Université de Montréal

Systematic analysis of protein complexes involved in the human RNA polymerase II machinery

par:
Racha Al-Khoury

Département de biochimie
Faculté de médecine

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de Maitre ès Sciences
en biochimie
Fèvrier, 2009

Université de Montréal
Faculté des études supérieures

Ce mémoire intitulé:
Systematic analysis of protein complexes involved in the human RNA polymerase II machinery

présenté par:
Racha Al-Khoury

a été évalué par un jury composé des personnes suivantes:

Pascale Legault
président-rapporteur

Benoit Coulombe
directeur de recherche

Jacques Archambault
membre du jury

Résumé

La transcription, la maturation d'ARN, et le remodelage de la chromatine sont tous des processus centraux dans l'interprétation de l'information contenue dans l'ADN. Bien que beaucoup de complexes de protéines formant la machinerie cellulaire de transcription aient été étudiés, plusieurs restent encore à identifier et caractériser.

En utilisant une approche protéomique, notre laboratoire a purifié plusieurs composantes de la machinerie de transcription de l'ARNPII humaine par double chromatographie d'affinité "TAP". Cette procédure permet l'isolement de complexes protéiques comme ils existent vraisemblablement *in vivo* dans les cellules mammifères, et l'identification de partenaires d'interactions par spectrométrie de masse. Les interactions protéiques qui sont validées bioinformatiquement, sont choisies et utilisées pour cartographier un réseau connectant plusieurs composantes de la machinerie transcriptionnelle. En appliquant cette procédure, notre laboratoire a identifié, pour la première fois, un groupe de protéines, qui interagit physiquement et fonctionnellement avec l'ARNPII humaine. Les propriétés de ces protéines suggèrent un rôle dans l'assemblage de complexes à plusieurs sous-unités, comme les protéines d'échafaudage et chaperonnes.

L'objectif de mon projet était de continuer la caractérisation du réseau de complexes protéiques impliquant les facteurs de transcription. Huit nouveaux partenaires de l'ARNPII (PIH1D1, GPN3, WDR92, PFDN2, KIAA0406, PDRG1, CCT4 et CCT5) ont été purifiés par la méthode TAP, et la spectrométrie de masse a permis d'identifier de nouvelles interactions.

Au cours des années, l'analyse par notre laboratoire des mécanismes de la transcription a contribué à apporter de nouvelles connaissances et à mieux comprendre son fonctionnement. Cette connaissance est essentielle au développement de médicaments qui cibleront les mécanismes de la transcription.

Mots clés : Machinerie transcriptionnelle, ARN polymérase II humaine, purification TAP, partenaires d'interactions, réseau d'interactions.

Abstract

Genomes encode most of the functions necessary for cell growth and differentiation. Gene transcription, RNA processing, and chromatin remodeling are central processes in the interpretation of the information contained in genomic DNA. Although many protein complexes forming the cellular machinery that interprets mammalian genomes have been studied, a number of additional complexes remain to be identified and characterized.

Using proteomic approaches, Dr. Benoit Coulombe's laboratory purified many components of the RNAPII transcription machinery using tandem affinity purification (TAP), a procedure that allows the isolation of protein complexes as they likely exist in live mammalian cells, and the identification of interaction partners using mass spectrometry. High confidence interactions were selected computationally and used to draw the map of a network connecting many components of the mRNA transcriptional machinery. By applying this procedure, our lab has identified, for the first time, a group of proteins, that interacts both physically and functionally with human RNAPII, and whose properties suggest a role in the assembly of multi-subunit complexes, acting as RNAPII-specific scaffolding proteins and chaperones.

The aim of my project was to continue the characterization of the network of protein complexes involving transcription factors, and thus, further pursuing our survey of protein complexes in whole cell extracts. Eight novel RNAPII interaction partners (PIH1D1, GPN3, WDR92, PFDN2, KIAA0406, PDRG1, CCT4 and CCT5) were purified using the tandem affinity purification (TAP) method, and their interaction partners were identified by mass spectrometry.

Over the years, our lab's analysis of transcriptional regulation and mechanisms has contributed novel and important knowledge that provided better understanding of mRNA synthesis. This knowledge is paramount to the development of therapeutics that will target transcriptional mechanisms.

Key words: Transcription machinery, human RNA polymerase II, TAP purification, interaction partners, interaction networks.

Table of contents

List of tables

List of figures

Abbreviations

Å	Ångström
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphate hydrolase
AT rich	Adenine-Thymine rich
BiP	Immunoglobulin heavy chain-binding protein
bps	Base pairs
BRE	TFIIB recognition element
BRE^u	TFIIB recognition element upstream
BRE^d	TFIIB recognition element downstream
BRET	Bioluminescence resonance energy transfer
CBP	Calmodulin binding protein
CCT	Chaperonin containing TCP1
CDK7	Cyclin dependant kinase 7
cDNA	coding deoxyribonucleic acid
CHX	cycloheximide
CFP	Cyan fluorescent protein
Co-IP	Co-immunoprecipitation
CTD	C-terminus domain
C- terminus	carboxyl-terminus
DCE	Downstream core element
DMEM	Dulbecco's modified eagle's medium

DNA	Deoxyribonucleic acid
DPE	Downstream promoter element
DSIF	DRB sensitive inducing factor
EcR	Ecdysone response
FACT	Facilitates chromatin transcription
FBS	Fetal bovine Serum
Fmol	femtomole
FRET	Fluorescence resonance energy transfer
GAL	Galactose
G- protein	Guanine nucleotide-binding protein
GTFs	General transcription factors
HAT	Histone acetyltransferases
HEK	Human embryonic kidney
HDAC	Histone deacetylases
H1	Histone 1
H2A	Histone 2 A
H2B	Histone 2 B
H3	Histone 3
H4	Histone 4
HSP	Heat shock protein
i.e	Example
IgG	Immunoglobulin G
Inr	Initiator
IR	Interaction reliability

ISWI	Imitation switch
Kb	Kilobase
LUMIER	Luminescence-based mammalian interactome mapping
MAT1	ménage à trois 1
m^7G cap	7-methyl guanosine cap
mRNA	messenger ribonucleic acid
mRNPs	mRNA ribonucleoprotein particles
MS	Mass spectrometry
MTE	Motif ten element
NELF	Negative elongation factor
nm	Nanometer
N- terminus	Amino-terminus
NTPs	Nucleoside triphosphates
LC- MS/MS	Liquid chromatography tandem mass spectrometer
OD	Optical density
PCA	Protein-fragment complementation assay
PDRG1	p53 and DNA damage regulated 1
PIC	Pre-initiation complex
PIH1D1	PIH1 domain containing 1
PFDN	Prefoldin
pmol	Picomole
Poly(A)	Poly(Adenosine)
Pre- mRNAs	Precursor-messenger ribonucleic acids
P-TEFb	Positive transcription elongation factor b

R2TP	Rvb1-Rvb2-Tah1-Pih1
RAP	RNA polymerase II associated factor
RNA	Ribonucleic acid
RNAP	RNA polymerase
RNAPI	RNA polymerase I
RNAPII	RNA polymerase II
RNAPIIA	RNA polymerase II in the A form
RNAPIIO	RNA polymerase II in the O form
RNAPIII	RNA polymerase III
RPAP	RNA polymerase II associated protein
rRNA	Ribosomal ribonucleic acid
SDS	Sodium dodecyl sulfate
SDS- PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SNF2	Sucrose non fermenting 2
snoRNAs	Small nucleolar ribonucleic acid
SnRNAs	Small nuclear ribonucleic acids
SWI/SNF	Switch/ Sucrose non fermenting
TAF	TBP associated factor
Tah1	Tpr-containing protein associated with Hsp90
TAP	Tandem affinity purification
TBP	TATA box binding protein
TCP-1	T-complex polypeptide 1
TEV	Tobacco etch virus
TNF-α	Tumor necrosis factor-α

TRAP	TNF receptor-associated protein
TRNAs	Transfer ribonucleic acids
UV	Ultraviolet
VP16	Herpes simplex virion protein 16
WDR92	WD repeat domain 92
YFP	Yellow fluorescent protein

Acknowledgements

First and foremost, I would like to thank GOD for his grace and guidance throughout my life, and for giving me the strength and courage to do this work.

I would also like to thank my supervisor Dr. Benoit Coulombe for providing me with the opportunity to work in his lab, as well as guiding and advising me throughout my work towards this thesis.

Thanks to all of the members of our laboratory with whom I have enjoyed working these past two years, especially Annie Bouchard and Philippe Cloutier for their help and advice, you both have taught me a lot.

A huge "thank you" goes to my sister Diala, my brothers, Firas and Nawras, and friends for all of their support and encouragement. I love you guys and I am so grateful for having you in my life.

I dedicate my Master's thesis to my parents, Wafaa Butros and Imad Al-Khoury, who have always provided me with their support, constant encouragement and their unconditional love, and who have always given me the strength to follow my dreams. Thank you and I love you dearly.

Thank you all…

Section 1: Introduction

1-1) From genes to proteins:

Our genome, made of DNA (deoxyribonucleic acid), is the storehouse that contains all of the required information necessary to build an organism. It is arranged into 24 distinct chromosomes each of which contains many genes which encode proteins. Less than 2% of our genome encodes for genes, the equivalent of about 30000 genes, while the rest of our genome consists of non-coding regions which are thought to play a role in maintaining the integrity of chromosomes and regulating the expression of proteins.

DNA consists of 4 different nucleobases, adenine, guanine, cytosine and thymine, and one DNA molecule can contain hundred million nucleotides. Nucleotides polymerize together, via phosphodiester bonds, to form nucleic acids, the linear representation of which constitutes the primary structure of nucleic acids. In fact, DNA is made up of two associated polynucleotide strands forming a double helix, where the two strands are complementary and oriented antiparallel to each other (Harvey Lodish *et al.,* 4th edition).

This genetic material, DNA, carries the necessary information to specify the amino acid sequences of proteins each of which will have a specific function in the cell (Harvey Lodish *et al.,* 4th edition). To interpret the information encoded by DNA, a transcription process occurs in which an enzyme, RNA polymerase, is recruited to the DNA and transcribes the information into RNA (ribonucleic acid). This RNA can then serve as template for the synthesis of proteins since each set of three ribonucleotides encodes an amino acid, the building block of proteins. This process occurs with the help of ribosomes which consist of ribosomal RNA (rRNA) associated to a set of proteins (Harvey Lodish *et al.,* 4th edition). All of these processes will be

explained in greater details next, starting with the transcription of DNA to RNA.

1-1.1) Transcription process:

Transcription is the process by which DNA is translated into RNA with the help of RNA polymerase enzymes. In eukaryotes, there exist three kinds of RNA polymerases (RNAP), RNAPI, RNAPII and RNAPIII; this is why genes are classified into three classes depending on the RNA polymerase enzyme that transcribes them. RNAPI, localized within the nucleoli, is the enzyme responsible for the transcription of ribosomal RNAs (rRNAs). On the other hand, RNAPII, which is localized within the nucleoplasm, transcribes all protein-encoding genes and certain small nuclear RNAs (snRNAs) genes. Finally, RNAPIII, which is also localized within the nucleoplasm, is the enzyme responsible for the transcription of transfer RNAs (tRNAs) and 5S rRNA (reviewed by Archambault *et al.*, 1993).

RNAPII transcription is the first step in gene expression and a focal point of cell regulation. It is a target of many signal transduction pathways, and a molecular switch for cell differentiation in development (Cramer *et al.,* 2001).

1-1.1.1) RNA polymerase II (RNAPII):

RNAPII is the enzyme responsible for the transcription of mRNA in eukaryotes. It is composed of 12 subunits, often referred to as Rpb1 to Rpb12 by decreasing order of their molecular mass. Five of these subunits are common to all three RNA polymerases and they are Rpb5, 6, 8, 10 and 12, and only Rpb4, 7, 9 and the CTD of Rpb1 (see below) are unique to RNAPII. Rpb1 and Rpb2 consist the two largest subunits and are responsible for most of the catalytic activity of polymerase (reviewed by *et al.,*, 2006).

Rpb1 also contains the CTD which consists of a tandem repeat of a heptapeptide: Tyr-Ser-Pro-Thr-Ser-Pro-Ser which appears 52 times in humans.

The CTD is an essential feature of RNAPII and is involved in transcription and mRNA processing. The CTD is disordered in RNAPII structures and tends to be degraded by proteases (reviewed by Thomas and Chiang, 2006). Depending on its phosphorylation state, RNAPII can exist in two forms; either the RNAPIIA, with an unphosphorylated CTD, involved in pre- initiation complex (PIC) assembly and transcription initiation, or RNAPIIO form, with a phosphorylated CTD, and is implicated in transcript elongation and termination (reviewed by Thomas *et al.*, 2006).

Rpb4 and Rpb7 are two subunits of the RNAPII known to form a heterodimer required for the formation of the pre- initiation complex (PIC), the initiation of transcription, and RNA chain elongation (reviewed by Thomas *et al.*, 2006).

Finally Rpb9, unique to RNAPII, plays an important role in transcription elongation, transcription-coupled DNA repair, and helps in the selection of the correct transcription start site and in maintenance of transcription fidelity (Chen *et al.,* 2007).

1-1.1.2) Recruitment of RNAPII to the promoter:

On its own, RNAPII is unable to initiate transcription but rather requires other proteins, such as the general transcription factors (GTFs), involved in the recognition of promoter sequences, the response to regulatory factors, and conformational changes essential to the activity of the polymerase during the transcription cycle (reviewed by Hahn, 2004). The general transcription factors TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH assemble the RNAPII onto promoter DNA to form the Pre-Initiation complex (PIC). Refer to table 1 for a summary about the general transcription factors, their composition and function (adapted from a review by Thomas *et al.*, 2006).

GTF	Protein composition	Function
TFIIA	P35 (α), P19 (β), P12 (γ)	-Stabilizes TBP- TATA complex -Anti repressor and a coactivator
TFIIB	P33	--Start site selection; stabilizes TBP-TATA complex, pol II/TFIIF recruitment
TFIID	TBP, TAFs (TAF1-TAF14)	-Core promoter- binding factor -Coactivator and a protein kinase -Ubiquitin- activating/ conjugating activity and Histone acetyltransferases
TFIIE	P56 (α), p34 (β)	-Recruits TFIIH and is involved in promoter clearance -Facilitates formation of an initiation-competent RNAPII
TFIIF	RAP30, RAP74	-Binds RNAPII and facilitates its recruitment to the promoter -Recruits TFIIE and TFIIH -Functions with TFIIB and RNAPII in start site selection -Promoter wrapping against RNAPII -Facilitates RNAPII promoter escape -Enhances the efficiency of RNAPII elongation
TFIIH	P89/XPB, p80/XPD, p62, p52, p44, p40/CDK7, p38/CyclinH, p34, p32/MAT1, p8/TFB5	-ATPase activity for transcription initiation and promoter clearance -Helicase activity for promoter opening -Transcription- coupled nucleotide excision repair -Kinase activity for phosphorylating pol II CTD -E3 ubiquitin ligase activity

Table I: Summary of the general transcription factors

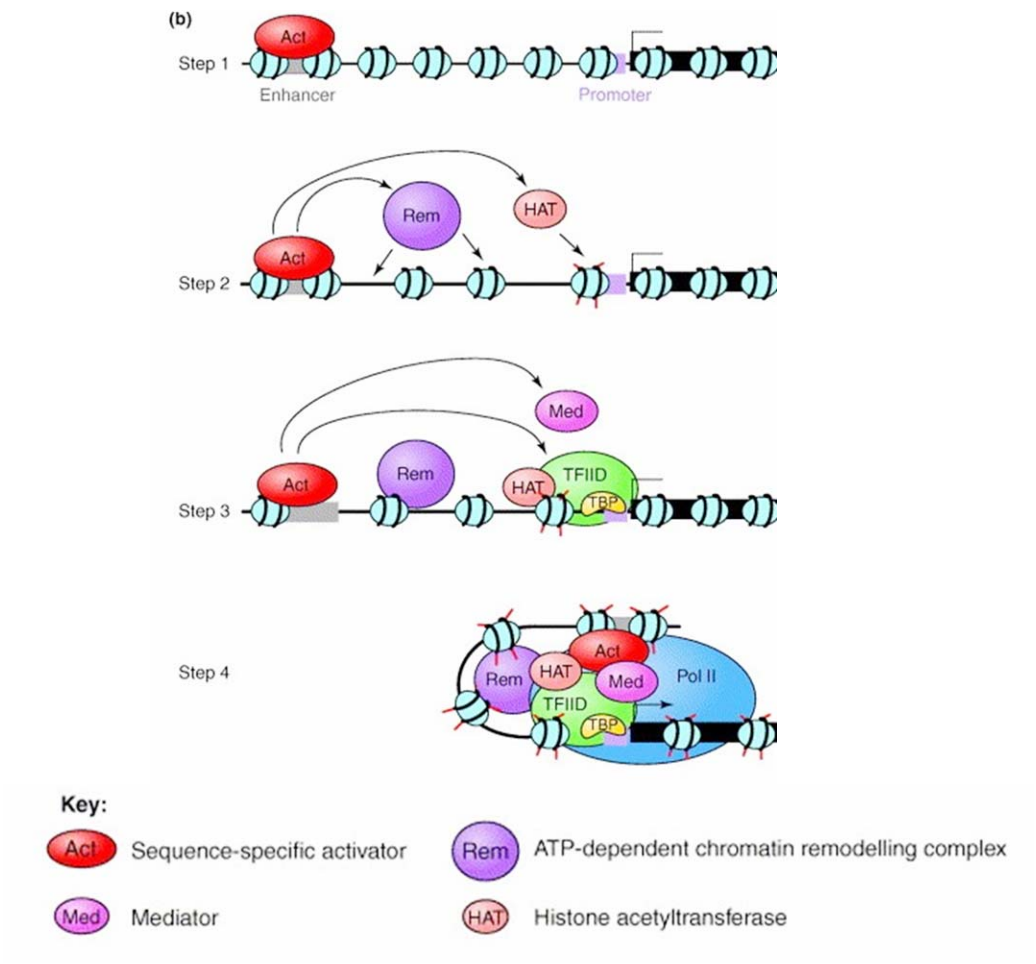
Other proteins also regulate transcription, such as the mediator which is a large multiprotein complex that communicates directly with many gene-specific regulators (Blackwell *et. al.,* 2006), and sequence-specific DNA-binding transcription regulators (i.e., activators and repressors) (Martinez, 2002).

Transcription initiation proceeds through a number of specific steps. First, the packed chromatin must be de-condensed so that the transcriptional machinery could have access to the promoter region of a gene. Second, the binding of transcriptional activators to enhancer elements located close to or many thousands of base pairs away from the transcription start site. Finally, co-activators are recruited to bring the RNAPII PIC to the site where the transcription activators are bound to enhancers (reviewed by Gross *et al*, 2006).

1-1.1.2.1) The de-condensation of chromatin:

Inside the non-dividing eukaryotic nuclei, huge DNA molecule is packaged, into filament with the help of proteins to form the 30 nm chromatin. This chromatin structure plays an important role in gene regulation since it presents an obstacle that must be overcome by transcription machineries to access the underlying DNA (reviewed by Ruthenberg *et al.*, 2007).

The first step in the transcription process is the binding of activators to enhancers which will recruit histone- modifying enzymes helping in the displacement of nucleosomes, as well as target the basal transcription machinery to the core promoter (discussed later on). The mediator, a co-activator, is also involved and will serve as a bridge between the enhancer-bound activators and the pre- initiation complex at the core promoter (reviewed by Szutorisz *et al.*, 2005). The mediator is a multi- subunit complex, discovered by Roger D. Kornberg and R. Young at the same time, which is essential for transcription of class II genes. Figure 1 summarizes these processes (adapted from a review by Szutorisz *et al.*, 2005).



<u>Figure 1: Schematic representation of a proposed model for the activation of class II gene transcription.</u> In the first step, the activator binds to the enhancer element and then recruits chromatin remodeling complexes to decondense the chromatin structure in step two. This is followed by the recruitment of the transcription machinery to the core promoter in step three. Finally in step 4, the enhancer- bound activator is brought into closer proximity, with the help of the mediator complex, to the transcription machinery at the core promoter.

1-1.1.2.1.1) Histones:

Histones are the main protein components of chromatin and play a major role in the structure of chromatin. There are five major types of histones, called H1, H2A, H2B, H3 and H4, and these histones are present in almost all cell types. Four of these five types of histones, H2A, H2B, H3 and H4, specifically bind to DNA and they are the called "core" histones (reviewed by Lusser *et al.*, 2003). Two of each assemble to form an octameric structure around which 146 base pairs of DNA are wrapped to form a structure called the nucleosome core, which is separated from other nucleosome cores by a thread of DNA referred to as the linker DNA. Together, one nucleosome core with a linker DNA, they form a complete chromatin subunit. H1, on the other hand, interacts with the nucleosomal core and the adjoining linker DNA (reviewed by Lusser *et al.*, 2003). Histones protect the DNA, wrapped around them, from cleavage by endonucleases, whereas the linker DNA region is susceptible to endonucleolytic cleavages.

Chromatin condensation creates an important obstacle for transcription, therefore, chromatin remodeling complexes and histone-modifying enzymes have to be recruited, by gene-specific regulatory factors, to de-condense the region where transcription is to be potentiated.

1-1.1.2.1.2) Chromatin remodeling

1-1.1.2.1.2.1) Chromatin remodeling complexes:

One class of chromatin remodeling complexes is the ATP-dependent molecular machine, which uses the hydrolysis of ATP to modulate the interactions between DNA and histones in the chromatin. Although they are very different in composition and in function, these complexes share an ATPase subunit which belongs to the Snf2-like family (reviewed by Lusser *et al.*, 2003).

ATP-dependent chromatin remodeling complexes are divided into three groups depending on their biochemical properties and the sequence similarity of their ATPase subunits.

The first group is the SWI/SNF group, which plays a role in the activation of transcription. Then there are the ISWI and the Mi-2/CHD groups, both of which are involved in the repression of transcription (reviewed by Peterson, 2002).

The most studied of these groups is the SWI/SNF group, which can modulate chromatin structure by helping DNA-bending proteins facilitate nucleosomal sliding, as well as disrupting high-order chromatin folding and nucleosomes (reviewed by Peterson, 2002, and by Vignali *et al.*, 2000).

1-1.1.2.1.2.2) Histone modifications:

Histones can also undergo posttranslational modifications that can alter their interaction with DNA. Different histone modifications can have a role in different biological processes such as, gene regulation, DNA repair or DNA chromatin condensation. Some of these modifications include the acetylation of lysines, the methylation of lysines and arginines, the phosphorylation of serines and threonines, the ubiquitination of lysines, the sumoylation of lysines, and the ADP-ribosylation of glutamic acids (reviewed by Khorasanizadeh, 2004). The enzymes responsible for these modifications are the histone acetyltransferases (HAT), histone deacetylases (HDAC), histone methyltransferases, kinases for phosphorylation, ubiquitinases and poly-(ADP-ribose) polymerases (reviewed by Davie, 1996). Most modifications localize to the amino- and carboxy-terminal histone tails, and a few localize to the histone globular domains (reviewed by Berger, 2007). As to the modifications that are involved in the regulation of transcription, they are divided into two groups, either activating or repressing transcription. These modifications are summarized in table 2 (adapted from a review by Berger, 2007).

Histone post-translational modification	Transcriptionally relevant sites	Transcriptional role
Acetylated Lysine	H3 (9,14,18,58), H4 (5,8,13,16), H2A, H2B	Activation
Phosphorylated Serine/ Threonine	H3 (3,10,28), H2A, H2B	Activation
Methylated Argenine	H3 (17,23), H4 (3)	Activation
Methylated Lysine	H3 (4,36,79) H3 (9, 27), H4 (20)	Activation Repression
Ubiquitylated Lysine	H2B (123 in yeast,120 in mammals) H2A (119 in mammals)	Activation Repression
Sumoylated Lysine	H2B(6/7), H2A (126)	Repression

Table II: **Histone post- translational modifications affecting transcription**

1-1.1.2.2) Assembly of the RNAP II Pre- Initiation Complex (PIC):

Once the chromatin has de-condensed to a more accessible structure, the gene- specific regulatory factors, already bound to the site of transcription initiation, will now recruit the RNA polymerase II to the core promoter where the pre-initiation complex can assemble.

1-1.1.2.2.1) Core promoter:

A core promoter (Figure 2) is defined as the minimal portion of DNA sequence, located upstream of a gene, required to properly initiate transcription. It most commonly consists of elements such as the TATA box, which is present 25- 30 bps upstream of the transcription start site (+1), and is an AT rich sequence that is recognized by the TBP protein (TATA-binding protein),

Figure 2: **The core promoter**

Schematic representation of the core promoter needed for basal transcription by the RNAPII enzyme. What is shown are the elements that most commonly consist the core promoter (TATA box, BRE^u and BRE^d, Inr, DPE, MTE and finally DCE).

a subunit of the TFIID general transcription factor. The initiator element (Inr), which has a conserved sequence, is located at the +1 site. The Inr is functionally similar to the TATA box (-25 to -30 bps away from the Inr position at +1), and can function independently of the TATA box when they are more than 30 bps apart, otherwise, these two elements function synergistically (reviewed by Smale *et al.*, 2003). The Inr element is also recognized by TFIID, through its TAF (TBP associated protein) subunits, TAF1 and TAF2 (reviewed by Smale, 1997, and by Thomas and Chiang, 2006). Another element is the DPE (downstream promoter element), present at +28 - +32 relative to the +1 position. This element is conserved from *Drosophila* to humans and is required for the binding of purified TFIID (through its TAF6 and TAF9 subunits) to a subset of TATA-less promoters. The DPE element also acts in conjunction with the Inr element (reviewed by Smale *et al.*, 2003, and by Thomas and Chiang, 2006).

In addition to DPE, two other core promoter elements are present downstream of the transcription start site, and they are the MTE and DCE

elements. The MTE (Motif Ten Element) is positioned between +18 to +29, and normally functions in conjunction with Inr to enhance transcription by RNAPII. It can also work synergistically, in an Inr-dependent manner, with the TATA box and DPE, or without them (when they are not present), to strengthen the promoter activity. DCE (Downstream Core Element), on the other hand, is found between +6 to +34 and is recognized by the TAF1 subunit of TFIID (reviewed by Thomas *et al.*, 2007, and by Thomas and Chiang, 2006). The DCE contributes to the transcriptional activity and binding of TFIID (reviewed by Smale *et al.*, 2003).

The core promoter also consists of a BRE (TFIIB recognition element) which could be present at two positions, upstream of the TATA box (BRE^u) and downstream (BRE^d) of the TATA box. BRE^d, bound to TFIIB, was shown to modulate promoter strength and act as a positive element for transcription when the promoter contains only the BRE^d element. However, in the presence of BRE^u, BRE^d has a negative effect on transcription. BRE^u, on the other hand, might have a positive or negative effect on transcription of TATA-less promoters likely to contain BRE^u, rather than the TATA-containing promoters (Deng *et al.*, 2006).

Although most core promoter studies were conducted on TATA box containing promoters, it turns out that the majority of mammalian genes have TATA-less promoters within which are present multiple start sites, generating diversity and complexity. This indicates that TATA-driven PIC assembly is the exception, rather than the rule, in eukaryotic transcription (reviewed by Sandelin *et al.*, 2007).

1-1.1.2.2.2) Recruitment of RNAPII to the core promoter:

This process starts with the formation of the PIC, which consists of the RNAPII and several general transcription factors including TFIIA, TFIIB,

TFIID, TFIIE, TFIIF and TFIIH. The assembly of the PIC is thought to proceed in two different pathways.

The first pathway is called the "sequential assembly pathway" where the TBP protein of TFIID first binds to the promoter region at the TATA box, resulting in the bending of the DNA at an 80° angle (Kim J. L. *et al.*, 1993, Kim Y. *et al.*, 1993). This event is followed by the entry of TFIIA and TFIIB, both of which will help in stabilizing the TFIID bound to promoter. TFIIB will help in facilitating the recruitment of the RNAPII- TFIIF complex (reviewed by Thomas *et al.*, 2007). The entry of TFIIF induces a second bending of the DNA, close to the +1 site, causing the wrapping of the promoter region around the RNAPII (Robert *et al.*, 1998, reviewed by Coulombe and Burton, 1999). This event is finally followed by the entry of TFIIE and TFIIH.

The second pathway is known as the "RNAPII holoenzyme pathway", where the RNAPII is present in a pre-assembled holoenzyme complex containing the RNAPII and suppressors of RNAPII mutations, with or without a subset of general transcription factors. In addition, other proteins associated with chromatin remodeling, DNA repair, and mRNA processing, are involved. This holoenzyme complex would be recruited to a transcription site by TFIID bound to a promoter (reviewed by Thomas *et al.,* 2007). Evidences supporting both models have been reported. Thus, it is likely that both assembly pathways exist *in vivo*.

In both pathways, it is clear that each general transcription factor involved plays an essential role in transcription. For example, when the TBP subunit of TFIID binds to the TATA box of the promoter, it bends the DNA around the RNAPII bringing it closer to the transcription start site. Furthermore, when TFIIH is recruited, it introduces negative superhelical tension in the DNA, through its ATPase/Helicase subunit, producing a transient bubble.

Finally initiation and RNA synthesis can be potentiated with the addition of two nucleoside triphosphates (NTPs) (reviewed by Kornberg, 2007).

1-1.1.3) Initiation:

The initiation process does not start smoothly; up until positions +8/+10 multiple rounds of abortive initiation occur resulting in the production of short (2- 9 nucleotides) transcripts. These rounds proceed until the complex reaches a point where it is stabilized, usually at position +9, and this transition is referred to as "promoter escape". Soon after promoter escape, RNAPII pauses again and this phenomenon is called the "promoter-proximal pausing", in which a large number of genes have been shown to be regulated (reviewed by Margaritis *et al.*, 2008).

1-1.1.3.1) mRNA processing prior to productive elongation:

The transcription of mRNA is tightly coupled to the recruitment of the mRNA processing machinery. The initial transcripts synthesized by the RNAPII, called pre-mRNAs, have to be processed to form the mRNA before getting exported to the cytoplasm. The first processing step in all eukaryotic cells occurs as soon as the transcript is ~25 bases long. An m^7G cap is added to the 5'-end of the transcript by a capping enzyme bound to the phosphorylated CTD. This cap modification has been shown to occur during the arrest of the transcription process, prior to its entry in the productive elongation phase. This suggests that the capping enzyme might play an important role as a checkpoint, ensuring that uncapped transcripts are not elongated. Once the 5'-end is capped, introns are excised with the help of the spliceosome, one of the largest macromolecular machines in the cell, also brought to the transcription site through interaction with the CTD of RNAPII (reviewed by Almeida *et al.*, 2008).

1-1.1.4) Elongation:

Once the RNAPII has escaped the promoter it enters the elongation phase. At this stage all of the general transcription factors dissociate to form the "transcription elongation complex", except for TFIIF. This phenomenon is referred to as the "Promoter clearance step".

Promoter clearance has been shown to coincide with phosphorylation of the CTD of RNAPII, which plays an important role in recruiting protein factors involved in elongation, as well as in mRNA maturation, surveillance, and export. Other elongation factors are recruited at this stage to the RNAPII to modulate its catalytic activity and form the transcription elongation complex. This includes factors such as Elongins, P-TEFb, DSIF and NELF, which can all influence the pausing of the RNAPII, either positively or negatively. FACT, a histone chaperone, is another elongation factor recruited to help facilitate the elongation process by remodeling the chromatin (reviewed by Sims *et al.*, 2004).

1-1.1.5) Termination and 3'-end mRNA processing:

Termination is the last step of transcription by the RNAPII and it is a process that depends on pre-mRNA 3'-end processing signals, such as the poly(A) signal. Therefore, once the poly(A) signal has been read by the RNAPII, it is recognized by a multi-component cleavage/polyadenylation complex recruited to the RNAPII through the CTD of Rpb1. This complex will direct the endonucleolytic cleavage and the polyadenylation of the free 3'-end (reviewed by Lykke-Andersen *et al.*, 2007, and Kim *et al.*, 2003). Poly(A)-binding proteins are then recruited to the 3'end of the processed transcript to help protect it from exonucleolytic degradation. Other signals also exist on different transcripts, such as the snRNAs and the snoRNAs, and are recognized

by other protein complexes, also recruited to RNAPII through the CTD of Rpb1, which contain the RNA binding proteins Nrd1p, Nab3p, and additional co-factors (reviewed by Lykke-Andersen *et al.*, 2007, and Kim *et al.*, 2003).

