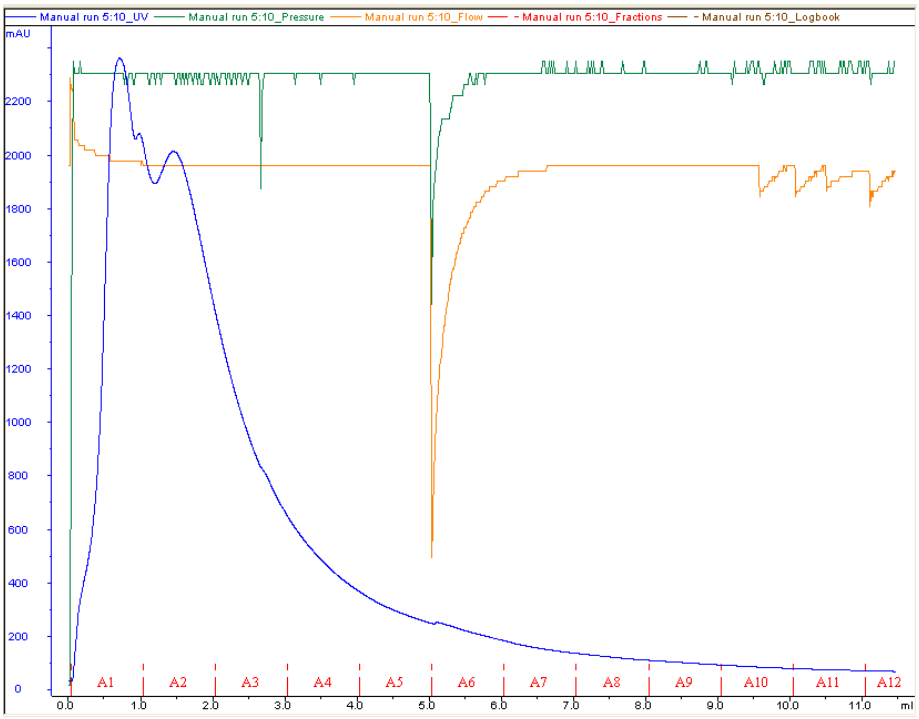

Figure 38 – Chromatogram of elution with HEPES buffer after on-column cleavage by thrombin.

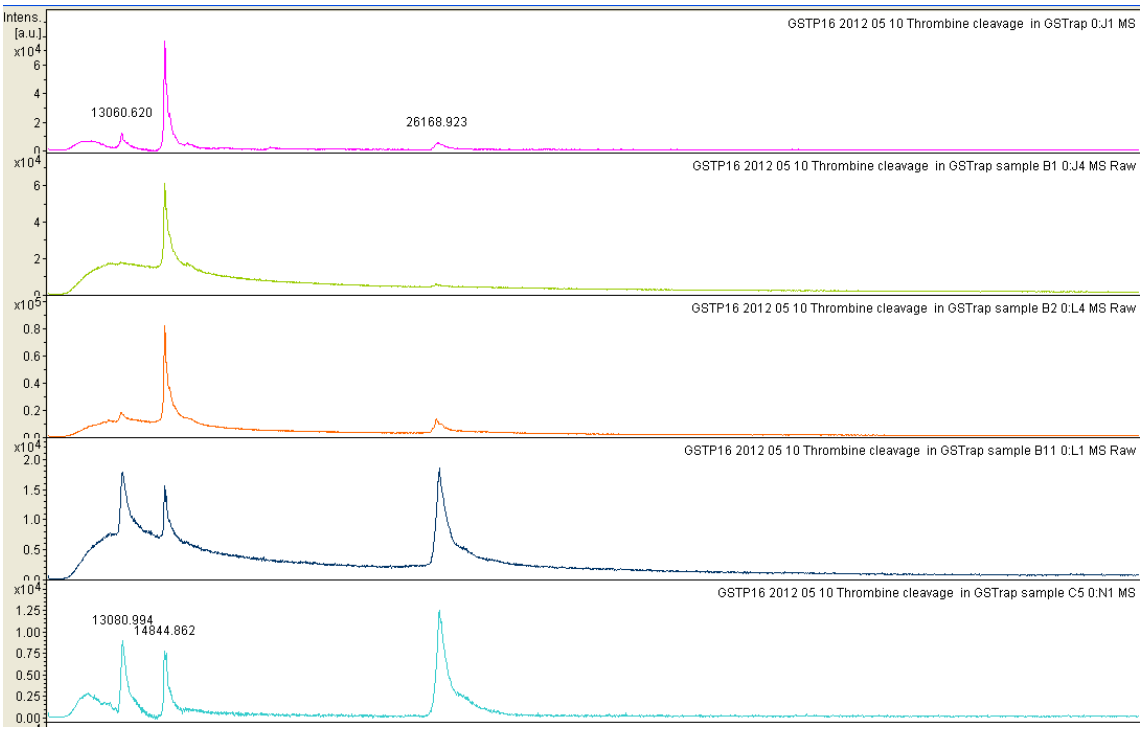

Figure 39 – MALDI-TOF of GSTrap elution samples with NMR buffer after on-column cleavage by thrombin. With exception of the first two samples, all eluted samples contain GST tag (26 kDa) just by eluting with NMR buffer.

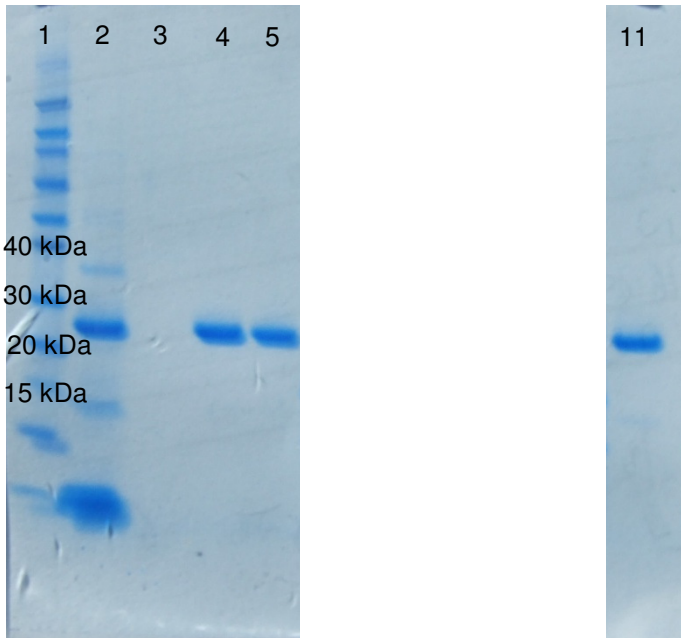

Figure 40 – SDS-PAGE of a purification of GST-p16 with on-column cleavage by thrombin and using NMR buffer as elution buffer. Lane 1 – MW standards. Lanes 2-5 – Eluted fractions of cut p16 by NMR buffer. Lane 11 – Some eluted fractions of cut p16 by NMR buffer joined together.

