universität wien

DISSERTATION

Titel der Dissertation

Finding RNA targets for the multidomain cyclophilin AtCyp59 by Genomic SELEX in *Arabidopsis thaliana*

angestrebter akademischer Grad

Doktor/in der Naturwissenschaften (Dr. rer.nat.)

Verfasserin / Verfasser:	Dipl. Biochem. Olga Bannikova
Matrikel-Nummer:	0709600
Dissertationsgebiet(lt. Studienblatt):	Molekulare Biologie
Betreuerin / Betreuer:	ao.Univ.-Prof. Dr. Andrea Barta

Wien, am 15. Dezember 2010

Посвящается

Моим родителям, за то, что благодаря им я смогла осуществить то, что хотела, и которые были со мной всегда рядом, давали мне уверенность в себе и в завтрашнем дне.

Косте, который делил со мной дни и ночи в лаборатории, мои радости и печали, всегда находил слова поддержки.

Моим двум котам, Семену и Дионисию, таким разным, но одинаково горячо любимым.

Acknowledgments

I would like to expresse my deepest gratefulness for all of you who were here with me these years:

My group leader, PhD committee members and collegues for encouragement and advices.

My friends for the great time we had together in Vienna

My family for being here for me always

And I owe my special gratitude to those without whom this project would not be accomplished:

Dr. Andrea Barta for being a great boss and giving me enough freedom to try all creasy ideas.

Dr. Marek Zywicki for bioinformatics analysis of the SELEX data and for being always curious to find something new. Without this part my work would not exist.

Mag. Andrij Belokurov for the establishing cell suspensions, protoplasts systems and for willingness to new discovery.

Abstract

Cyclophilins are ubiquitous proteins found in archea, bacteria, and eukarya, with the largest family described so far in the plant species *A. thaliana.*

Cyclophilins possess peptidyl–prolyl cis–trans isomerase (PPIase) activity, they catalyze cis–trans isomerization of peptide bonds preceding proline. The majority of cyclophilins (PPIases) are small proteins containing only a PPIase domain of about 120 amino acids. However, several multidomain cyclophilins from different organisms have been also described. Among them, the most complex multidomain cyclophilin characterized so far is *A. thaliana* Cyp59 (AtCyp59). It has a unique, conserved from yeast to human, domain organization, consisting of N-terminal PPIase domain followed by an RNA recognition motif (RRM) and a C-terminal domain enriched in charged amino acids and serines or RS/RD dipeptide repeats. As a consequence of its multidomain organization, AtCyp59 may be involved in several cellular processes such as splicing, RNA processing, protein trafficking and maturation, etc. AtCyp59 has been shown to localize in the nucleus where it interacts with the C-terminal domain of RNA Polymerase II. Via its SR/RD domain it interacts with majority of SR proteins. Up-to-date, the exact biological functions of RRM domain are still remaining unclear. Here we present genomic SELEX, the genome-wide screen method allowing us to search for RNA targets of RRM domain of AtCyp59. By random priming we constructed the representative genomic library of *A. thaliana,* consisting of 50-300nt long overlapping sequences. The library was accordingly used for genomic SELEX, allowing identification of the 7nt-long RNA binding consensus which was shown to bind to RRM domain of AtCyp59 in a sequence specific manner *in vitro* and *in vivo*. Mutations in either RRM domain or consensus sequence prevented formation of RNA-protein complex. Bioinformatics analysis has shown that this binding motif represents a global signature located near the translational stop-codon in 70% of *A. thaliana* mRNAs. Comparative analysis of RNA-binding activity AtCyp59 and its RRM domain indicated contribution of PPIase and RS/RD-rich domains to binding to structured RNAs. Presented data allowed us to shed a light on function of RRM domain of AtCyp59, global regulator protein lying on interconnection between two important cellular processes, transcription and splicing.

Zusammenfassung

Die Cyclophiline gehören zu den weitverbreitesten Proteinen in Archeen, Bakterien und Eukarionten und bilden auch die größte beschriebene Proteinfamilie der Spezies *A. thaliana.* Die Cyclophiline weisen eine peptidyl–prolyl *cis–trans* Isomerase (PPIase) Aktivität auf und katalysieren die *cis–trans* Isomerisierung der Prolin vorangehender Peptid-Bindungen. Die Mehrheit der Cyclophiline (PPIasen) ist eher klein, sie beinhalten lediglich eine einzige PPIase Domäne bestehend aus 120 Aminosäuren. Dennoch sind wenige Cyclophilin Proteine mit Mehrfachdomänen bekannt. Zu den höchst komplexen Cyclophilin Vertretern, gehört die *A. thaliana* Cyp59 (AtCyp59). Das Protein hat eine, in der Evolution von Hefe zu Menschen konservierte, besondere Domänen-Organisation. Diese besteht aus einer N-Terminal PPIase gefolgt von einem RNA-Erkennungs-Motiv (RNA recognition motif, RRM) und der C-Terminus Domäne, welche wiederum geladene Aminosäuren wie auch Serinen oder RS/RD Dipeptid Wiederholungen beinhaltet. Aufgrund ihrer multidomän Struktur könnte AtCyp59 in vielen zellulären Prozessen, wie beispielsweise Splicing, RNA processing, Proteintransport und Zellreifung etc., involviert sein. Darüberhinaus wurde bereits gezeigt dass AtCyp59 Proteine sich im Nukleus aufhalten und dort mit der C-Terminus Domäne von RNA Polymerase II interagieren. Die SR/RD Domäne hingegen wird zur Bindung an SR Proteine verwendet. Die exakte biologische Funktion des RRM Motives ist immer noch unklar. In dieser Arbeit präsentiere ich ein genomisches SELEX, eine genomweite analyse Methode, die eine Suche des RNA-bindenden Motivs der RRM Domäne von AtCyp59 ermöglicht. Mit Hilfe von random priming haben wir eine genomische Bibliothek von *A. thaliana,* bestehend aus 50-300nt langen überlappenden Sequenzen, erstellt. In Verbindung mit dem SELEX Verfahren waren wir im Stande eine 7nt lange RNA-bindende Konsensusequenz zu identifizieren, die eine spezifische Bindung an die AtCyp59 und dessen RRM Domäne *in vitro* und *in vivo* aufweist. Mutationen an der RRM Domäne so wie auch der Konsensusequenz verhinderten eine Bildung des RNA-Protein Komplexes. Erweiterte bioinformatischen Analysen zeigten, dass dieses Bindung–Motiv ein globales Sequenzmerkmal repräsentiert, welches sich in der Nähe des transnationalen stop-codon bei 70% aller *A. thaliana* mRNAs befindet. Untersuchungen der Bindungsaktivität zwischen RNA und der AtCyp59 der RRM Domäne wiesen eine Beteiligung der PPIase und

RS/RD- Reichen Domäne in die Bindung an die strukturierte RNA. Die präsentierten Daten erlauben uns einen Einblick in die Funktion der RRM Domäne von AtCyp59, eines regulatorischen Proteins welches zwischen 2 wichtigen zellulären Prozessen; Transkription und Splicing steht, zu gewähren.

Contents

Abbreviations

4E-BP	The eukaryotic initiation factor 4E (eIF4E) binding protein
9G8	splicing factor of 35 kDa
AS	Alternative Splicing
ASF/SF2	alternative splicing factor/splicing factor 2
CDK	Cyclin-dependent kinase
CF	Cleavage factor
ChIP	Chromatin immunoprecipitation
CLD	Cyclophilin-like domain
CLIP	Cross-linking immunoprecipitation
CLIP-HITS	HIgh Throughput Sequencing to CLIP
Clk/Sty	CDC-2 like kinase/ serine/threonine/Tyrosine kinase
CPSF	Polyadenylation specificity factor
CsA	Cyclosporine A
CstF	Cleavage stimulation factor
CTD	RNA Polymerase II C-terminal domain
CypA	Cyclophilin A
Dscam	Down Syndrome Cell Adhesion Molecule
EMSA	ElectroMobility gel-Shift Assay
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
Fcp1	Carboxy-terminal domain (CTD) phosphatase
FKBPs	The immunosuppressant drug FK506 binding proteins
Fox-1	Ataxin 2-binding protein 1
GST	Glutathione S-transferase
HIV	Human immunodeficiency virus
hnRNP	Heterogeneous nuclear ribonucleoprotein
Hsp	Heat-shock protein
IPAR	Isopentipyl adenosine
IRES	Internal ribosome entry site

ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
MAP	Mitogen-activated protein
MBNL1	Muscleblind-like (Drosophila) 1 gene
MD	Multidomain cycplophilins
mRNA	Mature messenger RNA
mTOR	The mammalian target of rapamycin
ninaA	neither inactivation nor afterpotential A protein
NMD	Nonsense-mediated decay
PAGE	Polyacrylamide gel electrophoresis
PAGE	Polyacrylamide gel electrophoresis
PAS	Polyadenylation signal
PcfI	polyadenylation cleavage factor I
PEG	Polyethylenglycol
Pol II	DNA dependent RNA polymerse II
PPIases	Peptidyl-prolyl cis/trans isomeraases
PPT	Polypyrimidine Tract
Pre-mRNA	Premature messenger RNA
Pro	Proline
PTB	Polypyrimidine tract binding protein
PTC	Premature termination codon
RIP	RNA immunoprecipitation
RNP	Ribonucleoprotein
RRM	RNA binding motif
RT	Room temperature
SDS	Sodium Dodecyl Sulfate
SELEX	Systematic Evolution of Ligands by EXponential enrichment
Ser	Serine
SF1	Splicing Factor 1
SF3b155	splicing factor 3B subunit 1
snRNP	Small nucleoprotein particles

SR	Serine Argenine Rich
SRPK	SR protein kinase family
SWAP	Suppressor of White Apricot Protein Homolog
TFIIH	Transcription factor II H
TPP	Thiamine pyrophosphate metabolite
U1-70K	U1 small nuclear ribonucleoprotein-70K
U1C	U1 snRNP-specific protein C
U2AF	U2 Auxiliary Factor
U2AF65	U2 small nuclear ribonucleoprotein auxiliary factor 65-kilodalton subunit
Wt	Wild type

1. Introduction

1.1 Pre-mRNA splicing

Exploring the 'RNA World' hypothesis (Cech, 2009) is one of the most developing fields nowadays. RNA is not only an intermediate between DNA as storage of genetic information and protein as a functional substance but rather a high complex and differentially regulated molecule.