1-1.2) mRNA export:

The transcription process is also coupled to the nuclear export of mRNAs, which is an essential process for the expression of genes in all eukaryotic cells. Therefore, as soon as the mRNA is synthesized, proteins are recruited forming the mRNA ribonucleoprotein particles (mRNPs). Export machineries will recognize signals within the proteins of mRNPs rather than within the mRNA itself, and will export the mRNA to the cytoplasm where the proteins synthesis occurs (reviewed by Iglesias *et al.*, 2008).

1-1.3) Protein synthesis:

Translation is the process by which mRNAs are used to direct the synthesis of polypeptide chains in three different steps, including initiation, elongation, and termination. This process is catalyzed by a complex called the ribosome which is, in fact, a ribozyme with RNA at the heart of its enzymatic activity (reviewed by Steitz, 2008 and by Culver, 2001). Mammalian ribosomes are only found in the cytosol and in the mitochondria (reviewed by Hebert *et al.*, 2007).

The eukaryotic ribosome is a complex that consists of two asymmetric ribonucleoprotein subunits, the 60s and the 40s subunits, which will bind the mRNA and use it as a template to catalyze the correct assembly of amino acids into polypeptide chains. tRNAs are the adaptor molecules used by ribosomes to decode the information in the mRNA. They are present at the interface between the mRNA and the growing amino acid chain.

There exits three tRNA-binding sites in a ribosome, the A, P and E sites. The A site is the site of entry of a cognate aminoacyl tRNA. The P site is the site of the peptidyl tRNA, where the elongation of the polypeptide chain occurs. The amino acid bound to the tRNA in the A site is transferred to the growing polypeptide chain bound to the tRNA in the P site. The E site is the exit site, occupied by the deacylated tRNA. An exception occurs during the initiation step where the cognate aminoacyl tRNA carrying a methionine enters the P site rather than the A site and base pairs with the start codon of the mRNA (reviewed by Culver, 2001). The translation ends once a stop codon is read by the tRNA; at this point the polypeptide chain separates form the tRNA, which itself separates from the ribosome. Finally the ribosomal subunits separate from the mRNA.

1-1.4) Protein folding:

Following the synthesis of a new polypeptide chain, its correct folding will convert it into the mature, active protein which can then be localized to its appropriate location (reviewed by Hebert *et al.*, 2007). This process will normally take place either in the cytoplasm, for cytoplasmic proteins, or in the endoplasmic reticulum, for transmembrane and secretory proteins (reviewed by Paulsson *et al.*, 2003). The information needed for the correct folding of a protein is encoded by its amino acid sequence. The protein folding process requires most of the time the assistance of a network of molecular chaperones to support the folding *in vivo.* This is important to prevent aggregation in the crowded intracellular environment, and prevent inappropriate inter- and intra-molecular interactions (reviewed by Deuerling *et al.*, 2004). As soon as the nascent polypeptide emerges from the ribosome, the ribosome-associated chaperones guide its folding. Finally, once a protein has been properly folded by a chaperone, it must be localized to its appropriate subcellular compartment.

Chaperones will help these proteins find their way to the place where they can display their appropriate function (Lund *et al.*, 2003).

1-1.4.1) Molecular Chaperones:

Molecular chaperones exist in all organisms and in all cellular compartments. They are defined as a large group of proteins that share the property of assisting the folding, unfolding, assembly and disassembly of other macromolecular structures (reviewed by Ellis, 2006). Furthermore, chaperones have also been shown to play a role in the translocation of newly synthesized proteins (reviewed by Nicoll *et al.*, 2005). The main chaperone classes are the Hsp40 (the DnaJ family), Hsp60 (includes GroEL and T-complex polypeptide 1 (TCP-1) ring complexes), Hsp70, Hsp90 (reviewed by Fink, 1999) and the AAA+ superfamily of ATPases.

1-1.4.1.1) Hsp40 family:

The best defined role for Hsp40 is as a cochaperone for the Hsp70. This family of proteins consists of over 100 members, defined by the presence of three distinct domains. The first is a highly conserved J- domain of ~70 amino acids near its N-terminus, responsible for the interaction with Hsp70. The presence of a J- domain in a protein defines it as a DnaJ-like protein, and is involved in regulating the ATPase activity of the Hsp70 (Hennessy *et al.,* 2000). The second domain is a glycine and phenylalanine rich region that acts as a flexible linker. Finally, the third domain is a cysteine-rich, zinc finger-containing C-terminus domain (Cheng *et al.*, 2008, and reviewed by Fink 1999).

1-1.4.1.2) Hsp70 family:

Members of this family help to protect cells from stress, aside from playing an important part in protein folding. They are characterized by a very

weak ATPase activity which is stimulated by the binding of the Hsp40 cochaperone. They are composed of two major functional domains; an N-terminal ATPase domain and a C-terminal domain that is responsible for the binding of polypeptides. Some of the better known mammalian members of this family include the constitutive cytosolic member HSC70, the stress- induced cytosolic form HSP70, the ER form BiP and the mitochondrial form mHSP70 (Morano KA, 2007, and reviewed by Fink 1999).

1-1.4.1.3) Hsp90 family:

Hsp90 is one of the most abundant chaperones in the cell found to be upregulated in response to stress. It also pays an important role, under normal conditions, in protein folding, intracellular transport, and stability of proteins many of which are critical for signal transduction. Mammalian Hsp90 exist as dimers and is often found in complexes with other chaperones. Members of this family of proteins are highly conserved in all organisms from bacteria to humans and are essential. Examples of the Hsp90 family include the cytosolic form, HSP90A, the endoplasmic reticulum form, HSP90B and the mitochondrial form, TRAP (reviewed by Fink, 1999, and by Pearl *et al.*, 2008).

Hsp90 has an ATPase activity and contains three functional domains. The first is the ATP binding domain near the N- terminus. The other two domains are the protein binding and the dimerization domains located near the C-terminus of the Hsp90 protein. All of these domains play an important role in the function of the protein (Southworth *et al.,* 2008, Barginear *et al.,* 2008 reviewed by Fink, 1999).

1-1.4.1.4) Hsp60 family/ Chaperonin family:

This family of proteins is the best studied of the chaperones, and is divided into two families/ groups; the GroEL (also called group I) and the TCP-1 ring complex families (also called group II), both of which are represented

schematically in Figure 3 (Figure 3 is adapted from Martin- Benito *et al.,* 2007, Martin- Benito *et al.,* 2002, and Roseman *et al,.* 1996). Members of this family are involved in the assembly of large multiprotein complexes (reviewed by Fink, 1999). Hsp60 genes are highly conserved, indicating a central role in cell viability (Lund *et al.*, 2003).

1-1.4.1.4.1) GroEL family:

This family of proteins is found in prokaryotes, chloroplasts, and mitochondria. The vast majority of organisms contain at least one gene for this protein. GroEL requires the assistance of a cochaperone called GroES "lid" and the presence of ATP to facilitate the folding of newly synthesized polypeptides. Substrates of this group of chaperones include several metabolic enzymes, RNA polymerase II subunits, and other proteins involved in transcription and translation (reviewed by Fink, 1999). GroEL chaperones have also been shown to play a role in conformational maintenance of pre-existing proteins, secretion, and proteolysis (Houry *et al.,* 1999, Kusukawa *et al.,* 1989, Kandror *et al.,* 1994, and reviewed by Deuerling *et al.*, 2004). GroEL is composed of two heptameric rings, of the large subunit GroEL, stacking back to back and forming a 14-subunit hollow cylinder (Lund, 2001). Binding of ATP and GroES induces large conformational changes in GroEL providing a polar environment favorizing the folding of the encapsulated protein. Then ATP hydrolysis primes the GroEL to release GroES "lid", allowing the folded substrate to exit the chaperone (Lund *et al.*, 2003, reviewed by Fink, 1999 and reviewed by Deuerling *et al.*, 2004).

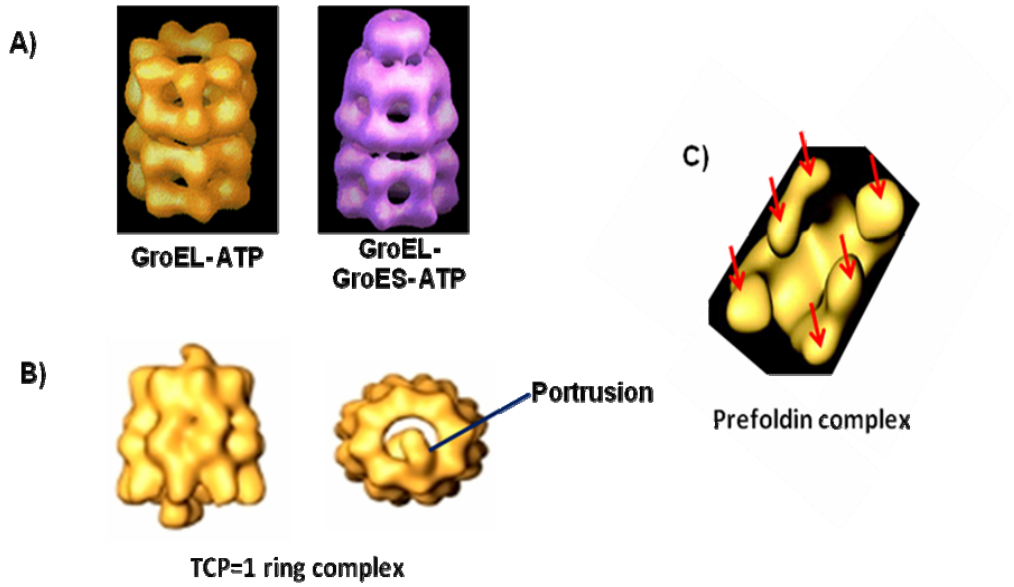


Figure 3: The GroEL and the TCP-1 ring complex families

A) Left: A side view of the GroEL chaperone in the ATP bound state. This view shows the two heptameric ring. Right: A side view of the GroEL chaperone in the ATP bound state in the presence of the GroES cochaperone.

B) Left: A side view of the TCP-1 ring complex chaperone showing the two octameric rings. Right: A top view of the TCP-1 ring complex chaperone showing the protrusion "lid".

C) The prefoldin complex with its six subunits (red arrows point to the six subunits).

1-1.4.1.4.2) TCP-1 ring complex family:

The TCP-1 complex is an ATP-dependent complex found in the eukaryotic cytosol, and is composed of two identical stacked rings, each composed of eight different proteins referred to as CCT1 (also referred to as TCP-1) to CCT8 encoded by different genes. This family of proteins does not require the assistance of a cochaperone such as the GroES, but is rather characterized by the presence of a portrusion "lid" on the TCP-1 ring complex, which is actually part of the CCT protein itself. TCP-1 ring complexes are chaperones that assist in the folding of a small number of proteins, mainly actin and tubulin (Yam *et al.,* 2008, Lund *et al.*, 2003, reviewed by Fink, 1999, and Llorca *et al.,* 1999).

1-1.4.1.4.2.1) The prefoldin complex:

The TCP-1 ring complex does not function alone in the folding of actin and tubulin, but rather requires the assistance of another molecular chaperone named prefoldin. The prefoldin complex itself is a heterohexameric complex found in archaeabacterial and eukaryotic organisms (Martin- Benito *et al.,* 2002). In eukaryotes, the prefoldin complex is an oligomer composed of six different proteins (PFDN1- PFDN6). The archaeal prefoldin chaperone, on the other hand, is an oligomer composed of two different proteins, two α-subunits and four β-subunits (Martin- Benito *et al.,* 2002).

Actins and tubulins are very abundant and highly conserved proteins involved in processes which are essential and unique to eukaryotes. These processes include muscle contraction, segregation of chromosomes, stabilization and alteration of cell shape, endocytosis and exocytosis to mention a few (Leroux *et al.*, 2000, Gu *et al.,* 2008). Once the N-terminus (~145 amino acids) of the actin or tubulin nascent chains have been synthesized, the prefoldin complex binds to it and keeps the polypeptide chain in an unfolded

state until its synthesis is complete. After which, the polypeptide is delivered to the TCP-1 complex for proper folding (Hansen *et al.*, 1999, Gu *et al.,* 2008).

1-1.4.1.4.3) The AAA+ superfamily of ATPases:

AAA stands for ATPases Associated with diverse cellular Activities (Kunau *et al.,* 1993). The AAA+ superfamily is a large and functionally diverse superfamily of NTPases that are characterized by a conserved nucleotide-binding and catalytic module, the AAA+ module (Snider *et al.,* 2008). Members of the AAA+ superfamily of ATPases are found in all three kingdoms of life and function in diverse cellular processes, often via chaperone-like activities (Zhang *et al.,* 2002; Ogura and Wilkinson, 2001). AAA+ proteins typically assemble into hexameric ring complexes that are involved in the energy-dependent remodeling of macromolecules (Iyer *et al.,* 2004, Snider *et. al.* 2008). The defining feature of AAA+ proteins is a structurally conserved ATP-binding module of 200-250 amino acids, that oligomerizes into active arrays (reviewed by Erzberger *et al.,* 2006, Ogura and Wilkinson 2001). ATP hydrolysis by the AAA+ protein is coupled to physical contact with the target substrate molecule and requires that specific interaction for remodeling of the substrate to take place (Zhang *et al.,* 2002).

1-2) Protein complexes:

Many cellular functions are performed by multiprotein complexes. For example, the RNA pol II enzyme which itself is composed of 12 subunits, that need to come together and interact with each other to form the enzyme. Once formed, this enzyme requires the assistance of the general transcription factors which will be recruited, and interact with it, to facilitate the transcription process. Another example of multiprotein complexes is the chaperones themselves. Chaperones are often composed of different proteins that interact

with each other to form the active chaperone; for example, the TCP-1 ring complex with its 8 subunits that need to interact with each other to generate the active chaperone activity. Furthermore, these chaperones will often require the assistance of cochaperones to facilitate the folding process of proteins.

All of this point to the fact that, proteins do not function on their own but are usually grouped into larger complexes to perform a specific function in an organism (Gingras *et al.,* 2005). Therefore, it is clear that the study of protein complexes is important, since it allows placing proteins of an unknown function into a functional context provided by their interactions with other proteins in the complex (reviewed by Bauer *et al.*, 2003). Therefore, the network of interacting partners in which a certain protein is present and its position within the network provide important information in defining its function (Jaeger *et al.,* 2008). This also applies to the analysis of proteins of known function which often play a role in different complexes and subcellular compartments. Thus, large-scale functional proteomics projects, which build interaction networks of protein complexes, can help in understanding the cross-talk that goes on between the different complexes of unconnected cellular activities (reviewed by Bauer *et al.*, 2003).

From a pharmacological point of view, numerous human diseases are caused by defects in cellular signal transduction pathways. These signaling pathways, themselves, are regulated by protein- protein interactions which often involve the assembly of large signaling complexes. Therefore, while studying protein complexes, if a protein is found to interact with proteins of a specific pathway, it can represent a new potential drug target of that pathway (reviewed by Shiota *et al*., 2008).

In the past, drug discovery efforts focused on a relatively limited number of proteins against which compounds could be developed such as, G-protein coupled receptors. The identification and validation of new drug targets

was a very difficult, time consuming and expensive process, discouraging investments in drug discovery programs (reviewed by Kramer *et al.*, 2004, and by Ruffner *et al.*, 2007). Nowadays, large-scale, genome-wide protein-protein interaction screens have a great potential in identifying novel therapeutic targets and are key in understanding the functions and molecular mechanisms of diseases (Ruffner *et al.*, 2007).

1-2.1) Methods for studying protein- protein interactions and protein complexes in eukaryotes:

Several different methods have been developed for studying protein-protein interactions such as affinity chromatography technique, phage display, co-immunoprecipitation experiments, and the BRET and FRET methods. Recently new approaches, such as LUMIER and PCA, have also been developed, but the main two methods used currently are the yeast two- hybrid method and the tandem affinity-purification procedure coupled to mass spectrometry.

1-2.1.1) The affinity chromatography technique:

This is a technique that requires the fusion of the protein of interest to an affinity tag. Cells are transfected with the plasmid coding for the protein of interest fused to the affinity tag (Berggård *et al.,* 2007). After an appropriate expression period, the cells are lysed and the tagged protein of interest is purified together with all its interaction partners using a specific ligand linked to a solid support (Berggård *et al.,* 2007). Eluted proteins are then separated on gel-electrophoresis and the proteins are identified by mass spectrometry (Berggård *et al*., 2007).

Although this technique could be used for the identification of protein-protein interactions, it is prone to generate higher backgrounds and thus, the small amounts of specific interaction partners may be masked by more abundant non- specific binding proteins (Berggård *et al.,* 2007). Another disadvantage of affinity chromatography is the fact that the protein of interest must be fused to a tag. The tag may cause a problem if it gets buried inside the complex and thus, the purification of the complex may not be achieved (Berggård *et al.,* 2007). An exogenous tag might also perturb the interaction of the fused protein of interest with other proteins in the complex (Berggård *et al.,* 2007).

1-2.1.2) Phage display:

Phages are viruses that infect bacterial cells (Smith *et al.,* 1997). In the phage display technique, the protein of interest are fused to a coat protein of a phage particle (Goodyear *et al.,* 2008) so that the protein of interest is displayed on the outer surface of a phage particle once released from the transformed bacterial cell (Smith *et al.,* 1997), thus making it accessible to other proteins for subsequent binding interactions (Goodyear *et al.,* 2008). This characteristic enables the selection of specific binding interactions from a mixture of nonbinding particles (Goodyear *et al.,* 2008). Selected phage particles are then eluted and used to infect fresh bacteria and the process is repeated until the desired enrichment is reached (Goodyear *et al., 2008*). Therefore, during the selection process, specific phage clones are progressively enriched on the basis of their specificity and affinity for a protein.

Although Phage display can be used in to identification of new protein-protein interactions, the technique itself suffers from several limitations such as, non-specific binding, elution efficiency, and differences in the ability of the eluted phages to re-infect and propagate in bacterial cells (Rhyner *et al.,* 2004). Recently, display systems, such as the bacterial 2-hybrid display system, have been developed and have shown to produce less background, than the classical

phage display system, in studying protein-protein interactions (Bair *et al.,* 2008).

1-2.1.3) Co-ImmunoPrecipitation (Co-IP) method:

Co-IP is a technique that uses the specificity of antibodies to purify a target protein together with its interaction partners (Kaboord *et al.*, 2008). To do so, whole cell extracts are prepared under nondenaturing conditions to maintain any interactions that occur. Cell extracts are then incubated with an antibody specific to the protein of interest "bait", provided that the antibody-protein interaction does not interfere in the interaction of the bait with other proteins within the cell extract (Yaciuck, 2007, and Miernyk *et al.*, 2008). The antibody-bound protein complex is then isolated on protein A or protein G sepharose beads. Proteins that did not bind to the beads are removed by a series of washes, and the bound protein complex is eluted from the beads. Members of the complex are then dissociated from each other by SDS sample buffer and then run on an SDS-PAGE followed western blotting with antibodies specific to the bait or interaction partners (Yaciuck, 2007, and Miernyk *et al.*, 2008).

1-2.1.4) BRET and FRET methods:

BRET (Bioluminescence Resonance Energy Transfer) is a method that takes the advantage of the resonance energy transfer between a luminescent donor (usually a luciferase) and a fluorescent acceptor (such as YFP). In this approach, the protein of interest must be fused to the bioluminescent donor and the suspected interacting partner must be fused to the fluorescent acceptor (note: the inverse is also plausible where the bait is fused to the fluorescent acceptor and the interacting partner is fused to the bioluminescent donor, since this methods only checks to see if an interaction occurs between two proteins). When the two proteins do not interact, only one signal, emitted by the luciferase, can be detected. It is only when the proteins are brought into close proximity of each other, typically at a distance of 100Å or less, that the

energy transfer occurs, and in this case an additional signal, emitted by the fluorescent protein, can be detected. This distance (100Å) is generally indicative of interaction between the two fused proteins of interest, either directly or as part of a complex (reviewed by Pfleger *et al.*, 2006, Boute *et al.,* 2002).

FRET (Fluorescence Resonance Energy Transfer), on the other hand, is a method that also exploits the advantage of resonance energy transfer, but this time between a fluorescent donor (example CFP) and a fluorescent acceptor (example YFP), provided that the two fluorophores are different. FRET is observed when the sample is exited at the wavelength of excitation of the donor (Tramier *et al.,* 2002, Evans *et al.,* 2006, and Gandía *et al.,* 2007). Now excited, the donor will emit energy at a wavelength that superimposes with the wavelength at which the acceptor is excited. This would only occur when the two proteins are in a close proximity of each other, 10-100Å in the case of FRET compared 100Å in the case of BRET, indicating an interaction between the two fused proteins of choice (Tramier *et al.,* 2002, Evans *et al.,* 2006, and Gandía *et al.,* 2007).

The advantages of the previous two approaches (BRET and FRET) are that they both measure protein-protein interactions (either direct or indirect interactions) *in vivo,* and do not require the lysis of cells or the purification of proteins. Their disadvantages, on the other hand, is that they require ectopic expression and/or overexpression of the fusion proteins which may cause artifacts that can either inhibit or induce protein- protein interactions (Bhat *et al.,* 2006). Furthermore, expensive high- resolution microscopes are required to measure both BRET and FRET.

1-2.1.5) LUMIER:

LUMIER (LUminescence-based Mammalian IntERactome) is an automated high-throughput technology aimed at mapping protein-protein

interaction networks systematically in mammalian cells (Barrios-Rodiles *et al.,* 2005). This approach uses *Renilla* luciferase enzyme fused to proteins of interest, which are then coexpressed with individual Flag-tagged partners in mammalian cells. Cell extracts are then prepared and the interactions are then determined by performing an *Renilla* luciferase enzymatic assay on the immunoprecipitates using an antibody against Flag (Barrios-Rodiles *et al.,* 2005).

An advantage of LUMIER is the fact that it can detect, in mammalian cells, interactions involving transmembrane receptors which play an important role in signaling networks (Barrios-Rodiles *et al.,* 2005). Previously, studying transmembrane receptors has shown to be a difficult task using other high-throughput approaches. However, the disadvantages are that the LUMIER technique is not able to measure absolute protein-protein interaction affinities, and is prone to generate false positives due to overexpression of the fused proteins (Barrios-Rodiles *et al.,* 2005).

1-2.1.6) The PCA approach:

PCA (protein-fragment complementation assays) is a technique that measures protein- protein interactions by fusing each of the proteins of interest to two fragments of a reporter protein that has been dissected into two fragments using protein-engineering strategies (Michnick *et al.,* 2007). If the interaction does indeed occur, the reporter-protein fragments are brought into proximity allowing them to fold together into the unique three-dimensional structure of the reporter protein and reconstitute its activity (Michnick *et al.,* 2007).

The advantage of PCA is that it can be created with many reporter proteins and thus provides for different types of readouts depending on the desired application (Michnick *et al.,* 2007). Furthermore, proteins are expressed in the relevant cellular context, and thus subcellular localizations and translocations of protein complexes can be determined (Michnick *et al.,* 2007).

On the other hand, PCA does require the fusion of the proteins of interest to unfolded fragments of the PCA reporter protein which could perturb the function of the proteins of interest.

LUMIER and PCA are both approaches that have not been widely used nor reported until now, but they for sure promise to serve in enhancing the confidence of protein-protein interactions by helping to describe the local topology of protein interaction networks (Figeys *et al.,* 2008).

1-2.1.7) The yeast two-hybrid method:

The yeast two- hybrid method is a sensitive, *in vivo* assay, which has proved, over the past several years, to be extremely effective in studying protein-protein interactions (Brachmann *et al.*, 1997, Ito *et al.,* 2001). In fact, the first interactome maps were obtained using this method (reviewed by Gingras *et al.,* 2007).

The method consists of developing yeast strains that carry a reporter gene, the most common of which is *LacZ*, with a unique promoter structure. The most common system used for studying protein-protein interactions is the GAL4 system which relies on a transcriptional readout for the detection of protein- protein interactions through the reconstitution of a functional transcriptional activator (Luban *et al.*, 1995, Brachmann *et al.*, 1997, and reviewed by Causier, 2004). GAL4 is a transcriptional activator containing two separate domains, the DNA-binding domain and the transcription activation domain. In two- hybrid assays, these two domains are separated from each other and fused to two proteins which are suspected to interact. In the case where an interaction does occur between the proteins, then the two domains are brought close together, reconstituting the GAL4 transcriptional activator. GAL4 can now potentiate the transcription of the reporter gene which is under its control (Luban *et al.*, 1995, Brachmann *et al.*, 1997, and reviewed by Causier, 2004).

The advantages of the two-hybrid method for studying protein- protein interactions are that, it is an *in vivo* assay, which is simple to set up, requires little optimization, and is inexpensive to use (reviewed by Causier, 2004). On the other hand, the disadvantage of the yeast two-hybrid method is that it cannot detect interactions between more than two proteins (Ito *et al.,* 2001, and Yu *et al., 2008*).

1-2.1.8) The tandem affinity-purification method coupled to mass spectrometry (AP-MS):

As discussed previously, important cellular functions, such as transcription, involve many polypeptides that assemble into multiprotein complexes of specific structures and compositions. Tandem-affinity purification coupled to mass spectrometry is a method that has proved invaluable in advancing the understanding of protein complexes (Gingras *et al.,* 2007). This is a technique that was originally developed in yeast, but could be adapted to various organisms.

Tandem affinity purification (TAP) is a technique that allows for the purification of protein complexes under native conditions (Rigault *et al.,* 1999). This approach requires the fusion of a TAP-tag to the protein of interest either at the C- or N-termini. The tag consists of two IgG-binding units of protein A of the *Staphylococcus aureus*, a cleavage site for tobacco etch virus (TEV) protease, and a calmodulin binding peptide (Drakas e*t al.,* 2005). Cell lines, carrying the tagged protein of interest, have to be created, from which macromolecular complexes can be isolated (Puig *et al.,* 2001). The tagged protein must be expressed, in these cell lines, near physiological levels to minimize the rate of false positives (non-specific protein-protein interactions). This could be achieved by the use of an ecdysone-inducible expression system.

The protein purification occurs in two steps using two different affinity columns (the IgG- and calmodulin-binding columns) under conditions that leave

proteins intact (Shevchenko *et al.,* 2002). Purified protein complexes are then analyzed by mass spectrometry to determine the interacting partners.

[Note: Details about how the cell lines are created, the ecdysone-inducible system, and how the protein complexes are purified are explained in section 3]

The main advantage of the AP-MS strategy, other than the fact that the purification, itself, occurs under native conditions, is that it is an easy and a very efficient large-scale purification technique. On the other hand, the main disadvantage is that this technique does require the fusion of a tag to the protein of interest which might obscure binding of the protein to its interacting partners. In addition, the tag might not be sufficiently exposed to allow for the binding of the protein complex to the affinity beads. Another problem that might be encountered is that the tagged protein or its interacting partners might be cut by the TEV protease, although this is unlikely to be frequent (Puig *et al.,* 2001).

Section 2: The Project and its objective

2-1) Previous work in Dr. Coulombe's laboratory:

Over the past years, Dr. Benoit Coulombe's laboratory has been interested in describing protein complexes and mapping their interaction networks in human cells. The project started with the purification of components of the RNAPII machinery. The TAP method was chosen to achieve the purification of the proteins of interest, coupled to mass spectrometry analysis and computational tools to build the interaction maps.

The project started with the affinity purification of 32 TAP- tagged proteins, resulting in a network of 820 protein- protein interactions in which we have high confidence (refer to Figure 4). This network includes many transcription factors, as expected, RNA processing factors, which is also expected since the RNA processing machinery is known to interact with the RNAPII machinery. The network also included several proteins involved in modulating the formation of protein complexes amongst which were four previously uncharacterized proteins that were named RNAPII associated proteins (RPAPs), RPAP1, RPAP2/C1ORF82, RPAP3/FLJ21908 and XAB1 (refer to Figure 5). These proteins are positioned at the interface between the RNAPII and a group of complexes such as the prefoldin complex, and the AAA+ chaperone-like ATPase RUVBL1 and RUVBL2, amongst other regulatory complexes. These RPAPs- XAB1 (XAB1 is also referred to as GPN1) proteins have also been observed to interact with some previously uncharacterized proteins (presented as yellow squares in Figure 5) such as MGC14560/ GPN3, LOC116143/ WDR92, NOP17/ PIH1D1, and PDRG1 (Jeronimo *et, al.,* 2007).

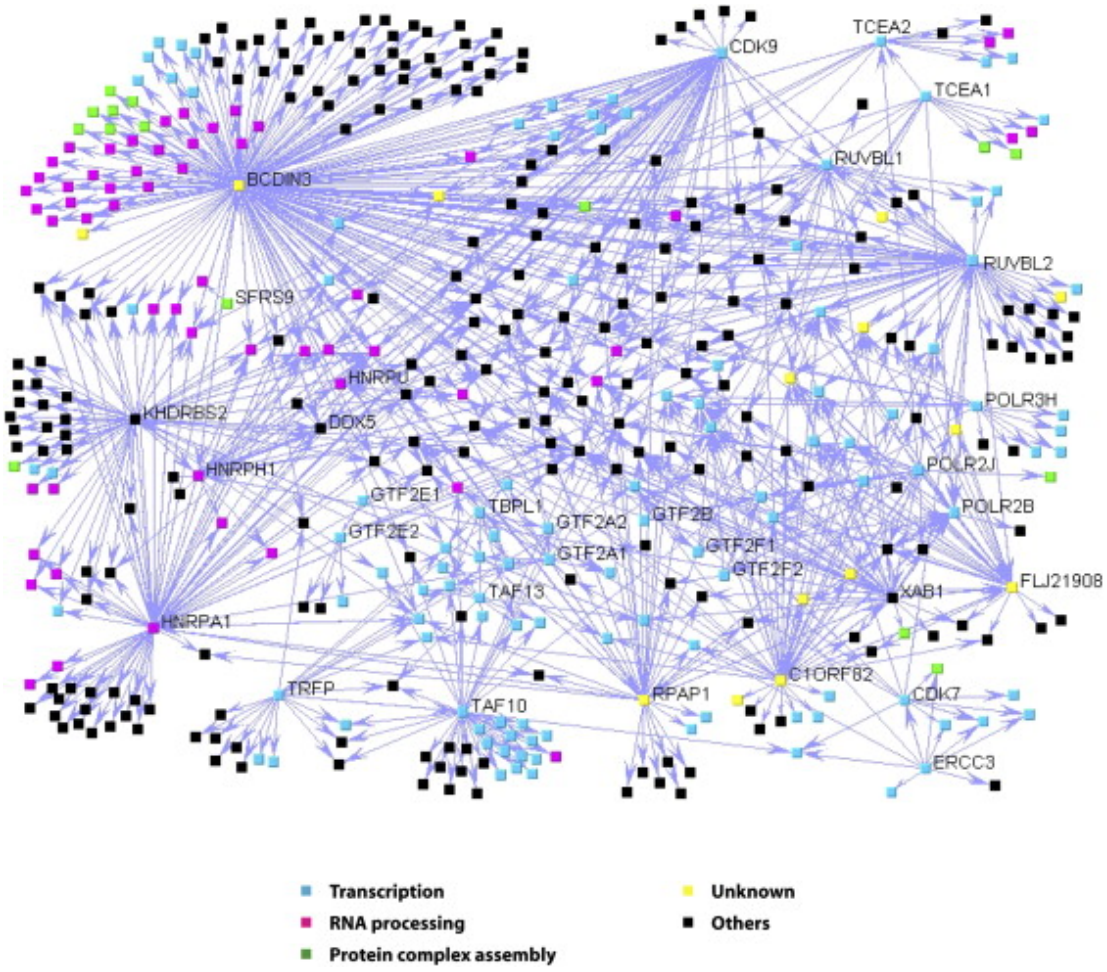


Figure 4: The network of protein interactions involving the RNAPII basal transcription machinery according to Jeronimo *et al.*, 2007.

Overview of the 820 high-confidence interactions obtained from 32 affinity- purified proteins.