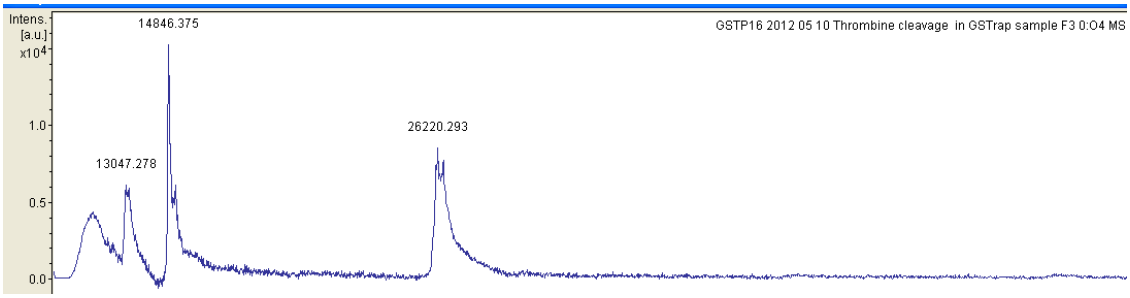

Figure 41 - MALDI-TOF of a fraction eluted with glutathione to remove GST tag and uncut GST-p16, after on-column cleavage by thrombin and elution with NMR buffer. The eluted fraction contains cut p16 besides the expected GST tag.

4.2.3 Gel Filtration of p16 by S-100 column

Cut p16 was not effectively eluted of S-100 column, as Figure 42 and Figure 43 clearly show. This phenomenon happened exclusively to p16 because GST tag was completely eluted, with a Gaussian-like peak in Figure 44. As a matter of fact, the peaks that appeared during the cleaning of the column (Figure 45) after usage might be due to removal of precipitated p16.

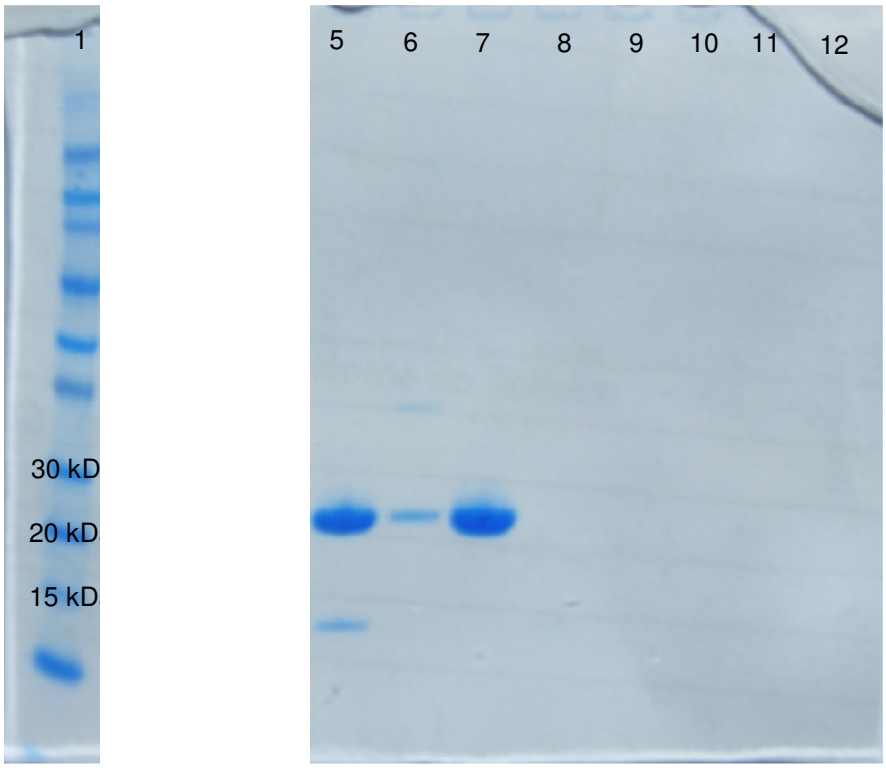

Figure 42 – SDS-PAGE of a purification of p16, with cleavage after elution and load of cut p16 and GST into a S-100 column. Lane 1 – MW standards. Lane 5 – Cut p16 and GST tag after thrombin removal and before gel filtration. Lanes 6-12 – Gel filtrations fractions.

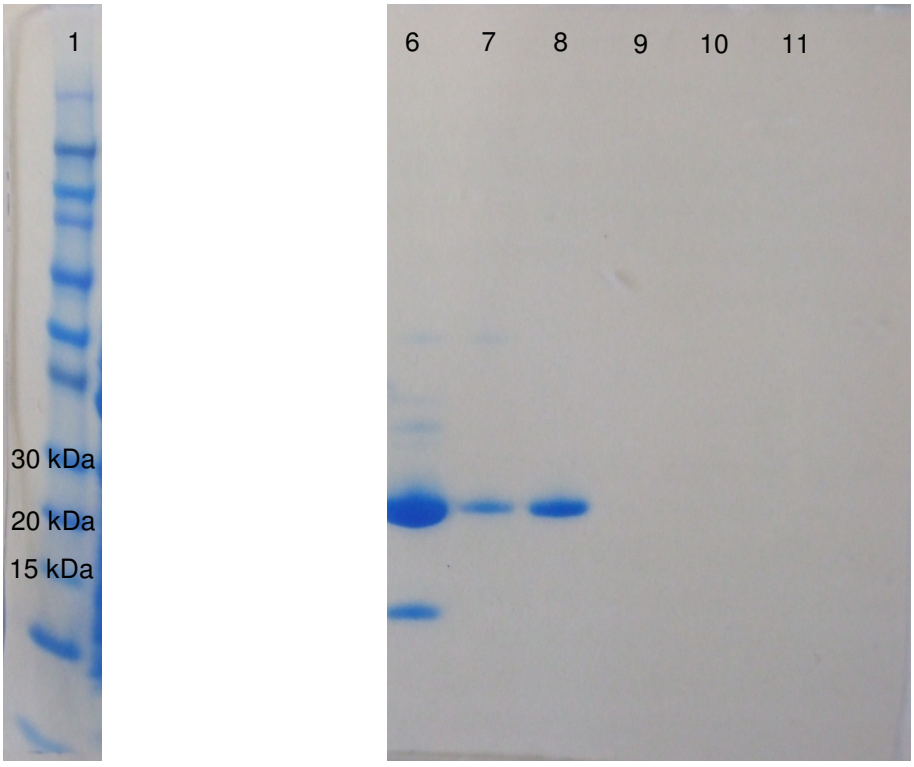

Figure 43 – SDS-PAGE of a purification of p16, with cleavage after elution and load of cut p16 and GST into a S-100 column. Lane 1 – MW standards. Lane 6 – Cut p16 and GST tag after cleavage by thrombin and before gel filtration. Lanes 7-11 – Gel filtrations fractions.