In eukaryotes, transcribed pre-mRNA is undergoing maturation through several rounds of modification such as 5'-cap structure addition (Shuman, 2001), splicing (Gornemann, 2005), (Lacadie, 2005) and 3'-end processing (Proudfoot et al., 2002). All steps are connected and strictly regulated via formation of correct messenger RNA and linked back to transcription (Buratowski, 2008), RNA export, further translation to a protein and RNA surveillance (Moore, 2009).

1.1.1 Spliceosomal assembly and splice-sites recognition

Splicing is being recognized as a fundamental gene regulation tool and basis for the proteome variability through generation of alternative RNA forms as an alternative to genes expansion (Kramer, 1996). This is a process of excision of non-coding genetic information (introns), and ligation of future coding parts (exons) from transcribed pre-mRNA (Zhou, 2002). On the structure, pre-mRNA consists of few elements which distinguish future exons from introns; those are a 5'- and 3'-splice sites, a branch point sequence and a polypyrimidine-rich tract (Figure 1.1. A). As it has been shown in the figure these sequence elements are possessed several key features important for a splicing process. For instance, 5'-splice site should contain GU sequence; 3'-splice site should have AG sequence and branch point sequence usually has a consensus comprised from U and AC nucleotides. As it well studied for decades two guanidines from splice sites and one adenosine from branch point sequence are the main players in splicing catalysis.

Chemical mechanism of splicing employs two coherent nucleophilic attacks resulting in transesterification reactions followed by formation of the lariant intron product and joint exons (figure 1.1 B) (Wachtel, 2009). To undertake and catalyze these two reactions, living cell creates a massive ribonucleoprotein (RNP) catalytic complex named spliceosome. This machinery (spliceosome) is composed of five small nucleoprotein particles (snRNP) and around 200 associated proteins (Jurica, 2003). The assembly of the spliceosome as well as the fundamental chemical steps shows a high level of conservation from yeast to humans and may be divided by a number of discrete steps (Barrass, 2003).

Firstly, U1 snRNP binds to the nascent pre-mRNA at the 5'-splice site which leads to the formation of E complex. Further, to build a pre-initiation spliceosomal A complex, 3'-splice site along with the polypyrimidine tract has to be recognized by the complex of U2 snRNP with secondary factor U2AF and SF1 protein (Gaur, 2000) (Figure 1.1 C). Next, the U4-U5-U6 tri-snRNP complex identifies the A complex and displaces U1 snRNP from 5'-splice site to produce B complex, which resembles active spliceosome by the composition but not by structure. Finally, to produce mature C complex, U4 snRNP has to disassemble from tri-snRNP allowing base pairing between U2 and U6 snRNPs (Brow, 2002). This step finalizes the formation of catalytically active spliceosome which promotes two step transesterification process.

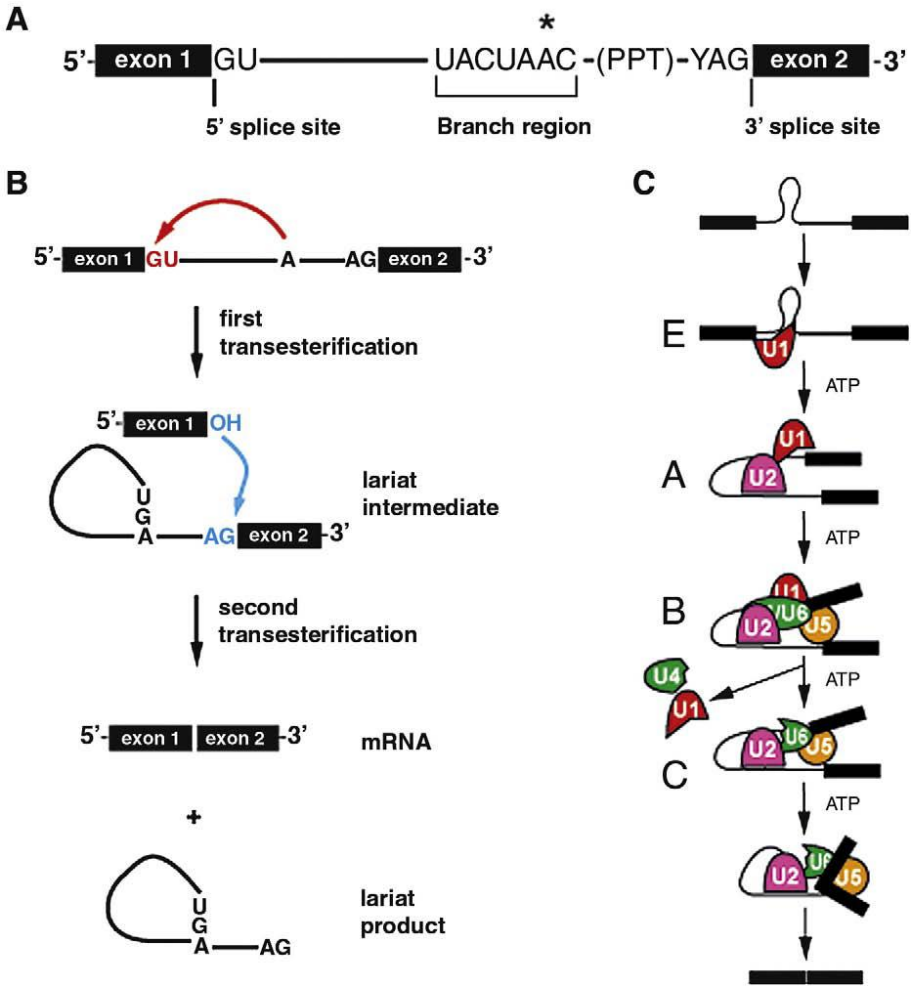

Figure 1.1: Pre-mRNA splicing. (A).Schematic structure of the pre-mRNA with features important for the splicing – 5' and 3' splice sites and branch point sequence with * - preferred adenosine, PPT- polypyrimidine tract. (B) Stepwise chemical nucleophilic attacks during the splicing and formation of the lariant product. (C) Spliceosomal assembly and complex progression through splicing reaction (Ritchie, 2009).

Although, the steps of the splicing are conserved through evolution, the exon surrounding (so-called splicing code) is very diverse. Remarkably, despite of differences in the genome complexity and exon-introns length mRNA splicing element of all organisms are recognized and spliced correctly. For example, *S. cerevisias* has a very small proportion of the genes containing introns and majority of those hold only one intron of relatively small size of approximately 200-300 bases (Barrass, 2003). In contrast, human genes are very large in size and multi-intronic where introns could amount to a thousand of nucleotides. And in general, it is known that in higher eukaryotes pre-mRNA sequence

elements are less conserved but most of their genes are spliced meaning of higher and more complex levels of regulation (Wahl, 2009).

Therefore the crucial step in splicing initiation is a precise splice-site determination which separates exons from introns. It takes place when U1 snRNP binds to the 5'-splice site and supports identification of the 3'-splice site through exon body (Sharma, 2008).

Several proteins associated with the U1 snRNP facilitate 5'-splice site recognition while a protein components of U2 snRNPs help to recognize 3'-splice sites on the pre-mRNA. In more details, at the first step 3'-splice site is recognized by U2AF65 large subunit of the heterodimeric U2AF complexes binding to the U2snRNP through its protein component SF3b155. Remarkably, protein U2AF65 consists of three RRM domain structures and RS (Argenine - Serine rich) region (Selenko, 2003). RRM stands for RNA-recognition motif and is important for RNA-binding activity whereas RS region is responsible for protein-protein interactions. The second key step in 3' splice site recognition is determination of the position of polypyrimidine tract (Py-tract). Small subunit of U2AF heterodimer–U2AF35 brings together Py-tract and 3'-splice site facilitating their interaction. Despite there is no direct structural evidence of U2AF35/U2AF65 interaction it's believed to take place through RNA participation (Wu, 1999), (Zorio, 1999).

From the other site of pre-spliceosomal complex, 5'-splice site is recognized by U1 snRNP where U1C protein component guides U1 snRNA to the 5'-end of the exonic sequence (Oubridge, 1994). It's known that in yeast U1C directly attaches to the 5'-splice site whereas human U1C plays intermediate role in 5'-splice site recognition. It is integrated in U1 snRNP through N-terminal part of U1-70K and Sm core proteins which do bind to 5'-splice site. It's believed that this intermediate interaction is facilitated by zinc-finger motif in human U1C. U1-70K containing a RRM domain forms an area for U1C binding while role of Sm proteins is concluded to create a platform for composite RNA structure (Muto, 2004).