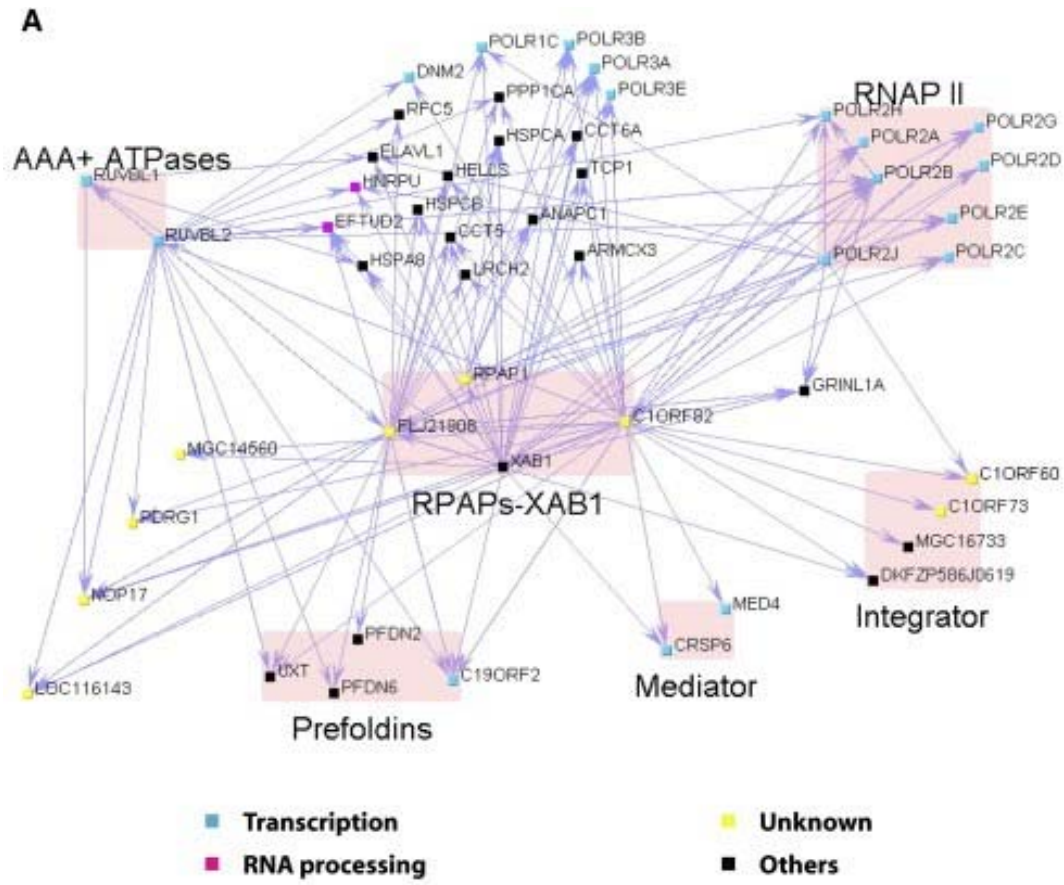


Figure 5: Network highlighting the interactions of RPAPs-XAB1 with RNAPII, the regulatory complexes integrator and mediator and a group of proteins with chaperone/scaffolding activity (Jeronimo *et, al.*, 2007) **Arrows point from bait to target. The TCP-1 ring chaperone complex is not shown in this graph.**

2-2) Objective of my project:

The aim of my project is to pursue the laboratory's survey of soluble human protein complexes containing components of the transcription machineries. This will be achieved by purifying eight newly-identified RNAPII interaction partners using the TAP method and further defining their interaction network and their connections with RNAPII. The data will serve to increase the precision of our protein interaction map of the RNAPII machinery, thereby providing a clearer picture of the molecular pathways that regulate this important enzyme. This knowledge is important since the deregulation of RNAPII activity could lead to diseases such as cancer and genetic disorders.

Four of the chosen RNAPII interaction partners are proteins of unknown function, namely MGC14560/GPN3, LOC116143/WDR92, PDRG1, NOP17/PIH1D1, which were observed to be interaction partners of the RPAPs-XAB1/GPN1. It was also observed that components of the prefoldin complex along with components of the TCP-1 ring complex also interacted with RPAPs-XAB/GPN1, mainly RPAP2, RPAP3 and XAB1/GPN1. Therefore, PFDN2 of the prefoldin complex, and CCT4 and CCT5 of the TCP-1 ring complex, were chosen for TAP purification and mass spectrometry analysis. What was also of interest to us is the fact that four of the chosen proteins for TAP purification, mainly WDR92, PIH1D1, PDRG1, and PFDN2, were reported to be members of the URI/Prefoldin complex (Sardiu *et al.,* 2007), which by itself has no specific function until now (The URI/Prefoldin complex will be discussed later on in section 4). The last protein chosen for TAP purification was the KIAA0406 protein observed to be an interaction partner of the AAA^+ chaperone-like ATPases RUVBL1 and RUVBL2.

My project mainly consisted of generating human cell lines expressing Tandem Affinity Peptide (TAP)-tagged versions of these eight RNAPII-interacting partners, purifying the protein complexes using tandem affinity

chromatography, identifying the interaction partners using mass spectrometry, and applying specific computational tools to build the interaction networks.

Information such as these will help to shed some light not only on the role these chosen proteins might playing in mammalian cells and the effects they might have on regulating the activity and assembly of RNAPII, but also on the role the URI/Prefoldin complex might be playing in human cells by purifying four of its members (WDR92, PIH1D1, PDRG1, and PFDN2) using the TAP method.

Section3: *Materials and Methods*

3-1) The TAP purification method:

TAP (Tandem Affinity Peptide) purification is a double affinity purification approach used to purify protein complexes under native conditions in cells. This technique has been optimized in our lab for the purification of soluble protein complexes in mammalian cells. The TAP method (summarized in Figure 6) consists of four main steps;

1) Constructing a vector encoding the protein of interest fused at its C-terminus to the double affinity tag.
2) Creating human cell lines carrying the tagged protein.
3) Purifying the complex in which the tagged protein is present by double affinity purification.
4) Identifying the interaction partners of the protein of interest by mass spectrometry.

Once the interacting partners have been identified, computational tools are applied to select high confidence interactions. It is worth mentioning that, it is crucial to achieve near physiological level of the tagged protein in the human cell line, therefore, avoiding an overexpression of the protein that can lead to non-specific protein interactions, and thus, measuring the rate of false positives. This is why a system is needed in which the expression of the tagged protein is under the control of an inducer. The system used in our lab is the ecdysone-inducible expression system (Figure 7), where the EcR293s cells containing the plasmid, pVgRXR, constitutively express the retinoic acid receptor (RXR), the ecdysone receptor (VgEcR) that contains the transactivation domain of Vp16, and the bleocin resistance gene. In the presence of the inducer (Ponasterone A), RXR and VgEcR heterodimerize and bind to the specific ecdysone response element (E/GRE) found on the pMZI vector, therefore, inducing the expression of the fused protein of interest.

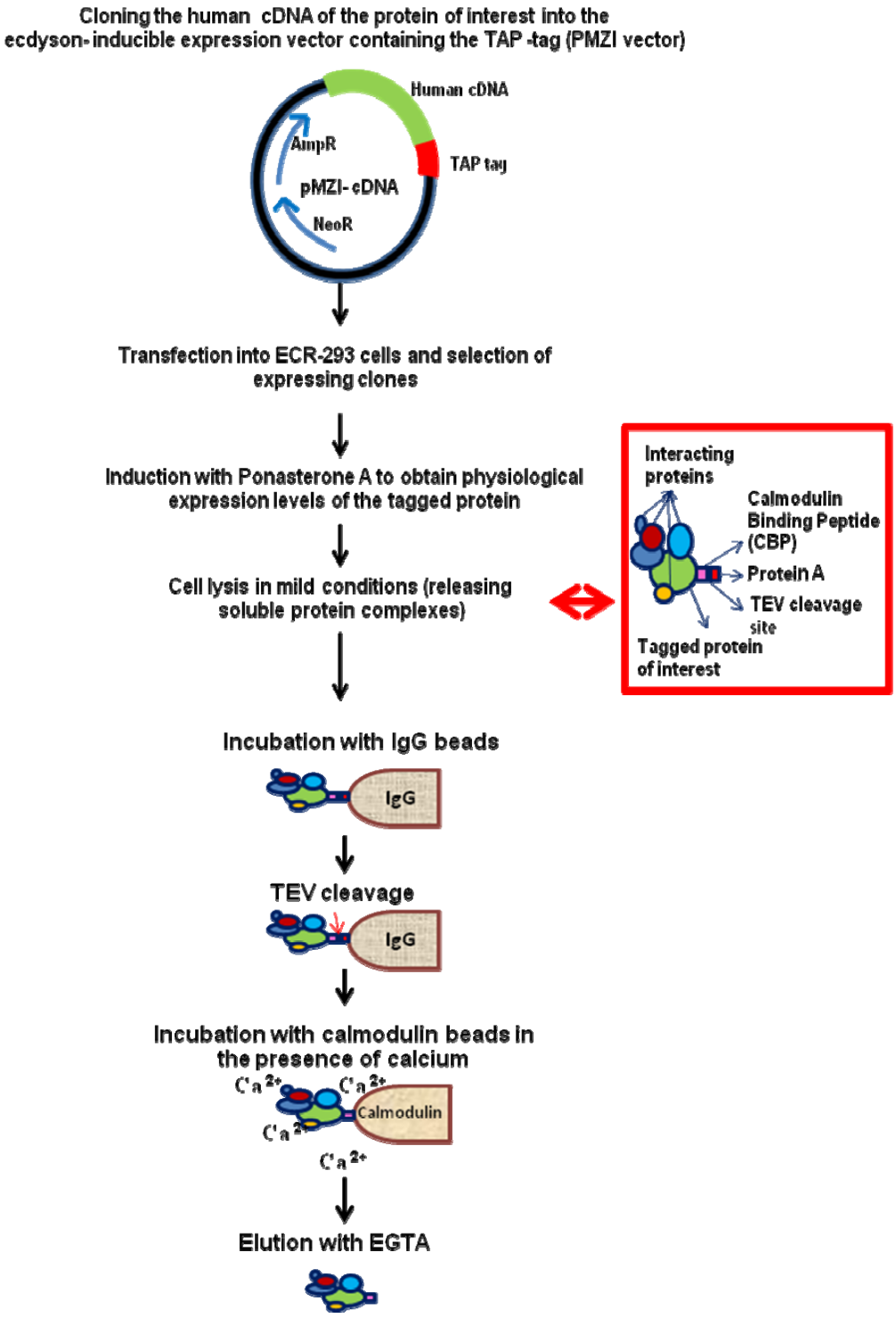


Figure 6: Schematic representation of the TAP procedure

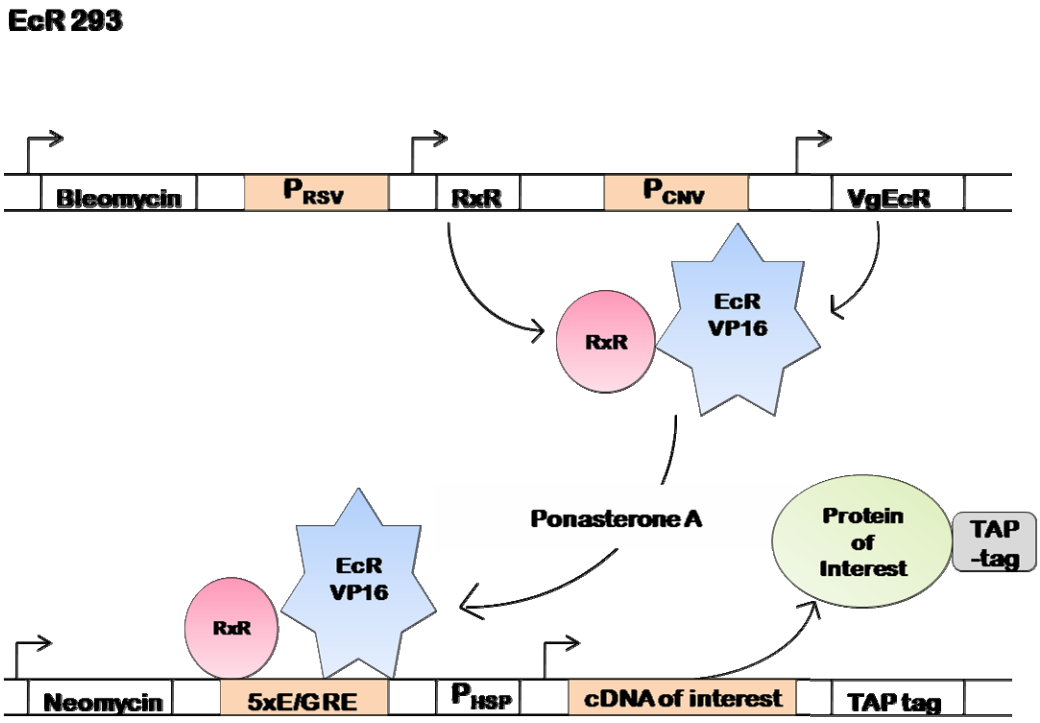


Figure 7: Schematic representation of the ecdysone- inducible system in the mammalian EcR293 cells

The upper part represents the vector, present in the EcR293 cells, is the vector encoding the retinoic acid receptor (RXR) and the ecdysone receptor (VgEcR), both of which are constitutively expressed. The lower part shows the vector is the pMZI vector carrying the gene of interest fused, C-terminally, to the TAP- tag vector, transfected into the mammalian EcR293 cells.

3.1.1) Constructing a vector encoding the protein of interest fused at its C-terminus to the tag:

The aim of this step is to create a vector in which the protein of interest is fused to the double affinity tag. To do so, the pMZI vector (refer to Figure 8) is used. This vector encodes a tandem affinity purification tag (TAP- tag) at its C-terminus, preceded by a multiple cloning site to which the cDNA of the protein of interest will be introduced. The TAP- tag itself consists of two binding sites, the *Staphylococcus aureus* protein A responsible for binding to the IgG beads, and the calmodulin binding peptide required for the binding to calmodulin beads, separated by the viral TEV protease cleavage site (refer to Figure 9).

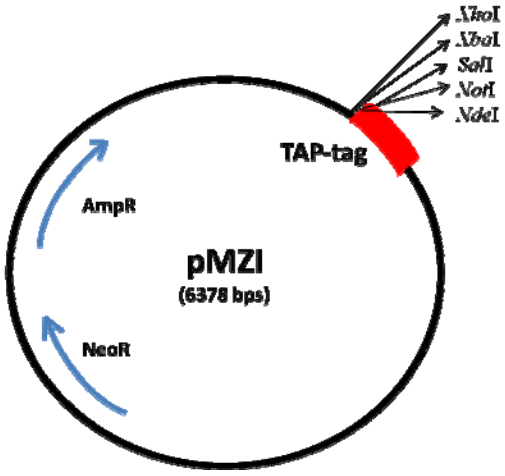


Figure 8: The pMZI vector

This plasmid consists of two resistance sites to neomycin and ampicillin, a multiple cloning site and the TAP-tag (shown as a red rectangle)

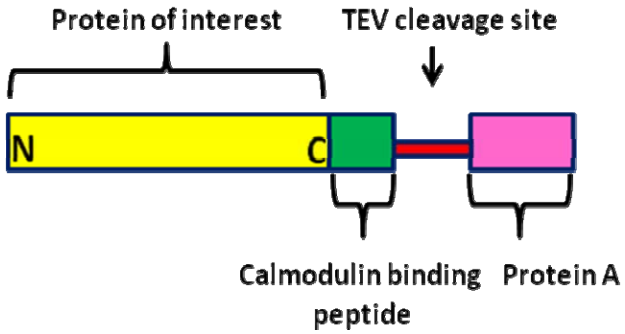


Figure 9: Schematic representation of the tandem-affinity purification tag fused to a protein at its C-terminus

In this project, the cDNA of eight different proteins were obtained from Invitrogen (for PFDN2, GPN3, PIH1D1 and WDR92), and Open Biosystems (for KIAA0406, PDRG1, CCT4 and CCT5), and had to be incorporated into the pMZI vector as follows:

3-1.1.1) PCR

PCR allows for the amplification of the cDNA of interest and the insertion of restriction sites at its 3' and 5' extremities. The restriction sites introduced were Xho1 and Not1 for PFDN2, PIH1D1, KIAA0406, PDRG1, WDR92, CCT4 and CCT5, and Sal1 and Not1 for GPN3. The PCR reactions were performed using either the Pfu turbo DNA polymerase (Stratagene) or the Phusion Hot Start DNA polymerase (Finnzymes). For PFDN2, PIH1D1 and WDR92, their PCR reactions (of a total volume of 50µl each) were performed using the Pfu turbo DNA polymearase and consisted of: 1ng of DNA, 1.75 µl of each forward and backward (10 µM) primers specific to the cDNA of interest, 0.75 µl of (10 mM) dNTPs, 1 µl of the Pfu turbo DNA polymerase, 5 µl of the Pfu turbo (10x) buffer. As for KIAA0406, PDRG1, GPN3, CCT4 and CCT5, their PCR reactions (of a total volume of 50 µl each) were performed using the Phusion Hot Start DNA polymerase, and consisted of: 1ng of DNA, 1.75 µl of

each forward and backward (10 µM) primers specific to the cDNA of interest, 0.75 µl of (10 mM) dNTPs, 1 µl of the Phusion Hot Start DNA polymerase, 10 µl of the Phusion buffer (5x). The PCR was run under the following conditions; For the Phusion Hot start DNA polymerase: 98°C for 30 secs, 98°C for 10 secs, (Tm- 5)°C for 30 secs, 72°C for 30 secs/ Kb, then the last three steps are repeated 25 times, and then 72°C for 10 mins, and finally hold at 12°C. As for the Pfu turbo DNA polymerase: 95°C for 2 mins, 95°C for 30 secs, (Tm- 5)°C for 30 secs, 72°C for 1 min/Kb. The last three steps are repeated 30 times, and then 72°C for 10 mins, and finally hold at 12°C.

3-1.1.2) Enzymatic digestion

The PCR product is then purified using the Qiagen kit and digested with Xho1 and Not1 or Sal1 and Not1 from the New England Biolabs. Purified pMZI is also digested with Xho1 and Not1 or Sa1 and Not1. The digestion is performed for approximately sixteen hours using the buffer recommended by New England Biolabs. Finally the digested samples are run on a 1% agarose gel stained with ethidium bromide. Then a gel extraction is performed using the Qiagen gel extraction kit. A dilution of 1(DNA): 25 (water) for each sample, and the concentration is then measured using a spectrometer at an OD of 260 nm.

3-1.1.3) Ligation

The ligation reaction is performed overnight at 4°C and consisted of 1 µl of the T4 DNA ligase (Invitrogen), 4 µl of the T4 ligase buffer (5X), 250 pmol of the digested cDNA of interest and 5 fmol of the digested pMZI vector.

3-1.1.4) Transformation into XL-1 cells

100 µl of competent bacterial cells (XL-1 cells) are added to 10 µl of the ligation sample and incubated for 30 mins on ice. Heat shock is then performed on the samples for 90 secs at 42°C and the samples are then incubated on ice

for 2 mins. 900 µl of LB medium is then added to the sample and the reaction is incubated for 1 hr, at 37°C, with agitation. The sample is then centrifuged to allow for the sedimentation of the cells and only a 50 µl of the supernatant is kept to suspend the cells and plate them on ampicillin plates (100 µg/ml of ampicllin). The plates are incubated overnight at 37°C. Minipreps are then prepared by adding to 5 ml of LB medium, 1 µl of ampicillin and one picked colony, this will allow the plasmid to grow to a bigger quantity. These minipreps are incubated overnight at 37°C with agitation and then miniprep purification is performed using the Qiagen kit and the samples are sequenced.

3-1.2) Creating human cell lines carrying the tagged protein of interest:

The goal of this step is to isolate stable human cell line that expresses the tagged protein, near physiological level, when induced with a specific concentration of ponasterone A, the inducer.

EcR293s cells are grown in DMEM medium containing 10% FBS (fetal bovine serum), 2 mM glutamine, 100 µg/ml penicillin, 100 µg/ml streptomycin and 30 µg/ml of bleocin. Cells are grown 20-24 hrs in an incubator at 37°C and 5% CO_2, before transfection with a mixture of 500 µl of (2x) HeBs (a 0.5 mM $Na_2HPO_4.7H_2O$, 274 mM NaCl, 54.5 mM HEPES, at a pH of 7.0) to which 8 µg/µl of DNA, 61 µl of 2 M $CaCl_2$ and 430 µl of H_2O. This mixture is incubated for 10 mins before adding it to the EcR293 cells. 12-14 hrs later, the cells are harvested and transferred to a DMEM medium containing 10% FBS, 2 mM glutamine, 100 µg/ml penicillin, 100 µg/ml streptomycin, 30 µg/ml bleocin and 300 µg/ml of geneticin. Clones that grow are isolated and transferred to other plates, containing medium with geneticin, to grow to larger quantities.

3-1.3) Screening tests:

To test the expression of the tagged protein in the selected cell lines, cells are grown in medium with geneticin and are induced with 3 µM ponasterone A for approximately 24 hrs, and then the cells are lysed with a lysis buffer (0.1% Triton x-100, 0.1 M Tris-HCl pH8.0) and loaded on a 4-12% SDS gel that will be used in western blot. The antibody used is Pol II (C-21) sc-900 from Santa Cruz, an affinity-purified rabbit polyclonal IgG raised against the CTD of the Rpb1 of RNA polymerase II of mouse origin. Clones that express the tagged protein are then frozen and kept at -80°C for later TAP purification.

3-1.4) Purifying the complex in which our protein of interest is present by double affinity purification:

Clones expressing the tagged protein of interest are grown in large quantities. These quantities are divided into two, a part which will be induced with 3 µM Ponasterone A for 42 hrs before harvest and a part which will not be induced, but will rather be used as a negative control to minimize false-positives (such as non-specific binding of soluble proteins to either the IgG beads or the calmodulin beads).

Harvested cells are homogenized on ice to lyse the cells releasing the whole cellular extract. Homogenization is performed, manually, using two different buffers, 10 strokes with each. The first round of strokes is with 4/3 volume of buffer A (4 ml of buffer A to 3 grams of cells) which consists of 10 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 10 mM of KCl, 0.5 mM of DTT, 0.5 mM of AEBSF and 1 tablet of complete EDTA-free (from Roche). The second round of strokes is with 1 volume of buffer B which consists of 50 mM Hepes pH 7.9, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM AEBSF, 1.26 M potassium acetate and glycerol. Ultracentrifugation is then performed on the samples for 3 hrs at a

speed of 37K and a temperature of 4°C, and then dialyses are performed overnight at 4°C in 3 liters of dialysis buffer (10 mM Hepes pH 7.9, 0.1 mM EDTA, 0.1 mM DTT, 0.1 M potassium acetate, 10% glycerol).

3-1.4.1) First affinity column:

The whole cell extract sample is added to IgG beads (50 µl per 1 g of cells) washed with IPP buffer (10 mM Tris pH 8, 100 mM NaCl, 0.1% Triton X100 and 10% glycerol) and incubated for 1 hr at 4°C with rotation. The beads are then washed few times with the IPP buffer and added to a column and then washed with the TEV buffer (10 mM Tris pH 8, 100 mM NaCl, 0.1% Triton X100, 0.5 mM EDTA, 10% glycerol and 1 mM DTT). Finally the column is sealed at the bottom and TEV buffer is added to the beads and then 30 units (3 µl for 1 g of cells) of TEV protease (Invitrogen) are added to liberate the proteins bound to the IgG beads. This mixture is incubated overnight at 4°C

3-1.4.2) Second affinity column:

The eluate from the first column is added to 50 µl of calmodulin beads washed with CBB buffer (10 mM Tris pH 8, 100 mM NaCl, 1 mM Immidazole, 1 mM magnesium acetate, 2 mM $CaCl_2$, 0.1% Triton X100, 10% glycerol and 10 mM β-mercaptoethanol). In the presence of $CaCl_2$ the mixture is incubated, with rotation, at 4°C for 2 hrs. The beads are then transferred to a second column and washed a few times with the CBB buffer and finally are eluted, in two steps, with the CEB buffer which consists of 2 mM EGTA, a calcium chelating agent that will break the binding of the CBP to the calmodulin beads, helping the elution process. CEB buffer also consists of 10 mM Tris pH 8, 100 mM NaCl, 1 mM Immidazole, 1 mM magnesium acetate and 10 mM β-mercaptoethanol.

3-1.5) Identifying the interaction partners of the protein of interest by mass spectrometry:

Once the screening tests are performed, the chosen cell lines, expressing the tagged proteins, are cultivated in larger quantities. For each cell lines, a quantity was induced with 3 μM Ponasterone A for 42 hours before harvest, and another quantity was cultivated in the absence of Ponasterone A, and was used as a negative control to eliminate false-positives, as discussed earlier in section 3. The cells are then harvested and lysed, and the tagged proteins "protein complexes" are purified, following the protocol described in section 3, in the absence of detergent and high salt concentrations to preserve the integrity of the purified protein complexes (Coulombe *et al.,* 2008). Once purified, the induced and non-induced samples are loaded on a 10% SDS gel stained with silver nitrate.

Finally the gel is cut into slices for both induced (ponasterone A induced) and non-induced samples. This will facilitate the elimination of non-specific interactions from the induced sample. Samples are then digested with trypsin, preparing them for analysis by mass spectrometry (MS). MS analysis for this project was performed at IRCM Proteomics Discovery Platform by Dr. Denis Faubert. The mass spectrometer used is the tandem mass spectrometer (LC-MS/MS) with microcapillary reversed-phase high- pressure liquid chromatography coupled to an LCQ DecaXP (ThermoFinnigan), LTQ or LTQ-Orbitrap (Thermoelectron) quadrupole ion trap mass spectrometer with a nanospray interface (Jeronimo *et al*., 2007). Mass spectrometry assigns a Mass Spectrometer (MS) score to each identified interacting partner.

3-1.6) Applying computational tools to select high confidence interactions:

The list of interaction partners identified by MS for a given bait (tagged protein) is likely to contain a certain fraction of false positives as a result of indirect interactions, incorrect peptide identifications, contamination of gel lanes, etc… (Cloutier *et al.,* submitted article). To select high- confidence interactions and minimize the number of false positives in our data set, our laboratory has developed a computational method that, first, filters out spurious interactions and, second, attributes an interaction reliability score (IRS) to each protein interaction detected by MS (Jeronimo *et al.,* 2007).

Therefore, for each purification, the non- specific interactions are filtered out by selecting proteins with an MS score that is at least 5 times higher in the induced sample compared to its equivalent non-induced sample (Jeronimo *et al.,* 2007). Once the spurious interaction partners have been filtered out, each putative protein interaction remaining is assigned an IRS based on the MS score of the interaction partner, and the local topology of the interaction (Jeronimo *et al.,* 2007). The IRS is computed using the Naïve Bayes classifier, trained on a set of 135 positive interactions validated from the literature and a set of 53 negative interactions that were judged likely to be false-positives (Jeronimo *et al.,* 2007). Finally, to select high confidence interactions, an IRS threshold (IRS above 0.6729) was chosen above which it is predicted that 17% of literature-supported interactions (false-negatives) will be missed, and 17% of interactions without literature support (false-positives) will be kept (Coulombe *et al.,* 2008). Using the web-based software VisANT, protein interaction maps could be built, with the selected high confidence interactions, to serve in representing the data in a comprehensive manner (Jeronimo *et al.,* 2007, Cloutier *et al.,* submitted article).

Section 4: Results

As mentioned earlier (in section 2), eight new RNAPII interaction partners have been purified using the tandem affinity purification methods to purify the complex in which they are present. Their interaction partners were then analyzed by MS and computational tools were finally applied to build their interaction maps. The eight proteins are GPN3, WDR92, PIH1D1, PDRG1, KIAA0406, PFDN2, CCT4 and CCT5. Following are the results of their purification.

Note:

a) The purification step, for three of the eight proteins (PIH1D1, CCT5 and PFDN2), was performed by Philippe Cloutier.

b) Under screening tests, the western blots performed to check for the expression of the tagged protein in the selected cell lines are presented. Each lane, on the western blots, represents a different cell line except for the last lane which was loaded with a control sample (a cell line that has been already tested for the expression of the tagged protein of interest). The red arrow points to the purified tagged protein, the blue arrow points to the purified tagged control protein, and the green arrow points to the cell line from which the tagged protein of interest was purified using the TAP method.

Furthermore, the highest molecular weight bands, which are observed in figures 11-B, 13-B (most clearly in lanes L2 and F3), represent Rpb1. This is due to the use of the C21 polyclonal antibody which is raised against the CTD of Rpb1 of RNAPII. As for the bands that appear at around 250 KDa in the same figures (also most clearly in lanes L2 and F3), these are unknowns but could be a non-specific protein that is recognized by the C21 antibody.

c) For each TAP gel (silver stained SDS gel) presented next, the main interaction partners, as identified by MS, for each purified tagged protein are presented. A more complete list of interaction partners, following the filtration of non-specific interactions, is represented in the "Mascot results" figures. Furthermore, the red box on each TAP gel encloses the purified tagged protein.

d) Whenever an interaction between two proteins is said to be "confirmed", this means that both proteins co-purified with each other by TAP, and had an interaction IRS above threshold (0.6729), in each case. Furthermore, a purified protein is said to be "previously observed" to interact with another protein, referring to previous TAP purification experiments performed in our laboratory.

4-1) Purification of GPN3/ MGC14560:

This protein is a GTPase that was observed to interact with RPAP2 and XAB1/GPN1 (Jeronimo *et al.,* 2007). GPN3 stands for GPN-loop GTPase 3 and thus, contains a GPN loop together with an "ATP-binding" domain (represented in Figure 10-A). Otherwise, nothing much is known about its function.

Figure 10-A: Linear representation of GPN3 cDNA showing the conserved domain.

MPRYAQLVMGPAGSGKSTYCATMVQHCEALNRSVQVVNLDPAAEHFNYSVMADIRELIEVQDVM
EDDSLRFGPNGGLVFCMEYFANNFDWLENCLGHVEDDYILFDCPGQIELYTHLPVMKQLVQQLE
QWEFRVCGVFLVDSQFMVESFKFISGILAALSAMISLEIPQVNIMTKMDLLSKKAKKEIEKFLD
PDMYSLLEDSTSDLRSKKFKKLTKAICGLIDDYSMVRFLPYDQSDEESMNIALQHIDFAIQYGE
DLEFKEPKEREDESSSMFDEYFQECQDE

Figure 10-B: Screening test

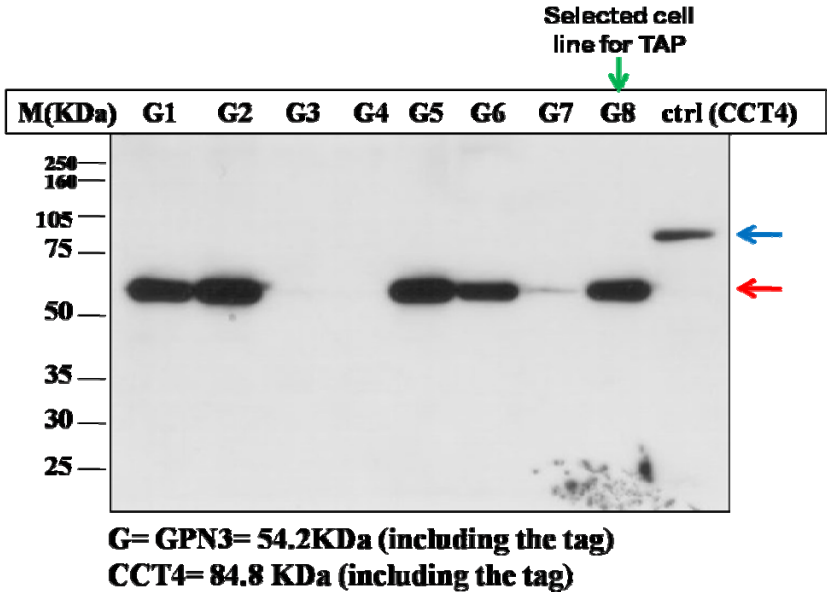


Figure 10-C: TAP gel

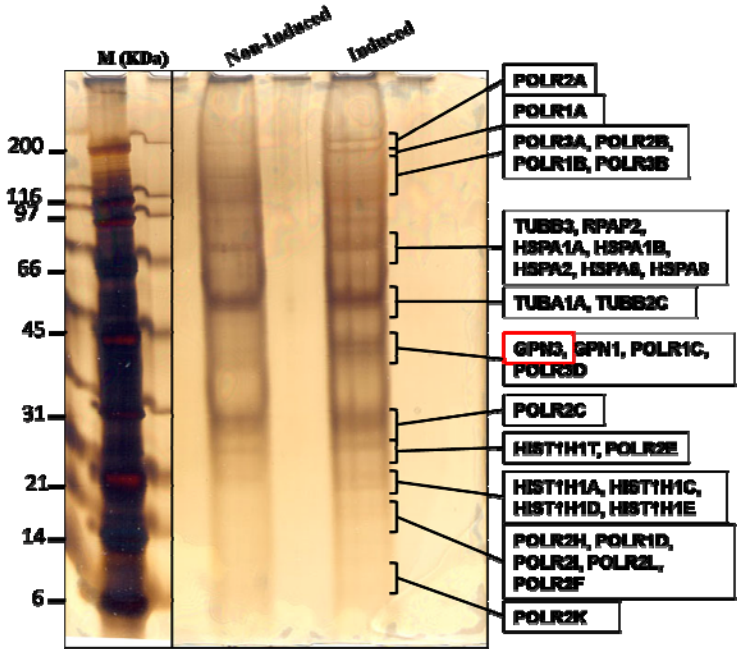


Figure 10-D: Mascot results

Target	IRS	Target	IRS	Target	IRS	Target	IRS
HSPA1A	0.0	ACTA1	0.035	HIST1H1E	0.099	TUBB3	0.999
HSPA1B	0.0	ACTA2	0.035	HIST1H1T	0.099	GPN1	1.0
113425742	0.026	ACTB	0.035	TCOF1	0.143	HNRNPU	1.0
GNPTG	0.026	ACTG1	0.035	POLR1D	0.165	POLR1A	1.0
SLC32A1	0.026	ACTG2	0.035	POLR2I	0.225	POLR2A	1.0
ZW10	0.026	A26C1A	0.037	HSPA2	0.523	POLR2B	1.0
119572092	0.027	A26C1B	0.037	HSPA8	0.523	POLR2C	1.0
MCM7	0.027	BAT2D1	0.040	POLR2F	0.890	POLR2H	1.0
POLR2K	0.027	POLR1B	0.041	POLR3A	0.902	RPAP2	1.0
POLR2L	0.027	PYGM	0.042	POLR3B	0.910	TUBA1A	1.0
TUBA3C	0.027	ACTC1	0.043	POLR1C	0.945	TUBB2C	1.0
TUBA3D	0.027	HIST1H1A	0.099	POLR2E	0.996		
TUBA3E	0.027	HIST1H1C	0.099	ATPBD1C	0.999		
POLR3D	0.028	HIST1H1D	0.099	HSPA6	0.999		

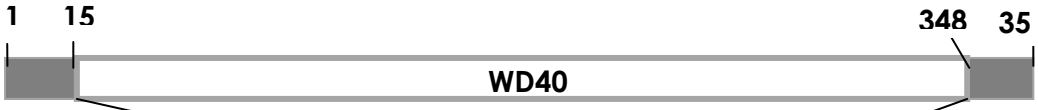
GPN3, as mentioned earlier, is an ATP-binding protein, of an unknown function that was previously picked up (co-purified) with purified RPAP2 and XAB1/GPN1 (also an ATP-binding protein) which, themselves, are proteins of an unknown function. Therefore, the purpose behind purifying GPN3 and identifying its interaction partners was not only to enrich our data set, but also to form an idea about the kind of a role this protein might be playing in regulating the human RNA polymerase II activity.