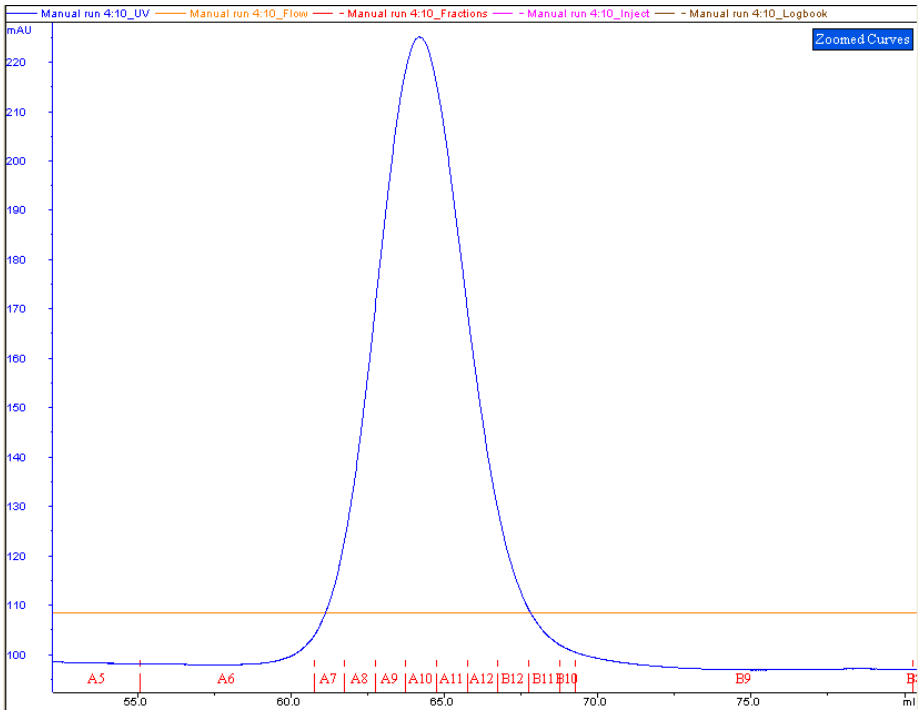

Figure 44 – Chromatogram of a gel filtration profile whose sole peak consists of GST tag.

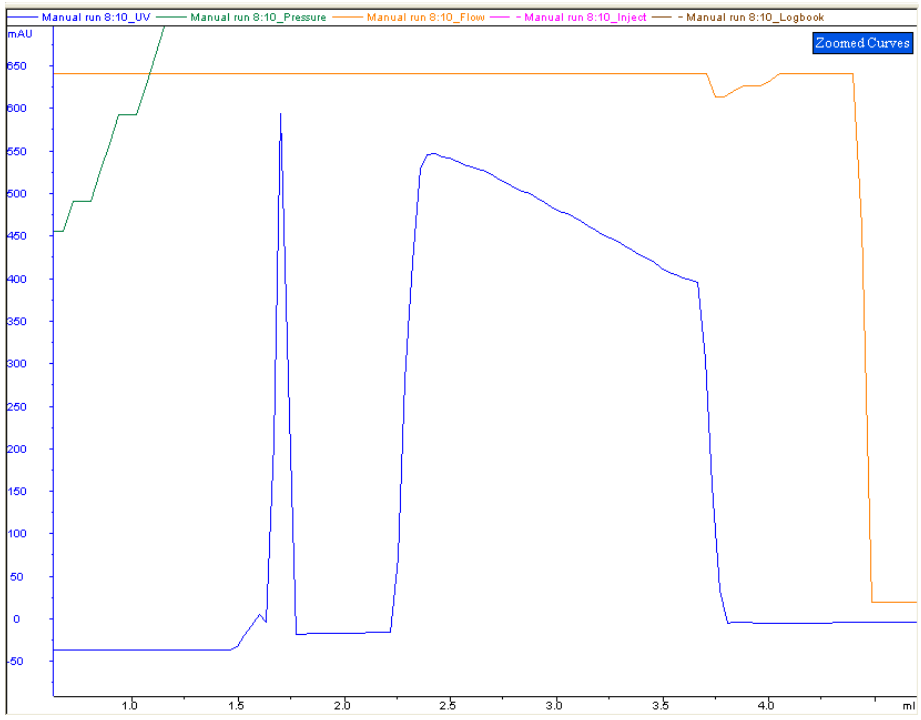

Figure 45 – Chromatogram of S-100 cleaning with 20% ethanol after usage. The absorbance signal at 280 nm is in blue.

4.2.4 Protein desalting

The results of desalting of protein were variable, sometimes the operation was harmless to protein mass (Figure 46) and sometimes the operation caused loss of protein (Figure 47). There are two operation variables possibly associated with the two different outcomes. One of the variable is the type of buffer used for dialysis. When the sample was dialysed against NMR

buffer, there was no loss of protein. On the other hand, when the sample was dialysed against HEPES buffer, there was on average a 40% of target protein loss. In both cases, the quantity of protein before desalting were on the same range: 2 ± 1 % of what would be needed for performing NMR experiments, therefore much below the solubility limit of p16 referred in the literature. Moreover, both NMR and HEPES contain reducing agents – β-mercaptoethanol and dithiothreitol, respectively – and so the formation of aggregates due to disulfide bonds formation was minimized.

The other variable was the protein. In the experiments where there was no loss of protein, the protein was in its original state – GST-p16 – while the ones where loss of protein occurred were performed with cut protein, i.e. – p16, GST tag and uncut GST-p16.

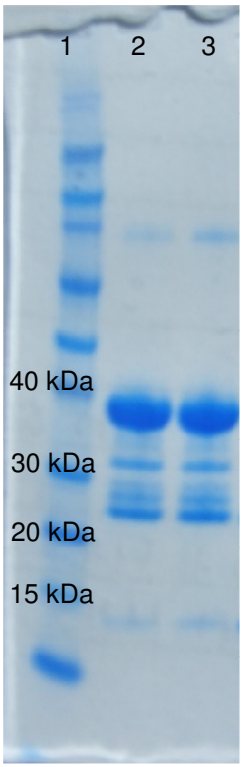

Figure 46 – SDS-PAGE showing desalting of GST-p16 in which the sample was dialysed against NMR buffer. Lane 1 – MW standards. Lanes 2 and 3 – Eluted fraction of GST-p16 by the glutathione solution before and after desalting.

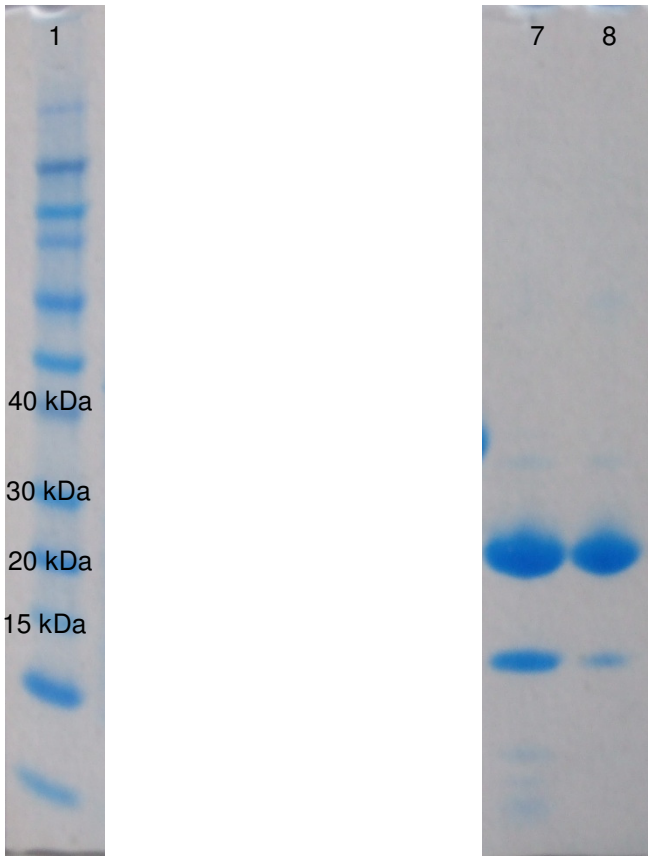

Figure 47 - SDS-PAGE showing desalting of p16 in which the sample was dialysed against HEPES buffer. Lane 1 – MW standards. Lane 7 – Eluted fraction after on-column cleavage by thrombin. Lane 8 – Eluted fraction after on-column cleavage by thrombin and after desalting.