Further transition from a pre-sliceosomal E complex to the A complex requires U2snRNP association on the branch-point sequence where it pushes away adenosine through imperfect complementation with pre-mRNA and positions this residue for further nucleophilic attack in splicing reaction (Query, 1994).

Since statistical occurrence of the splicing consensus sequences on pre-mRNA is very high and splicing doesn't take place on all of those sequences it's very important to choose splice sites in the right way. And the "right way" for particular mRNA can vary dependent on many external factors. As a consequence, different exons of the pre-mRNA can be chosen in the process called alternative splicing (AS). Preference to the specific splice site is determined by presence, number and competition between proteins factors associated with a pre-mRNA which affect spiceosome assembly in positive or negative way. One of the ideas how it could be regulated is highlighted in the work of Black and colleagues (Sharma, 2008). They explore a theory of division between the exon recognition complexes and the intron-recognition complexes within pre-mRNA based on differences in protein-binding components which leads to usage of a particular pattern of exons.

1.1.2 Alternative splicing

Due to the possibility to create more than one mRNA forms from one gene, alternative splicing (AS) has been implemented by eukaryotes as a source of genome complexity and protein diversity. Inclusion of exclusion part of the whole exon or intron could affect both mRNA and protein function (Moore, 2008). As for mRNAs AS could lead to changes in an mRNA surveillance and stability because majority of aberrant or alternatively spliced isoforms contained premature stop codones that direct such transcript to the nonsense mediated decay after the first round of translation (Moore, 2008). As for the proteins, changes in mRNA levels directly linked to the translational potential of the protein together with it further stability, localization and function.

Alternative splicing is regulated by activity of multiple RNA-binding proteins expressing in a particular tissue at the time (Matlin, 2005). They can be either activators or repressors of splicing which bind to the enhancer or silencer sequences around alternatively spliced exons or introns. Such sequences are called "cis-regulatory elements" and could be divided into exonic splicing enhancer or silencer (ESE or ESS) and intronic splicing enhancer or silencer (ISE or ISS). Silencers elements are usually occupied

by an hnRNP proteins (Mauger, 2008) while enhancers are bound to a SR proteins (Long, 2009), (discussed in section 1.1.3).

Decision of exact splicing site to take is usually being made at the stage of an exon recognition or early sliceosome E complex assembly (Black, 2003), but newest studies revealed that such decision could happen at literally every single step of splicing even at transesterification reaction (Lallena, 2002). Growing evidences suggested an influence of RNA transcription, for instance the rate of RNA polymerase II (discussed in section 1.3), on splicing regulation (de la Mata, 2006).

Since spliceosome assembly at premature stages occurs around exons (Sterner, 1996) which are much shorter then introns (50-250 bp length versus >1000bp in the human genome), AS regulation of splice-site selection can occur based on promotion or inhibition of the U1 and U2 snRNP binding to that splice site (Kotlajich, 2009), (Figure 1.2).

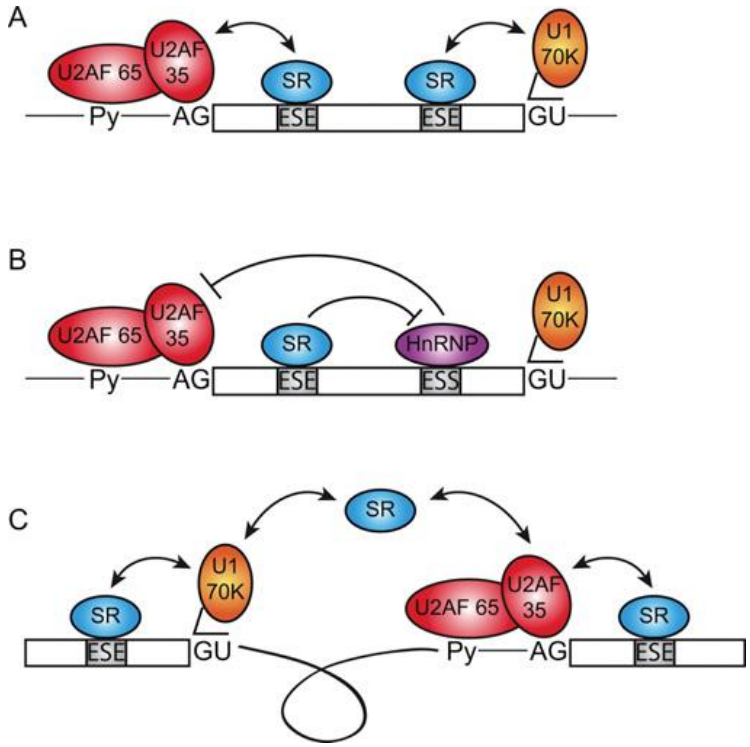

Figure 1.2: Splice site selection. (A) SR proteins factors (blue) directly interact with sequence elements on pre-mRNA (exonic splicing enhancers) and promote binding of snRNPs to the splice site. (B) Competition between splicing enhancers – SR proteins and splicing silencers – hnRNP for the splice site selection. (C) Bridging mechanisms of SR proteins action on splice site selection by bringing together U1 and U2 snRNPs across intron (Long, 2009).

Promotion of the AS site selection is modulated by action of SR proteins. For instance, they can facilitate “bridging” between U1 and U2 snRNPs. SR proteins may directly interact with ESE on target RNA. This interaction engages binding of U1 and U2 snRNPs to the 5’- and 3’-splice sites respectively (Bourgeois, 1999). A cross-talk between SR proteins and spliceosomal components is modulated by RS domain which undergoes rounds of intensive phosphorylation (Feng, 2008). As an example of that regulation, human SR protein SRp38 acts in promotion of spliceosomal A complex formation by bringing together U1 and U2 snRNPs and stabilization of 5’-splice site and branch point sequence recognition (Shin, 2004).

Inhibition of the alternative splicing is regulated by silencers. The first possible way of action is to sterically block positive regulatory elements or snRNPs via binding in a close proximity to the splice site. For instance, it’s well known that PTB protein when being bound to the polypyrimidine tract prevents an interaction of U2AF with exon (Sauliere, 2006). Second mechanism of inhibition is binding of silencer to the region over 100-200 bp upstream or downstream of the regulated exon followed by its sequestering via multiple interactions along RNA (Spellman, 2006). Another possible mechanism is to “loop out” an alternative exon through numerous interactions among RNA-binding proteins on the sites of neighboring exons which conformationaly hinder assembly of the spliceosomal complex on the first exon (Damgaard, 2002). Well known example is when nhRNP protein interacts with its own pre-mRNA near alternatively spliced exon 7 therefore preventing this exon to be spliced (Hutchison, 2002).

Nevertheless, splicing and alternative splicing of each individual pre-mRNA is highly dependent on number, activity and competition of splicing activators and inhibitors. Most frequently, decision of a particular exon to be included or excluded is undertaken by difference in SR and hnRNP occupancy on that exon (Zhu, 2000). For instance, splicing of human exon 2 gene encoding α-tropomyosin is dependent on winner in the competition for the same binding sequence between SR protein 9G8 and hnRNP F (Crawford, 2006).

There are a lot of evidences that alternative splicing is one of the determinations of tissue specification in higher eukaryotes. Interestingly, although over 90% of human genes are spliced alternatively giving in the average two-three mRNAs and they do so tissue-specifically (Wang, 2008). Recently, combined approached involved RNA-seq

method, bioinformatics analysis and microarrays of different tissues revealed that mammalian brain posses a highest number of tissue-specific different spliced variants of mRNAs (Xu, 2002). It is linked with the fact that brain is the most dynamic developing tissue and has thousand of different cell types (Li, 2007). Recently it has been shown that over 50% of the alternatively spliced mRNA isoforms are differentially expressed within the tissue which can be clarified by tissue-specific expression of the splicing factors (Wang, 2008). For instance, there are differences in expression of PTB and nPTB - specific for neural tissue form of PTB protein created from another mRNA isoform – in neuron development. PTB could only be expressed in the undifferentiated neurons whereas nPTB is upregulated in adult tissue. It is believed that such regulation occurs on transcriptional level (Boutz, 2007).

1.1.3 SR proteins.

The first discovery of SR proteins as splicing regulators came in 90-ties from research group of Fu (Fu, 1995). However, domain enriched in serine and arginine was previously found in *Drosophila melanogaster* associated within the genes SWAP, TRA and TRA-2 (Boggs, 1987). Structurally SR proteins are defined by presence of one or two an RRM domain which provides possibilities for RNA binding followed by RS domain which set for protein-protein interactions (Wu, 1993).

RNA recognition motif is a structure that sequence specifically binds nucleic acids, particularly single-stranded RNA. RRMs fall into a subgroup of the ferredoxin fold (Carte, 2008) that is featured by a β-α-β-β-α-β-structure. Usually RRM consists of two notable motifs: RNP1 and RNP2 which share conserved aromatic residues. Several X-ray structures of different RRM motifs alone or together with bound RNA were published (Oubridge, 1994), (Allain, 2000). The classical way of RNA molecule interacts with motif is that RNA lies on the β-sheet part between two RNP motifs while most conserved amioacids (tyrosine and phenylalanine) from the β-sheets stacks behind the bases (Maris, 2005). Recently the identification of and PTB and Fox-1 binding partners has brought a new level of complexity into RRM-RNA recognition. In this model aromatic-nucleic acid stacking are

substituted by range of hydrophobic interactions on RNA with loop structure of RRM motif (Auweter, 2006).