TAP purification results of GPN3 show that it interacts with heat shock proteins, which are molecular chaperones that play a role in protein-protein interactions such as protein folding and intracellular transport of proteins. Furthermore, TAP purification results of both GPN3 and previous purification of XAB1/GPN1 show that these proteins also interact with actins and tubulins. XAB1/GPN1 have also been shown to interact with few subunits of the TCP-1 ring complex, in previous TAP purification (Jeronimo *et al.,* 2007), which is mainly involved in the folding of actins and tubulins.

4-2) Purification of WDR92/ LOC116143:

WDR92 is a protein that has been observed to interact with RPAP2, RPAP3, XAB1/GPN1 and RUVBL2 to mention a few (Jeronimo *et al.,* 2007). WDR92 (WD Repeat domain 92) contains a WD40 repeat domain (represented in Figure 11-A), which are common structural module in eukaryotes. Proteins containing WD40 domains have a diverse range of biological functions, including signal transduction, cell cycle regulation, RNA splicing, and transcription (Saeki *et al.,* 2006). WDR92 has been suggested to play a role as a novel modulator of apoptosis pathway since its overexpression in HEK 293 cells potentiated apoptosis and caspase-3 activation induced by tumor necrosis factor-α (TNF- α) and cycloheximide (CHX) (Saeki *et al.,* 2006).

Figure 11-A: Linear representation of WDR92 cDNA showing the conserved domains



MSAFEKPQIIAHIQKGFNYTVFDCKWVPCSAKFVTMGNFARGTGVIQLYEIQHGDLKLLREIEK
AKPIKCGTFGATSLQQRYLATGDFGGNLHIWNLEAPEMPVYSVKGHKEIINAIDGIGGLGIGEG
APEIVTGSRDGTVKVWDPRQKDDPVANMEPVQGENKRDCWTVAFGNAYNQEERVVCAGYDNGDI
KLFDLRNMALRWETNIKNGVCSLEFDRKDISMNKLVATSLEGKFHVFDMRTQHPTKGFASVSEK
AHKSTVWQVRHLPQNRELFLTAGGAGGLHLWKYEYPIQRSKKDSEGIEMGVAGSVSLLQNVTLS
TQPISSLDWSPDKRGLCVCSSFDQTVRVLIVTKLNKI

Figure 11-B: Screening test (the red boxes enclose the screening tests of WDR92).

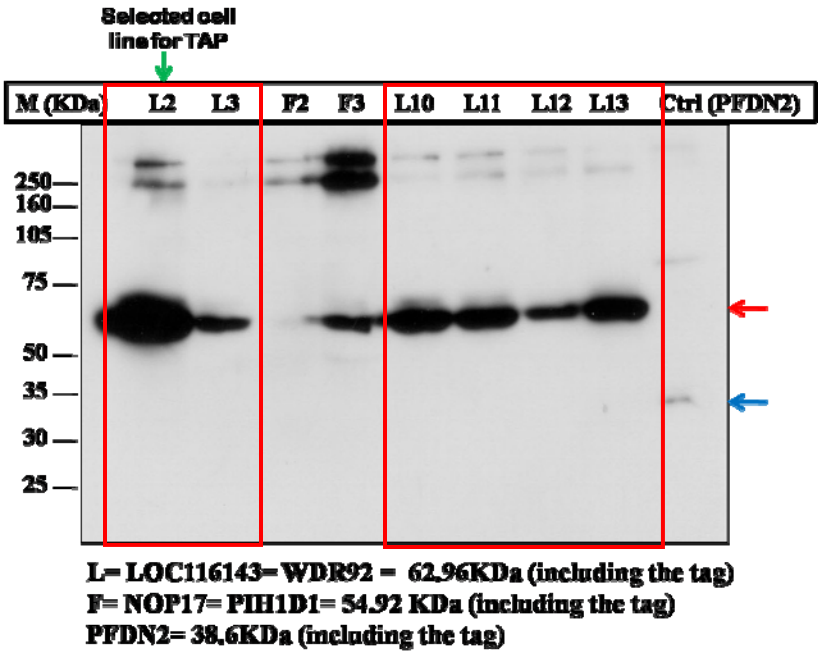


Figure 11-C: TAP gel

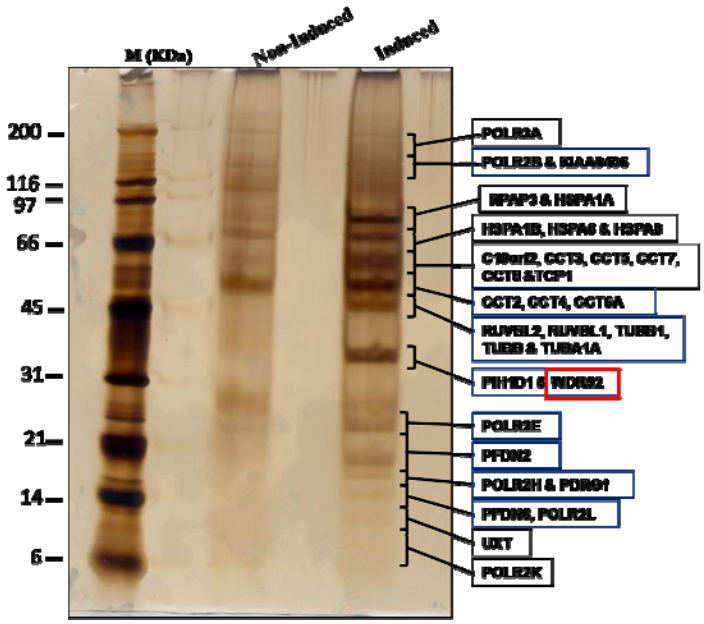


Figure 11-D: Mascot results

Target	IRS	Target	IRS	Target	IRS	Target	IRS
ALB	0.0	SAT1	0.026	C19orf2	1.0	POLR3A	1.0
HSPA1A	0.0	SUV420H2	0.026	CCT2	1.0	RPAP3	1.0
HSPA1B	0.0	LOC730826	0.104	CCT3	1.0	RUVBL1	1.0
HSPA6	0.0	TTC27	0.104	CCT4	1.0	RUVBL2	1.0
PDRG1	0.0	UTP14A	0.104	CCT5	1.0	TCP1	1.0
PFDN2	0.0	TTC6	0.106	CCT6A	1.0	TUBA1A	1.0
PIH1D1	0.0	POLR2K	0.110	CCT6B	1.0	TUBB	1.0
POLR2E	0.0	A26C1A	0.112	CCT7	1.0		
UXT	0.0	KIF3A	0.147	CCT8	1.0		
WDR92	0.0	TUBB1	0.185	HSPA8	1.0		
OPA1	0.025	GFAP	0.2	KIAA0406	1.0		
POLR2L	0.025	NCOA5	0.442	PFDN6	1.0		
SIGLEC11	0.025	PLG	0.442	POLR2B	1.0		
THEX1	0.025	CDO1	0.470	POLR2H	1.0		

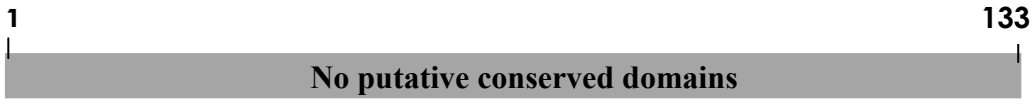
TAP purification of WDR92 did confirm its interaction with RPAP3 and RUVBL2 but not with RPAP2 and XAB1/GPN1. The interaction of WDR92 with RPAP3 was, furthermore, confirmed recently by a group in Osaka University, Japan, where V5-tagged WDR92 was subjected to affinity purification using V5 antibodies. Mass spectrometry analysis of their purified V5-tagged WDR92 identified RPAP3 as one of its interaction partners (Itsuki *et al.,* 2008).

What was also interesting to observe, in our results, was the fact that the eight subunits of the TCP-1 ring complex, co-purified with WDR92. On the other hand the TAP purification of both CCT4 and CCT5 (presented below), which are components of the TCP-1 ring complex, did not yield WDR92 as an interaction partner. Further interaction partners obtained will be discussed later on when talking about the URI/Prefoldin complex.

4-3) Purification of PDRG1:

PDRG1 stands for p53 and DNA damage regulated 1, and does not contain any conserved domain. It was observed to be an interaction partner of RPAP3, XAB1/GPN1 and RUVBL2 (Jeronimo *et al.,* 2007). PDRG1 mRNA has been shown to be upregulated by ultraviolet radiation (UV) and downregulated by tumor suppressor p53. In fact, analysis of the PDRG1 promoter revealed the presence of Oct-1-binding element recognized by oct-1, a transcription factor that is activated by UV. Furthermore, a putative head-to–tail type p53-binding site was also found in the PDRG1 promoter (Luo, Huang, and Sheikh, 2003).

Figure 12-A: Linear representation of PDRG1 cDNA



MLSPEAERVLRYLVEVEELAEEVLADKRQIVDLDTKRNQNREGLRALQKDLSLSEDVMVCFGNM
FIKMPHPETKEMIEKDQDHLDKEIEKLRKQLKVKVNRLFEAQGKPELKGFNLNPLNQDELKALK
VILKG

Figure 12-B: Screening test

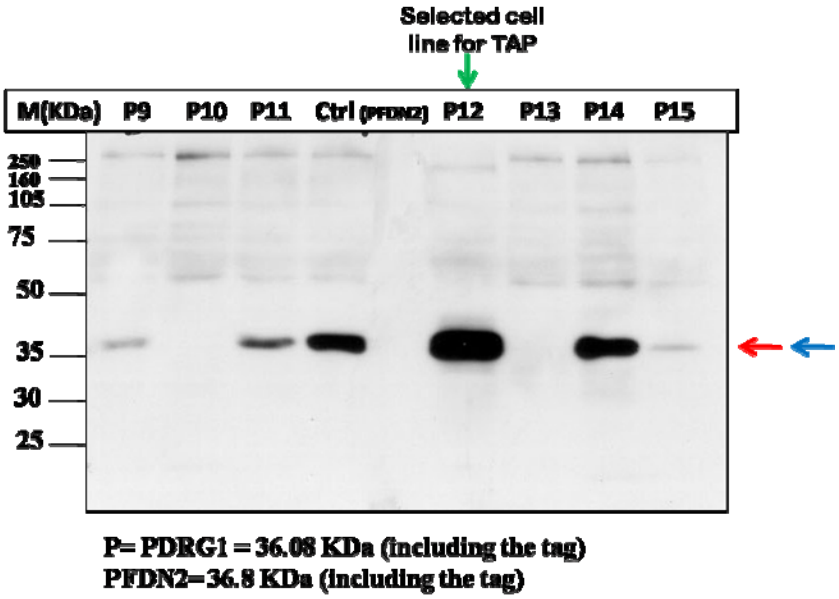


Figure 12-C: TAP gel.

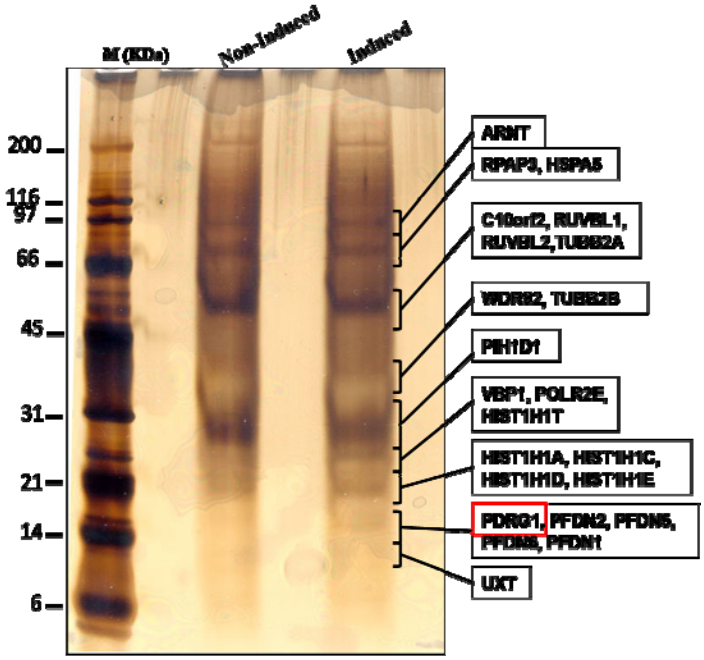


Figure 12-D: Mascot results

Target	IRS	Target	IRS	Target	IRS	Target	IRS
PDRG1	0.0	VBP1	0.041	HIST1H1C	0.436	119600484	1.0
SNRPB	0.0	RBM14	0.051	HIST1H1D	0.436	PFDN2	1.0
SNRPD3	0.0	RBM22	0.056	HIST1H1E	0.436	PFDN6	1.0
SNRPE	0.0	CYorf15A	0.104	ARHGEF12	0.442	PIH1D1	1.0
SNRPN	0.0	TARDBP	0.111	BRD4	0.443	POLR2E	1.0
CDH19	0.025	YTHDF2	0.119	SRI	0.500	RPAP3	1.0
SFRS11	0.025	KLC1	0.127	CSNK2B	0.589	RUVBL1	1.0
CYB	0.026	ARNT	0.144	LONP1	0.749	RUVBL2	1.0
119606377	0.026	DDX39	0.154	PFDN1	0.999	TUBB2A	1.0
SDCCAG1	0.026	HIST1H1T	0.280	HSPA5	1.0	TUBB2B	1.0
ISPF6484	0.026	HIST1H1A	0.280	C19orf2	1.0	UXT	1.0
HMG20B	0.027	PFDN5	0.430	LOC100131673	1.0	WDR92	1.0

Tap purification of PDRG1 confirmed its interaction with RPAP3, XAB1/GPN1 and RUVBL2. What was also interesting to observe is that several of its most abundant interaction partners were prefoldin subunits, amongst other proteins which will be discussed later on, in the URI/Prefoldin complex section.

4-4) Purification of NOP17/ PIH1D1:

PIH1D1 is an interaction partner of RPAP2, RPAP3, XAB1/GPN1, RUVBL1 and RUVBL2 (Jeronimo *et al.,* 2007). PIH1D1 stands for PIH1 domain containing 1, and thus contains a PIH1 domain. In yeast, the homologue of PIH1D1 is Pih1 which was reported to be a member of a complex called R2TP. The R2TP complex is composed of Pih1 of course, together with a protein called Tah1 and the helicases Rvb1 and Rvb2 (the yeast homologues of RUVBL1 and RUVBL2). This complex plays an important role in the correct accumulation of box C/D small nucleolar ribonucleoproteins.

Furthermore, it was observed that Pih1 is an unstable protein that is stabilized by the assistance of chaperones. The chaperone involved in stabilizing this protein in log-phase yeast cultures is still unknown. On the other hand, in the stationary-phase of cells (the stress-phase in yeast cells) Pih1 was shown to be stabilized by Tah1 and the Hsp90 chaperone (Zhao *et al.,* 2008).

Figure 13-A: Linear representation of PIH1D1 cDNA showing the conserved domain

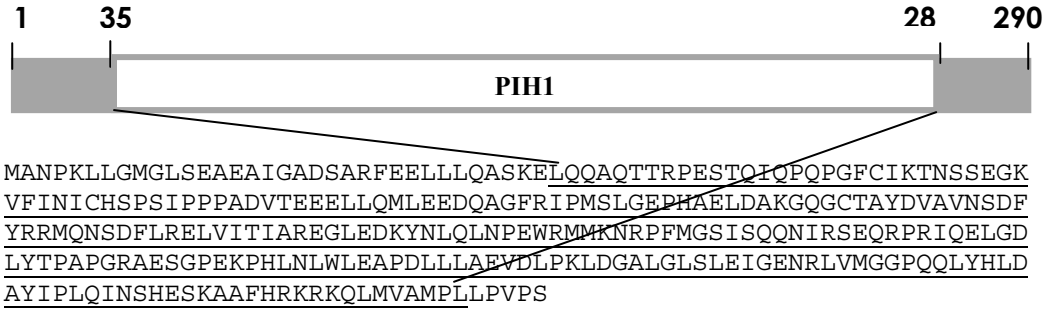


Figure 13-B: Screening test (the red box encloses the screening test of PIH1D1)

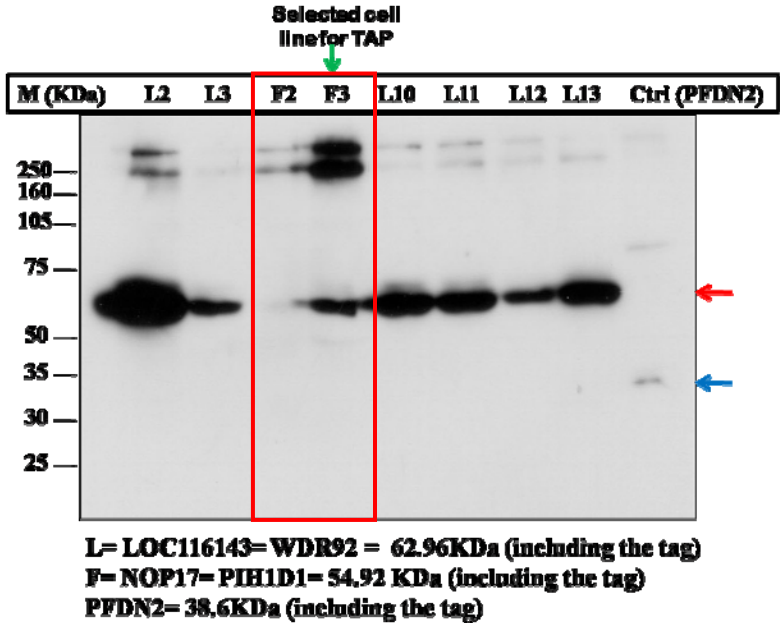


Figure 13-C: TAP gel.

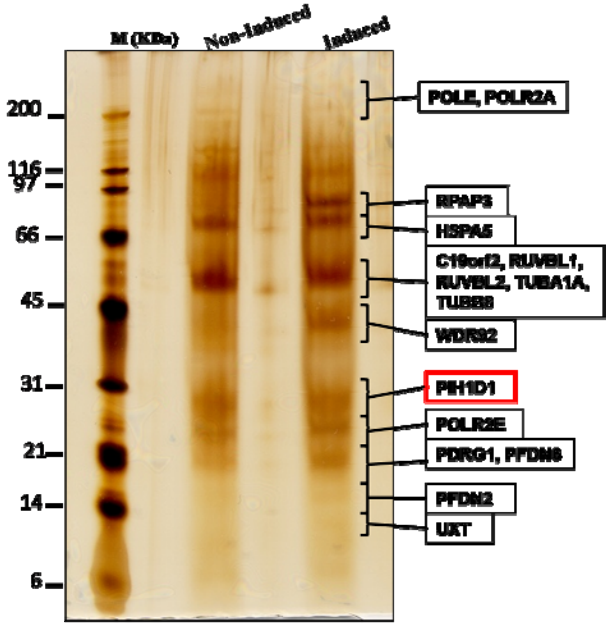


Figure 13-D: Mascot results

Target	IRS	Target	IRS	Target	IRS	Target	IRS
CCDC80	0.025	SPRR2D	0.031	A26C1A	0.165	TUBB8	1.0
DHX40	0.025	SPRR2F	0.031	INTS9	0.443	119600484	1.0
GCDH	0.025	SPRR2G	0.031	CHD3	0.472	LOC100131673	1.0
KIAA1602	0.025	CYorf15A	0.104	TMTC3	0.473	PDRG1	1.0
MYH10	0.025	POLE	0.104	POLR2A	1.0	PFDN2	1.0
SMARCC1	0.025	KIF3A	0.105	POLR2E	1.0	WDR92	1.0
CEP170	0.026	RNF103	0.105	HSPA5	1.0	PFDN6	1.0
KLC2	0.026	MYO5B	0.110	RPAP3	1.0	PIH1D1	1.0
RPL9	0.028	ACAA2	0.110	C19orf2	1.0	TUBA1A	1.0
SPRR2A	0.029	PLG	0.112	RUVBL1	1.0	TUBB4	1.0
SPRR2B	0.029	CKAP5	0.125	RUVBL2	1.0		
SPRR2E	0.029	NCKAP1	0.126	UXT	1.0		

By TAP purifying PIH1D1, its interactions with RPAP3, RUVBL1 and RUVBL2 were confirmed. On the other hand, an interaction with either RPAP2 or XAB1/GPN1 was not observed. Further identified interaction partners of PIH1D1 will be discussed next, in the URI/Prefoldin complex section

4-5) Purification of PFDN2:

PFDN2 interacts mainly with RPAP3 in the network of protein complexes involving the human RNAPII transcription machinery (Jeronimo *et al.,* 2007). As mentioned previously in this introduction, PFDN2 is a subunit of the prefoldin complex which plays a role as a cochaperone delivering cytosolic proteins to the TCP-1 ring complex for proper folding. PFDN2 contains a prefoldin_2 domain.

Recently, PFDN2, WDR92, PIH1D1 and PDRG1, have all been reported to be members of the URI/Prefoldin complex. This complex does not have any known function until now but as reported, it is composed of 10 proteins, including PFDN2, WDR93, PIH1D1, PDRG1, RPAP3, UXT, URI, PFDN6, Rpb5/POLR2E and HKE2 (Saridiu *et al.,* 2007).

Figure 14-A: Linear representation of PFDN2 cDNA showing the conserved domain

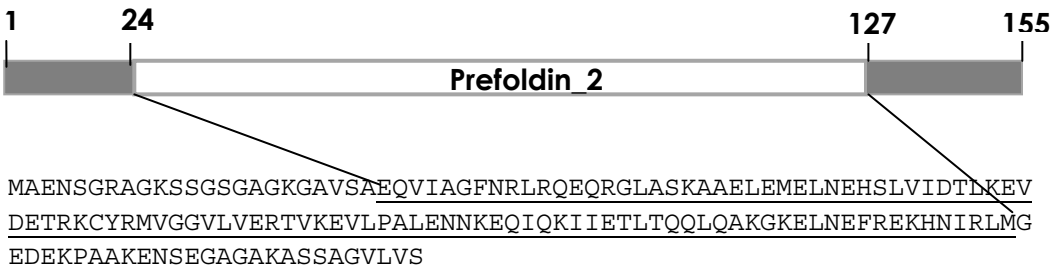
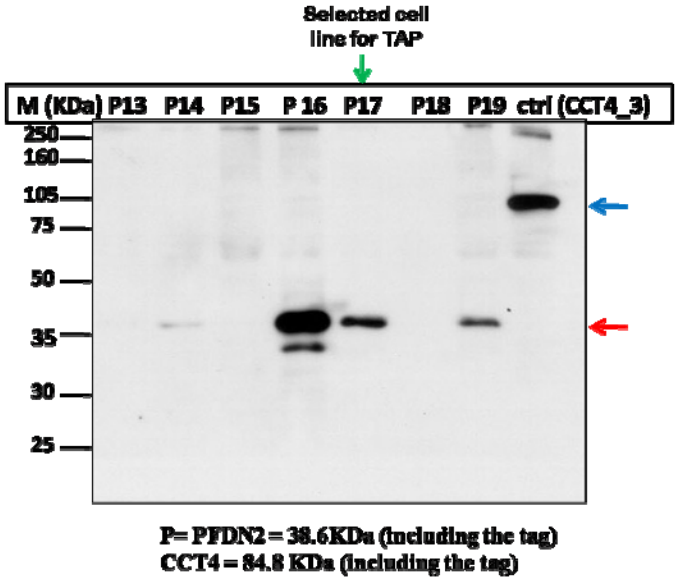


Figure 14-B: Screening test



The lowest molecular weight band that appears in lane P16 is an unknown which could be the result of degradation or splicing of the tagged protein of interest.

Figure 14-C: TAP gel.

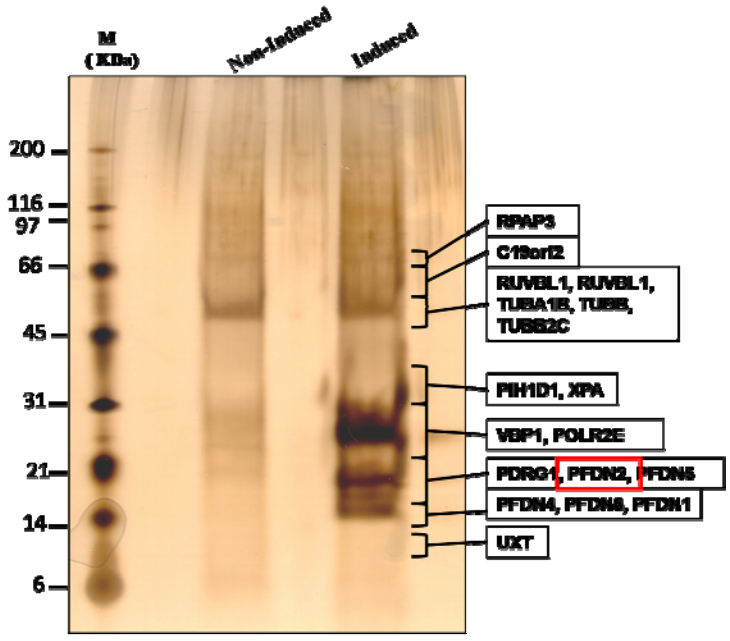


Figure 14-D: Mascot results

Target	IRS	Target	IRS	Target	IRS	Target	IRS
VBP1	0.0	TSPYL2	0.025	C15orf15	0.109	HNRNPA1	1.0
PFDN1	0.0	XPA	0.025	SYNE2	0.111	HSPA2	1.0
PFDN2	0.0	TRIP12	0.026	MORC1	0.112	LOC283767	1.0
PFDN5	0.0	DDX41	0.027	DNAH9	0.124	PCBP2	1.0
TUBA1B	0.0	NRBF2	0.027	SOS1	0.125	PDRG1	1.0
AP3B1	0.025	TET2	0.027	LOC641515	0.136	PFDN4	1.0
ATP9B	0.025	TRIM4	0.027	HNRNPR	0.143	PFDN6	1.0
C11orf63	0.025	C3	0.027	MYH11	0.290	PIH1D1	1.0
FAM133A	0.025	MYH14	0.038	KIN	0.391	POLR2E	1.0
ICA1	0.025	KIF5B	0.051	CKAP5	0.443	RPAP3	1.0
ITSN2	0.025	ANAPC1	0.104	Gcom1	0.447	RUVBL1	1.0
KIAA1826	0.025	CERKL	0.104	TCP1	0.470	RUVBL2	1.0
KIF3C	0.025	LOC730826	0.104	CUX1	0.498	TUBB4	1.0
LARP2	0.025	FLJ39660	0.104	HNRNPU	0.979	UXT	1.0
MRPL52	0.025	SFRS12IP1	0.104	HNRNPA2B1	0.989		
MYH3	0.025	RBM33	0.104	TUBB2C	0.999		
NSBP1	0.025	SH2D4B	0.104	TUBB	0.999		
PRKCSH	0.025	GBP2	0.106	C19orf2	1.0		
RPUSD2	0.025	GOLGA1	0.108	LOC100131673	1.0		
SND1	0.025	WDR87	0.109	CCT4	1.0		

As mentioned earlier, PFDN2 is a component of the prefoldin complex, the co-chaperone that helps deliver cytosolic proteins to the TCP-1 ring complex. The reason PFDN2 was chosen for TAP purification is because of the fact that it was previously observed to interact with RPAP3, and we wanted to confirm this interaction. TAP purification results of this protein did not yield any new interaction partners and, as expected, most of its abundant interaction partners are prefoldin subunits, amongst other proteins including RPAP3, thus, confirming the interaction between the two proteins.

4-6) Purification of the URI/Prefoldin complex:

WDR92, PIH1D1, PDRG1 and PFDN2 together with eight other proteins, mainly RPAP3, UXT, PFDN6, RUVBL1, RUVBL2, URI, RPB5, and POL3A, are all proteins reported to be members of a complex called the URI/Prefoldin complex (Sardiu *et al.,* 2007), which until now has no specific function. Tap purification of WDR92, PIH1D1, PDRG1, PFDN2 (results presented above), and previous TAP purification of RPAP3, PFDN6 (results reported in Jeronimo *et. al.,* 2007), and UXT (by Philippe Cloutier, unpublished data), resulted in the purification of almost all of the components of the URI/Prefoldin complex except one which is the POL3A. Several hypotheses could be made about the role the URI/Prefoldin complex might be playing in human cells, from the purification of several of its components in this project and other purifications performed in our laboratory. Most components of the URI/Prefoldin complex have been tagged in our laboratory, including PFDN6, UXT, RPAP3, RUVBL1, RUVBL2, and RPB5, and the hypotheses concerning this complex are discussed later in the "Discussion" section.

4-7) Purification of CCT4 and CCT5:

These two proteins are components of the TCP-1 ring complex discussed earlier which were observed to interact mainly with XAB1/GPN1 (Jeronimo *et al.,* 2007). The TCP-1 ring complex is a chaperone that assists in the folding of cytosolic proteins, mainly actin and tubulin. CCT5 contains a TCP1_epsilon conserved domain, and CCT4 contains a TCP1_delta conserved domain.