4.2.5 Thrombin removal by Streptavidin Agarose

Using Streptavidin agarose to remove thrombin from protein solution revealed to be a cause of loss of p16, of about 60% in mass on average, as can be concluded by analyzing Figure 48 and Figure 49.

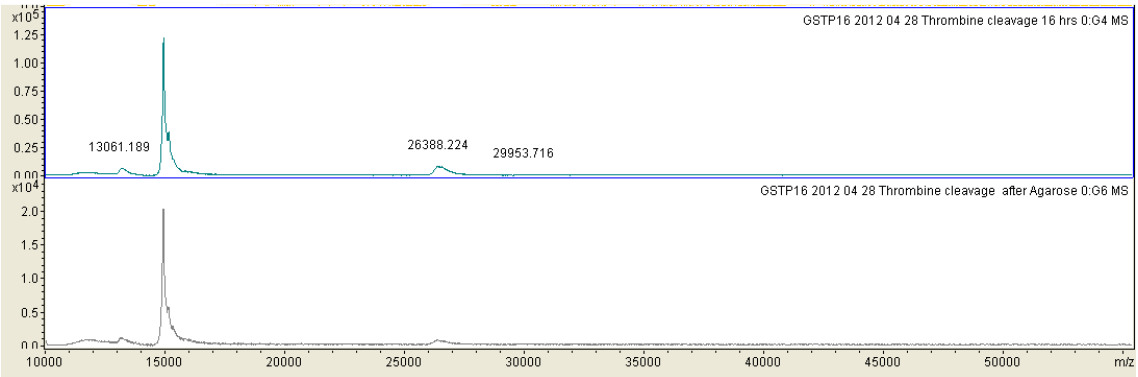

Figure 48 - MALDI-TOF of a sample of cut p16 before and after thrombin removal with Streptavidin Agarose. The signal for cut p16 decreases one order of magnitude when Streptavidin Agarose is removed from sample by filtration.

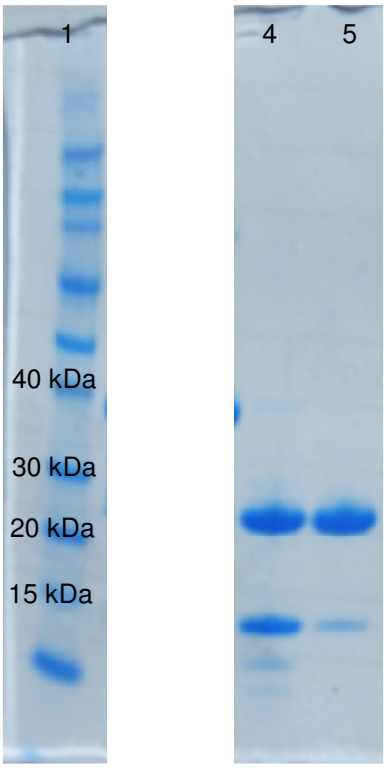

Figure 49 – SDS-PAGE of a purification experiment of GST-p16. After removal of thrombin by streptavidin agarose, the majority of p16 was lost. Lane 1 – MW standards. Lane 4 – Protein sample after thrombin cleavage, with Strepatvidin Agarose added to it. Lane 5 – Protein sample after removal of Streptavidin Agarose.

5 Discussion

Naked mole-rat $p16^{INK4a}$ has a truncated structure, unique in the animal kingdom, and metabolically behaves in a different manner than the human and mouse orthologous. Moreover, $p16^{INK4a}$ is a tumor suppressor in a animal which has no reported cases of spontaneous cancer. Thus, there is high interest in knowing better the protein. Structure determination is usually a requisite for deeper understanding of a protein. The objective of the work was to purify and concentrate the protein in order to perform NMR experiments.

The purification procedure adopted was affinity purification, as that normally results in high yields (often over 90%) making it economically viable, besides allowing a reduction in chromatographic steps which reduces the number of unit operations needed and operation time (Arnau et al, 2006). The first choice of affinity tag for $p16^{INK4a}$ was the standard his-tag – it was used for more than 60% of structural studies (Arnau et al, 2006).

Thus, p16 was produced fused to a histidine tail through the expression system *E.coli* BL21 (DE3) transformed with pET28b plasmid. The mass production of p16 was insufficient to satisfy the requirements for NMR experiments: a protein sample with a minimum volume of 0.4 mL at a minimum protein concentration of 0.2 mM. The elution yield was also very low, though p16 does not contain metal ions neither has many cysteines and histidines in its aminoacid sequence which would result in unwanted protein binding during IMAC purification (Westra et al, 2001). The histidine tail was N-terminal which is associated with better binding properties than C-terminal (Goel et al, 2000), therefore changing the position is not recommended. Suboptimal elution of his-tag proteins is also sometimes due to the low imidazole concentrations used (Arnau et al, 2006), but the concentration used for p16 equals the recommended by the manufacturer of Ni-NTA tough concentrations of four times more have been used in the literature (Arnau et al, 2006). Furthermore, the purification procedure for his-p16 resulted in very low purity which was prohibitive for NMR experiments. This is a likely consequence of the use of β-mercaptoethanol (Lichty et al, 2005) in binding, washing and elution solutions. To reduce this problem, higher concentrations of imidazole in washing should have been used (Westra et al, 2001), at least as high as recommended in the manufacturer's instructions and as used in the literature (Arnau et al, 2006) (20 mM of imidazole). Alternatively, if p16 had been eluted after thrombin cleavage in a subtractive manner, the purity would probably have been higher as his-tag process enzymes and other unspecific binders would have been retained in the column (Arnau et al, 2006). Consequently, another expression system was attempted: *E.coli* BL21 (DE3) transformed with pGEX-4T1 plasmid, with p16 tagged with GST protein. The production of this expression system was able to accomplish the requirements for NMR experiments.