As it was mentioned above (section 1.1.2) main role of SR proteins is to facilitate splice-site recognition. There are two theories how this facilitation could be accomplished. The first way is to provide a "bridge" between U1 and U2 snRNps to convey further spliceosomal assembly (Robberson, 1990) and the second is through SR proteins competition for binding to the ESE elements in pre-mRNA with negative acting factors such as hnRNPs (Martinez-Contreras, 2007).

Besides primary role of SR proteins in promoting spliceosomal assembly they are also involved in mRNA transcription (discussed in section 1.2.1), translation, nuclear export and NMD (nonsense-mediated decay) (Huang, 2005). If to look at SR protein localization, majority of SR proteins have pattern localization in nucleus's compartment-splicing speckles. Speckles as believed serve as a reservoir for splicing factors (Lamond, 2003). However, there are other SR proteins like SF2/AFS, Srp20, and 9G8 in human, which show a shuttling profile. They move continuously between nucleus and cytoplasm engaged in mRNA isoforms transport from nucleus to the cytoplasm (Huang, 2001).

SR proteins are involved in regulation of NMD for the mRNA isoforms containing a PTC (premature termination codon). Recently, it has been discovered that overexpression of several SR proteins (SF2/ASF, SC35, SRp40) strongly amplifies NMD pathway independently of their nucleus-cytoplasm shuttling activity (Zhang, 2004), (Sato, 2008).

SR proteins could also regulate mRNA on a translational level. It has been shown that SF2/ASF proteins while accompanying polyribosome promote translation of the mRNAs containing ESE elements *in vivo* and *in vitro* (Sanford, 2004). The mechanism of is based on mTOR pathway activation leading to activation of the 4E-BP protein – a main inhibitor of a cap-dependent translation (Michlewski, 2008). Another example of the participation of SR proteins in translational regulation is SRp20. This protein could bind IRES (internal ribosome entry site) and promotes translation of viral RNAs (Bedard, 2007).

Activity of SR proteins is highly dependent on their phosphorylation status (Mermoud, 1994). Presence of arginine-serine repeats within RS domain allows tuning activity and localization of SR proteins (Lin, 2007). As an example, only hyperphospahorylated RS domain of ASF/SF2 protein is able to interact with other proteins of spliceosomal assembly for instance U1-70K (Xiao, 1997), although,

hypophosphorylated status of RS domain is required for a splicing catalysis (Tazi, 1993). There are several kinase families in the cell which are responsible for phosphorylation of SR proteins. These include SRPK (SR protein kinase family) (Wang, 1998), the Clk/Sty (Colwill, 1996) and topoisomerase I (Rossi, 1996). It is known that for example SRPK1 kinase could only phosphorylate “docking motif” of RS domain which prevents other kinase activities (Ngo, 2008), whereas Clk/Sty family is able to add Phospho-group to the whole C-terminus of SR proteins (Ngo, 2005).

1.1.4 pre-mRNA structure in splicing

Influence of pre-mRNA structure is a further layer of complexity in assembly of the spliceosome and regulation of alternative splicing. It has been known for a long-time that RNA structures are capable of immediate splicing regulation. There are several aspects how this question could be approached: the first, presence of internal RNA structures within pre-mRNA with intervenient or promoting effect on splicing; the second, occurrence of protein-RNA or small metabolite–RNA complexes creating special RNA structures affecting splicing (Chen, 1999).

There are several examples of how splicing is facilitated or inhibited by presence of local RNA structure (figure 1.3). One is when internal RNA structure is present in 5’ splice site of branch-point sequence leads to the U1 and U2 snRNPs incapability of unwinding the structure. Also it has been discovered that presence of local RNA structure in polypyrimidine tract is hindering U2AF65 interaction (Warf, 2009). When splicing is being promoted by RNA structure, pre-mRNA forms an area locally in which important splicing factors are coming together closer in space. One of the famous examples is how splicing of *Drosophila melanogaster* Dscam gene is taking place (Graveley, 2005). Another example is an existence of cryptic splice sites (sites that are corresponded the criteria of splice sites but not involved in splicing). Local RNA structures could conceal cryptic sites but upon distraction of such structures these sites become available for splicing which typically resulted in an erroneous mRNA form.

Anyway, in understanding importance of internal pre-mRNA structures, it should be taken into account that splicing takes place co-transcriptionally (Singh, 2007).

Therefore low-energy and slow-forming structures could not have time to appear in a pre-mRNA when splicing reaction is taking place involving speed of transcription in splice-site selection. Even more, competition between proteins preferred to be attached to the single-stranded RNA and formation of a local RNA structure in this place at particular time point should be taken into consideration.

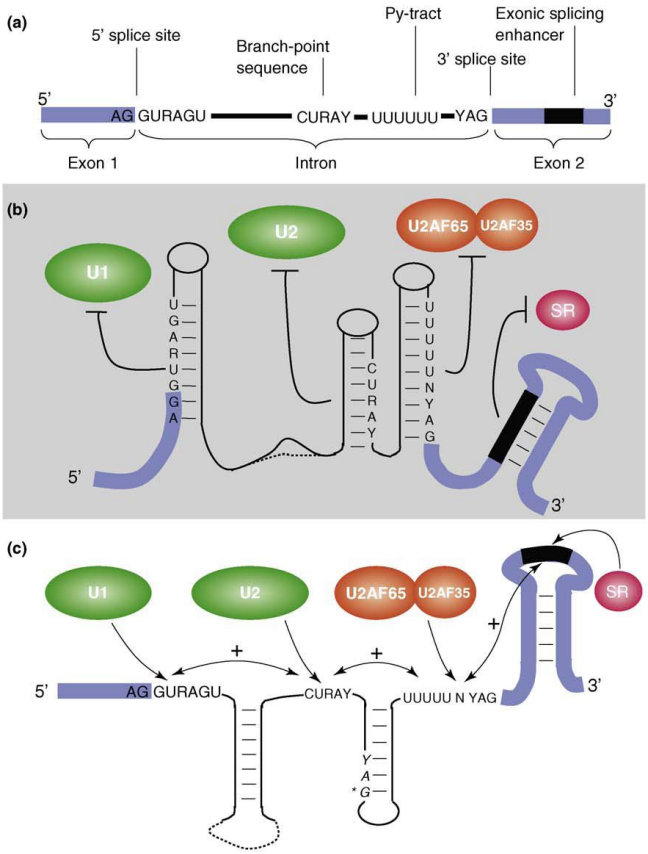

Figure 1.3: Examples of pre-mRNA structural elements that have an influence on splicing. (A) Schematic presentation of the sequence elements on the primary pre-mRNA structure which are important for splicing. (B) Pre-mRNA structure that inhibits splicing. Loop formation at 5' and 3' splice sites prevents binding of U1 and U2 (green) snRNP to them and thereby interferes with binding of U2AF65 and SR proteins. (C) Example of pre-mRNA structure that promotes splicing. This loop-structure forms an environment in which all important splicing signals are brought together to the close proximity. Also cryptic splice site (*) is masked in the loop which leads to SR protein binding and increase in usage of this splice site (Warf, 2010).

Recently, some proteins have been found which are able to interact with local structural elements on a pre-mRNA therefore influence the splice-site selection. One of such examples is an alternative splicing protein MBNL1. It has been shown that upon

binding to the structured RNA MBNL1 could suppress or boost alternative splicing on particular cassette exons (Warf, 2007). Another example deals with RNA helicases - proteins which could unwind inhibitory structures on pre-mRNA. For instance, activity DEAD box RNA helicase p72 increases frequency of insertion of exon 4 in CD44 minigene (Honig, 2002).

Small metabolites recently have brought a new complexity to the field. They could specifically bind to the pre-mRNA local structures and by doing that regulate splicing. For example, thiamine pyrophosphate metabolite (TPP) binds to the aptamer structure in the 3'UTR of the pre-mRNA of genes implicated in its own biosynthesis and creates a negative loop of regulation of vitamine B1 synthesis (Croft, 2007), (Wachter, 2007) .

1.1.5 Plant-specific aspects of splicing

The first hint of why splicing in plants are separated from other eukaryotes came from the observation that mammals introns could not be processed in plants (Hartmuth, 1986). This observation is perhaps consistent with the fact that there are several differences in genes organization between metazoan and vertebrates for example, in plants length of genes are smaller and introns are much shorter (Lander, 2001).

Studying splicing in plants is associated with some complications such as absence of *in vitro* splicing extract. This fact led to difficulties in accomplishment of such a great success in plants for studying roles of splicing factors and RNA-binding proteins as in vertebrates. However development of the "invertase mini-exon system" for plants lines has given a broad overview on splicing (Brown, 2002).

As it was discovered afterwards mammals and plants share a lot of similarities in how splicing process is taking place. For example, it seems that branch point sequence is very important for splice-site recognition in both branches. However, plants do not contain strong signal for a branch point sequence. Instead, they created U-rich sequence between branch point sequence and 3' splice site. It has been shown the U or UA-rich sequences are critical for splicing in plants and mutations in this region lead to mistakes in splicing (Goodall, 1989). As for the plant SR and hnRNP proteins over 60% from them have a high order of similarity to mammals splicing factors suggesting a general

conservation in mechanism of splicing (Simpson, 2008). Rest of the proteins appeared to have only partial similarity to vertebrate factors. However, there were found several proteins including proteins belonging to SR family with unique organization. Function of these proteins is not fully understood till now (Kalyna, 2004), (Kalyna, 2006).