Figure 15-A.1: Linear representation of CCT4 cDNA showing the conserved domain

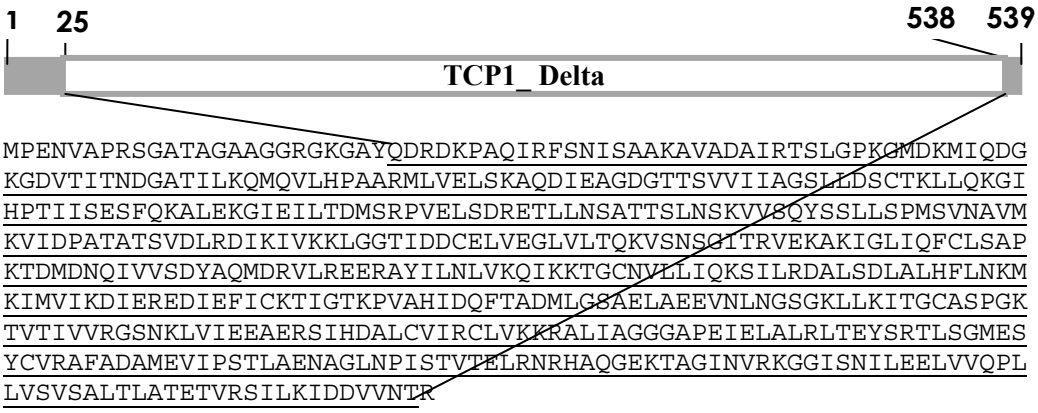


Figure 15-A.2: Linear representation of CCT5 cDNA showing the conserved domain

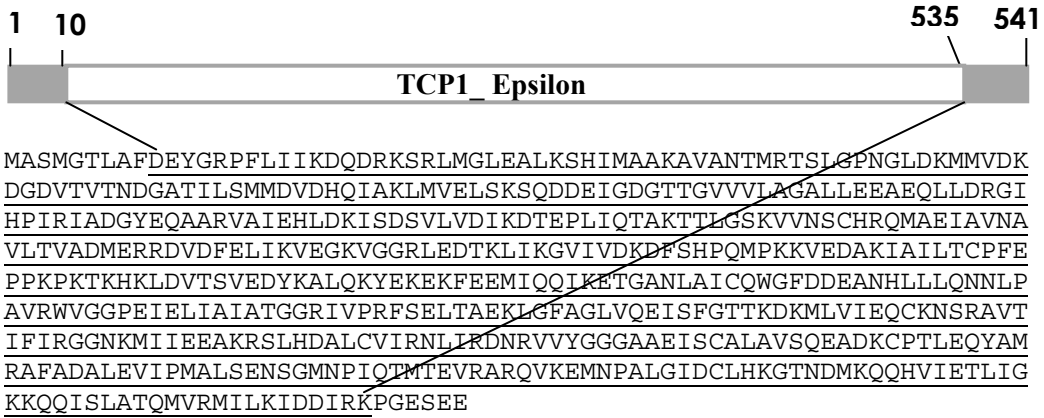


Figure 15-B.1: Screening tests for CCT4.

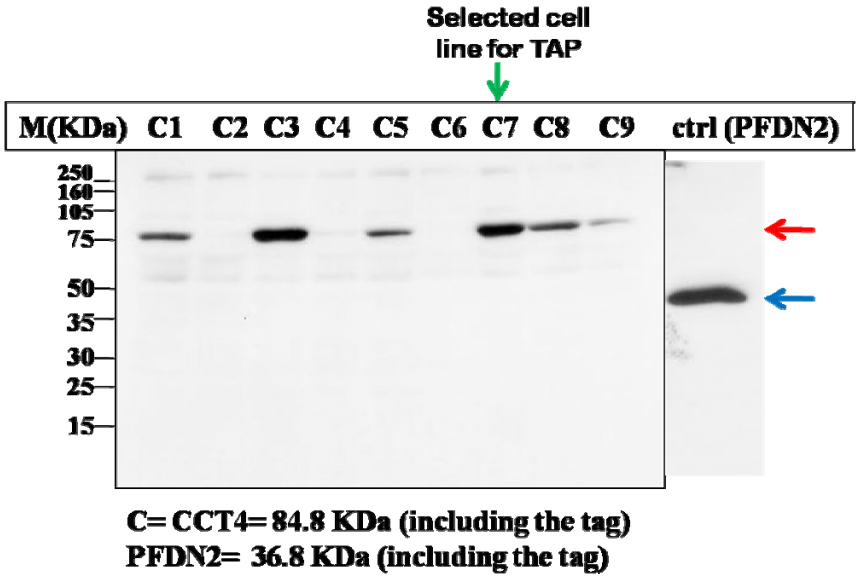
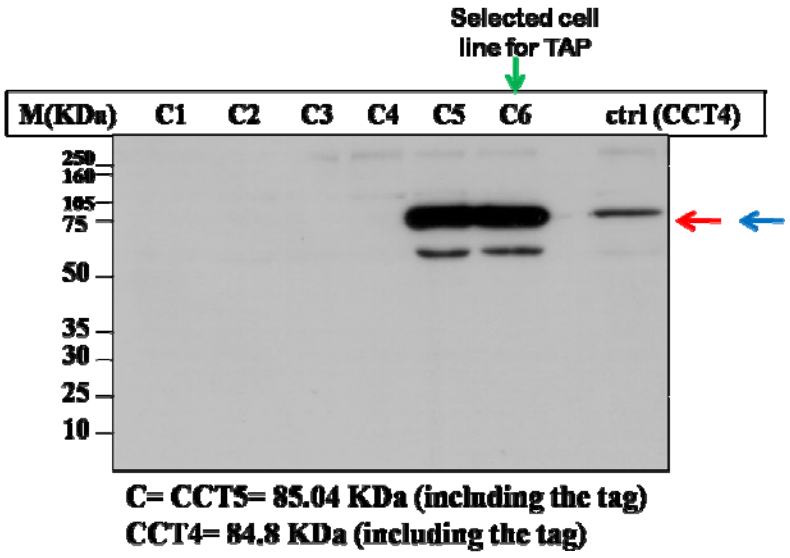


Figure 15-B.2: Screening tests for CCT5.



The lowest molecular weight bands which appear in lanes C5 and C6 are unknowns, but could be the result of degradation or splicing of the tagged protein of interest.

Figure 15-C.1: TAP gel for CCT4.

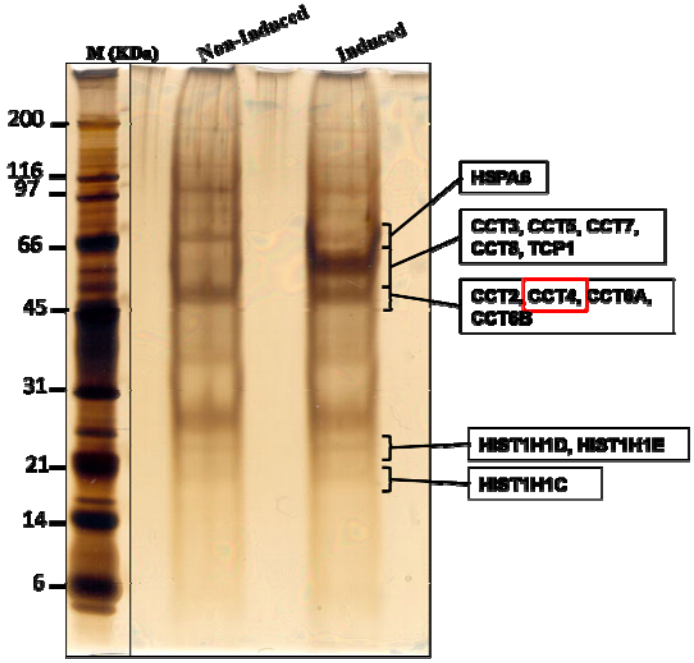


Figure 15-C.2: Tap gel for CCT5.

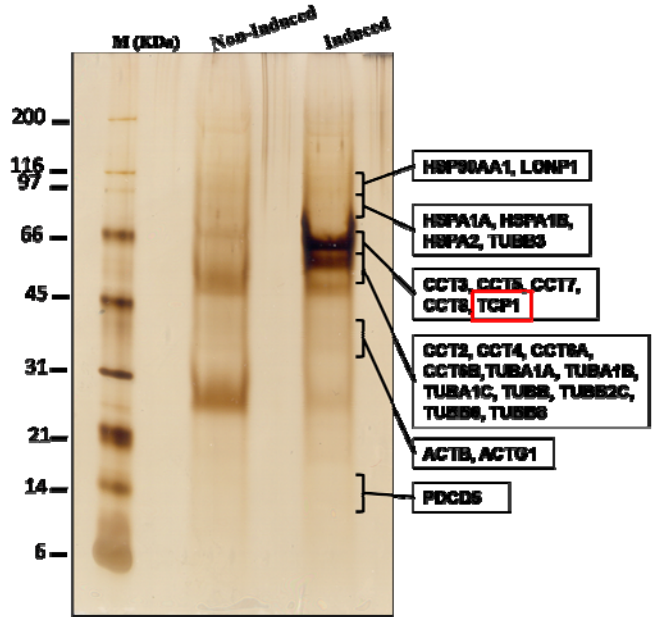


Figure 15-D.1: Mascot results for CCT4

Target	IRS	Target	IRS	Target	IRS	Target	IRS
CCT4	0.0	HAS1	0.1263	NCOA5	0.893	LONP1	1.0
HRNR	0.029	CDO1	0.1264	RPL22	0.941	CCT6A	1.0
PLG	0.029	LOC391656	0.167	CCT6B	0.9902	PCBP2	1.0
ACTB	0.0446	HIST1H1D	0.198	HSPA6	0.999	CCT7	1.0
ACTG1	0.0446	HIST1H1C	0.533	CCT2	1.0	CCT8	1.0
PCBP1	0.122	HIST1H1E	0.720	CCT3	1.0	TCP1	1.0
DLK1	0.1260	ARHGEF12	0.8901	CCT5	1.0		

Figure 15-D.2: Mascot results for CCT5

Target	IRS	Target	IRS	Target	IRS	Target	IRS
CCT5	0.0	PRIC285	0.029	TRAP1	0.264	DNM2	1.0
TUBA1A	0.0	TFDP1	0.029	PHGDH	0.292	HNRNPU	1.0
TUBA1B	0.0	TXNDC9	0.029	KIF5A	0.299	HSP90AA1	1.0
ACTB	0.0	CTR9	0.030	XPO5	0.353	HSPA1A	1.0
ACTG1	0.0	DYNC1H1	0.030	TUBB8	0.375	HSPA1B	1.0
TUBA1C	0.0	KCTD19	0.031	TUBB2A	0.457	KHSRP	1.0
TUBB	0.0	MOBKL1A	0.031	CDO1	0.463	KPNA2	1.0
BLM	0.025	MOBKL1B	0.031	C8orf42	0.565	KPNB1	1.0
LOC150759	0.025	TIMM13	0.033	CSE1L	0.606	LONP1	1.0
LYST	0.025	PSMA5	0.042	TUBA3C	0.641	PCBP1	1.0
PDCL3	0.025	KIF5B	0.086	XPO1	0.641	PDCD5	1.0
RC3H2	0.025	KIF5C	0.086	EIF4A2	0.822	RQCD1	1.0
SPAG17	0.025	NUDT21	0.091	CAND1	0.901	RUVBL1	1.0
THBS1	0.025	DIAPH1	0.104	EIF4A1	0.970	RUVBL2	1.0
TRPA1	0.025	KLHL3	0.104	HSPA2	0.977	TUBB2C	1.0
113422833	0.026	THOC4	0.104	INF2	0.999	TUBB3	1.0
ABCB1	0.026	TRIP12	0.104	BAT2D1	1.0	TUBB4	1.0
C3orf26	0.026	SAC	0.105	TCP1	1.0	TUBB6	1.0
RBM19	0.026	119583598	0.108	CCT2	1.0		
SEPX1	0.026	TUBB2B	0.109	CCT3	1.0		
SMC2	0.026	FANC1	0.111	CCT4	1.0		
CARS	0.027	EEF1A2	0.125	CCT6A	1.0		
CHPF	0.027	LOC391656	0.134	CCT6B	1.0		
RPL9	0.027	HAS1	0.144	CCT7	1.0		
UCK1	0.027	RBM14	0.201	CCT8	1.0		
UCK2	0.027	TUBA3D	0.207	CKAP5	1.0		
ZC3H12C	0.027	TUBA3E	0.207	DDX42	1.0		

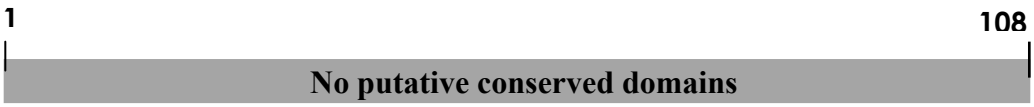
As mentioned earlier, CCT4 and CCT5 are components of the TCP-1 ring complex responsible for the folding of mainly actin and tubulin. Considering our hypothesis of the role of both XAB1/GPN1 and GPN3 in the formation of microtubules (discussed earlier in "Purification of GPN3" section), it was of interest to us to confirm the interaction of XAB1/GPN1 with TCP-1 ring complex. CCT4 and CCT5 were the chosen subunits since they both were shown to co-purify with the XAB1/GPN1 protein. TAP purification of both CCT4 and CCT5 did confirm their interaction with XAB1/GPN1.

It was of no surprise, of course, to find that the most abundant interaction partners of both CCT4 and CCT5 were subunits of the TCP-1 ring complex. On the other hand, the Tap purification of CCT5 did yield a new interaction partner, the PDCD5 protein.

4-8) Purification of KIAA0406:

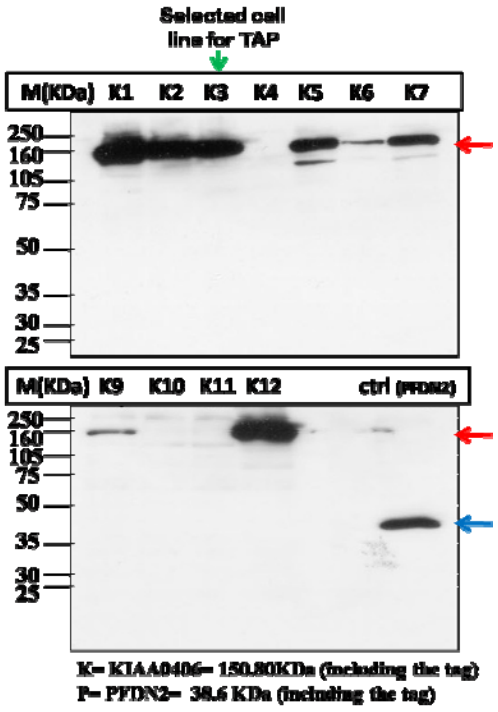
KIAA0406 does not contain any identifiable conserved domains. It has been observed to be an interaction partner of mainly RUVBL1 and RUVBL2 (Jeronimo *et al.,* 2007), but nothing much is known about the kind of a function it might be playing in human cells.

Figure 16-A: Linear representation of KIAA0406 cDNA



MAVFDTPEEAFGVLRPVCVQLTKTQTVENVEHLQTRLQAVSDSALQELQQYILFPLRFTLKTPG
PKRERLIQSVVECLTFVLSSTCVKEQELLQELFSELSACLYSPSSQKPAAVSEELKLAVIQGLS
TLMHSAYGDIILTFYEPSILPRLGFAVSLLLGLAEQEKSKQIKIAALKCLQVLLLQCDCQDHPR
SLDELEQKQLGDLFASFLPGISTALTRLITGDFKQGHSIVVSSLKIFYKTVSFIMADEQLKRIS
KVQAKPAVEHRVAELMVYREADWVKKTGDKLTILIKKIIECVSVHPHWKVRLELVELVEDLLLK
CSQSLVECAGPLLKALVGLVNDESPEIQAQCNKVLRHFADQKVVVGNKALADILSESLHSLATS
LPRLMNSQDDQGKFSTLSLLLGYLKLLGPKINFVLNSVAHLQRLSKALIQVLELDVADIKIVEE
RRWNSDDLNASPKTSATQPWNRIQRRYFRFFTDERIFMLLRQVCQLLGYYGNLYLLVDHFMELY
HQSVVYRKQAAMILNELVTGAAGLEVEDLHEKHIKTNPEELREIVTSILEEYTSQENWYLVTCL
ETEEMGEELMMEHPGLQAITSGEHTCQVTSFLAFSKPSPTICSMNSNIWQICIQLEGIGQFAYA
LGKDFCLLLMSALYPVLEKAGDQTLLISQVATSTMMDVCRACGYDSLQHLINQNSDYLVNGISL
NLRHLALHPHTPKVLEVMLRNSDANLLPLVADVVQDVLATLDQFYDKRAASFVSVLHALMAALA
QWFPDTGNLGHLQEQSLGEEGSHLNQRPAALEKSTTTAEDIEQFLLNYLKEKDVADGNVSDFDN
EEEEQSVPPKVDENDTRPDVEPPLPLQIQIAMDVMERCIHLLSDKNLQIRLKVLDVLDLCVVVL
QSHKNQLLPLAHQAWPSLVHRLTRDAPLAVLRAFKVLRTLGSKCGDFLRSRFCKDVLPKLAGSL
VTQAPISARAGPVYSHTLAFKLQLAVLQGLGPLCERLDLGEGDLNKVADACLIYLSVKQPVKLQ
EAARSVFLHLMKVDPDSTWFLLNELYCPVQFTPPHPSLHPVQLHGASGQQNPYTTNVLQLLKEL
Q

Figure 16-B: Screening test



The lowest molecular weight bands which appear in lanes K5 and K7 are unknowns, but could be due to degradation or splicing of the tagged protein of interest.

Figure 16-C: TAP gel.

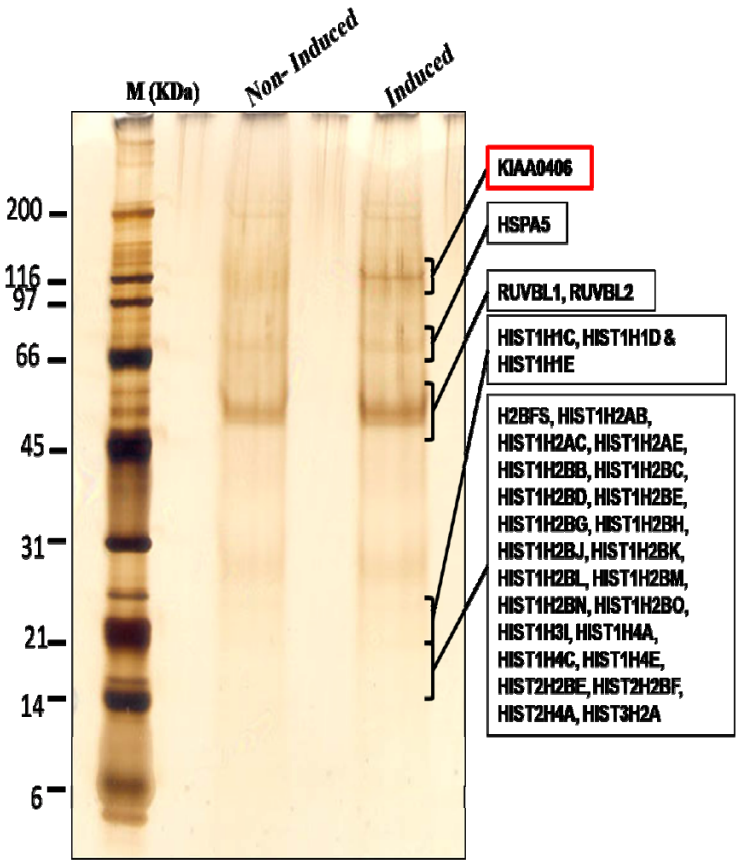


Figure 16-D: Mascot results

Target	IRS	Target	IRS	Target	IRS	Target	IRS
KIAA0406	0.0	TCOF1	0.029	HIST1H2AB	0.958	HIST1H2BH	1.0
113426227	0.025	ELAVL2	0.035	HIST1H2AC	0.958	TUBB4	1.0
113426576	0.025	LOC653781	0.044	HIST1H2AE	0.958	HIST1H2BK	1.0
CAMK2G	0.025	HNRNPA3	0.081	HIST3H2A	0.958	HIST1H2BL	1.0
ARID5B	0.025	HNRPA3	0.081	HIST1H1C	0.999	HIST1H2BM	1.0
BARD1	0.025	TOP2A	0.082	HIST1H1D	0.999	HIST1H2BN	1.0
CAMK2B	0.025	HSP90AA2	0.112	HIST1H1E	0.999	HIST1H4A	1.0
TTBK1	0.025	HSP90AB1	0.112	HIST1H2BB	0.999	HIST1H4C	1.0
hCG1779729	0.025	DCD	0.162	HIST1H2BJ	0.999	HIST1H4E	1.0
METTL3	0.025	HNRPH1	0.175	HIST1H2BO	0.999	RUVBL1	1.0
AZGP1	0.026	HNRPH2	0.175	HIST1H3I	0.999	RUVBL2	1.0
C1orf163	0.027	EIF4A1	0.264	HIST2H2BE	0.999	HIST2H2BF	1.0
CCDC102A	0.027	EIF4A2	0.264	H2BFS	1.0	HIST2H4A	1.0
GATAD2A	0.027	HSPA5	0.359	HIST1H2BC	1.0		
GATAD2B	0.027	HSP90AA1	0.463	HIST1H2BD	1.0		
DLK1	0.027	MATR3	0.559	HIST1H2BE	1.0		
HSP90AB2P	0.027	H2AFV	0.719	HIST1H2BG	1.0		

KIAA0406 is a protein of an unknown function, as mentioned earlier, which was shown to interact with RUVBL1 and RUVBL2, both of which are members of the AAA^+ family (ATPases associated with different cellular activities), in a previous purification performed in our lab. RUVBL1 and RUVBL2 are proteins usually found in a number of nuclear complexes, such as chromatin modifying complexes, transcription-activating complexes, and small nucleolar ribonucleoprotein complexes (snoRNPs) (Puri *et al.,* 2007).

By purifying KIAA0406, its interaction with both RUVBL1 and RUVBL2 was confirmed, since both of these proteins co-purified with it. It was also observed that the most abundant interaction partners that co-purified with this protein are mainly histones.

4-9) Graphical representation of the presented TAP results in this project:

Finally, using the web-based software VisANT, an interaction map showing all of the validated interactions for each purified protein can be generated and is shown in Figure 17 (next page). This figure represents only the interaction for which the IR score exceeded the threshold 0.6729 (validated interactions). The RPAP-GPN1 complex, the URI/Prefoldin complex, subunits of all three RNA polymerases amongst other complexes are also presented on this graph.

Note for Figure 17: Yellow nodes: TAP purified proteins. Blue nodes: Protein TAP purified previously in the laboratory. Green nodes: Identified interaction partners. Grey arrows: Points to targets of the previously purified proteins in our laboratory. Blue arrows point to the interaction partners of the 8 TAP purified proteins mentioned in this project. Red arrows are interactions amongst the URI/Prefoldin complex.

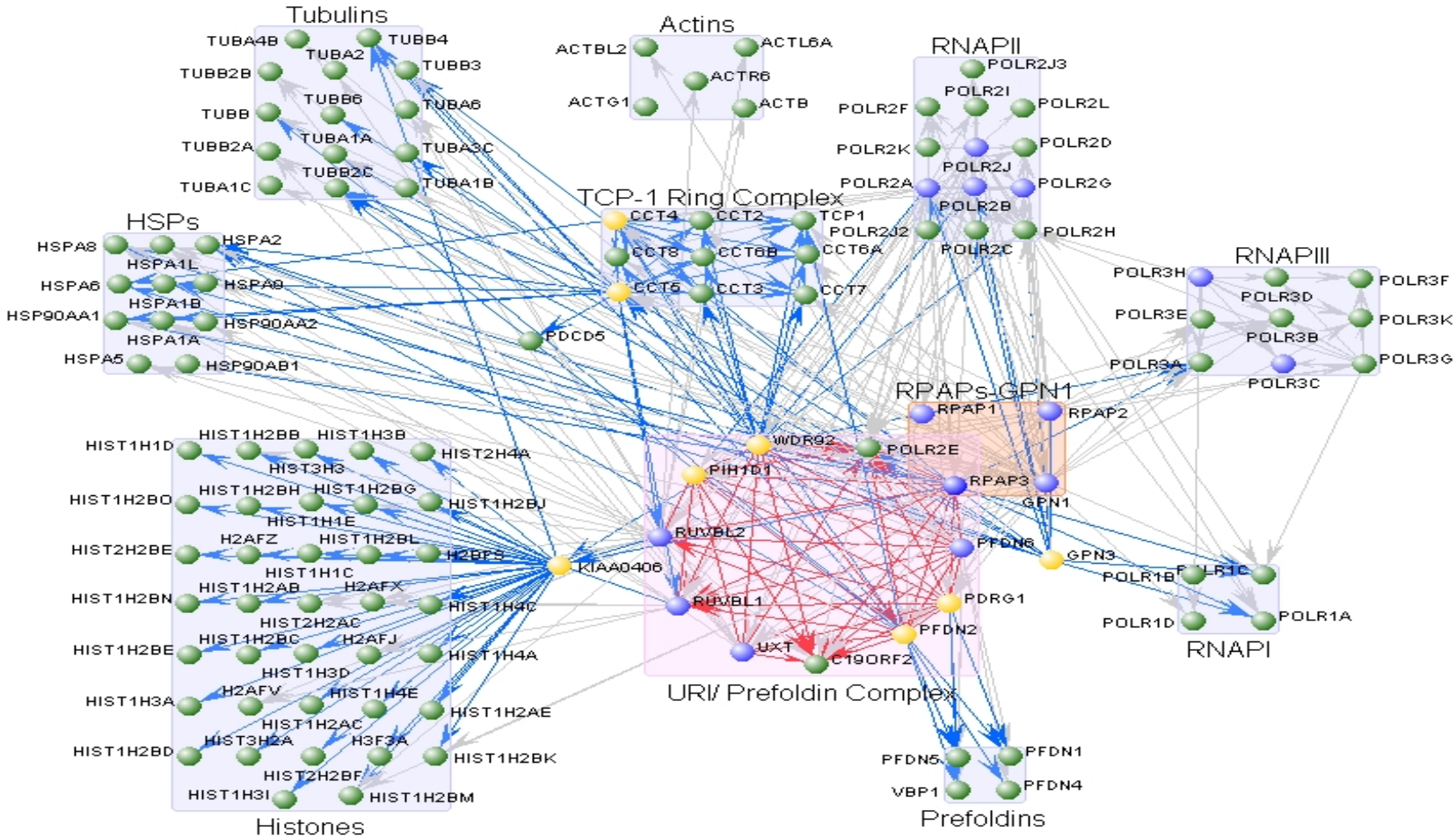


Figure 17: Interaction map showing the eight purified proteins, in this project, and their high confidence interactions

Section 5: Discussion

5-1) The AP-MS method:

The AP-MS method has proven to be an efficient tool for the characterization of protein complexes and the identification of protein interaction partners. Amongst various possible strategies of conducting AP-MS, the Tandem Affinity Purification (TAP) technique coupled to mass spectrometry has been the method of choice in our laboratory for the purification of protein complexes in mammalian cells (Cloutier *et al.,* submitted article).

As discussed earlier, in section 1-2.2.2, the AP-MS method does have few limitations that must be overcome or resolved to obtain the most accurate results. For example, it is known that affinity chromatography is a technique that is prone to generate high levels of contaminants. By using TAP as a method of purification in our laboratory, we are able to reduce the amount of contaminants from the affinity purified eluate, but the caveat is that weak, transient interactions will more readily be disrupted, leading to the loss of interesting interaction partners (Cloutier *et al.,* submitted article). Furthermore, the use of sensitive, high- accuracy mass spectrometry, in our laboratory, coupled to the TAP technique also lead to a significant decrease in the rate of false positives that are normally generated. Also by making sure of using an expression system, such as the ecdysone- inducible system used in our laboratory, that avoids the overexpression of the tagged protein, has been shown to minimize the occurrence of spurious protein- protein interactions in our datasets (Cloutier *et al.,* submitted article). Finally, to increase the confidence in the protein- protein interaction datasets in our laboratory, the development of a computational algorithm trained to minimize the rate of false positive and false negative

interactions has been shown to be very powerful and useful (Cloutier *et al.,* submitted article).

5-2) Purification of the 8 newly-identified RNAPII interaction partners:

The study of protein-protein interactions and protein complexes has proved to be essential for understanding how cells function in normal conditions; and, consequently, is useful to reveal novel targets which are paramount to the development of new drugs to treat diseases (Al-Khoury R, and Coulombe B, 2009).

In efforts of pursuing our navigation on the network of protein complexes forming the transcription machinery in the soluble fraction of human cells, eight RPAPs-XAB1 newly-identified interaction partners were chosen for TAP purification, mainly GPN3, WDR92, PDRG1, PFDN2, PIH1D1, KIAA0406, CCT4 and CCT5.

As can be observed in the section 4, purification of these eight proteins helped us in further enriching our data set, and confirming previously identified interactions. For most of proteins of an unknown function such as, GPN3, and KIAA0406, PDRG1, PIH1D1 and WDR92, identifying their interaction partners helped us in forming an idea about the kind of a role they might be playing in mammalian cells.

For example, in the case of GPN3's purification, presented in section 4-1, and previous purification of RPAP2 and XAB1/GPN1, presented in Jeronimo *et al.,* 2007, most of their abundant interaction partners included subunits of RNAPI, RNAPII and RNAPIII (except for RPAP2 which was only observed to interact with subunits of RNAPII and RNAPIII). Knowing that it is

still unclear until now where and in what manner do the subunits of each of the three RNA polymerases come together to form the different enzymes, it is reasonable to suggest that GPN3, together with RPAP2 and XAB1/GPN1, could be playing a role in the assembly of the RNA polymerases. As mentioned earlier, in section 4-1, both GPN3 and XAB1/GPN1 were observed to interact with actins and tubulins, and subunits of the TCP-1 ring complex (in the case of XAB1/GPN1). It is, therefore, also reasonable to assume a role for these proteins (GPN3 and XAB1/GPN1) in, perhaps, the formation of microtubules. The role of XAB1/GPN1 is currently being investigated in our laboratory.

As for the KIAA0406 protein, the fact that it was observed to interact with both RUVBL1 and RUVBL2 which, as mentioned previously, are found in chromatin modification complexes, and the fact that most of its interaction partners were observed to be histones, it is reasonable to assume that this protein might be involved somehow in chromatin modification.

CCT4 and CCT5 are two proteins of known function, as subunits of TCP-1 ring complex chaperone, which have been chosen for TAP purification. Although CCT4 did not yield any unexpected interaction partners other than the subunits of the TCP-1 ring complex, CCT5 did yield an unexpected interaction partner, the PDCD5 protein.

PDCD5 (programmed cell death 5) is a protein that is expressed at higher levels in tumor cells, compared to normal cells, during apoptosis. Although its exact function is still not defined, it is thought to regulate both apoptotic and non-apoptotic programmed cell death (Liu *et al.,* 2005). The PDCD5 protein is distributed in both the nucleus and the cytoplasm, but once apoptosis is induced, its level increases, and by relocation from the cytoplasm, it accumulates in the nucleus (Chen *et al.,* 2001). The mechanism by which PDCD5 crosses the nuclear envelope and functions in the nucleus is still

unclear. But it is thought that the N-terminal α-helix of this protein may play some role in its translocation to the nucleus of target cells. The N-terminal residues of PDCD5 tend to form a stable α-helical structure independently of the core of the protein. Apoptosis activity assay indicates that the deletion of the N-terminal α-helix of PDCD5 attenuates the apoptosis-promoting effects of this protein (Liu *et al.,* 2005).

Having obtained PDCD5 as an interaction partner of CCT5, it would be reasonable to assume that perhaps the TCP-1 ring complex could be playing a role in either the folding of the PDCD5 protein as a whole, or simply folding its N-terminal region into the stable α-helical structure. Once folded properly, the TCP-1 ring complex could be the molecular chaperone responsible for the translocation across the nucleus when needed. If that was to be true, we would have expected to observe an interaction between PDCD5 and the purified CCT4, which is unfortunately not the case. This does not mean that the interaction does not occur; the interaction could simply not be abundant.

TAP purification of PDCD5 is currently being done in our laboratory to hopefully, confirm its interaction with CCT5, and to determine if other TCP-1 ring complex subunits might be identified with it.

5-3) PFDN2, PDRG1, PIH1D1, WDR92 and the URI/Prefoldin complex:

Our results show that, using the TAP approach to purify PFDN2, PDRG1, PIH1D1 and WDR92, we were able to purify the URI/Prefoldin complex. As we mentioned earlier, the URI/Prefoldin complex has not been assigned any specific function until now, but from what is known about its components and by purifying most of them, either in this project or by work done in our laboratory (for RPAP3, RUVBL1, RUVBL2, UXT, RPB5 and

PFDN6), and identifying their interaction partners, hypotheses could be made about the role this complex might be playing in mammalian cells.

5-3.1) **First Hypothesis**: URI/Prefoldin complex might be involved in the proper folding of Rpb5 or in the assembly of RNA polymerases:

Several components of the URI/Prefoldin complex are subunits of the prefoldin complex such as PFDN2 and PFDN6, and prefoldin- like proteins such as UXT, URI and PDRG1 (a PFDN4-like protein).

It is worth pointing out that, aside from PFDN2 and PFDN6, PDRG1 is the only component of the URI/Prefoldin complex which picked up all of the prefoldin subunits except PFDN4, when purified. The questions raised now are; does the prefoldin complex exist in two different states, one with PFDN4 and another with PDRG1 instead? If so, how do these two complexes differ in activity? Furthermore, is the PDRG1-Prefoldin complex unique to the URI/Prefoldin complex?

Knowing that the URI/Prefoldin complex consists of several prefoldins and prefoldin-like proteins, and that Rpb5 is a component of the complex, the first assumption that could be made is that, perhaps, this complex might be playing a role in the proper folding of Rpb5.