However, the followed protein purification strategy appropriate for structure determination projects (Dian et al, 2002) did not work out well, for both situations of cleavage after elution of GST-p16 and of on-column cleavage of GST-p16. In the former, the elution yield of GST-p16 was strangely still very low (less than 10%), and almost as low as elution of cut p16. Despite generally following manufacturer's instructions, the cause may be non optimized conditions in the elution column, because precipitation of protein due to excessive concentration was

determined to be unlikely, either in cell lysate or in loading into GSTrap. Assuming p16 production is similar to the what is reported in the literature for p15 in LB medium (2-4 mg/L) and a final cell density of 3 g of cells per 500 mL of culture, and knowing that cell suspension buffer is in a proportion of 5 mL to every gram of cell, the protein concentration yields 0.133 mg of p16 per mL of lysate. This is equivalent to 9×10^{-3} mM of p16 in the cell lysate, well below the solubility limit referred by the literature (0.2-0.4 mM). In addition, the maximum binding capacity of column is 10 mg of tagged protein per mL of resin which means a maximum concentration of 0.25 mM of p16 in GSTrap column, still below the solubility limit referred by the literature. Unless that concentration is already enough to induce precipitation of nmr's p16 and therefore diminish drastically the yield of elution. In that case, the purification of p16 in the lysate should be divided in several batches to prevent full use of column binding capacity. Despite non-binding fractions revealed that 40% of GST-p16 flowed through the column (maybe due to the slow binding kinetics of GST to glutathione-Sepharose resin (Hunt, 2005) which required several times more GSH resin than expected (Wilharm et al, 2003) and (Mercado-Pimentel et al, 2002)), the remaining protein in the column was still in much higher quantity than the eluted protein. Thus, the low elution yield could be due to instability of GST-p16, as it frequently happens. As a matter of fact, overwhelming evidence has shown that GST is, at best, a poor solubility enhancer (Esposito and Chatterjee, 2006). This instability means possible precipitation of the fusion protein in the column. Hydrophobic binding is not a possible explanation because low ionic buffers have been tried for elution and the elution yield kept low.

The low purity of GST-p16 eluted is also relevant – the target protein was on average eluted with 50% of not wanted protein. The impurity could be due to intrinsic susceptibility of GST-p16 to bacterial proteases resulting in the co-purification of some amounts of cleaved GST. To tackle this purity issue, a dual affinity tag approach (also referred to as tandem affinity purification, TAP) could be adopted. In this approach, p16 would be fused in-frame with an N- or C-terminal tag comprised of two affinity tags surrounding an endoprotease cleavage site. By applying two different affinity purification regimes sequentially, highly purified p16 would be obtained (Cass et al, 2005).

Thrombin cleavage of GST-p16 was effective with a yield of 97% and showing specific cleavage. This means that thrombin has good access to the cleavage site and there is no need for improvements in designing the aminoacid sequence between GST and p16 sequences. However, after cutting the GST tag off eluted GST-p16, the protein seemed to be very susceptible to precipitation, despite the many different procedures and conditions tried. It seems that pH and ionic strength of buffers are factors without any influence in preventing precipitation, because precipitation has occurred with values above and below the isoelectric point of the protein, and with high and low ionic strengths, in desalting, in thrombin removal by streptavidin agarose and in gel filtration. Thus, it is assumed that the precipitation is due to the inherent instability of p16 after cleavage. As a matter of fact, that might explain why GST-p16 is considerably more stable than his-p16, as GST accounts for 64% of fused protein versus 12% of the histidine tail. Similarly to p16, his-p16 also precipitated in buffers with pH above and

below its isoelectric point, and ionic strength of buffers of elution did not seem to influence yield in a significant extent, during dialysis and concentration operations. Moreover, the increase of stability of target protein by GST may be the reason for higher concentrations of protein in the lysates than those obtained with his-p16.

Analogously, elution of p16 after on-column cleavage was also very low in yield, despite seeming that the thrombin cleavage was complete. This is one more evidence supporting the major hypothesis that cut p16 is inherently unstable, aggregating in the column after on-column cleavage. Once more, the pH and ionic strengths of the buffers did not improve the operation yield, meaning that hydrophobic binding was not the cause of low yield (in fact, p16 grand average of hydropathy is negative). Moreover, eluting p16 in buffers with low ionic strength resulted in co-elution of uncut GST-p16 and GST tag, which could have been prevented if an auxiliary GSTrap FF column had been mounted downstream of the primary column. Besides capturing any unbound GST-p16 and GST-tag contaminating the eluting p16, the column would have served two additional purposes: first it would have buffered any start-up flow pressure peaks that may alter the baseline established prior to the cleavage reaction following proteolytic incubation; and second, it would have created a small void volume allowing for precise fraction collecting (Dian et al, 2002).

Formation of inclusion bodies (aggregates of incorrectly folded protein) because of overexpression of protein (Sørensen & Mortensen, 2005) would explain the low elution yield of the fusion protein, and is somewhat consistent with the fact that the fusion tag is solely cut completely by thrombin after elution of soluble fusion protein. Simultaneously, the observed instability of p16 may be due to the fact that a large fraction of tagged p16 is not properly folded, and after cleavage naturally precipitates (Esposito and Chatterjee, 2006), making of the used GST-p16 a soluble inclusion body (Sørensen & Mortensen, 2005). Consequently, this precipitation could lead to aggregate formation, supported by the fact that protein losses of p16 and his-p16 were not accompanied by loss of purity. This aggregation hypothesis could be tested by the method described by (Kapust and Waught, 2000) in MBP fusion proteins and applying in the expression system used in this work, and concerning the non-native p16 folding hypothesis, p16 activity could be tested by interaction with its binding partners. Nevertheless, it could have been tried performing fusion protein purification under denaturing conditions, after a inclusion body purification and/or preparation of cleared lysates under denaturing conditions, as described in Ni-NTA his-bind® resins' manual of Novagen. After elution of the inclusion bodies, fusion protein could be refolded through a number of strategies available and subsequently had its fusion tag removed. However, nothing guarantees that p16 will be refolded or even folded in its native conformation, as for most proteins, spontaneous *in vitro* refolding of denatured proteins to native forms is unsuccessful (Cole, 1996).

The difficulties on purifying soluble p16 are not, however, unusual. The soluble expression of heterologous proteins in *Escherichia coli* still remains a serious bottleneck in protein production (Esposito and Chatterjee, 2006), as protein expression in that host often results in insoluble and/or nonfunctional proteins (Sørensen & Mortensen, 2005). The basis for

this deficiency is unclear and probably multifactorial (Cole, 1996). Nevertheless, changing the expression host is not really an option, as *E. coli* remains the standard in the field (Esposito and Chatterjee, 2006), a position held because of its association with hardiness, rapid growth, ease of genetic manipulation and simplicity (Cole, 1996).

Therefore, the first standard procedure for improving solubility is by engineering the target protein by: changing the fusion tag or by stabilizing mRNA (Sørensen & Mortensen, 2005). Concretely, the fusion tag of p16 could be switched to MBP (maltose-biding protein) which would very likely enhance the solubility of tagged p16, assist correct protein folding and may prevent proteolytic degradation *in vivo* (Kapust and Waugh, 1999) and (Sørensen & Mortensen, 2005). This would increase protein production and hopefully, assuming an elution yield comparable to that of GST-p16, would mean more eluted p16 and therefore increased concentration for NMR experiments. Another potent solubility enhancing fusion partner is NusA, and both MBP and NusA are especially suited for the expression of proteins prone to form inclusion bodies (Sørensen & Mortensen, 2005). However, these fusion partners are relatively large so the protein yield would be eventually better if a small and highly soluble partner were used, like the N-terminal fragment of translation initiation factor IF2. Thus, the amount of energy required to obtain a certain number of molecules would be reduced, and steric hindrance would be diminished (Sørensen & Mortensen, 2005). Fortunately, due to the fact that modern cloning technologies have made parallel cloning into multiple vectors with different tags a routine matter, several different solubility-tagged protein expression vectors can be compared in a single experiment (Dyson et al, 2004) and thus the protein tag conferring the biggest solubility can be then chosen. This is the best option, as the enhancing of solubility by a tag depends heavily on the target protein (Esposito and Chatterjee, 2006).