Alternative splicing occurs less frequently in plants compare to vertebrates (35% versus at least 74%) but recent studies have shown that this number is highly underestimated (Barbazuk, 2008). Nevertheless, AS plays as same important role in plants as in animals. Genes undergoing alternative splicing are involved in all developmental processes in plants as well as in stress responses and flowering time (Reddy, 2007), (Hirose, 1993). Unfortunately, little is known about the functions of proteins produced or whether they are existed from alternatively spliced mRNA isoforms (Simpson, 2010). Interestingly, plants possessed the highest number of SR genes, 24 in rice and 19 in *Arabidopsis thaliana* (Lopato, 1999), (Lopato, 2002), (Golovkin, 1998). Plants SR proteins could be divided in several groups, (figure 1.4). Among them there are true homologues for human SR proteins such as SF2/ASF, SC35 and 9G8 whereas other SR proteins are unique for plant kingdom and therefore considered to have plant-specific functions (Kalyna, 2003). Several groups are intensively studying such subfamilies (Isshiki, 2006), (Ali, 2007), (Lorkovic, 2008), (Brown, 2009). Like in animals SR proteins activity and interaction with other proteins in plant SR families are highly dependent of their phosphorylation status. It's been shown that 13 out of 19 SR proteins could be phosphorylated *in vivo* (de la Fuente van Bentem, 2006). Same kinase families as in mammals are responsible for RS domain phosphorylation in plants, however it's been discovered that MAP kinases could also be involved in SR proteins regulation (Feilner, 2005).

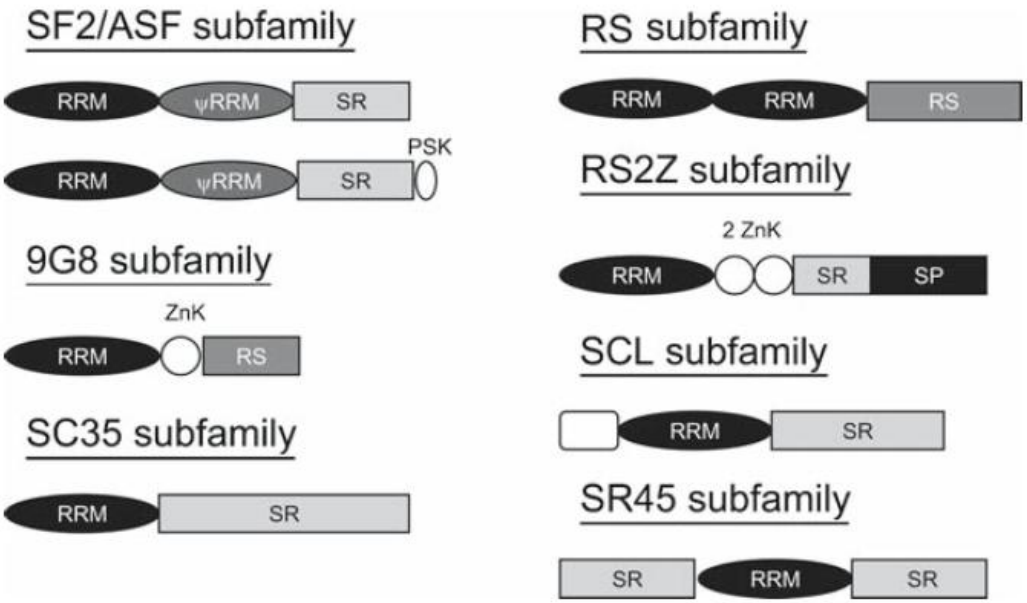

Figure 1.4: Subfamilies of plant SR proteins. Left column represents the SR proteins similar to the mammals and on the right – specific for the plant kingdom. Structural elements are named according to the aminoacid differences: RRR – RNA recognition domain; ψRRM – Specific for ASF/SF2 homologs; RS, and SR – domain rich in argenins and serins according to majority; ZnK – zinc knuckle domain of CCHC type; SP and PSK – domains rich in serines, prolines and lysines; SCL – have plant-specific extension on the N-terminal domain (Barta, 2008)

1.2 Co-transcriptional RNA processing

Current understanding of the pre-mRNA processing is based on the view that majority of mRNA are spliced co-transcriptionally. Especially, a lot of studies were performed on how alternative splicing influences on transcription. Recently it has been shown that rate of Polymerase II transcription determines what type of splicing factors on the particular binding site near the alternative exon would have the major influence on splicing (Kolasinska-Zwierz, 2009), (Loomis, 2009). There are two theories how RNA Pol II regulates alternative splicing (Kornblihtt, 2006). One is “recruitment model” and the other is “kinetic model”. In the “recruitment model” RNA polymerase along with transcription factors is directly or indirectly involved which interaction with splicing factors influences on the rate of transcription. Thus, alterations in occupancy of splicing factors on their binding sites lead to inclusion or exclusion of the alternative exon (Das,

2007). Such examples were intensively studied by groups of O'Malley (Auboeuf, 2004) and Kornblihtt (Cramer, 1999). In the “kinetic model” splicing on particular alternative site is dependent on whether spliceosome could be assembled fast enough on this site. This assemble relies on the speed of RNA Pol II in this region on the gene. For instance, it has been shown that mutations in Pol II which slow down the rate of transcription also lead to the inclusion of the exons with predominantly weak splice sites (de la Mata, 2003). The way how RNA polymerase can change its pace lies in modulation of the phosphorylation status of the biggest C-terminal domain (Phatnani, 2006).

1.2.1 C-terminal domain of Polymerase II

RNA polymerase II is accountable for transcription of the multitude of genes and is specialized on mRNA genes. It has a unique C-terminal region organization, named CTD domain. Beside C-terminal domain, other parts of polymerase are conserved within all kingdoms and between other DNA dependent RNA polymerases (Allison, 1985). CTD consists of Tyr-Ser-Pro-Thr-Ser-Pro-Ser repeated heptamers of different length which is dependent on the organism with maximum of 52 repeats in humans (Corden, 1990). CTD remains an essential domain because full or partial deletion of C-terminal domain is lethal as it has been shown in yeast and drosophila (Gerber, 1995). Nevertheless, for the *in vitro* transcription systems, Pol II could proceed without C-terminal domain (McCracken, 1998), meaning that CTD is served as modulator of the process. The modern understanding of the role of this domain in transcription represents as a platform for the multiple interactions of the variety of factors influencing mRNA processing and further translation (Phatnani, 2006). During transcription CTD expands from the enzyme core to the nascent mRNA site which allows it to influence on mRNA processing by recruiting specific factors responsible for modifications (Bentley, 2005).

The largest domain of RNA polymerase undergoes multiple modifications to successfully accomplish mRNA synthesis which consists of following steps: initiation, elongation and termination (figure 1.5) (Buratowski, 2003).