RPAP3 is another member of the complex which has previously been shown to interact with several subunits of RNAPI, RNAPII and RNAPIII (Jeronimo *et al.,* 2007). Since Rpb5 is a common subunit to all three polymerases, it is plausible to speculate that the URI/Prefoldin complex might be playing a role in the assembly of RNA polymerases.

5-3.2) **Second Hypothesis**: The URI/Prefoldin complex might be involved in the regulation of transcription by the RNAPII:

URI (Unconventional Prefoldin Rpb5 Interactor) is a prefoldin-like protein, also referred to as RMP for Rpb5-mediating protein, that has been reported to negatively modulates transcription through its interaction with the RNAPII subunit 5 "Rpb5", a common subunit of all three RNA polymerases (Wei *et al.,* 2003, Dorjsuren *et al.,* 1998). URI's negative modulation of transcription was first observed by its ability to interact with Rpb5, through its Rpb5-binding domain, and counteract transactivation by HBx, the multifunctional viral regulator protein of hepatitis B virus, which was previously shown to directly interact with Rpb5 and TFIIB and facilitate transcription (Wei *et al.,* 2003, Dorjsuren *et al.,* 1998). URI has also been shown to regulate transcription through interaction with TFIIF, a general transcription factor that functions in both transcription initiation and elongation. This could be another pathway in which URI is able to regulate transcription. It is also possible that TFIIF may cooperate with Rpb5 and TFIIB for the corepressor function of URI (Wei *et al.,* 2003).

Functional analysis of URI in yeast has revealed that it is a downstream substrate of TOR which acts to control nutrient- sensitive gene expression. In mammalian cells, URI was found to be in a phosphorylated state, influenced by signals that affect the activity of mTOR, and positively contributes to rapamycin- sensitive transcription (Gstaiger *et al.,* 2003, Djouder *et al.,* 2007).

Thus, in both cases, involving either interaction with RPB5 or the rapamycin- sensitive signaling pathway of mTOR, URI is involved in modulating transcription by RNAPII. Therefore, it can be postulated that the URI/Prefoldin complex might be involved in the regulation of transcription by RNAPII.

5-3.3) **Third Hypothesis:** The URI/Prefoldin complex might be involved in RNA processing:

PIH1D1 is another component of the URI/Prefoldin complex. Recently, Pih1, the yeast homologue of the human PIH1D1, was identified as a member of a newly discovered complex, the R2TP complex, together with Tah1, Rvb1 and Rvb2 (the yeast homologues of the human RUVBL1 and RUVBL2) (Zhao *et al.,* 2008).

Pih1 was shown to be an unstable protein of the R2TP complex under normal conditions (log phase of yeast), but is not prone for degradation. Therefore, it was assumed that chaperones, which have not yet been identified, might be involved in its stabilization under normal conditions. On the other hand, under stress condition (stationary phase of yeast), it was demonstrated that Hsp90 together with Tah1 stabilize Pih1 (Zhao *et al.,* 2008). The R2TP complex, in fact, was found to interact with the Hsp90 chaperone, and this interaction is required for the correct accumulation of box C/D snoRNPs, which consist of box C/D snoRNAs associated with a set of core proteins, especially under stress conditions (Zhao *et al.,* 2008). snoRNPs, themselves, are trans-acting regulators responsible for cleavage and modifications of snRNAs, rRNAs and tRNAs (Matera *et al.,* 2007, Zhao *et al.,* 2008).

The URI/Prefoldin complex does contain the human homologs of Pih1, Rvb1 and Rvb2, which are PIH1D1, RUVBL1 and RUVBL2 respectively. Tah1, on the other hand, is a protein that contains two tetratricopeptide repeats (TPRs), and until now, its human orthologue has not been identified yet; however it is believed that other TPR domain-containing proteins might serve as an alternative for Tah1 in human cells (Zhao *et al.,* 2008). One component of the URI/Prefoldin complex is a TPR domain-containing protein, RPAP3, which, in fact, contains two TPR domains. Previous TAP purification of RPAP3 did yield Hsp90 as an interaction partner of the purified RPAP3 complex (results

reported in Jeronimo *et al.,* 2007). Therefore, it is reasonable to assume that RPAP3 could be the human orthologue of the yeast Tah1, and thus, the URI/Prefoldin complex might be playing an important role in RNA processing.

5-3.4) **Fourth Hypothesis:** The URI/Prefoldin complex might play an important role in apoptosis:

mTOR is a protein that coordinately controls cell growth in response to growth factors and the availability of nutrients. One of its downstream effector proteins is S6K1, which promotes protein synthesis and cell survival (Sabatini 2006, Wullschleger *et al.,* 2006, and Djouder *et al.,* 2007). In response to growth factors, S6K1 was found to phosphorylate URI (at Ser-371) found associated with PP1γ at the mitochondria. This phosphorylation leads to the dissociation of URI/PP1γ. PP1γ now released contributes to the downregulation of S6K1 activity *in vivo*, thus, enhancing a cell's susceptibility to undergo apoptosis (Djouder *et al.,* 2007).

Furthermore, WDR92, a component of the URI/Prefoldin complex, has been suggested to play a role as a novel modulator of apoptosis pathway induced by TNF-α, as discussed earlier in section 1 (Saeki *et al.,* 2006). In 2007, it was reported that the knockdown of UXT, which is a prefoldin-like component of the URI/Prefoldin complex, sensitizes cells to apoptosis induced by TNF-α (Sun *et al.,* 2007). It is therefore reasonable to assume that the URI/Prefoldin complex might be playing an important role in apoptosis.

As can be observed, the output of the AP-MS procedure not only helps in forming an idea or inferring a function to the proteins of previously unknown function, but also in identifying new interaction partners that are either unexpected or that have not been previously characterized. It is, however, these proteins that are unexpected that are often the most promising in terms of proteomic discovery (Cloutier *et al.,* 2007). Furthermore and as discussed earlier, the output of the AP-MS procedure did help us in forming an idea or

coming up with hypotheses about the role some of these proteins might be playing in mammalian cells. This information is crucial and will help, later on, directing us in choosing the kind of functional assays needed to study the function of these proteins and to confirm or reject the presented hypotheses some of which are currently being studied in our laboratory.

It is evident that large-scale, genome-wide protein-protein interaction screens and the mapping of protein interaction networks are now showing a great potential in the identification of novel therapeutic targets and putative biomarkers to be used as diagnostic tools, and are key in understanding the molecular mechanisms of diseases (Oltersdorf *et al.,* 2005, Al-Khoury R. and Coulombe B. 2009). The resulting protein interaction maps represent in many ways the fingerprint of the physiological status of a cell and their modulation is predicted to represent the signature of specific disease conditions, including those observed in cancers, viral infections and other pathological conditions (Coulombe *et al.,* 2008, Al-Khoury R. and Coulombe B. 2009).

***Section 6:* Conclusion and Perspectives**

In efforts of defining the interaction network of soluble protein complexes in human cells, our laboratory was interested in defining the interaction network involving the human RNAPII machinery using the TAP purification method. This approach has proven to be very powerful allowing for a large-scale study of protein interaction networks.

The aim of my project was to further pursue our survey of soluble protein complexes containing components of the human transcription machinery, using the same TAP method. Eight newly identified RNAPII interaction partners were chosen for TAP purification (KIAA0406, WDR92, PIH1D1, PFDN2, GPN3, CCT4, CCT5 and PDRG1) to determine their own interaction partners and perhaps identify some novel interacting proteins. A novel interacting partner was only identified with CCT5, and it is PDCD5, which is a very interesting protein, and currently, work for its TAP purification is being pursued in the laboratory.

Several of the chosen proteins were of an unknown function and we hoped that by identifying their interaction partners it would be possible to infer putative functions to them. For all of these proteins, we were able to come up with assumptions and hypotheses about their function from their interaction partners, but biochemical and functional approaches still need to be applied to either validate or reject these hypotheses.

Using the TAP purification method, we were also able to purify the URI/Prefoldin complex which includes, in addition to other proteins, several of the proteins that were purified in my project. Using what is known in the literature about some of its components and having identified the interaction partners of most of its components, we were able to come up with few

interesting hypothesis, mentioned earlier in section 5, about the kind of a role this complex might have in human cells. These hypotheses will be further investigated in the laboratory in efforts of deciphering the function of the URI/Prefoldin complex.

References:

Al-Khoury R, Coulombe B. Defining protein interactions that regulate disease progression. Expert. Opin. Ther. Targets; 13(1): 13-7 (2009).

Almeida SF. and Carmo-Fonesca M. The CTD role in cotranscriptional RNA processing and surveillance. FEBS Lett; 582(14): 1971- 6 (2008).

Archambault J, Friesen JD. Genetics of eukaryotic RNA polymerases I, II, and III. Microbiological reviews 57(3):703- 24 (1993).

Bair C.L, Oppenheim A, Trostel A, Prag G, Adhya S. A phage display system designed to detect and study protein- protein interactions.

Barginear M. F, Van Poznak C, Rosen N, Modi S, Hudis C. A, Budman D. R. The heat shock protein 90 chaperone complex: and evolving therapeutic target. Curr. Cancer Drug Targets; 8 (6): 522- 532 (2008).

Barrios-Rodiles M, Brown K. R, Ozdamar B, Bose R, Liu Z, Donovan R. S, Shinjo F, Liu Y, Dembowy J, Taylor I. W, Luga V, Przulj N, Robinson M, Suzuki H, Hayashizaki Y, Jurisica I, Wrana J. L. High-throughput mapping of a dynamic signaling network in mammalian cells. Science; 307(5715): 1621-5 (2005).

Bauer A. and Kuster B. Affinity Purification- mass spectrometry: Powerful tools for the characterization of protein complexes. Eur. J. Biochem; 270: 570-578 (2003).

Berger S. L. The complex language of chromatin regulation during transcription. Nature; 447: 407- 412 (2007).

Berrgård T, Linse S, James P. Methods for the detection and analysis of protein- protein interactions. Proteomics; 7: 2833- 2842 (2007).

Bhat R. A, Lahaye T, Panstruga R. The visible touch: *in planta* visualization of protein- protein interactions by fluorophores- based methods. Plant Methods; 2:12 (2006).

Blackwell T. K. and Walker A. K. Transcription mechanisms (September 5, 2006), *WormBook*, ed. The *C. elegans* Research community, WormBook, doi/10.1895/wormbook.1.121.1, http://www.wormbook.org.

Boute N, Jokers R, Issad T. The use of resonance energy transfer in high-throughput screening: BRET versus FRET. Trends in pharmacological Sciences; 23(8): 251- 254 (2002).

Brachmann R. K, Boeke J. D. Tag games in yeast: the two-hybrid system and beyond. Curr. Opin. Biotechnol; 8(5): 561- 8 (1997).

Causier B. Studying the interactome with the yeast two-hybrid system and mass spectrometry. Mass Spectrom. Rev; 23(5): 350- 67 (2004).

Chen X, Ruggiero C, Li S. Yeast Rpb9 plays an important role in ubiquitylation and degradation of Rpb1 in response to UV-induced DNA damage. Mol. Cell. Biol.; 27(13): 4617- 25 (2007).

Chen Y, Sun R, Han W, Zhang Y, Song Q, Di C, Ma D. Nuclear translocation of PDCD5 (TFAR19): an early signal for apoptosis? FEBS Lett; 509(2): 191-6 (2001).

Cheng X, Belshan M, Ratner L. Hsp40 facilitates nuclear import of the human immunodeficiency virus type 2 Vpx-mediated preintegration complex. J. Virol; 82(3): 1229- 37 (2008).

Cloutier P, Al-Khoury R, Lavallée-Adam M, Faubert D, Jiang H, Poitras C, Bouchard A, Forget D, Blanchette M, Coulombe B. High- resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II- associated complexes. Submitted to Methods.

Coulombe B, Blanchette M, Jeronimo C. Steps towards a repertoire of comprehensive maps of human protein interaction networks: the Human Proteotheque Initiative (HuPI). Biochem Cell Biol; 86(2): 149- 56 (2008).

Coulombe B, Burton Z. F. DNA bending and wrapping around RNA polymerase: a "Revolutionary" model describing transcriptional mechanisms. Microbiol. Mol. Biol. Rev; 63: 457- 478 (1999).

Coulombe B, Jeronimo C, Langelier M-F, Cojocaru M, Bergeron D. Interaction networks of the molecular machines that decode, replicate, and maintain the integrity of the human genome. Mol. Cell Proteomics; 3(9): 851- 6 (2004).

Cramer P, Bushnell D. A, Kornberg R. D. Structural Basis of Transcription: RNA Polymerase II at 2.8 Ångstrom Resolution. Science; 292(5523): 1863- 1876 (2001).

Culver G. M. Meanderings of the mRNA through the ribosome. Structure; 9: 751- 758 (2001).

Davie J. R. Histone modifications, chromatin structure, and the nuclear matrix. Journal of Celluar Biochemistry; 62: 149- 157 (1996).

Deng W. and Roberts S. G. E. Core promoter elements recognized by transcription factor IIB. Biochemical Society Transactions; 34(6): 1051- 53 (2006).

Deuerling E. and Bukau B. Chaperone- assisted folding of newly synthesized proteins in the cytosol. Crit. Rev. Biochem. Mol. Biol; 39(5- 6): 261- 77 (2004).

Djouder N, Metzler S. C, Schmidt A, Wirbelauer C, Gstaiger M, Aebersold R, Hess D, Krek W. S6K1-Mediated Disassembly of Mitochondrial URI/PP1γ Complexes Activates a Negative Feedback Program that Counters S6K1 Survival Signaling. Mol. Cell; 28(1): 28- 40 (2007).

Dorjsuren D, Lin Y, Wei W, Yamashita T, Nomura T, Hayashi N, Murakami S. RMP, a Novel RNA Polymerase II Subunit 5-Interacting Protein, Counteracts Transactivation by Hepatitis B Virus X Protein. Mol Cell Biol; 18(12): 7546- 55 (1998).

Drakas R, Prisco M, Baserga R. A modified tandem affinity purification tag technique for the purification of protein complexes in mammalian cells. Proteomics; 5(1): 132- 7 (2005).

Ellis R. J. Molecular chaperones: assisting assembly in addition to folding. Trends in Biochemical Sciences; 31(7): 395- 401 (2006).

Erzberger J. P, Berger J. M. Evolutionary relationships and structural mechanisms of AAA+ proteins. Annu. Rev. Biophys. Biomol. Struct; 35:93-114. Review (2006).

Evans S. K, Aiello D. P, Green M. R. Fluorescence resonance energy transfer as a method for dissecting *in vivo* mechanisms of transcriptional activation. Biochem. Soc. Symp; 73: 217- 224 (2006).

Figeys D. Mapping the human protein interactome. Cell Research; 18: 716- 724 (2008).

Fink A. L. Chaperone- mediated protein folding. Physiological reviews; 79(2): 425- 449 (1999).

Gandía J, Luís C, Ferré S, Franco R, Ciruela F. Light resonance energy transfer- based methods in the study of G protein- coupled receptor oligomerization. BioEssays; 30: 82- 89 (2007).

Gingras A-C, Aebersold R, Raught B. Advances in protein complex analysis using mass spectrometry. J. Physiol; 563(1): 11-21 (2005).

Gingras A-C, Stager S, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. Nat. Rev. Mol. Cell Biol; 8(8): 645- 54 (2007).

Goodyear C. S. and Silverman G. J. Phage-Display Methodology for the Study of Protein-Protein Interactions: Overview. CSH Protocols; 2008; doi:10.1101/pdb.top48.

Gross P. and Oelgeschäger T. Core promoter- selective RNA polymerase II transcription. Biochem. Soc. Symp.; 73: 225- 236 (2006).

Gstaiger M, Luke B, Hess D, Oakeley E. J, Wirbelauer C, Blondel M, Vigneron M, Peter M, Krek W. Control of Nutrient-Sensitive Transcription

Programs by the Unconventional Prefoldin URI. Science; 302(5648): 1208- 12 (2003).

Gu Y, Deng Z, Paredez A. R, BeBolt S, Wang Z-Y, Somerville C. Prefoldin 6 is required for normal microtubule dynamics and organization in *Arabidopsis*. Proc. Natl. Acad. Sci. U. S. A; 105 (46): 18064- 9 (2008).

Hahn S. Structure and mechanism of the RNA polymerase II transcription machinery. Nature Structural and Molecular Biology; 11(5): 394- 403 (2004) review.

Hansen W. J, Cowan N. J, and Welch W. J. Prefoldin-nascent chain complexes in the folding of cytoskeletal proteins. J. Cell Biol; 145(2): 265- 77 (1999).

Hebert D. N, Molinari M. In and out of the ER: Protein folding, Quality control, Degredation and related human Disease. Physiol. Rev; 87: 1377- 1408 (2007).

Hennessy F, Cheetham M. E, Dirr H. W, Blatch G. L. Analysis of the levels of conservation of the J domain among the various types of DnaJ- like proteins. Cell Stress Chaperones; 5(4): 347- 358 (2000).

Houry W. A, Frishman D, Eckerskorn C, Lottspeich F, Hartl F. U. Identification of *in vivo* substrates of the chaperonin GroEL. Nature; 402(6758): 147-54 (1999).

Iglesias N, and Stutz F. Regulation of mRNP dynamics along the export pathway. FEBS Lett; 582(14): 1987- 96 (2008).

Ito T, Chiba T, Yoshida M. Exploring the protein interactome using comprehensive two hybrid projects. Trends Biotechnol; 19: S23- S27 (2001).

Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y. A comprehensive two- hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. U. S. A; 98: 4569- 4574 (2001).

Itsuki Y, Saeki M, Nakahara H, Egusa H, Irie Y, Terao Y, Kawabata S, Yatani H, Kamisaki Y. Molecular cloning of novel Monad binding protein containing tetratricopeptide repeat domains. FEBS Lett; 582(16): 2365- 70 (2008).

Iyer L. M, Leipe D. D, Koonin E. V, Aravind L. Evolutionary history and higher order classification of AAA+ ATPases. J. Struct. Biol; 146(1-2):11-31 (2004).

Jeronimo C, Forget D, Bouchard A, Li Q, Chua G, Poitras C, Thérien C, Bergeron D, Bourassa S, Greenblatt J, Chabot B, Poirier G. G, Hughes T. R, Blanchette M, Price D. H, Coulombe B. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol. Cell; 27(2): 262- 74 (2007).

Jaeger S, Gaudan S, Leser U, Rebholz-Schuhmann D. Integrating protein-protein interactions and text mining for protein function prediction. BMC Bioinformatics; 9(Suppl 8):S2 (2008).

Kaboord B, Perr M. Isolation of proteins and protein complexes by immunoprecipitation. Methods Mol Biol; 424: 349- 64 (2008).

Kandror O, Busconi L, Sherman M, Goldberg AL. Rapid degradation of an abnormal protein in Escherichia coli involves the chaperones GroEL and GroES. J. Biol. Chem; 269(38): 23575- 82 (1994).

Kim Y, Geiger J. H, Hahn S, Sigler P. B. Crystal structure of a yeast TBP/TATA-box complex. Nature; 365: 512-520 (1993).

Kim S. J, and Martinson HG. Poly(A)-dependent transcription termination: continued communication of the poly(A) signal with the polymerase is required long after extrusion in vivo. J. Biol. Chem; 278(43): 41691- 701 (2003).

Kim J. L., Nikolov D. B., Burley S. K. Co-crystal structure of TBP recognizing the minor groove of a TATA element. Nature; 365: 520-527 (1993).

Khorasanizadeh S. The nucleosome: From genomic organization to genomic regulation. Cell; 116: 259- 272 (2004).

Kornberg R. D. The molecular basis of eukaryotic transcription. PNAS; 104 (32): 12955- 12961. (2007).

Kramer R and Cohen D. Functional genomics to new drug targets. Nature Reviews; 3: 965- 972 (2004).

Kunau W. H, Beyer A, Franken T, Götte K, Marzioch M, Saidowsky J, Skaletz-Rorowski A, Wiebel F. F. Two complementary approaches to study peroxisome biogenesis in *Saccharomyces cerevisiae*: Forward and reversed genetics. Biochimie; 75(3-4): 209-24 (1993).

Kusukawa N, Yura T, Ueguchi C, Akiyama Y, Ito K. Effects of mutations in heat-shock genes groES and groEL on protein export in Escherichia coli. EMBO J; 8(11): 3517- 21 (1989).

Leroux M. R. and Hartl F. U. Protein folding: versatility of the cytosolic chaperonin TRiC/CCT. Curr. Biol; 10(7): R260- 4 (2000).

Liu D, Yao H, Chen Y, Feng Y, Chen Y, Wang J. The N-terminal 26-residue fragment of human programmed cell death 5 protein can form a stable alpha-helix having unique electrostatic potential character. Biochem J; 392(Pt 1):47-54 (2005).

Llorca O, McCormack E. A, Hynes G, Grantham J, Cordell J, Carrascosa J. L, Willison K. R, Fernandez J. J, Valpuesta J. M. Eukaryotic type II chaperonin CCT interacts with actin through specific subunits. Nature; 402(6762): 693- 6 (1999).

Lodish H, Berk A, Zipursky S. L, Matsudaira P, Baltimore D, Darnell J. E. Molecular Cell Biology- 4th edition. Pages 100-105, 109 and 116. W. H. Freeman and Company, New York. Third printing 2001.

Luban J, Goff S. P. The yeast two-hybrid system for studying protein-protein interactions. Curr. Opin. Biotechnol; 6(1): 59- 64 (1995).

Lund P. A. Molecular Chaperones in the Cell. Published by Oxford University Press, 2001.

Lund PA, Large AT and Kapatai G. The chaperonins: perspectives from the Archaea. Biochem. Soc. Trans; 31(pt3): 681- 5 (2003).

Luo X, Huang Y, Sheikh M. S. Cloning and characterization of a novel gene PDRG that is differentially regulated by p53 and ultraviolet radiation. Oncogene; 22(46): 7247- 57 (2003).

Lusser A. and Kadonaga J. T. Chromatin remodeling by ATP- dependent molecular machines. BioEssays; 25:1192- 1200 (2003).

Lykke-Andersen S, and Jensen TH. Overlapping pathways dictate termination of RNA polymerase II transcription. Biochimie; 89(10): 1177- 1182 (2007).

Margaritis T. and Holstege F. C. P. Poised RNA polymerase II gives pause for thought. Cell; 133: 581- 584 (2008).

Martinez E. Multi- protein complexes in eukaryotic gene transcription. Plant Mol. Biol.; 50: 925- 947 (2002).

Martín-Benito J, Boskovic J, Gómez-Puertas P, Carrascosa J. L, Simons C. T, Lewis S. A, Bartolini F, Cowan N. J, Valpuesta J. M. Structure of eukaryotic prefoldin and of its complexes with unfolded actin and the cytosolic chaperonin CCT. EMBO J; 21(23): 6377- 86 (2002).

Martín-Benito J, Grantham J, Boskovic J, Brackley K. I, Carrascosa J. L, Willison K. R, Valpuesta J. M. The inter-ring arrangement of the cytosolic chaperonin CCT. EMBO Rep; 8(3): 252- 7 (2007).

Matera A. G, Tycowski K. T, Steitz J. A, Ward D. C. Organization of small nucleolar ribonucleoproteins (snoRNPs) by fluorescence in situ hybridization and immunocytochemistry. Mol. Biol. Cell; 5(12): 1289- 99 (1994).

Miernyk J. A, Thelen J. J. Biochemical approaches for discovering protein-protein interactions. Plant J; 53(4): 597- 609 (2008).

Michnick S. W, Ear P. H, Manderson E. N, Remy I, Stefan E. Universal strategies in research and drug discovery based on protein-fragment complementation assays. Nature Reviews Drug Discovery; 6: 569-582 (2007).

Morano KA. New tricks for an old dog: the evolving world of Hsp70. Ann. N. Y. Acad. Sci; 1113: 1- 14 (2007).

Nicoll WS, Boshoff A, Ludewig MH, Hennessy F, Jung M, Blatch JL. Approaches to the isolation of molecular chaperones. Protein Expr. Purif; 46(1): 1- 15 (2006).

Ogura T, Wilkinson A. J. AAA+ superfamily ATPases: common structure—diverse funcion. Genes Cells; 6(7): 575- 97 (2001).

Oltersdorf T, Elmore S. W, Shoemaker A. R, Armstrong R. C, Augeri D. J, Belli B. A, Bruncko M, Deckwerth T. L, Dinges J, Hajduk P. J, Joseph M. K, Kitada S, Korsmeyer S. J, Kunzer A. R, Letai A, Li C, Mitten M. J, Nettesheim D. G, Ng SC, Nimmer P. M, O'Connor J. M, Oleksijew A, Petros A. M, Reed J. C, Shen W, Tahir S. K, Thompson C. B, Tomaselli K. J, Wang B, Wendt M. D, Zhang H, Fesik S. W, Rosenberg S. H. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature; 435(7042): 677-81 (2005).

Paulsson K., Wang P. Chaperones and folding of MHC class I molecules in the endoplasmic reticulum. Biochim. Biophys. Acta; 1641(1): 1-12 (2003).

Pearl LH, Prodromou C, Workman P. The Hsp90 molecular chaperone: an open and shut case for treatment. Biochem. J; 410(3): 439- 53 (2008).

Peterson C. L. Chromatin remodeling enzymes: taming the machines. EMBO reports; 3(4): 319- 322 (2002).

Pfleger K. D. G, Eidne K. A. Illuminating insights into protein-protein interactions using bioluminescence resonance energy transfer (BRET). Nat. Methods; 3(3): 165- 74 (2006).

Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, Bragado-Nilsson E, Wilm M, Séraphin B. The Tandem Affinity Purification (TAP) Method: A General Procedure of Protein Complex Purification. Methods; 24(3): 218- 29 (2001).

Puri T, Wendler P, Sigala B, Saibil H, Tsaneva I. R. Dodecameric Structure and ATPase Activity of the Human TIP48/TIP49 Complex. J. Mol. Biol; 366(1): 179-92 (2007).

Rigault G, Shevchenko A, Rutz B, Wilm M, Mann M, Séraphin B. A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol; 17(10): 1030- 2 (1999).

Rhyner C, Weichel M, Flückiger S, Hemmann S, Kleber-Janke T, Crameri R. Cloning allergens via phage display. Methods; 32(3): 212- 218 (2004).

Robert F., Douziech M., Forget D., Egly J. M., Greenblatt J., Burton Z. F., Coulombe B. Wrapping of promoter DNA around the RNA polymerase II initiation complex induced by TFIIF. Mol. Cell; 2: 341-351 (1998).

Roseman A. M, Chen S, White H, Braig K, Saibil H. R. The Chaperonin ATPase Cycle: Mechanism of Allosteric Switching and Movements of Substrate-Binding Domains in GroEL. Cell; 87(2): 241- 51 (1996).

Ruffner H, Bauer A, Bouwmeester T. Human protein-protein interaction networks and the value for drug discovery. Drug Discov. Today; 12(17- 18): 709- 16 (2007).

Ruthenburg A. J, Li H, Patel D. J, Allis C. D. Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol.; 8(12): 983- 94 (2007).

Saeki M, Irie Y, Ni L, Yoshida M, Itsuki Y, Kamisaki Y. Monad, a WD40 repeat protein, promotes apoptosis induced by TNF-alpha. Biochem. Biophys. Res. Commun; 342(2): 568- 72 (2006).

Sabatini D. M. mTOR and cancer: insights into a complex relationship. Nat. Rev. Cancer; 6(9): 729- 34 (2006).

Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y and Hume DA. Mammalian RNA polymerase II core promoters: insights from genome-wide studies. Nat. Rev Genet; 8(6): 424- 36 (2007).

Sardiu M. E, Cai Y, Jin J, Swanson S. K, Conaway R. C, Conaway J. W, Florens L, Washburn M. P. Probabilistic assembly of human protein interaction networks from label- free quantitative proteomics. PNAS; 105(5): 1454- 1459 (2007).

Séraphin B. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. Methods; 24(3): 218- 29 (2001).

Shevchenko A, Schaft D, Roguev A, Pijnappel W. W. M. P, Stewart A. F, Shevchenko A. Deciphering protein complexes and protein interaction networks by tandem affinity purification and mass spectrometry: analytical perspective. Mol. Cell Proteomics; 1(3): 204- 12 (2004).

Shiota M, Kusakabe H, Hikita Y, Nakao T, Izumi Y, Iwao H. Pharmacogenomics of cardiovascular pharmacology: Molecular network analysis in Pleiotropic effects of statin – an experimental elucidation of the pharmacologic action from protein- protein interaction analysis. J. Pharmacol. Sci; 107: 15- 19 (2008).

Sims RJ 3rd, Belotserkovskaya R, and Reinberg D. Elongation by RNA polymerase II: the short and long of it. Genes Dev; 18(20): 2437- 68 (2004).

Smale S. T. Transcription initiation from TATA- less promoters within eukaryotic protein- coding genes. BBA; 1351: 73- 88. (1997).

Smale S. T. and Kadonaga J. T. The RNA polymerase II core promoter. Annu. Rev. Biochem; 72: 449- 79 (2003).

Smith G. P. and Petrenko V. A. Phage Display. Chem. Rev; 97: 391- 410 (1997).

Snider J, Thibault G, Houry W.D. The AAA+ superfamily of functionally diverse proteins. Genome Biol; 9(4): 216 (2008).

Southworth D. R, Agard D. A. Species- Dependent Ensembles of Conserved Conformational States Define the Hsp90 Chaperone ATPase Cycle. Mol. Cell; 32: 631- 640 (2008).

Steitz TA. A structural understanding of the dynamic ribosome machine. Nat. Rev. Mol. Biol; 9(3): 242- 53 (2008).

Sun S, Tang Y, Lou X, Zhu L, Yang K, Zhang B, Shi H, Wang C. UXT is a novel and essential cofactor in the NF-κB transcriptional enhanceosome. J. Cell Biol; 178(2): 231- 44 (2007).

Szutorisz H, Dillon N, Tora L. The role of enhancers as centres for general transcription factor recruitment. Trends Biochem. Sci.; 30(11): 593- 9 (2005).

Thomas M. C. and Chiang C-M. The general transcription machinery and general cofactors. Critical reviews in Biochemistry and Molecular Biology; 41:105- 178 (2006).

Tramier M, Gautier I, Piolot T, Ravalet S, Kemnitz K, Coppey J, Durieux C, Mignotte V, Coppey-Moisan M. Picosecond-hetero-FRET microscopy to probe protein-protein interactions in live cells. Biophys. J; 83(6): 3570- 7 (2002).

Vignali M, Hassan A. H, Neely K. E, Workman J. L. ATP- dependent chromatin remodeling complexes. Mol. Cell Biol; 20(6): 1899- 1910 (2000).

Wei W, Gu J. X, Zhu C. Q, Sun F. Y, Dorjsuren D, Lin Y, Murakami S. Interaction with general transcription factor IIF (TFIIF) is required for the suppression of activated transcription by RPB5-mediating protein (RMP). Cell Res; 13(2): 111-20 (2003).

Wullschleger S, Loewith R, Hall M. N. TOR Signaling in Growth and Metabolism. Cell; 124(3): 471- 84 (2006).

Yaciuk P. Co- immunoprecipitation of protein complexes. Methods Mol. Med; 131: 103- 12 (2007).

Yam A. Y, Xia Y, Lin H-T J, Burlingame A, Gerstein M, Frydman J. Defining the TRiC/CCT interactome links chaperonin function to stabilization of newly made proteins with complex topologies. Nat. Struct. Mol. Biol; 15(12): 1255- 62 (2008).

Yu H, Braun P, Yildirim M. A, Lemmens I, Venkatesan K, Sahalie J, Hirozane- Kishikawa T, Gebreab F, Li N, Simonis N, Hao T, Rual J-F, Dricot A, Vazquez A, Murray R. R, Simon C, Tardivo L, Tam S, Svrzikapa N, Fan C, de Smet A- S, Motyl A, Hudson M. E, Park J, Xin X, Cusick M. E, Moore T, Boone C, Snyder M, Roth F. P, Barabási A-L, Tavernier J, Hill D. E, Vidal M. High-quality binary protein interaction map of the yeast interactome network. Science. DOI: 10.1126/science.1158684 (21 August 2008).

Zhang X, **Chaney M**, **Wigneshweraraj SR**, **Schumacher J**, **Bordes P**, **Cannon W**, **Buck M**. Mechanochemical ATPases and transcriptional activation. Mol. Microbiol; 45(4): 895-903 (2002).

Zhao R, Kakihara Y, Gribun A, Huen J, Yang G, Khanna M, Costanzo M, Brost R-L, Boone C, Hughes T. R, Yip C. M, Houry W. A. Molecular chaperone Hsp90 stabilizes Pih1/Nop17 to maintain R2TP complex activity that regulates snoRNA accumulation. J. Cell Biol; 180(3): 563- 78 (2008).

Publications

This section presents two manuscripts that will be used as references in the section 5.