Subsequent strategies would be reducing culture temperatures, as the aggregation reaction is determined by the strong temperature dependence of hydrophobic interactions (Kiefhaber et al, 1991). In addition to this, reducing temperatures eliminates partially the action of heat shock proteases that are induced under overexpression conditions (Chesshyre and Hipkiss, 1989). On the other hand, low temperature means lower transcription and translation rates. To overcome this, systems with different promoters have been designed and optimized (Sørensen & Mortensen, 2005). Another strategy that is relevant to the problem of p16, is the co-overexpression of molecular chaperones for the prevention of aggregation and inclusion body formation (Sørensen & Mortensen, 2005) and (Cole, 1996). Some reports also suggest that the co-expression of interaction partners potentially favours *in vivo* solubility of target proteins (Sørensen & Mortensen, 2005). In the case of p16, the interaction partners to be used could be CDK4 and CDK6.

As a last resource, instead of removing the fusion tag off p16, one could try to perform NMR experiments of GST-p16 or to perform x-ray crystallography of GST-p16. The former has the advantage that, with the help of software tools, the spectra signals for the already determined structure of GST could be identified. The latter is somehow an emergent approach as a way to crystallize problematic proteins (Smyth et al, 2003), tough only MBP fusion proteins

had its crystal structure determined (in the same range of aminoacid number of p16) (Ke and Wolberger, 2003), (Kobe et al, 1999) and (Liu et al, 2001). Nevertheless, preliminary X-ray diffraction results have also been reported for crystals of GST fusion proteins (Kuge et al, 1997). However, the structures obtained might not be exactly the same as of *in vivo* due to the unpredictable changes that adding a tag may introduce in a protein and its behavior (Arnau et al, 2006) and (Cole, 1996).

6 Conclusion

The production and purification of nmr's p16 with the needed concentration and purity for NMR experiments have failed, when using either polyhistidine tail or GST fusion partners. In both cases, the fusion protein seemed to form insoluble inclusion bodies leading to precipitation and subsequent loss of the protein, aggravated after cleavage of the fusion partner. Thus, alternatives should be adopted in order to solve this bottleneck: use fusion tags known to promote higher solubility and better folding, co-express the protein with chaperones and interaction partners, and experiment other expression conditions such as low temperature during cell culture.

Overall, the far more pronounced instability of nmr's p16 in comparison to the reported for human's p16 might be a consequence of the diminished size of the protein by the absence of the third exon. This structural difference might be related to the metabolic differences of p16 existent between nmr and human.

7 References

Altieri, A. S., & Byrd, R. A. (2004). Automation of NMR structure determination of proteins. *Current Opinion in Structural Biology* (14), pp. 547-553.

Arnau, J., Lauritzen, C., Petersen, G. E., & Pedersen, J. (2006). Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein Expression and Purification* (48), pp. 1-13.

Buffenstein, R. (2008). Negligible senescence in the longest living rodent, the naked mole-rat: insights from a successfully aging species. *Journal of Comparative Physiology B* (178), pp. 439-445.

Buffenstein, R. (2005). The Naked Mole-Rat: A New Long-Living Model for Human Aging Research. *Journal of Gerontology: Biological Sciences , 60* (11), pp. 1369-1377.

Buffenstein, R., Remmen, H., Chaudhiri, A., Pierce, A., & E.M., D. (2006). High oxidative damage levels in the longest -living rodent, the naked mole-rat. (B. P. Limited, Ed.) *Aging Cell* (5), pp. 463-471.

Byeon, I.-J. L., Li, J., Ericson, K., Selby, T. L., Tevelev, A., Kim, H.-J., et al. (1998, February). Tumor Suppressor p16INK4A: Determination of Solution Structure and Analysis of Its Interaction with Cyclin-Dependent Kinase 4. *Molecular Cell , 1*, pp. 421-431.

Byeon, I.-J. L., Tsai, M.-D., Selby, T. L., Li, J., & Yuan, C. (1999). Tumor Suppressor INK4: Comparisons of Conformational Properties Between p16INK4A and p18INK4C. *Journal of Molecular Biology* (294), pp. 201-211.

Byeon, I.-J., Tsai, M.-D., Yuan, C., Selby, T. L., & Li, J. (2000). Tumor Suppressor INK4: Refinement of p16INK4A structure and determination of p15INK4B structure by comparative modeling and NMR data. *Protein Science* (9), pp. 1120-1128.

Cass, B., Pham, P. L., Kamen, A., & Durocher, Y. (2005). Purification of recombinant proteins from mammalian cell culture using a generic double-affinity chromatography scheme. *Protein Expression and Purification* (40), pp. 77-85.

Chesshyre, J. A., & Hipkiss, A. R. (1989). Low temperatures stabilize interferon a-2 against proteolysis in Methylophilus methylotrophus and Escherichia coli. *Applied Microbiology and Biotechnology* (31), pp. 158-162.

Cole, P. A. (1996). Chaperone-assisted protein expression. *Structure* (4), pp. 239-242.

Dian, C., Eshaghi, S., Urbig, T., Seán, M., Heijbel, A., Salbert, G., et al. (2002). Strategies for the purification and on-column cleavage of glutathione-S-transferase fusion target proteins. *Journal of Chromatography B* (769), pp. 133-144.

Dyson, M. R., Shadbolt, S. P., Vincent, K. J., Perera, R. L., & McCafferty, J. (2004). Production of soluble mammalian proteins in Escherichia coli: identification of protein features that correlate with successful expression. *BMC Biotechnology , 4* (32).

Esposito, D., & Chatterjee, D. K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Current Opinion in Biotechnology* (17), pp. 353-358.

Gladyshev, V., Wang, J., Guojie, Z., Park, T., & Sunyaev, S. (2011). Genome Sequencing reveals insights into physiology and longevity of the naked mole rat. *Nature* .

Goel, A., Colcher, D., Koo, J. S., Booth, B. J., Pavlinkova, G., & Batra, S. K. (2000). Relative position of the hexahistidine tag effects binding properties of a tumor-associated single-chain Fv construct. *Biochimica and Biophysica Acta* (1523), pp. 13-20.

Holak, T. A., Engh, R. A., Huber, R., Winoto, A., Fernandez-Catalan, C., & Baumgartner, R. (1998). Structure of human cyclin-dependent kinase inhibitor p19INK4d; comparison to known ankyrin-repeat-containing structures and implications for the dysfunction fo tumor suppressor p16INK4a. *Current Biology* (6), pp. 1279-1290.