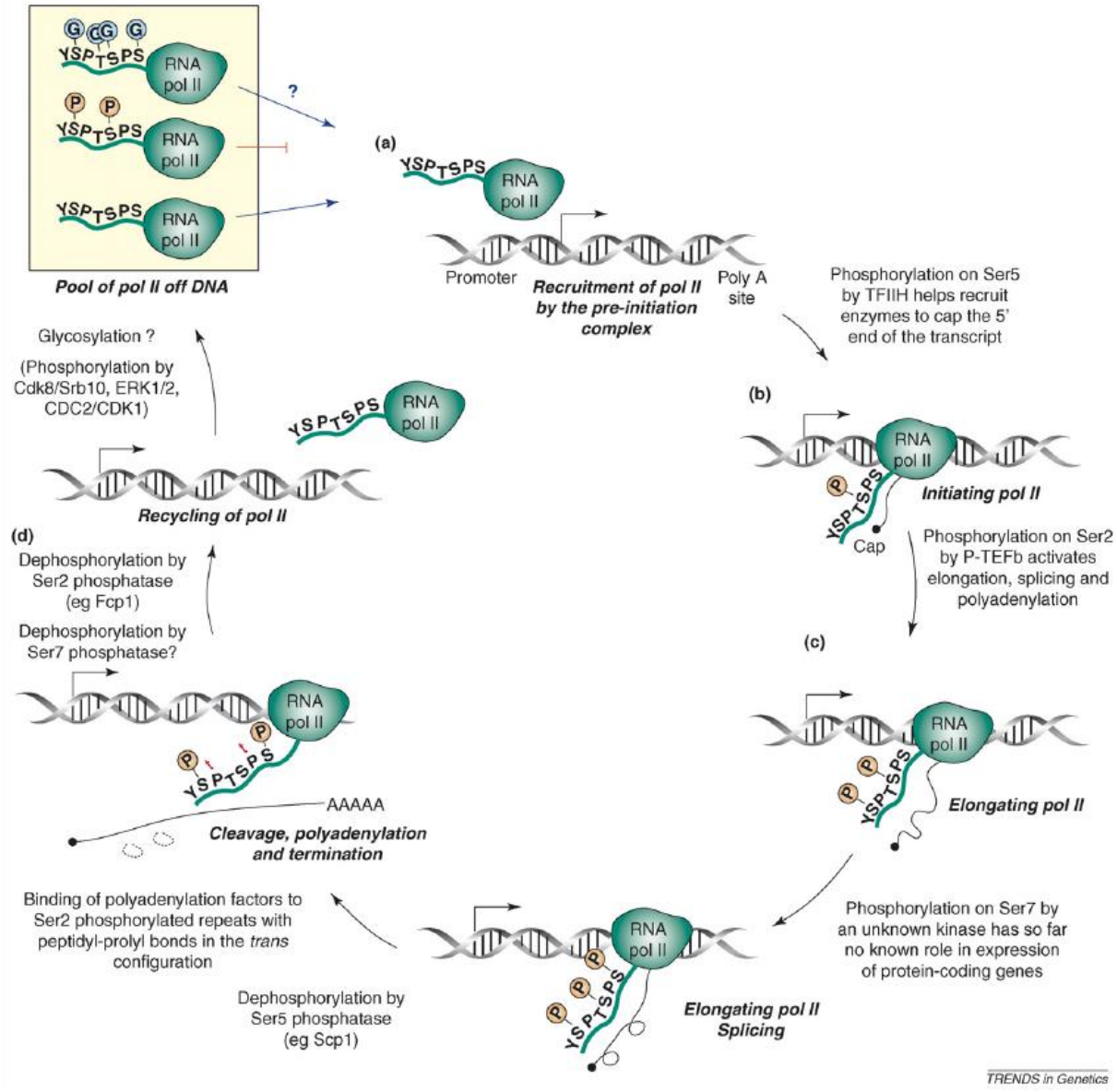

Figure 1.5: C-terminal domain phosphorylation during transcription. (A) CTD of Pol II at transcription initiation is in unphosphorylated stage which is believed to block transcription start. (B) Beginning of the transcription. Ser 5 is phosphorylated by CDK7 and CTD of the Pol II is located near the pre-mRNA exit which helps to recruit appropriate enzymes for 5'-mRNA end formation. (C) Efficient elongation complex of Pol II with Ser5 and Ser2 phosphorylated. Both serines should be with phosphor group to promote producative pre-mRNA formation. (D) After successful 3'-end mRNA formation CTD of Pol II becomes dephosphorylated which helps for Pol II recycling (Egloff, 2008).

Alterations on CTD of PolII are tightly connected to the enrollment of the specific transcription factors during mRNA synthesis; predominantly phosphorylation of the two serines is of the great importance (Buratowski, 2003). Other *in vivo* proved sites of

modifications include: tyrosine phosphorylation, serines and threonines glycosylation, isomerization of the two prolines. These modifications at particular point in time could be recognized as the “CTD code”. This recognition provides range of signals affecting interaction of the transcriptional and splicing factors with the polymerase (Corden, 2007). Phosphorylation of the serine 2 and serine 5 is the most well deliberated alteration in the CTD (Buratowski, 2003). Dependent on phosphorylation status, Pol II could form two main stages: hypo- and hyperphosphorylated (IIa and IIo) which migrate differently in SDS PAGE (Baskaran, 1993). Hypophosphorylated CTD is associated with transcription initiation and hyperphosphorylated CTD is essential for efficient elongation on the transcription cycle (Zhang, 1991). It is still remain unclear how many phosphates are correlated with Pol II at given point in time. However it’s documented that in the hyperphosphorylated stage there is approximately one phosphate residue per repeat (Payne, 1993). Cycline-dependent kinase (Cdk7) is responsible for the Ser5 phosphorylation in vivo (Komarnitsky, 2000). Cdk7 is a part of the universal transcription factor TFIIH which interacts with the 5’-end of the transcribed gene. Phosphorylation of Ser 5 helps to recruit enzymes responsible for the 5’-cap addition to the nascent mRNA (Gomes, 2006). Other cycline-dependent kinase - Cdk9 – acts on Ser2 phopshorylation. This modification serves as a switch from initiation-early elongation to the efficient elongation form of the Pol II (Peterlin, 2006). Moreover, phosphorylation of both serines is required for involvement of splicing and poly-adenilation machineries (Hirose, 1999). During a cell-division Ser2 and Ser5 could be phosphorylated by another kinase, Cdk8. In the absence of DNA template Cdk8 could produce the highly phosphorylated stage of CTD that believed to be inactive (Bird, 2004). Removal of phosphates from both serines also plays an important role in transcription especially for mRNA 3’- end processing and Pol II recycling (Meinhart, 2005). Dephosphorylation of Ser2 and 5 requires two enzymes, Fcp1 and Ssu72 respectively (Reyes-Reyes, 2007).

Next crucial modification on the C-terminal domain of the Pol II is isomerization of the two prolines. The peptide bond preceding prolines could exist in two conformations *cis* and *trans.* It results in four possible arrangements of each CTD repeat (Stiller, 2004). The bond is naturally fixed in *trans* conformation and requires energy to switch to the *cis* conformation. Proteins are conscientious for the proline–bond change called peptidyl-prolyl cis/trans isomeraases (PPIases) (further discussed in section 1.3) and known to be

mainly involved in protein folding *in vivo* (Schiene, 2000). However there are mammalian Pin1 and yeast ESS1 isomerases which regulate structure of the CTD and by that also affect transcription (Xu, 2004). Recently, it has been discovered that Pin1 and ESS1 show high level of specificity to the phosphor-serine and phosho-treonine bonds and bind directly to the phosphorylated CTD (Verdecia, 2000). Since phopshorylation of the C-terminal domain of the Pol II is correlated with active elongation, interaction of these proteins with CTD may influence transcription and 3'-end mRNA processing. Indeed, it has been shown that mutation in ESS1 is linked to the abnormal mRNA 3'-end formation (Kops, 2002). Therefore Pin1/ESS1 represents bright example of how CTD repeat conformation regulates transcription (Wilcox, 2004).

1.2.2 Advantages of co-transcriptional RNA processing

As it was mentioned above pre-mRNA are processed co-transcriptionally. Giving a closer look to the "recruitment model" of co-transcriptional splicing, one can discover an order of spliceosomal factors assembly on the actively transcribed gene. For instance, the U1 snRNP binding to the first 5' splice site is followed by interaction of U2 snRNP with the 3' splice site during the intron transcription (Gornemann, 2005). Also, there are evidences that genes transcribed by Pol II in vitro are more efficiently spliced than those transcribed by T7 polymerase. This fact further proves the existence of interaction of spliceosomal factors with the CTD of Pol II during transcription (Das, 2007). Advantage of linkage between transcription and splicing is clearly favorable for fast and efficient processing of the long introns especially in human where otherwise would be difficult to bring together spliceosomal factors on the distant exons (Dye, 2006, Lacadie, 2006). The approach used by cellular systems to deal with large introns is a co-transcriptional excision of them. *In vivo* studies supported that spliceosome could form even if intronic sequence is cleaved and degraded as long as exons are attached to the transcriptional unit (Kim, 2007, Morlando, 2008).

Role of SR proteins in co-transcriptional splicing couldn't be underestimated. SR proteins are located in the nuclear speckles and it has been shown, that they are directly engage in the nascent RNA transcript by direct interaction with CTD of the Pol II (Misteli,

1997), (Bauren, 1994), (Beyer, 1991). Recent studies revealed that the SR proteins not only stimulate the snRNP assembly on the alternative splice site during transcription, but also are dynamically loaded to the transcription elongation complex. The reason for that is that they prefer phosphorylated status of CTD Pol II which resembles efficient elongation stage (Sapra, 2009), (Saunders, 2006).

1.2.3 mRNA 3'-end processing

Modifications of the CTD of Pol II play an important role in processing of the last exon and the 3'-end of the mRNA. It has been shown by ChIP analysis that polymerase loses almost all the Ser5 phosphorylations when it arrives at the polyadenylation signal of the protein coding gene (Komarnitsky, 2000). In fact, protein responsible for the polyadenylation cleavage PcfI could only interact with CTD phosphorylated on Ser2 (Licatalosi, 2002). Even more, recently it has been discovered that this factor binds only *trans* isomer of the proline preceding Ser2 from mixture of *cis-trans* isomers. This is the further indication that recognition of the CTD accomplished through proline isomerization (Noble, 2005).

Formation of the correct 3'-end of mRNA is important because it promotes mRNA transport to the cytoplasm (Vinciguerra, 2004), influences on stability of mRNAs (Wickens, 1997) and enhances the translation (Sachs, 1997). In general the 3'-end mRNA formation is exaggerated upon mutations, deletions or loss of phosphorylation of the CTD of the Pol II (Wahle, 1999). 3'-end processing is a two step mechanism involving cleavage of the mRNA and poly (A) accumulation on the precursor product. This formation is performed by combination of the poly(A) polymerase together with complex of proteins, CF1A, CF1B and CFII in yeast or CstF-cleavage stimulation factor, CPSF-polyadenylation specificity factor, CFI_m and $CFII_m$ – cleavage factors in higher eukaryotes (McCracken, 1997). Remarkably, CFI_m protein has number of unique features. It's exclusively expressed in mammals as a heterodimer (Takagaki, 1989). Its N-terminal domain consists of RNP-type RNA recognition motif which has been revealed to interact with splicing factor U2AF (Selenko, 2003) in humans. Its C-terminal domain is enriched with RS, RD and RE repeats and possesses the same structure as SR proteins. It has been shown that this protein is co

purified with the spliceosome *in vitro* (Rappsilber, 2002). As it was revealed by SELEX experiment (discussed in section 1.5) this protein tends to bind UGUAA motif (Brown, 2003) collated just upstream polyadenylation signal (PAS) on pre-mRNA. The initial role of this protein is to interact with pre-mRNA near the PAS. In general, CFI_m acts as mediator for interaction between CTD and the spliceosome and enhances proper recognition of the polyadenylation site (Venkataraman, 2005).