1- **Racha Al-Khoury, Benoit Coulombe.** Defining protein interactions that regulate disease progression. Expert. Opin. Ther. Targets; 13(1): 13-7 (2009).

2- **Cloutier, Racha Al-Khoury, Mathieu Lavallée- Adam, Denis Faubert, Heng Jiang, Christian Poitras, Annie Bouchard, Diane Forget, Mathieu Blanchette, Benoit Coulombe.** High- resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II- associated complexes. Submitted to Methods.

Expert Opinion

informa
healthcare

Defining protein interactions that regulate disease progression

Racha Al-Khoury & Benoit Coulombe†
†*Université de Montréal, Institut de recherches cliniques de Montréal, Département de biochimie, Laboratory of Gene Transcription and Proteomics, 110 avenue des Pins Ouest, Montréal (Québec), Canada H2W 1R7*

Over the past few years, the study of protein–protein interactions and protein complexes has shed more light on cellular processes and cell function. Because alterations in protein–protein interactions perturb the normal sequence of events in the cell and contribute to diseases such as cancer, the understanding of both the normal cellular protein–protein interaction networks and their modulation during the establishment of disease is a crucial issue in biomedical research, as it will facilitate the development of drugs to fight these diseases. In this article, the most commonly used approaches for studying protein–protein interactions are discussed as well as the direction in which the field of systematic characterization of protein interaction networks is progressing. We also discuss some success stories in the modulation of disease-related protein–protein interactions using small molecules.

Keywords: diseases, drug discovery, interaction networks, protein–protein interactions

Expert Opin. Ther. Targets (2009) 13(1):13-17

1. Introduction

Once synthesized, properly folded and directed to its appropriate location, a polypeptide rarely works on its own but is rather assembled into larger complexes of a specific structure and composition where it can exert its specific function. In fact, many cellular processes, such as gene transcription, DNA replication and repair, and others are performed by multiprotein complexes composed of several proteins interacting with each other. Understanding the properties of cells at a systems level, therefore, requires that cellular protein interaction networks are described as comprehensively as possible.

The study of protein–protein interactions and protein complexes has proved to be essential not only for understanding how cells function in normal conditions, but also to reveal novel targets that are paramount for the development of new drugs to treat diseases. These studies also allow placing proteins of previously unknown function, which represent a large proportion of the human proteome, into a functional context provided by their interactions with other proteins, helping in many cases to infer a putative function to proteins [1]. Therefore, defining the network of interacting partners of a given protein and its position within the network provides important information in explaining its function [2]. This also applies to the analysis of proteins of known function, which often play roles in different complexes and/or subcellular compartments. Consequently, large-scale functional proteomics projects that map protein interaction networks are the key in understanding the cross-talk that goes on between the different proteins and protein complexes [1]. The resulting protein interaction maps represent in many ways the fingerprint of the physiological status of a cell and their modulation is predicted to represent the signature of specific disease conditions, including those observed in cancers, viral infections and other pathological conditions [3].

10.1517/14728220802631876 ISSN 1472-8222

Table 1. Some methods used for the study of protein–protein interactions.

Method	Description	Ref.
BRET (Bioluminescence resonance energy transfer)	Takes advantage of the resonance energy transfer between a luminescent donor and a fluorescent acceptor	[8]
FRET (Fluorescence resonance energy transfer)	Takes advantage of resonance energy transfer between a fluorescent donor and a fluorescent acceptor, provided that the two fluorophores are different	[9]
Co-immunoprecipitation	Uses the specificity of antibodies to purify a target protein with its interaction partners	[7]
Yeast two-hybrid	Relies on a transcriptional readout for the detection of protein–protein interactions through the reconstitution of a functional transcriptional activator	[12]
AP-MS (Affinity purification coupled to mass spectrometry)	Allows for the purification of protein complexes under native conditions and the identification of all direct or indirect interacting partners without previous knowledge of complex composition	[13,14]
LUMIER (Luminescence-based mammalian interactome mapping)	An automated high-throughput technology to map protein–protein interaction networks systematically in mammalian cells. It uses *Renilla* luciferase enzyme fused to proteins of interest, which are then coexpressed with individual flag-tagged partners in mammalian cells	[5]
PCA (Protein-fragment complementation assay)	Comprises a reassembly of separate fragments of a protein reporter molecule fused to the two proteins of interest. The reporter will reconstitute its activity only when the two fused proteins are in proximity of each other	[18]
Affinity chromatography	Based on the highly specific interaction of a protein of interest immobilized on a solid support with a protein that specifically adsorbs to it while passing through the column	[10]

From a pharmacological point of view, numerous human diseases are caused by defects in cellular signal transduction pathways, as exemplified by the case of the TNF-α [4] and TGF-β [5] pathways, and others by defects targeting transcription factors (see below for examples). Signaling pathways are regulated by protein–protein interactions that often involve the assembly of large protein complexes. Therefore, the study of protein–protein interactions in specific signaling pathways or in transcriptional responses, some being perturbed in disease conditions, is considered important for the discovery of new potential drugs that target these pathways and responses [6].

2. Methods for studying protein–protein interactions

Several different experimental methods have been developed to identify protein–protein interactions such as co-immunoprecipitation experiments [7], the BRET (bioluminescence resonance energy transfer) [8] and FRET (fluorescence resonance energy transfer) [9] methods, affinity chromatography [10] and phage display [11], but the methods that have been most popular in recent years are the yeast two-hybrid system [12] and the protein affinity purification procedure coupled to mass spectrometry (AP-MS) (see [13] for a review and [14] for an example) (refer to Table 1 for a summary of the methods used for studying protein–protein interactions). Co-immunoprecipitation, BRET, FRET, affinity chromatography and phage display approaches have mainly been used to confirm direct, pair-wise interactions between already known partners. The yeast two-hybrid assay is now the most standardized technique [12] in identifying and mapping protein–protein interactions. Although the yeast two-hybrid is a method that is known to be prone to generate high rates of false positives [12], it has been recently reported that the quality of the high-throughput yeast two-hybrid datasets can be substantially ameliorated when measured against a set of high-confidence physical binary interactions [15]. This set would include direct physical interactions within well-established complexes as well as conditional interactions such as those that are dependent on post-translational modifications [15]. The second major technique is AP-MS that has helped tremendously in advancing the understanding of protein complexes and their composition [13,14]. The AP-MS method allows for the purification of protein complexes under native conditions, at near physiological levels [16,17]. It is, nonetheless, a fact that to map the human interactome, the combination and standardization of various approaches, such as the yeast two-hybrid and the AP-MS, will be required in the future [12].

Novel approaches, such as LUMIER (luminescence-based mammalian interactome mapping) [5], PCA technique

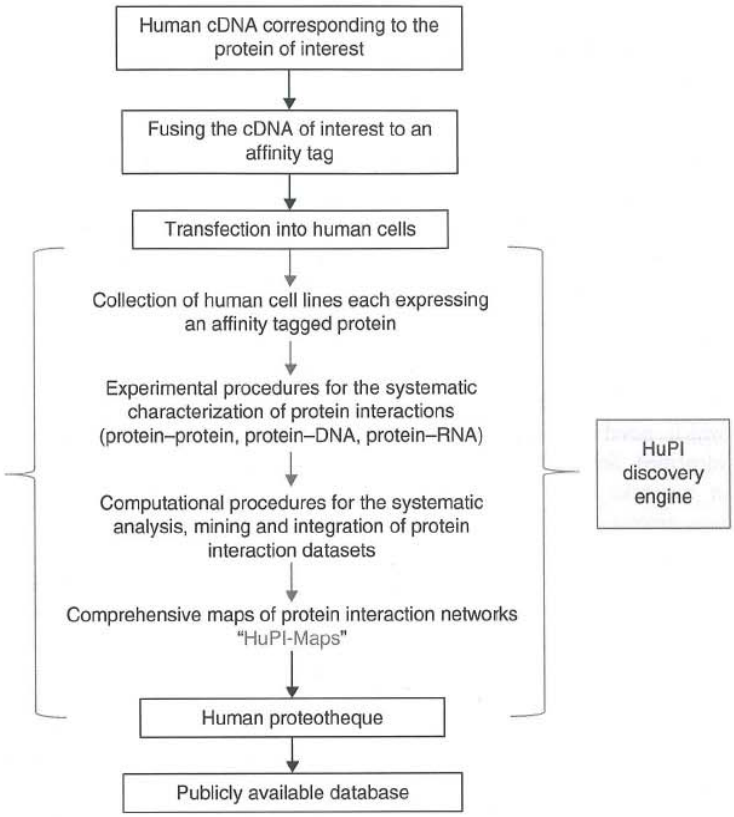


Figure 1. Overview of the Human Proteotheque Initiative. This flow chart describes the series of steps required to build the Human Proteotheque, a repertoire of comprehensive maps of human protein interaction networks (the HuPI-Maps). The HuPI discovery engine refers to the core of the procedure that generates datasets made publicly available through the internet.

(protein-fragment complementation assay) [18] and high-throughput imaging of protein localization [19], have also been developed to help map protein–protein interactions in space and time in mammalian cells. Although these approaches have not been widely used until now, they will for sure serve in enhancing the confidence of protein–protein interactions by helping to describe the local topology of protein interaction networks [12].

3. HuPI

Because defining protein interaction networks in cells of various normal and disease conditions is a titanic challenge, a number of large-scale projects, grouping together many investigators from different institutions, were launched in different countries. As an example, the Human Proteotheque Initiative (HuPI) is a continuing project that has been developed to generate comprehensive maps of protein–protein interaction networks describing some fundamental human molecular networks that regulate cell growth, differentiation and disease progression (refer to **Figure 1** for a schematic representation). The generated 'HuPI-map' is deposited in a repertoire, the human proteotheque, which is made publicly available as a resource for the scientific community. In building this repertoire, which focuses in its first phase of development on the protein complexes and networks involving the molecular machines that regulate the transcription, replication and repair of the human genome, the HuPI is conducted in such a way that data acquisition and analysis are performed in a highly systematic manner, favoring automation when possible, preventing any bias that could occur owing to human decisions and using computational tools to ensure that comprehensive interaction networks are generated. Doing more so will further ensure that the protein interaction maps deposited in the proteotheque are as accurate and complete as possible. As the HuPI project progresses, new HuPI-maps will be made available online to be used by the scientific community [3].

4. The potential of studying protein–protein interactions in drug discovery

In the past, efforts in drug discovery have focused almost exclusively on targets with enzymatic activities such as G-protein-coupled receptors and protein kinases [20]. The

identification and validation of new drug targets is a very difficult, time consuming and expensive process. Nowadays, the scope of drug discovery is shifting from a 'protein-centric' view towards a more global 'network-based' view for the treatment of complex diseases.

Large-scale, genome-wide protein–protein interaction screens and the mapping of protein interaction networks are now showing a great potential in the identification of novel therapeutic targets and putative biomarkers to be used as diagnostic tools, and are the key in understanding the molecular mechanisms of diseases [20].

Protein–protein interaction studies involving specific groups of proteins implicated in particular diseases such as ataxia [21,22], xeroderma pigmentosum [23] and faconi anemia [24] have been reported in which novel proteins involved in these diseases have been identified. Scientists hoped that by defining more interaction networks it would be possible to identify novel interacting protein partners and link the interacting pairs to known cellular pathways [20].

Furthermore, protein–protein interaction studies involving proteins of the death pathway have been reported. For example, members of the B-cell chronic lymphocytic leukemia/lymphoma 2 (Bcl-2) family of proteins are known to play a central role in programmed cell death by inhibiting apoptosis. In fact, members such as Bcl-2 and Bcl-x, have been found to be overexpressed in many cancers and contribute to tumor initiation, progression and resistance to therapy [25]. Therefore, developing an inhibitor to these proteins would serve as potential anticancer therapeutics and has already been explored, but obtaining a potent small-molecule effector remains a difficult task due to the necessity of specifically targeting a protein–protein interaction [25]. Protein–protein interaction studies have facilitated the development of a Bcl-2 homology 3 (BH3) domain mimetics that bind Bcl-x family members, thereby relieving their inhibitory interaction with the death-inducers Bcl-2–associated X protein (Bax) and Bcl-2 homologous antagonist/killer (Bak) [25].

Another interesting example concerns the case of protein 53 (p53), which is a tumor suppressor gene that lacks an enzymatic activity and is mutated or deleted in 50% of human cancers [26,27]. In the remaining cancers, the specific interaction of p53 with the E3 ubiquitin ligase murine double minute (MDM2) was shown to negatively regulate its activity [26,27]. A class of small-molecule inhibitors of the p53–MDM2 interaction, the nutlins, has been reported in 2004, which leads to the stabilization and activation of p53 thereby inducing the apoptosis of p53-positive cancer cells [26,27]. At present, in clinical trials, both the BH3 domain mimetics and nutlins are in fact showing very good antitumor activity.

5. Conclusion

The study of protein interaction networks represents a relatively young field of investigation but success stories such as nutlins, the BH3 domain mimetics and others demonstrate the importance of the systematic characterization of protein interaction networks in drug discovery.

By comprehensively describing the dynamic topology of the protein interaction networks underlying physical connections between disease-causing proteins and their regulatory interaction partners, large-scale, systematic projects such as the HuPI promise to serve as a tool for understanding how the mis-wiring of protein networks that underlie complex diseases such as cancer can be exploited in the design of new classes of network-based therapeutics.

6. Expert opinion

Integrative systems biology is the natural next step of research in biology and medicine, so much so that scientists who adopt systems biology approaches in their research are the most competitive now and will be increasingly so in the future.

Defining the interaction networks that underlie cell functions is one of the most important yet challenging aspects of systems biology research that will require a coordinated international effort. The literature contains numerous papers reporting on protein–protein interactions [3], and what would be more valuable is to integrate these interaction data and efforts into a public dataset that is accessible to the scientific community.

A number of projects have indeed emerged around the world, one of which is the HuPI. As discussed, this project is aimed at building a repertoire of comprehensive maps of protein interaction networks that underlie cellular functions in humans and make it publicly available through the Internet together with the protein interaction datasets. Doing so will hopefully help the users find all the protein interaction information relevant to their research. Expanding this repertoire of human protein interaction maps now requires a concerted initiative involving key research centers around the world and a multi-site discovery platform aimed at defining protein–protein, protein–DNA, protein–RNA and protein–lipid interactions, and integrating the resulting data into comprehensive protein interaction maps through bioinformatics. Furthermore, because protein interactions are controlled by post-translational modifications, protein stability, subcellular localization and gene expression, a multi-center, concerted initiative is also required to capture the dynamic nature of mammalian networks and the re-wiring of networks in response to normal and disease-related signals. It is expected that funding agencies around the world and industrial stakeholders would actively support such an integrated project because of the tremendous contribution it promises to have both to drug discovery and technology development.

Declaration of interest

The authors state no conflict of interest and have received no payment in preparation of this manuscript.

Bibliography

1. Bauer A, Kuster B. Affinity Purification- mass spectrometry: Powerful tools for the characterization of protein complexes. Eur J Biochem 2003;270:570-8
2. Jaeger S, Gaudan S, Leser U, Rebholz-Schuhmann D. Integrating protein- protein interactions and text mining for protein function prediction. BMC Bioinformatics 2008;9(Suppl 8):S2
3. Coulombe B, Blanchette M, Jeronimo C. Steps towards a repertoire of comprehensive maps of human protein interaction networks: the Human Proteotheque Initiative (HuPI). Biochem Cell Biol 2008;86(2):149-56
4. Bouwmeester T, Bauch A, Ruffner H, et al. A physical and functional map of the human TNF-alpha/NF-kappa B signal transduction pathway. Nat cell Biol 2004;6:97-105
5. Barrios-Rodiles M, Brown KR, Ozdamar B, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. Science 2005;307(5715):1621-5
6. Shiota M, Kusakabe H, Hikita Y, et al. Pharmacogenomics of cardiovascular pharmacology: molecular network analysis in Pleiotropic effects of statin – an experimental elucidation of the pharmacologic action from protein- protein interaction analysis. J Pharmacol Sci 2008;107:15-19
7. Miernyk JA, Thelan JJ. Biochemical approaches for discovering protein- protein interactions. Plant J 2008;53(4):597-609
8. Fleger KDG, Eidne KA. Illuminating insights into protein- protein interactions using bioluminescence resonance energy transfer (BRET). Nat Methods 2006;3(3):165-7
9. Tramier M, Gautier I, Piolot T, et al. Picosecond- hetero- FRET microscopy to probe protein- protein interactions in live cells. Biophys J 2002;83(6):3570-7
10. Wilchek M, Chaiken I. An overview of affinity chromatography. Methods Mol Biol 2000;147:1-6
11. Smith GP. Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. Science 1985;228(4705):1315-7
12. Figeys D. Mapping the human protein interactome. Cell Res 2008;18:716-24
13. Gingras A-C, Stager M, Raught B, Aebersold R. Analysis of protein complexes using mass spectrometry. Nat Rev Mol Cell Biol 2007;8(8):645-54
14. Jeronimo C, Forget D, Bouchard A, et al. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. Mol Cell 2007;27(2):262-74
15. Yu H, Braun P, Yildirim MA, et al. Science 2008. DOI:10.1126/science.1158684
16. Rigault G, Shevchenko A, Rutz B, et al. A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 1999;17(10):1030-2
17. Ruffner H, Bauer A, Bouwmeester T. Human protein-protein interaction networks and the value for drug discovery. Drug Discov Today 2007;12(17-18):709-16
18. Michnick SW, Ear PH, Manderson EN, et al. Universal strategies in research and drug discovery based on protein-fragment complementation assays. Nat Rev Drug Discov 2007;6:569-82
19. Rosochacki SJ, Matejczyk M. Green fluorescent protein as a molecular marker in microbiology. Acta Microbiol Pol 2005;51(3):205-16
20. Kramer R, Cohen D. Functional genomics to new drug targets. Nature 2004;3:965-72
21. Lim J, Hao T, Shaw C, et al. A protein–protein interaction network for human inherited ataxias and disorders of purkinje cell degeneration. Cell 2006;125(4):801-14
22. Chen H-K, Fernandez-Funez P, Acevedo SF, et al. Interaction of Akt-phosphorylated ataxin-1 with 14-3-3 mediates neurodegeneration in spinocerebellar ataxia type 1. Cell 2003;113(4):457-68
23. Giglia-Mari G, Coin F, Ranish JA, et al. A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. Nat Genet 2004;36(7):714-9
24. Meetei AR, de Winter JP, Medhurst AL, et al. A novel ubiquitin ligase is deficient in Fanconi anemia. Nat Genet 2003;35(2):165-70
25. Oltersdorf T, Elmore SW, Shoemaker AR, et al. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature 2005;435(7042):677-81
26. Vassilev LT, Vu BT, Graves B, et al. Liu. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. Science 2004;303(5659):844-8
27. Vassilev LT. MDM2 inhibitors for cancer therapy. Trends Mol Med 2007;13:23-31

Affiliation

Racha Al-Khoury[1] & Benoit Coulombe[†1,2]
[†]Author for correspondence
[1]Institut de recherches cliniques de Montréal,
Montréal (Québec),
Canada H2W 1R7
[†2]Université de Montréal,
Institut de recherches cliniques de Montréal,
Département de biochimie,
Laboratory of Gene Transcription and Proteomics,
110 avenue des Pins Ouest,
Montréal (Québec),
Canada H2W 1R7
Tel: +514 987 5662; Fax: +514 987 5663;
E-mail: benoit.coulombe@ircm.qc.ca

Elsevier Editorial System(tm) for Methods
Manuscript Draft

Manuscript Number:

Title: High-resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II-associated complexes

Article Type: Invited Article

Keywords: Protein affinity purification; RNA polymerase II; protein interaction networks; protein complexes; proteomics; bioinformatics

Corresponding Author: Dr. Benoit Coulombe,

Corresponding Author's Institution: Institut de recherches cliniques de Montréal

First Author: Philippe Cloutier

Order of Authors: Philippe Cloutier; Racha Al-Khoury; Mathieu Lavallée-Adam; Denis Faubert; Heng Jiang; Christian Poitras; Annie Bouchard; Diane Forget; Mathieu Blanchette; Benoit Coulombe

Abstract: Thirty years of research on gene transcription has uncovered a myriad of factors that regulate, directly or indirectly, the activity of RNA polymerase II (RNAPII) during mRNA synthesis. Yet many regulatory factors remain to be discovered. Using protein affinity purification coupled to mass spectrometry (AP-MS), we recently unraveled a high-density interaction network formed by RNAPII and its accessory factors from the soluble fraction of human cell extracts. Validation of the dataset using a machine learning approach trained to minimize the rate of false positives and false negatives yielded a high-confidence dataset and uncovered novel interactors that regulate the RNAPII transcription machinery, including a new protein assembly we named the RNAPII-Associated Protein 3 (RPAP3) complex.

Cover Letter

December 8th, 2008

Dear Dr. Price,

Please find attached our invited article entitled ''High-resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II-associated complexes'' that I am submitting for publication in Methods.

Sincerely.

Benoit Coulombe

High-resolution mapping of the protein interaction network for the human transcription machinery and affinity purification of RNA polymerase II-associated complexes

Philippe Cloutier[a], Racha Al-Khoury[a], Mathieu Lavallée-Adam[b], Denis Faubert[c], Heng Jiang[a], Christian Poitras[a], Annie Bouchard[a], Diane Forget[a], Mathieu Blanchette[b] and Benoit Coulombe[a,c,d]

[a] : Institut de recherches cliniques de Montréal
Montréal (Québec) Canada

[b] : McGill Centre for Bioinformatics
McGill University, Montréal (Québec) Canada

[c] : Proteomics Discovery Platform
Institut de recherches cliniques de Montréal
Montréal (Québec) Canada

[d] : Département de biochimie
Université de Montréal, Montréal (Québec) Canada

Corresponding Author:
Benoit Coulombe
Laboratory of Gene Transcription and Proteomics
Institut de recherches cliniques de Montréal
110 avenue des Pins Ouest
Montréal (Québec) Canada H2W 1R7
Tel.: (514) 987-5662
Fax: (514) 987-5663
E-mail: benoit.coulombe@ircm.qc.ca

Abstract

Thirty years of research on gene transcription has uncovered a myriad of factors that regulate, directly or indirectly, the activity of RNA polymerase II (RNAPII) during mRNA synthesis. Yet many regulatory factors remain to be discovered. Using protein affinity purification coupled to mass spectrometry (AP-MS), we recently unraveled a high-density interaction network formed by RNAPII and its accessory factors from the soluble fraction of human cell extracts. Validation of the dataset using a machine learning approach trained to minimize the rate of false positives and false negatives yielded a high-confidence dataset and uncovered novel interactors that regulate the RNAPII transcription machinery, including a new protein assembly we named the RNAPII-Associated Protein 3 (RPAP3) complex.

Keywords: Protein affinity purification, RNA polymerase II, protein interaction networks, protein complexes, proteomics, bioinformatics

1. Introduction

Protein affinity purification coupled to mass spectrometry (AP-MS) is a method of choice for the characterization of protein complexes and the identification of protein interaction partners. Among the various possible strategies for conducting AP-MS, the use of an affinity tag attached to the protein of interest has been widely used. For example, the Tandem Affinity Purification (TAP) procedure was used by many different groups in yeast and mammalian cells to characterize protein interaction networks and complexes involved in various cellular processes [1,2]. The output of the AP-MS procedure is usually a list of proteins in which each putative interactor is attributed a mass spectrometry (MS) score through the use of specialized software such as MASCOT or SEQUEST. Making sense of these lists is often the major challenge of the procedure. Even if a negative control is used to identify proteins that contaminate the affinity purified eluate (*e.g.* proteins found both in the control and affinity purified eluates), the challenge remains significant, especially when dealing with interactors that are either unexpected (proteins known to function in a different pathway or process) or that have not been previously characterized and have not been assigned a function. However, these proteins that are unexpected in a list of candidate interactors are often the most promising in terms of proteomic discovery.

As the experimental validation of the complete list of putative interaction partners (preys) for a tagged protein (bait) is not an option because it may contain many proteins, more manageable solutions need to be defined. Four specific solutions are most convenient. First, the use of additional purification steps that would ''clean'' the affinity purified eluate from contaminating proteins have been used in many reports [3]; one caveat of this solution is that weak, transient interactions will more readily be disrupted, leading to the loss of putatively interesting interactors. Secondly, the use of high-accuracy mass spectrometers leads to a significant decrease in the rate of false positives generated by the inaccuracy in the molecular mass of peptides used in protein identification. Thirdly, the use of expression systems that avoid the overexpression of the tagged proteins has been shown to minimize the occurrence of spurious interactions in protein-protein interaction datasets. Finally, the development of computational procedures that help to increase the confidence in protein-protein interaction datasets is often powerful and very useful [4].

In previous work, we reported the development and use of a multistep procedure for the systematic analysis of the protein interaction network for the human RNAPII transcription machinery [5,6]. This proteomic procedure couples (*i*) affinity purification of tagged proteins (TAP-tagging) from the soluble cell fraction in gentle conditions in order to preserve weak, transient interactions, (*ii*) the identification of co-purified proteins using sensitive, high-accuracy mass spectrometry, (*iii*) the validation of protein-protein interactions using a computational algorithm trained to minimize the rate of false positive and false negative interactions and (*iv*) the schematic representation of protein-protein interaction networks to visualize protein connectivity. A key aspect of our procedure relies on the reciprocal tagging of preys identified in our experiments. This step is important to confirm some interactions and to expand the dataset. In this procedure, the accuracy of the network that we end up mapping increases proportionally with the number of baits used in our AP-MS experiments. As compared to the dataset we

published in 2007 [6], which was built using 32 baits, the dataset used in the current report uses 77 baits. Coupled to other technical improvements of the procedure, especially regarding MS accuracy and sensitivity, the dataset used in the current report permits a higher resolution in mapping the human RNAPII transcription network. A schematic representation of our procedure is presented in Fig. 1.

Our previous 32-bait dataset allowed the identification of two novel factors that regulate the activity of the positive transcription elongation factor P-TEFb, a factor that is recruited to the RNAPII elongation complex where it functions by phosphorylating the RNAPII CTD and negative elongation factors such as NELF and DSIF to stimulate transcriptional elongation. The newly identified factors favor the sequestration of P-TEFb away from chromatin DNA. Indeed, the methylphosphate capping enzyme MEPCE and the RNA-binding protein LARP7 associate with and stabilize the 7SK snRNA which, in association with inhibitory proteins termed HEXIMs, binds to P-TEFb and prevents its recruitment to transcribing RNAPII complexes [6,7]. The formation of transcriptionally active P-TEFb requires its dissociation from the HEXIM-7SK inhibitory complex.

The 32-bait dataset also identified a set of proteins that are tightly connected to RNAPII [5,6]. Accordingly, these proteins were named RNAPII-Associated Proteins (RPAPs). Although the exact function of these factors remains elusive, the 77-bait dataset now reveals that one of these RPAPs, RPAP3, is part of an 11-subunit protein complex akin to that described by Gstaiger and colleagues [8] (see figures for details). Of note, and in contrast to the aforementioned report, we have not found STAP1 in any of our purifications, and our complex is more similar in composition with that described in a recent report [9], with the exception that, unlike POLR2E (RPB5), our data do not support the idea that the RNAPIII subunit POLR3A (RPC1) is a *bona fide* component of this complex. In addition to POLR2E, this complex contains human homologues of the yeast R2TP complex [10,11], including RPAP3 (Tah1, Spag), PIH1D1 (Pih1, Nop17), RUVBL1 (Tip49) and RUVBL2 (Tip48). WDR92 (Monad), whose interaction with RPAP3 was recently published [12], is also present but unlike R2TP, this complex comprises PFDN2 and PFDN6 (HKE2), two subunits shared with the canonical prefoldin complex, and three prefoldin-like proteins, UXT (ART-27), PDRG1 and C19orf2 (URI, RMP). The latter is a well-characterized protein known to bind POLR2E [13] and has a conserved role in TOR signaling [8,14]. Other proteins co-purified with some tagged subunits of the RPAP3 (R2TP/Prefoldin-like) complex, notably POLR2A (RPB1), RPAP2, RPAP4 and RP11-529I10.4 (DPCD), confirming that these are likely transient interactors, but not components of the stable complex. The unidirectionality of these interactions also supports this conclusion.

2. Affinity purification of RNA polymerase II and its associated proteins

2.1. Generation of stable human cell lines expressing TAP-tagged proteins

Full-length human cDNAs (Open Biosystems) were amplified by PCR and cloned into the pMZI vector using either *Xho*I, *Xba*I, *Sal*I, *Not*I or *Nde*I sites. pMZI drives the expression of proteins with a C-terminal TAP-tag in mammalian cells under the control of an ecdysone inducible promoter [1]. The resulting plasmids were transfected into HEK 293 cells (ATCC) previously transfected with pVgRxR (Invitrogen), which encodes the

ecdysone receptor heterodimer. Stable clones were selected and grown in DMEM media supplemented by 10% fetal bovine serum, 2 mM glutamine, 30 µg/ml Bleocin (Calbiochem) and 300µg/ml G418 (Invitrogen). Once the culture reached a volume of about one-hundred 150 mm plates and a confluency of 20%, the cells were induced in half of these plates in 3 µM ponasterone A (Invitrogen). The other half is maintained as is to serve as a negative control. Two days following induction, the cells were harvested in ice-cold PBS by tapping plates and centrifuged at 3,000 rpm for 10 min at 4 °C in a SLA-3000 rotor (Sorvall). Supernatant was discarded and pellets were dislodged with 20 ml ice-cold PBS, transferred to 50 ml Falcon tubes and spun once more for 10 min at 4 °C in a table top centrifuge at 3,000 rpm. Following decantation, the pellets were usually frozen at -70 °C before proceeding the next day with extraction of protein complexes.

2.2. Preparation of whole cell extracts

All of the following steps were performed on ice to limit protein precipitation and proteolysis. Buffer A (10 mM Hepes, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM DTT, 0.5 mM AEBSF, complete EDTA-free protease inhibitor cocktail (Roche)) was added to the pellet with a ratio of 4/3 (ml/g of cells). The mixture was transferred to a 10 or 30 ml glass homogenizer (Wheaton) and the pellet was broken up by 10 gentle strokes while being careful not to create foam in the sample. By pipetting up and down, buffer B (50 mM Hepes, pH 7.9, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM AEBSF, 1.26 M K acetate, glycerol 75%) was mixed in the lysate with a ratio of 1/1 (ml/g of cells). Membranes were further disrupted by 10 extra pestle strokes. The lysate was poured into an ultracentrifuge tube and incubated for 30 min at 4 °C on a mixer and then spun at 37,000 rpm for 3 hours at 4 °C in a 50.2TI rotor (Beckman Coulter). The soluble fraction was separated from insoluble materials and dialyzed overnight using 18 mm Spectra/Por 3 membranes (Spectrum Laboratories) in 3 liters dialysis buffer (10 mM Hepes, pH 7.9, 0.1 mM EDTA, 0.1 mM DTT, 0.1 M K acetate, 10% glycerol). The following day, the cell extracts were transferred to 15 ml Corex tubes and centrifuged once more at 14,000 rpm for 30 min at 4 °C in a SS-34 rotor (Sorvall).

2.3. Tandem-affinity purification

For each gram of cells harvested, 50 µl of the IgG sepharose 6 Fast Flow beads (GE Healtcare) used in the first affinity purification were washed by centrifugation at 3,000*g* for 2 min at 4 °C once in 1 ml of Tris-saline Tween-20 (TST) buffer (50 mM Tris-HCl, pH 7.6, 150 mM NaCl, 0.05% Tween-20) and then twice in 500 µl immunoprecipitation (IPP) buffer (10 mM Tris-HCl, pH 8, 100 mM NaCl, 0.1% Triton X-100, 10% glycerol). The whole cell extract was incubated with the washed beads for 1 hour at 4 °C on a mixer. The resin was then washed twice in 500 µl IPP before loading onto a Bio-spin disposable chromatography column (Bio-Rad). Five hundred µl of TEV buffer (10 mM Tris-HCl, pH 8, 100 mM NaCl, 0.1% Triton X-100, 0.5 mM EDTA, 10% glycerol, 1 mM DTT) were allowed to drip out by gravity before plugging the column and adding 30 units of AcTEV protease (Invitrogen) in 200 µl of TEV. The column was incubated over night at 4 °C on a mixer to allow cleavage of the protein A component of the tagged polypeptide, thereby freeing the protein complexes from the IgG beads.