Hornsby, P., Buffenstein, R., Wu, Y., Mele, J., & Liang, S. (2010). Resistance to experimental tumorigenesis in cells of a long-lived mammal, the naked mole-rat (Heterocephalus glaber). *Aging Cell* (9), pp. 626-635.

Horton, H. R., Moran, L. A., Scrimgeour, K. G., Perry, M. D., & Rawn, J. D. (2006). *Principles of Biochemistry* (4th Edition ed.). New Jersey, USA: Pearson Prentice Hall.

Hunt, I. (2005). From gene to protein: a review of new and enabling technologies for multi-parallel protein expression. *Protein Expression and Purification* (40), pp. 1-22.
Kapust, R. B., & Waugh, D. S. (2000). Controlled intracellular processing of fusion proteins by TEV protease. *Protein Expression and Purification* (19), pp. 312-318.
Kapust, R. B., & Waugh, D. S. (1999). Eschericihia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which is fused. *Protein Science* (8), pp. 1668-1674.
Ke, A., & Wolberger, C. (2003). Insights into binding cooperativity of MATa1/MATα2 from the crystal structure of a MATa1 homeodomain-maltose binding protein chimera. *Protein Science* (12), pp. 306-312.
Kiefhaber, T., Rudolph, R., Kohler, H. H., & Buchner, J. (1991). Protein aggregation in vitro and in vivo: a quantitative model of the kinetic competition between folding and aggregation. *Biotechnology (NY)* (9), pp. 825-829.
Kobe, B., Center, R. J., Kemp, B. E., & Poumbourios, P. (1999). Crystal structure of human T cell leukemia virus type 1 gp21 ectodomain crystallized as a maltose-binding protein chimera reveals structural evolution of retroviral transmembrane proteins. *Proceedings of the National Academy of Sciences of U.S.A.* (96), pp. 4319-4324.
Kuge, M., Fujii, Y., Shimizu, T., Hirose, F., Matsukage, A., & Hakoshima, T. (1997). Use of a fusion protein to obtain crystals suitable for X-ray analysis: Crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. *Protein Science* (6), pp. 1783-1786.
Li, J., Tsai, M.-D., Selby, T., O'Maille, P., Poi, M.-J., Ericson, K., et al. (1999). Tumor Suppressor INK4: Determination of the Solution Structure of p18INK4C and Demonstration of the Functional Significance of Loops in p18INK4C and p16INK4A. *Biochemistry* (38), pp. 2930-2940.
Li, M. (2010, January). Can We Determine a Protein Structure Quickly? *Journal of Computer Science and Technology , 25* (1), pp. 95-106.
Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., & Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein Expression and Purification* (41), pp. 98-105.
Liu, Y. F., Manna, A., Li, R. G., Martin, W. E., Murphy, R. C., Cheung, A. L., et al. (2001). Crystal structure of the SarR protein from Staphylococcus aureus. *Proceedings of the National Academy of Sciences of U.S.A.* (98), pp. 6877-6882.
Lowe, S. W., & Sherr, C. J. (2003). Tumor Suppression by Ink4a-ARF: progress and puzzles. *Current Opinion in Genetics & Development* (13), pp. 77-83.
Luh, F. Y., Archer, S. J., Domaille, P. J., Smith, B. O., Owen, D., Brotherton, D. H., et al. (1997). Structure of the cyclin-dependent kinase inhibitor p19INK4d. *Nature , 389*.
Mercado-Pimentel, M. E., Jordan, N. C., & Aisemberg, G. O. (2002). Affinity purification of GST fusion proteins for immunohistochemical studies of gene expression. *Protein Expression and Purification* (26), pp. 260-265.
Russo, A., Tong, L., Lee, J., Jeffrey, P., & Pavletich, N. (1998). Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16INK4a. *Nature* (395), pp. 237-243.
Schmidt, E., & Güntert, P. (2012, July 16). A New Algorithm for Reliable and General NMR Resonance Assignment. *Journal of the American Chemical Society* (134), pp. 12817-12829.
Seluanov, A., Gorbunova, V., Catania, K., Mao, Z., Bozzella, M., Feigenson, M., et al. (2009). Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. (E. Nevo, Ed.) *Proceedings of the National Academy of Sciences of U.S.A.*
Serrano, M., Hannon, G. J., & Beach, D. (1993). A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* (366), pp. 704-777.
Sharpless, N. (2005). INK4A/ARF: A multifunctional tumor suppressor locus. *Mutation Research* (576), pp. 22-38.

Smyth, D. R., Mrozkiewicz, M. K., McGrath, W., Listwan, P., & Kobe, B. (2003). Crystal structures of fusion proteins with large-affinity tags. *Protein Science* (12), pp. 1313-1322.
Sørensen, H. P., & Mortensen, K. K. (2005). Soluble expression of recombinant proteins in the cytoplasm of Escherichia coli. *Microbial Cell Factories , I*.
Stoll, V. S., Manohar, A. V., Gillon, W., MacFarlane, E. L., Hynes, R. C., & Pai, E. F. (n.d.). A thioredoxin fusion protein of VanH, a D-lactate dehydrogenase from Enterococcus faecium: Cloning, expression, purification, kinetic analysis, and crystallization. *Protein Science* (7), pp. 1147-1155.
Westra, D. F., Welling, G. W., Koedijk, D. G., Scheffer, A. J., The, T. H., & Welling-Wester, S. (2001). Immobilized metal-ion affinity chromatography purification of histidine-tagged recombinant proteins: a wash step with a low concentration of EDTA. *Journal of Chromatography B: Biomedical Sciences and Applications* (760), pp. 129-136.
Wilharm, G., Neumayer, W., & Heesemann, J. (2003). Recombinant Yersinia enterocolitica YscM1 and YscM2: homodimer formation and susceptibility to thrombin cleavage. *Protein Expression and Purification* (31), pp. 167-172.

8 Annex

8.1 Material

8.1.1 Equipment

Table 6 – Extensive list in alphabetic order of all equipment used for experiments, with respective brands and models.

Type	Brand, Model
Centrifuge	SORVALL® RC 5B Plus
	KUBOTA, 3500
	Eppendorf, Centrifuge 5403
Centrifugal Filters ("Spin-Column")	Amicon® UltraCel 10,000 MWCO
	Amicon® UltraCel 3,000 MWCO
Centrifugal Filter with Microporous Membrane	Ultrafree-MC, Merck Millipore
Centrifugal Rotors	SORVALL® SLA-1500
	SORVALL® SS-34
Chromatography Column	BIO-RAD, Poly-Prep®
	GSTrap HP, GE Healthcare
	Sephacryl S-100 HiPrep 16/60, GE Healthcare
Chromatography System	ÄKTA FPLC, Amersham Pharmacia Biotech
Current Supplier	Invitrogen, ZOOM® Dual Power
Dialysis Membrane	Spectra/Por® Dialysis Membrane MWCO: 6-8000
	ROTH® Zellu Trans MWCO: 3,5 kDa
Dialysis Capsule	ROTH® Micro Dialysis Capsules QuixSep®
Filter tips for protein desalting for MALDI-TOF	Thermo Scientific Pierce C18 tips
Holder for PAGE	Invitrogen, Novex® Mini-Cell
Incubator	Infors HT, Minitron
Magnetic Stirrer	Heidolph, MR 1000
Mass Spectrometry	Bruker, autoflex III smartbeam
Pipettes (0,5-10 µL; 10-100 µL; 100-1000 µL)	Eppendorf Research, PhysioCare concept
Pipette Tips (0.5-10 µL; 10-100 µL; 100-1000 µL)	Eppendorf

Sterile Syringe	BD Discardit™ II
Sonifier	Branson, Sonifier 250
Tubes (15 and 50 mL)	SARSTEDT
UV/Visible Spectrophotometer	Pharmacia Biotech, Ultrospec 2000
	Saveen Werner, NanoDrop® Spectrophotometer ND-100

8.1.2 Solutions and Reagents

Table 7 – Extensive list in alphabetic order of all solutions and reagents used for experiments, with their respective brand.