Intriguingly, factor CFI_m is absent in plants and yeast (Hunt, 2008). The exact composition of the plant 3'-end mRNA processing machinery remains unclear because many mammals single-gene factors such as CstF, CPSF are encoded by gene families in *Arabidopsis thaliana* which also could be developmental or organ specific (Zhao, 2009). And other functions are believed to be undertaken by plant-specific proteins because mammals and plant poly(A) signals are different (Bienroth, 1993).

Many 3'-end processing factors are tending to co-purify with CTD on affinity column. Phosphorylation on Ser2 but not on Ser5 is required for that interration (Ahn, 2004). For instance, 3-end processing is linked to the efficient transcription termination of Pol II and its release from transcript (Bauren, 1998). As it has been revealed, yeast Rtt103 protein directly interacts with CTD where Ser2 is phosphorylated. This interaction leads to involvement of 5'- and 3'- RNA exonucleases to release Pol II from the DNA template (West, 2004). This finalizes the connection of CTD modification to the all stages of pre-mRNA processing.

1.3 Proline *cis/trans* isomerases

The first discovery of the 'peptidyl-prolyl isomerase' (PPIase) or 'cyclophilin' are dated back to 1980s when Cyclophilin A (CypA) protein was identified in the bovine thymocytes cells (Fischer, 1984). This intracellular protein showed high affinity to the immunosuppressive drug cyclosporine A (CsA) (Handschumacher, 1984). Cyclophilins belong to the superfamily of the immunophilins. This group of proteins includes cycplophillins, FKBPs – the immunosuppressant drug FK506 binding proteins and parvulins (Galat, 2003). Common characteristic of all cyclophilins is the presence of the cyclophilin-like domain (CLD) enclosed with unique to each member domain which is

required for undertaking specific function or/and localization (Marks, 1996). This family of proteins is present in all branches of evolution including yeast and bacteria and all of them share structurally conserved PPIase domain. Up to date there have been identified 7 major cyclophilins in humans (Galat, 2003), nine – in *Drosophila* (Waldmeier, 2003) and as many as 29 genes in *Arabidopsis thaliana* (He, 2004).

Why the existence of the PPIases is so important and why prolyl bond is so different from others could be seen in the figure 1.6. In general, the peptide bond between aminoacid residues could exist in *trans* and *cis* conformation. It's known that thermodynamically *trans* conformation is far more favorable in terms of free energy compare to the *cis* conformation (Stewart, 1990). However, there is one exception which is the peptidyl-prolyl imid bond. Being trapped in rotation it could exist in two distinct stage *trans* or *cis* (Pahlke, 2005).

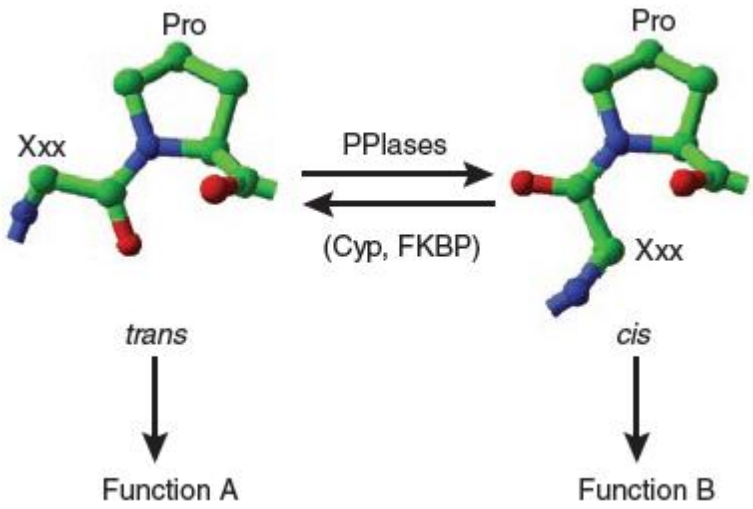

Figure 1.6: Proline cis/trans isomerization as modulation of the protein function. Based on substrate specificity (Phosphodependent or not) different classes of PPIases could act on promotion of either cis or trans form of proline bond that lead to the differences in function of the protein (Lu, 2007).

Switch in conformation of prolyl bond creates an important tool for the protein dynamic and could change functional stage of the protein by preventing or enhancing of the binding capacity for particular factor. Despite of low energy difference between *trans* and *cis*- conformation of imid bond rotation is associated with high-energy barrier of so

called ω=90^{o} *syn* transition conformation (Corey, 1953) (figure 1.7). The conformational exchange rate is very slow and could take couple of minutes which is inappropriate for biological systems *in vivo* and requires a catalyst. Fundamental PPIase function in catalysis is to lower down the high-energy barrier by binding to the each isomers and equilibrating free energies enzyme-*cis* and enzyme-*trans* to the ground state complexes (Lu, 2007). This action increases isomerization rate to the millisecond timescale which is more favorable for the biological interactions (Fanghanel, 2004).

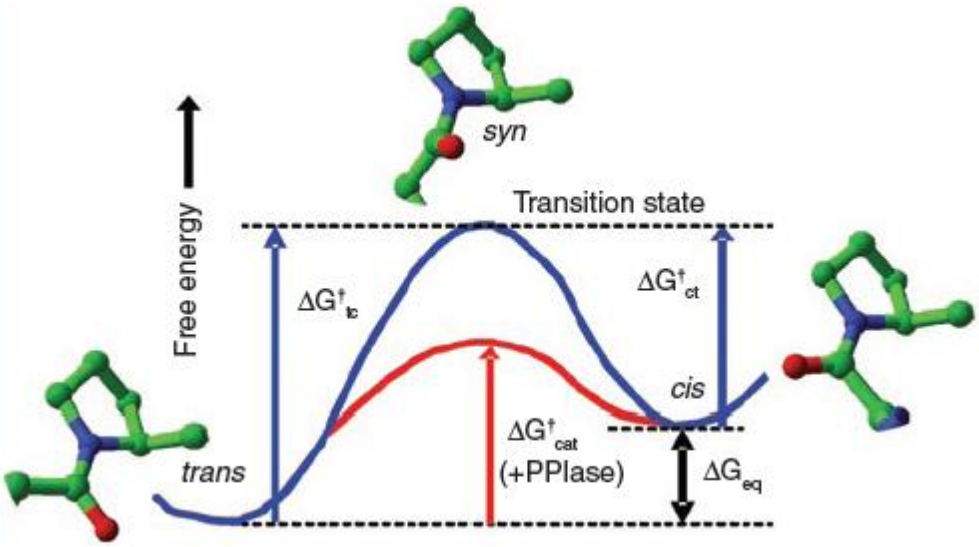

Figure 1.7: Energy diagram of cis/trans isomerization of the prolyl bond. Blue curve represents the normal transition from *trans* to *cis* conformation through *syn* conformation. Red curve shows gain in free energy with PPIase catalysis (Lu, 2007).

There is one exceptional case within *cis-trans* isomarases which is Pin 1 human PPIase. Pin 1 could promote proline isomerization of specific substrate Ser/Thr-Pro, where it could act only if serine is phosphorylated (Zhou, 2000). Pin1 could separate the binding and the catalytical active domains within the protein which makes it more flexible. It is known that this function is important in transcription regulation (Ranganathan, 1997).

Crystal structures of many cyclophilins have been solved (Dornan, 2003). They revealed high level of conservation. For example, structure of the CypA human cyclophilin has mixture of β-barrels with surrounded by two α-stretches and hydrophobic core for the CsA binding (Kallen, 2005).

By localization cyclophilins could be situated in all compartments of most tissues dependent on the specific domain organization or localization signal (Dornan, 2003). For instance, human cyclophilin CypD has a mitochondrial localization signal (Hamilton, 1998), CypE possessed an RNA recognition motif and localized in nucleus (Mi, 1996) and Cyp40 is situated in cytosol (Kieffer, 1993). Functions of the cyclophilins are also dependent on the additional domains and localization. For example, CypA besides formation of the CsA-CypA complex acts as the protein chaperon and upholds formation a virions of human immunodeficiency virus (HIV) (Luban, 1993). Other example is *Drosophila melanogaster* NinaA protein – retina-specific cyclophilin which is important for folding of rhodopsin isoforms (Stamnes, 1991). Cyclophilins could adjust protein functions, for instance, mammalian Cyp40 being a part of steroid receptor complex interacts with heat-shock protein Hsp90 (Ratajczak, 1993).

Despite of these activities recent studies revealed that proteomics analysis of the human spliceosome identified several splicing factors located in the nucleus – members of immunophilin family (Rappsilber, 2002). Some examples of that factors include PPIG (Bourquin, 1997) – binds to the Clk/Sty protein – member of the SR family, which regulates RNA splicing through rounds of phosphorylation; PPIE (Mi, 1996) – possesses N - terminal RNA binding domain and seems to prefer poly (A) and poly (U) stretches of RNA; PPIL1 (Pushkarsky, 2001) – part of the 45S snRNP complex; PPIL2 and PPILH (Horowitz, 2002) – intermingles with hPrp proteins on human spliceosome during the tri - U4/U6/U5 snRNP transition. These examples show the significance of the cyclophilins especially in coupling transcription to the splicing knowing that some of them interact with CTD of the Pol II (Horowitz, 2002).