The next morning, the column was opened and drained into a tube containing 50 µl of calmodulin sepharose 4B beads (GE Healthcare) that were previously washed twice in 500 µl calmodulin binding (CBB) buffer (10 mM Tris-HCl, pH 8, 100 mM NaCl, 1mM immidazole, 1 mM Mg acetate, 2 mM $CaCl_2$, 0.1% Triton X-100, 10% glycerol, 10 mM β-mercaptoethanol). The column was further washed with 50 µl of TEV and 2×300 µl of CBB all of which were eluted by gravity flow onto the calmodulin resin before adding 0.8 µl of 1 M $CaCl_2$ directly to the tube. The resulting mixture was incubated 2 hours at 4 °C on a mixer. Following incubation, the beads were washed twice in 500 µl CBB and loaded into another Bio-spin column. To push out remaining CBB from the beads, 40 µl of calmodulin elution (CEB) buffer (10 mM Tris-HCl, pH 8, 100 mM NaCl, 1mM immidazole, 1 mM Mg acetate, 2 mM EGTA, 10% glycerol, 10 mM β-mercaptoethanol) was added to the column. The protein complexes were then eluted from the beads by addition of two volumes of 100 µl and 150 µl of CEB. The eluate was frozen in liquid nitrogen and its volume was reduced to about 30 µl using a speed vac. The complexes were then separated on a 1 mm NuPAGE 4-12% Bis-Tris Gel (Invitrogen) followed by silver staining (see Fig. 2 for an example).

3. Analysis of protein complexes using mass spectrometry

3.1. Protein digestion with trypsin

The entire gel lane was excised into 18-20 bands and each band was cut in 1 mm^3 pieces. For the following steps, all volumes were adjusted according to the volume of gel pieces. Gel pieces were first washed with water for 5 min and destained twice with the destaining buffer (100 mM sodium thiosulfate, 30 mM potassium ferricyanide) for 15 min. An extra wash of 5 min was performed after destaining with a buffer of ammonium bicarbonate (50 mM). Gel pieces were then dehydrated with acetonitrile. Proteins were reduced by adding the reduction buffer (10 mM DTT, 100 mM ammonium bicarbonate) for 30 min at 40°C, and then alkylated by adding the alkylation buffer (55 mM iodoacetamide, 100 mM ammonium bicarbonate) for 20 min at 40°C. Gel pieces were dehydrated and washed at 40°C by adding ACN for 5 min before discarding all the reagents. Gel pieces were dried for 5 min at 40°C and then re-hydrated with the trypsin solution (6 ng/µL of trypsin sequencing grade from Promega, 25 mM ammonium bicarbonate). Protein digestion was performed at 58°C for 1 hour and stopped with 15 µl of 1% formic acid/2% acetonitrile. Supernatant was transferred into a 96-well plate and peptides extraction was performed with two 30-min extraction steps at room temperature using the extraction buffer (1% formic acid/50% ACN). All peptide extracts were pooled into the 96-well plate and then completely dried in vacuum centrifuge. The plate was sealed and stored at -20°C until LC-MS/MS analysis.

3.2. LC-MS/MS

Prior to LC-MS/MS, peptide extracts were re-solubilized under agitation for 15 min in 12 µl of 0.2% formic acid and then centrifuged at 2,000 rpm for 1 min. The LC column was a C18 reversed-phase column packed with a high-pressure packing cell. A 75 µm i.d. fused silica capillary of 100 mm long was packed with the C18 Jupiter 5 µm 300 Å reverse-phase material (Phenomenex). This column was installed on the nanoLC-2D system (Eksigent) and coupled to the LTQ Orbitrap (ThermoFisher Scientific). The buffers used for chromatography were 0.2% formic acid (buffer A) and 100 % acetonitrile/0.2% formic acid (buffer B). During the first 12 min, 5 µl of sample were loaded on column with a flow of 650 nL/min and, subsequently, the gradient went from 2-80% buffer B in 20 min and then came back to 2% buffer B for 10 min. LC-MS/MS data acquisition was accomplished using a four scan event cycle comprised of a full scan MS for scan event 1 acquired in the Orbitrap which enables high resolution/high mass accuracy analysis. The mass resolution for MS was set to 30,000 (at m/z 400) and used to trigger the three additional MS/MS events acquired in parallel in the linear ion trap for the top three most intense ions. Mass over charge ratio range was from 380 to 2000 for MS scanning with a target value of 500,000 charges and from ~1/3 of parent m/z ratio to 2000 for MS/MS scanning with a target value of 20,000 charges. The data dependent scan events used a maximum ion fill time of 100 msec and 1 microscan to increase the duty cycle for ion detection. Target ions already selected for MS/MS were dynamically excluded for 15 s. Nanospray, capillary and tube lens voltages were set to 0.9-1.6 kV, 5 V and 100 V, respectively. Capillary temperature was set to 200°C. MS/MS conditions were: normalized collision energy, 35V; activation q, 0.25; activation time, 30 msec.

In some experiments, we used two-dimensional (2D-) LC-MS/MS on peptide mixtures generated by trypsin digestion of TAP eluates that have not been submitted to SDS gel analysis. In most cases, the gel-free 2D-LC-MS/MS method produced results that were mainly confirmatory (and in some cases complementary) to the gel-based method described in the previous paragraph. For this reason 2D-LC-MS/MS will not be described in details here.

3.3. Protein identification

Protein database searching was performed with Mascot 2.1 (Matrix Science) against the human NCBInr protein database. The mass tolerances for precursor and fragment ions were set to 10 ppm and 0.6 Da, respectively. Trypsin was used as the enzyme allowing for up to 2 missed cleavages. Carbamidomethyl and oxididation of methionine were allowed as variable modifications.

4. MS data analysis

4.1. Reliability of protein-protein interactions

The list of candidate prey proteins identified by MS for a given bait is likely to contain a certain fraction of false-positives. These erroneous interactions could be the result of indirect interactions, incorrect peptide identification, contamination of gel lanes, etc. Previously, we developed an approach to assign a confidence score to each candidate interaction, based on the Mascot score of the prey and the local topology of the network [6]. Here, we describe an improved approach to the estimation interaction reliability score (IRS), following in part ideas originally proposed by Ewing et al. [4].

Several factors reflect, to various degrees, the probability that a candidate interaction is real. First are the output of the MS instrument and software, measuring the confidence in the identification of a given prey P from a bait B. The popular Mascot program [15] outputs various statistics supporting the protein identified. Two of them proved particularly useful at distinguishing true from false positive predictions: (*i*) the total Mascot score for P, and (*ii*) the highest Mascot score of all peptides found for P. In addition to MS-specific scores, our confidence in a particular interaction is reflected by properties of the protein-protein interaction (PPI) network surrounding the interaction of interest. In particular, the presence of proteins C_1, C_2, ... C_n interacting with both B and P may increase our belief in an interaction between B and P. We call n the number of common partners. Finally, other factors may affect our confidence in interaction (B,P): (*i*) whether the interaction is bidirectional (P is found as a prey when B is the bait, and vice-versa); (*ii*) the number of baits that found P as a prey.

The five features describing each interaction detected (Mascot score, best peptide Mascot score, number of common partners, bidirectionality, and number of baits per prey) are combined into a predictor using a logistic regression approach, which predicts the probability that an interaction is correct as a function of a weighted sum of these features and combinations thereof. Specifically, our logistic regression model includes 19 terms: each of the 5 features, the square of the value of each of the features (with the exception of the bidirectionality, which is a binary feature), and each of the products of the values of pairs of features (10 terms). Training the predictor, *i.e.* choosing the weight attributed to each of the 19 terms, requires a training set of interactions detected by mass spectrometry and deemed likely true positives, and a set of interactions deemed likely false positives. In Jeronimo et al. [6], we manually identified a set of 149 interactions that are strongly supported by the literature, and a set of 54 interactions that, on the basis of the function of the proteins alone (but without using our PPI data), seem likely false-positives. We call these interactions Literature-Likely and Literature-Unlikely, respectively. This high-quality training set, though small, provides an excellent basis for the evaluation of our approach. However, its limited size reduces its usefulness for training a complex predictor such as that proposed here. As an alternative, we used the protein Gene Ontology (GO) annotation [16] to label interactions as GO-Likely or GO-Unlikely. Let us call a GO category "x%-specific" if less than x% of the proteins annotated within the network have this annotation. An interaction was labeled GO-Likely if the bait and the prey share a 3%-

specific GO annotation. On the opposite, an interaction was labeled GO-Unlikely if the bait and the prey both have 10%-specific GO annotations but these do not overlap. It is important to note that our training procedure does not require our positive and negative training sets to be pure (*i.e.* to respectively contain only true positive and true negative interactions), nor do they need to be complete. The only requirement is that they are substantially enriched for a representative subset of these interactions. In our network of 5106 candidate interactions, 248 were labeled GO-Likely and 2403 were labeled GO-Unlikely, a sufficiently large training set to accurately learn the weights of our logistic regression and avoid over-fitting. Regression weights were chosen so as to minimize the cross-entropy between the prediction and the label. Weights assigned to each feature or feature combination (both standardized to have mean 0 and standard deviation 1) are listed in Supplementary Table 1. Because our training set is very noisy (in particular, many interactions labeled as GO-Likely are not real), the sensitivity and specificity of the predictor on the GO-based labels is relatively low (68% sensitivity and 69% specificity). However, when the same predictor is evaluated on the basis of its ability to separate Literature-Likely interactions from Literature-Unlikely ones (even though it has never used this type of labels for training), we obtain a more accurate estimate of its accuracy. As shown in Fig. 3 choosing an appropriate score threshold results in an 81% sensitivity (fraction of Literature-Likely predicted as positives), and 81% specificity (fraction of Literature-Unlikely predicted as negatives). In fact, given the likely presence of a few mis-annotated interactions in our literature-based set, the true accuracy of our prediction is likely to be higher than that. At the chosen threshold, 2355 of the 5106 candidate interactions are predicted as reliable. The Interaction Reliability Score (IRS) assigned to an interaction is the posterior probability of the interaction being real, given the score obtained from the logistic regression (assuming an equal priori probability for true and false interactions).

4.2. Graphic representation of protein-protein interactions

The assignment of IRS to individual protein-protein interactions and the selection of those interactions with IRS over a stringent threshold define high-confidence PPI datasets. Many different tools can be used to represent the data in a comprehensive manner. Here, we present two examples. Fig. 4 shows a graphical representation of part of our high-confidence interaction dataset in which edges that extend from a given bait are connected to the preys that have been confidently identified in our experiments. In this representation, which was generated using the VisANT software [17], the nodes have been clustered according to their GO annotation and/or the degree of connectivity between nodes. Fig. 5 shows a heat map where preys and baits are clustered based on the similarity of their sets of partners. Both Figs. 4 and 5 indicate the existence of the RPAP3 complex (R2TP/Prefoldin-like) and its tight connection to RNAPII and the CCT complex.

Acknowledgments

We are grateful to members of our laboratories for helpful discussions. This work was supported by grants from the Canadian Institutes of Health Research (CIHR) and the National Science and Engineering Council of Canada (NSERCC) to M.B and B.C

Figure legends

Figure 1. Schematic representation of our multi-step proteomics procedure. The procedure couples the regulated expression of affinity tagged proteins (baits) in human cells, the purification of putative interaction partners (preys) in the form of protein complexes, the identification of co-purified proteins using mass spectrometry (MS) and the computational validation and analysis of the data. Each step is referred to the appropriate section in the text.

Figure 2. Purification of the TAP-tagged human RPAP3 complex. TAP eluates of induced or non-induced HEK 293 cells expressing TAP-tagged RPAP3 (WDR92) were analyzed by SDS-PAGE. Gel lanes were cut from top to bottom, singling out discernable proteins, slices were digested with trypsin, and analyzed by LC-MS/MS. The position of molecular weight markers (left) and that of identified components of the RPAP3 and CCT complexes (right) are shown.

Figure 3. Sensitivity-specificity (ROC) curve for our logistic regression predictor on a set of literature-based interaction annotations (149 positive, 54 negative). Sensitivity = fraction of positive interactions that are predicted as positive. Specificity = fraction of negative interactions that are predicted as negative.

Figure 4. Network highlighting the composition of the RPAP3 (R2TP/prefoldin-like) complex (blue box). Subunit overlap with PFD (light blue box) and RNAP I, II and III (yellow box) is shown, as are interactions with RPAP2, RPAP4, RP11-529I10.4 (green nodes) and CCT complex (red box). Tagged proteins (baits) are in color, while untagged ones are represented in gray.

Figure 5. Heat map of a subset of our protein interaction network pertaining to the RPAP3 (R2TP/prefoldin-like) complex. For each pair bait/prey, the IRS is shown by color intensity (white being no interaction).

References

[1] G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann, B. Seraphin, Nat. Biotechnol. 17 (1999) 1030-1032.

[2] O. Puig, F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, B. Seraphin, Methods 24 (2001) 218-29.

[3] L. Florens, M.J. Carozza, S.K. Swanson, M. Fournier, M.K. Coleman, J.L. Workman, M.P. Washburn, Methods 40 (2006) 303-11.

[4] R.M. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M.D. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore, S. Zhang, O. Ornatsky, Y.V. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu, M. Abu-Farha, J.P. Lambert, H.S. Duewel, I.I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S.L. Adams, M.F. Moran, G.B. Morin, T. Topaloglou, D. Figeys, Mol. Syst. Biol. 3 89 (2007) 1-17.

[5] C. Jeronimo, M.F. Langelier, M. Zeghouf, M. Cojocaru, D. Bergeron, D. Baali, D. Forget, S. Mnaimneh, A.P. Davierwala, J. Pootoolal, M. Chandy, V. Canadien, B.K. Beattie, D.P. Richards, J.L. Workman, T.R. Hughes, J. Greenblatt, B. Coulombe, Mol Cell Biol 24 (2004) 7043-7058.

[6] C. Jeronimo, D. Forget, A. Bouchard, Q. Li, G. Chua, C. Poitras, C. Therien, D. Bergeron, S. Bourassa, J. Greenblatt, B. Chabot, G.G. Poirier, T.R. Hughes, M. Blanchette, D.H. Price, B. Coulombe, Mol Cell 27 (2007) 262-274.

[7] B.J. Krueger, C. Jeronimo, B.B. Roy, A. Bouchard, C. Barrandon, S.A. Byers, C.E. Searcey, J.J. Cooper, O. Bensaude, E.A. Cohen, B. Coulombe, D.H. Price, Nucleic Acids Res. 36 (2008) 2219-2229.

[8] M. Gstaiger, B. Luke, D. Hess, E.J. Oakeley, C. Wirbelauer, M. Blondel, M. Vigneron, M. Peter, W. Krek, Science 302 (2003) 1208-1212.

[9] M.E. Sardiu, Y. Cai, J.J. Jin, S.K. Swanson, R.C. Conaway, J.W. Conaway, L. Florens, M.P. Washburn, Proceedings of the National Academy of Sciences of the United States of America 105 (2008) 1454-1459.

[10] R.M. Zhao, M. Davey, Y.C. Hsu, P. Kaplanek, A. Tong, A.B. Parsons, N. Krogan, G. Cagney, D. Mai, J. Greenblatt, C. Boone, A. Emili, W.A. Houry, Cell 120 (2005) 715-727.

[11] R.M. Zhao, Y. Kakihara, A. Gribun, J. Huen, G.C. Yang, M. Khanna, M. Costanzo, R.L. Brost, C Boone, T.R. Hughes, C.M. Yip, W.A. Houry, Journal of Cell Biology 180 (2008) 563-578.

[12] Y. Itsuki, M. Saeki, H. Nakahara, H. Egusa, Y. Irie, Y. Terao, S. Kawabata, H. Yatani, Y. Kamisaki, FEBS Letters 582 (2008) 2365-2370.

[13] D. Dorjsuren, Y. Lin, W.X. Wei, T. Yamashita, T. Nomura, N. Hayashi, S. Murakami, Molecular and Cellular Biology 18 (1998) 7546-7555.

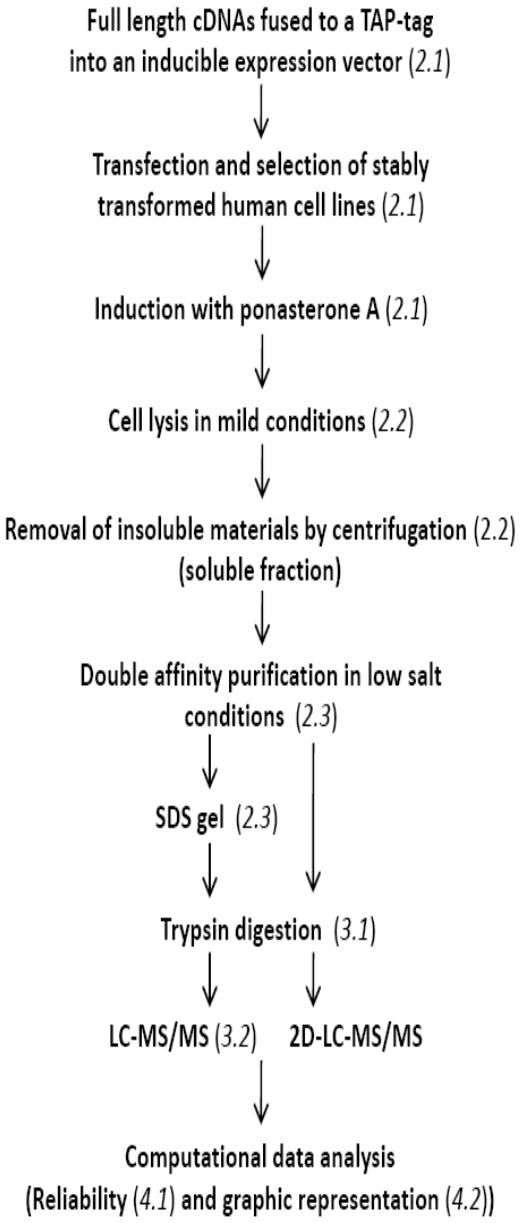
[14] N. Djouder, S.C. Metzler, A. Schmidt, C. Wirbelauer, M. Gstaiger, R. Aebersold, D. Hess, W. Krek, Molecular Cell 28 (2007) 28-40.

[15] D.N. Perkins, D.J. Pappin, D.M. Creasy, J.S. Cottrell, Electrophoresis 20 (1999) 3551-67.

[16] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Nat. Genet. 25 (2000) 25-9.

[17] Z. Hu, E.S. Snitkin, C. DeLisi, Brief Bioinform. 9 (2008) 317-25.

Figure 1


Full length cDNAs fused to a TAP-tag into an inducible expression vector (2.1)
Transfection and selection of stably transformed human cell lines (2.1)
Induction with ponasterone A (2.1)
Cell lysis in mild conditions (2.2)
Removal of insoluble materials by centrifugation (2.2) (soluble fraction)
Double affinity purification in low salt conditions (2.3)
SDS gel (2.3)
Trypsin digestion (3.1)
LC-MS/MS (3.2)
2D-LC-MS/MS
Computational data analysis (Reliability (4.1) and graphic representation (4.2))

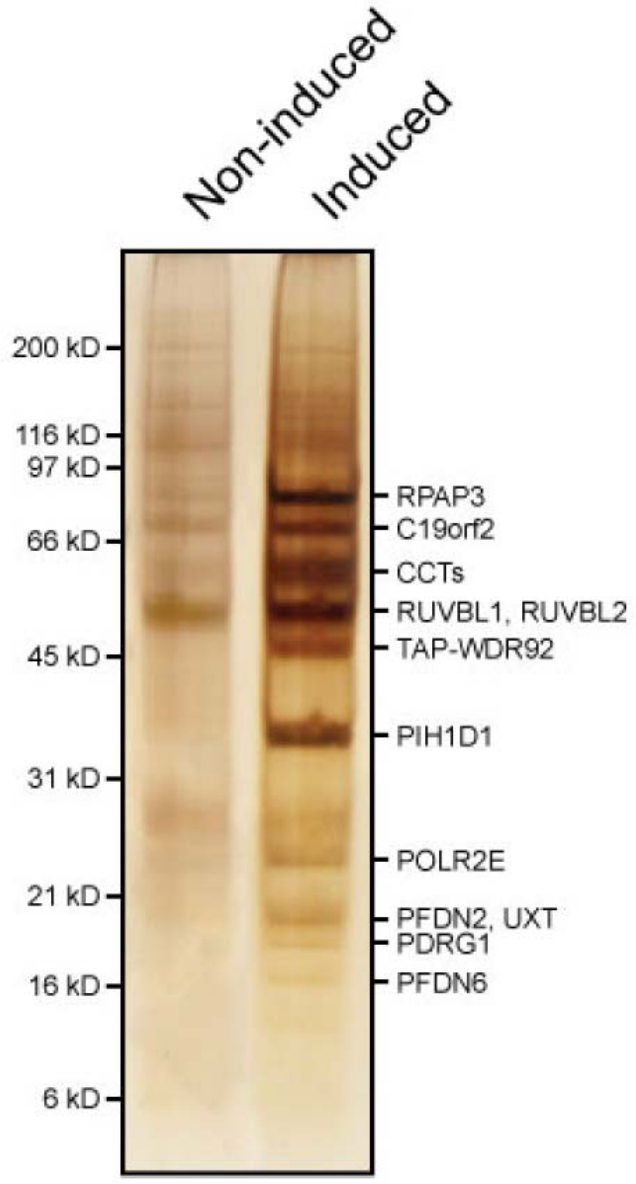

Non-induced
Induced
200 kD
116 kD
97 kD
66 kD
45 kD
31 kD
21 kD
16 kD
6 kD
RPAP3
C19orf2
CCTs
RUVBL1, RUVBL2
TAP-WDR92
PIH1D1
POLR2E
PFDN2, UXT
PDRG1
PFDN6

Figure 3

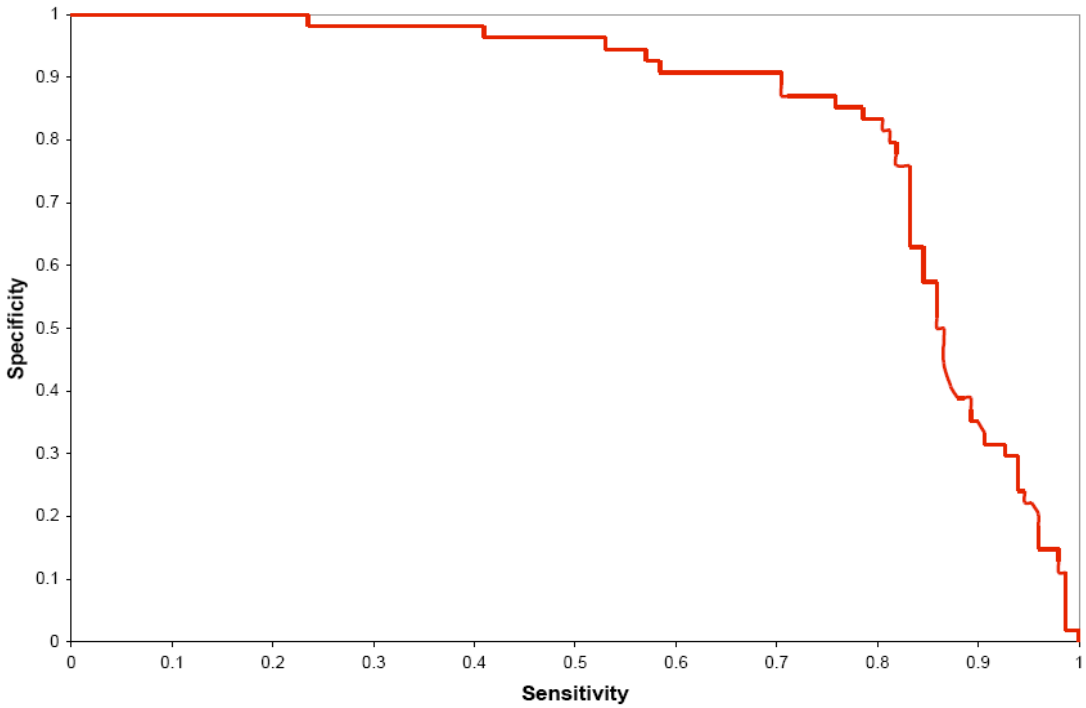

Specificity
Sensitivity
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 4

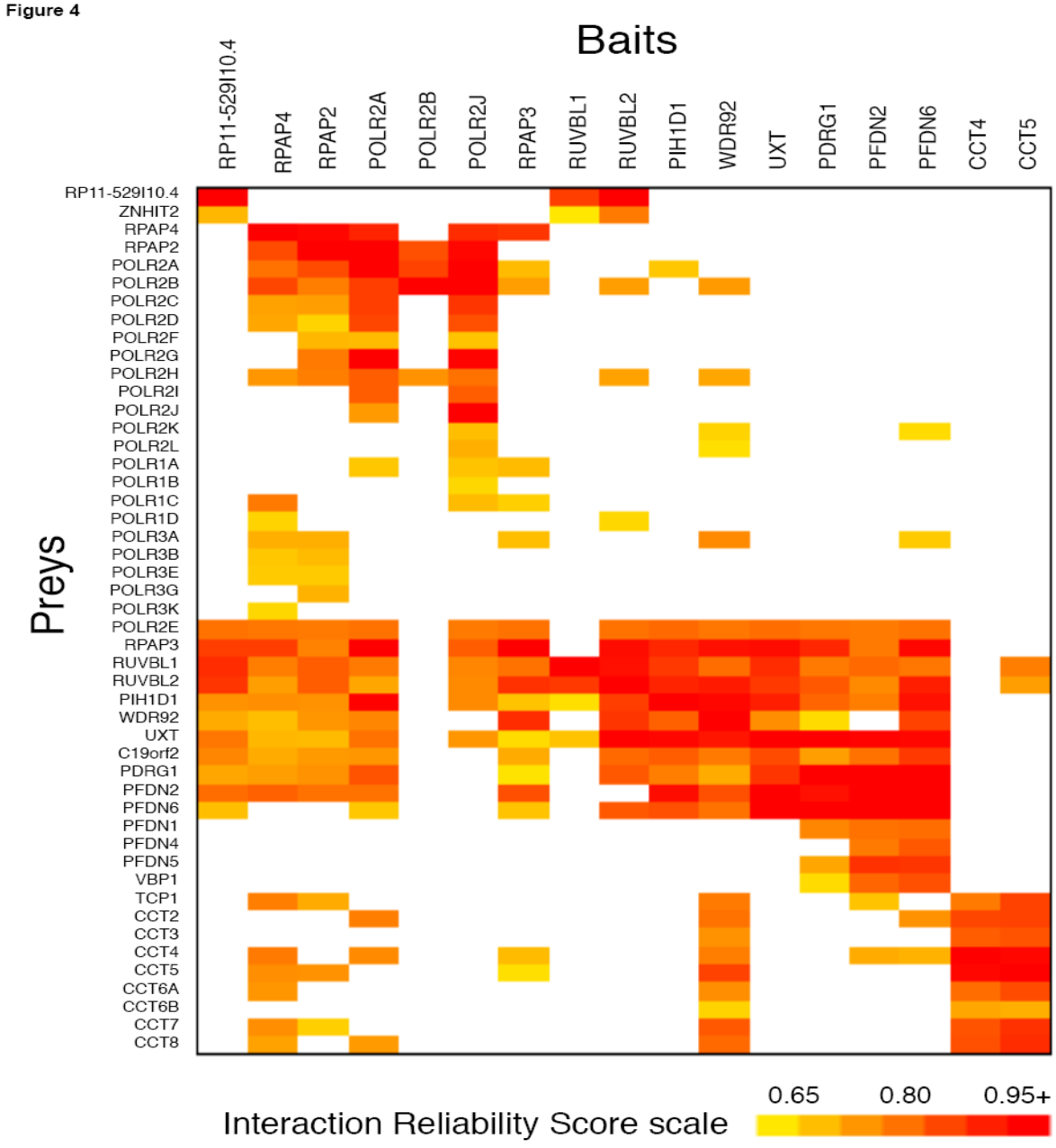
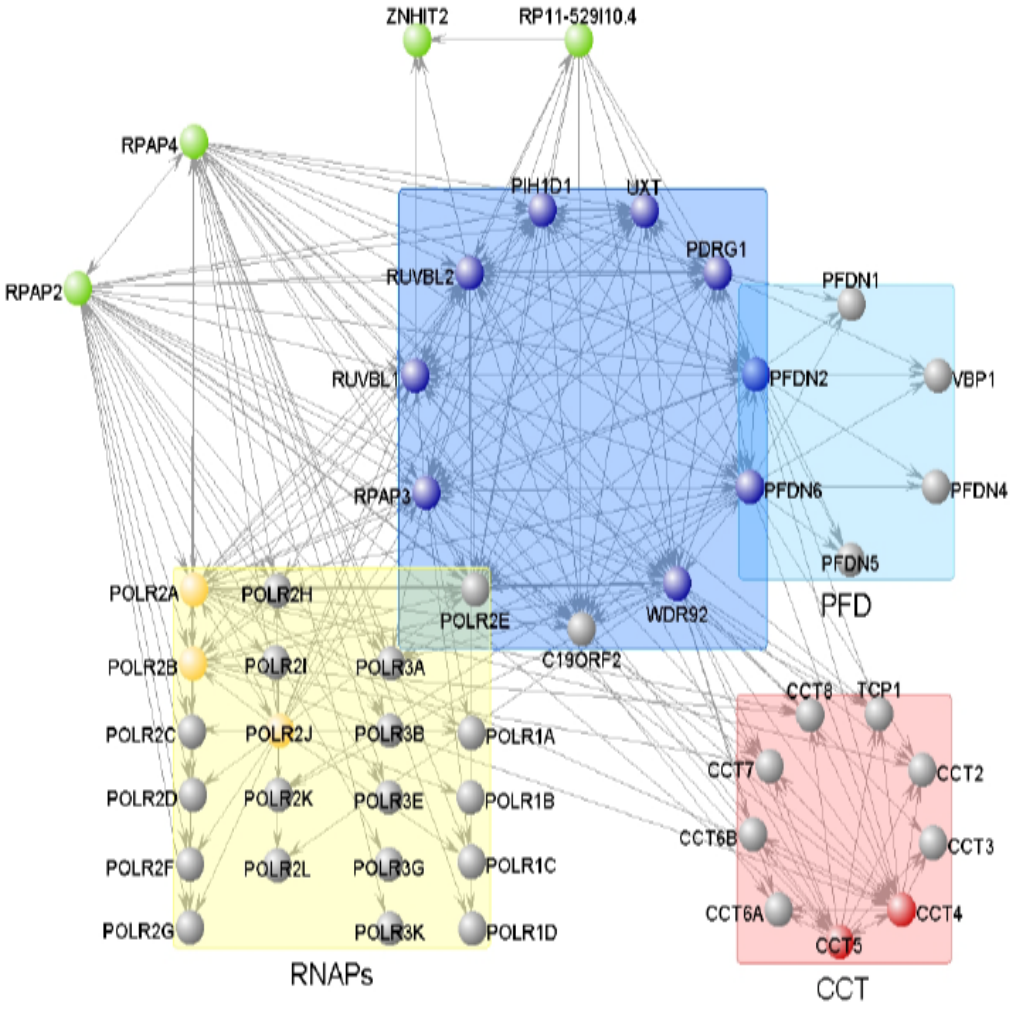

Baits
RP11-529I10.4
RPAP4
RPAP2
POLR2A
POLR2B
POLR2J
RPAP3
RUVBL1
RUVBL2
PIH1D1
WDR92
UXT
PDRG1
PFDN2
PFDN6
CCT4
CCT5
Preys
RP11-529I10.4
ZNHIT2
RPAP4
RPAP2
POLR2A
POLR2B
POLR2C
POLR2D
POLR2F
POLR2G
POLR2H
POLR2I
POLR2J
POLR2K
POLR2L
POLR1A
POLR1B
POLR1C
POLR1D
POLR3A
POLR3B
POLR3E
POLR3G
POLR3K
POLR2E
RPAP3
RUVBL1
RUVBL2
PIH1D1
WDR92
UXT
C19orf2
PDRG1
PFDN2
PFDN6
PFDN1
PFDN4
PFDN5
VBP1
TCP1
CCT2
CCT3
CCT4
CCT5
CCT6A
CCT6B
CCT7
CCT8
0.65
0.80
0.95+
Interaction Reliability Score scale

Figure 5


ZNHIT2
RP11-529I10.4
RPAP4
PIH1D1
UXT
PDRG1
RPAP2
RUVBL2
PFDN1
RUVBL1
PFDN2
VBP1
RPAP3
PFDN6
PFDN4
PFDN5
PFD
POLR2A
POLR2H
POLR2E
WDR92
C19ORF2
POLR2B
POLR2I
POLR3A
CCT8
TCP1
POLR2C
POLR2J
POLR3B
POLR1A
CCT7
CCT2
POLR2D
POLR2K
POLR3E
POLR1B
CCT6B
CCT3
POLR2F
POLR2L
POLR3G
POLR1C
CCT6A
CCT4
POLR2G
POLR3K
POLR1D
CCT5
RNAPs
CCT

Supplementary Table S1

Click here to download Supplementary Material: TableS1.Methods.2008.doc