Name	Brand
Biotinylated Thrombin	Novagen
Ni-NTA His-Bind® Superflow	Novagen
Novex® Sharp Pre-Stained Protein Standard	Invitrogen
NuPAGE® Antioxidant	Invitrogen
NuPAGE® 4-12% Bis-Tris Gel	Invitrogen
NuPAGE® LDS Sample Buffer (4x)	Invitrogen
NuPAGE® MED SDS Running Buffer (20x)	Invitrogen
NuPAGE® Reducing Agent	Invitrogen
SimplyBlue™ SafeStain	Invitrogen
Streptavidin Agarose	Novagen
Thrombin Cleavage Buffer 10x	Novagen

8.1.3 Buffers

8.1.3.1 Basic Buffer 5x

Table 8 – Composition to the buffer that was colloquially called "Basic Buffer" because it was the base solvent of other buffers. The pH of the buffer depended on each experiment, from pH = 5.1 to pH = 7.8, which was manipulated by addition of HCl or NaOH.

Imidazole	25	mM
NaCl	2,5	M
Na_2HPO_4	250	mM
Tween-20	0.01	% (v/v)

8.1.3.2 Cell Suspension Buffer, or "Lysis buffer"

Table 9 – Composition of Cell Suspension Buffer, used to suspend the cells before sonication. Sometimes EDTA or DTT was added to the solution. The pH was manipulated by addition of HCl or NaOH, usually to a set point of pH = 7.5.

Complete	1	Tablet/50 mL
β-Mercaptoethanol	5	mM
Tween	0.02%	(v/v)
Basic Buffer		

8.1.3.3 Imidazole Stock 1.5 M in Basic Buffer

8.1.3.4 Binding Buffer Stock: 2 mM Beta-Mercaptoethanol in Basic Buffer

8.1.3.5 Binding Buffer + Imidazole 1.5 M Solutions: 10, 13, 15 and 250 mM of Imidazole

8.1.3.6 PBS

Table 10 – Composition of PBS, used as binding buffer for GSTrap. The pH was adjusted to pH = 7.3 by addition of HCl or NaOH.

KCl	2.7	mM
KH_2PO_4	1.8	mM
NaCl	140	mM
Na_2HPO_4	10	mM

8.1.3.7 NMR Buffer 10x

Table 11 – Composition of what was colloquially called NMR Buffer, because this is the usual final buffer for target protein before being submitted to NMR experiments. The pH of the buffer depended on each experiment, since pH = 5.1 to pH = 7.8, which was manipulated by addition of HCl or NaOH.

β-Mercaptoethanol	20	mM
NaCl	100	mM
Na_2HPO_4	100	mM

8.1.3.8 HEPES Buffer

Table 12 – Composition of HEPES Buffer, used as a final buffer for target protein before being submitted to NMR experiments. The pH of the buffer was adjusted to pH = 7.5 by addition of HCl or NaOH.

EDTA	5	mM
HEPES	4	mM
DTT	1	mM

8.1.3.9 Running Buffer

Table 13 – Composition of Running Buffer, used for running an SDS-PAGE experiment. The buffer would be resused usually five times before making a new one.

NuPAGE MES 20x	50	mL
Distilled Water	950	mL

8.1.4 Solutions

8.1.4.1 LB

Table 14 – Composition of LB medium used for growing *E. coli* transfected with a vector that produces p16 fused with a tag protein and that confers resistance to an antibiotic to the host (kanamicin or ampicillin, depending on the vector). The medium was sterilized by autoclave and only after cooling, the antibiotic (to which bacteria is resistant to) was added.

NaCl	10	g/L
Triptone	10	g/L
Yeast Extract	5	g/L
Kanamicin	50	mg/L
Or Ampicillin	100	mg/L

8.1.4.2 LB Agar

Table 15 – Composition of LB medium used for growing *E. coli* transfected with a vector that produces p16 fused with a tag protein and that confers resistance to an antibiotic to the host (kanamicin or ampicillin, depending on the vector). The medium was sterilized by autoclave and only after cooling, the antibiotic (to which bacteria is resistant to) was added.

Agar 1.5% / 0.7%	15 / 10	g/L
NaCl	10	g/L
Triptone	10	g/L
Yeast Extract	5	g/L
Kanamicin	50	mg/L
Or Ampicillin	100	mg/L

8.1.4.3 M9 Salts 5x

Table 16 – Composition of M9 Salts 5x, a component for M9 growth medium (see 8.1.4.5). This solution would be set to a pH = 7.0 by addition of HCl or NaOH, sterilized by autoclave and kept at 4 ºC.

KH_2PO_4	15	g/L
NaCl	2.5	g/L
Na_2HPO_4	30	g/L

8.1.4.4 SL6 100x

Table 17 – Composition of SL6 100x, a component for M9 growth medium (see 8.1.4.5).

Cu^{2+}	1	mM
Co^{2+}	1	mM
H_3BO_3	1	mM
Mn^{2+}	5	mM
MoO_3^-	1	mM
Zn^{2+}	5	mM

8.1.4.5 M9

Table 18 – Composition of M9 growth medium, that is sterilized by a 0.22 µm filter because it contains temperature labile components.

B1 Vitamin 0.1%	0.032	% (v/v)
$CaCl_2$ 0.1 M	0.1	% (v/v)
$FeSO_4$ 0.01 M	0.1	% (v/v)
Glucose (labeled)	4	g/L
Kanamicin	50	mg/L
$MgSO_4$ 1 M	0.2	% (v/v)
MEM Vitamin 100x	1	% (v/v)
M9 Salts 5x	20	% (v/v)
NH_4Cl (labeled)	1	g/L
SL6 100x	1	% (v/v)

8.1.4.6 Kanamicin

Prepared with a concentration of 50 mg/mL and sterilized by a 0.22 µm filter before addition.

8.1.4.7 Ampicillin

Prepared with a concentration of 100 mg/mL and sterilized by a 0.22 µm filter before addition.

8.1.4.8 Denaturative Solution

Table 19 – Composition of Denaturative solution used for lysing cell samples before applying them in SDS-PAGE.

EDTA	**5**	mM
Lysozyme	0.2	mg/mL
Tween	0.2%	
Urea	4	M
Cell Suspension Buffer pH=7.8		