Looking at the plant immunophilins repertoire they have the largest family in the all kingdom of 29 cyclophilin genes and 23 FKBP isoforms (figure 1.8) (Romano, 2004). Among them 21 is a single-domain cyclophilins where nine are cytosolic and five – in the secretory pathway and two - in mitochondria (Peltier, 2002). Only three on the cyclophilin subclass are located in the nucleus and all three have multidomain organization.

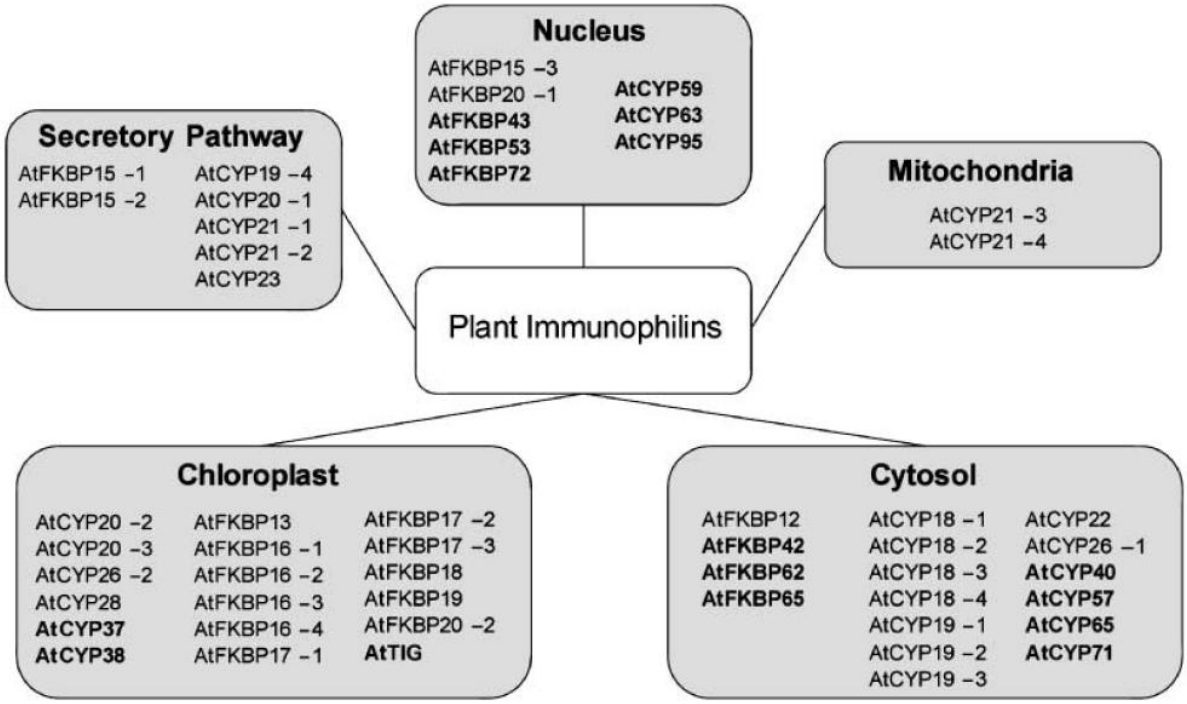

Figure 1.8: Schematic diagram of 53 immunophilins in *Arabidopsis Thaliana* genome distributed by their localization. In bold letters are the multidomain proteins (Romano 2004).

Majority of cyclophilins show ubiquitous expression pattern in the plants and are induced by various stress response such as biotic and abiotic, viral infection, salt, heat stress, cold, light and so on (Marivet, 1995). In *Arabidopsis* genome there are only 8 genes encoding the multidomain cycplophilins (MD) and among them AtCyp40 was proved to have *in vivo* function so far (Berardini, 2001). Other cyclophilin AtCyp38 has well characterized homolog in spinach TLP40 (Fulgosi, 1998). For the rest of MDs primary sequences were indentified. Four of them, AtCyp57, AtCyp59, AtCyp63 and AtCyp95, have RNA recognition motif along with other motifs that are rich in charged aminoacids (Birney, 1993), suggesting their involvement in pre-mRNA processing (Weighardt, 1999). Regardless of equal presence of the RNA-recognition motif in these proteins, they are probably contributing at different stages of the pre-mRNA processing (Weighardt, 1999). Only three of them, AtCyp59, AtCyp63 and AtCyp95, possess proper nuclear localization signal and are located in the nucleus. AtCyp57 has been shown to interact with hnRNP A1 and seems to shuttle between cytoplasm and nucleus along with hnRNP (Krecic, 1999). Since this protein is resembled the same functional characteristics as human parvulin

Par14 which is involved in translation, probably hnRNP domain of AtCyp57 could help to fold newly translated proteins (Scholz, 1997).

Other nuclear protein AtCyp59 posses an RRM domain composed of 80 to 90 amino acids and found to be conserved among species (Lorkovic, 2002). It is also established to be an interacting partner to the most of the SR proteins in Arabidopsis, but nonspecifically and through its C-terminal domain (Gullerova, 2006). AtCyp59 has pattern localization in the nucleus but does not resolve speckles. It has a spotted localization at the active sites of transcription and is co-localized with Polymerase II (Gullerova, 2006). It was found that this protein similar to its ortholog in *Schizosaccharomyces pombe,* Rct1, binds to the CTD of PolII and influences on its phosphorylation status. In conclusion it's believed that Atcyp59 is involved in connecting transcription to splicing regulation (Gullerova, 2006, 2007).

1.4 Genomic SELEX

Systematic evolution of ligands by exponential enrichment (SELEX) is a combination of combinatorial chemistry approach and experimental molecular biology techniques allowing determination of high affinity binding partner to a given molecular object (Djordjevic, 2007). The first SELEX experiment was performed in 1990 by Tuerk and Gold (Tuerk, 1990) where they created artificial RNA-aptamer library and found the ribosomal-binding site of the mRNA for T4 DNA polymerase. Typically, initial aptamer library contains around 10^{15} to 10^{16} oligonucleotides with randomized central part and fixed flanking regions (James, 2010). Such pool can be easily converted by in-vitro transcription to RNA and back to DNA via RT-PCR. Next round of selection based on incubation of the library with target molecule followed by separation of the unbound fraction from formed complex which is usually performed on nitrocellulose filters (Schneider, 1993). Selected complex is broken down and released nucleic acids are then amplified. Cycles are repeated several times to reach high affinity of binding between target molecule and selected library pool. Aim of selection is to isolate oligos which have the strongest binding affinity to target of interest.

Since SELEX procedure implicates in determination of the tight binding partner for the target of interest it finds a lot of applications in diagnostic and therapeutic applications (Bunka, 2007), for example, usage of strong binders as alternatives to antibodies in variety applications. Also several problems can appear during selection implementation. Few of the possible effects are losses of the bound oligonucleotides or enrichment of unspecific targets due to properties of nitrocellulose membrane during selection step. To overcome that problem few others possibilities of partition were described and overview in (Gopinath, 2007). Other complication is that the best winner could not be present in the genome where target molecule exists. This fact has positive and negative sides. On the one side such artificial aptamer could be used in therapeutic application from the inhibition of the particular protein to antibodies substitution which is now found a widely applications in clinical studies. On the other side best binder couldn't be corresponded to real binders occurring in vivo. In field of RNA–protein interactions other type of SELEX was developed named genomic SELEX. In contrast to randomized library, middle part of genomic library is derived directly from pieces of genome of interest, what allows searching for real-existing DNA or RNAs from particular organism. This procedure doesn't vary from aptamer-based selection with only one difference in initial library-development step. In case of genomic SELEX, library forms via Klenow-dependent addition of adaptor sequences to the genomic DNA.

First DNA library for genomic SELEX was developed in 1997 by B.S. Singer (Singer, 1997) for *E.coli, S.cerevisiae* (Gold, 1997) and human genomes (Gold, 1997). Then primary experiments on studying interaction of Cra-transcription factor with RNA-library in *E.coli* genome are made by T. Shimada (Shimada, 2005). Afterwards, using genomic SELEX experiments Kim (Kim, 2003) has been found RNA-targerts for the pre-mRNA splicing factor B52 in *Drosophila melanogaster*. Experiments showed ability of genomic SELEX to find already known targets for proteins and new targets which then are proved in various *in vivo* studies. Unsurprisingly, almost same problems appeared in genomic SELEX as in aptamer-based one. The global point for genomic approach is overselection and loss of weak but biologically significant binders during selection (Lorenz, 2006), (Lorenz, 2010). Solution could be to decrease stringency of selection conditions especially on the first rounds of SELEX and leave diversity in binding-partners in comparison to fewer winners (Zimmermann, 2010). As well as this means that high-strict selection as it applies for

aptamer-based selection could be performed in genomic variant (Zimmermann, 2010). Other interference in such experiment which is studying RNA-protein interactions is the possibility of formation secondary structures between middle part and adapter sequence in the library which could lead to unspecific or sometimes incorrect selection for protein of interest (Wen, 2001). Approach to handle this issue was suggested by Wen and Grey (Wen, 2004) in primer-free genomic SELEX. They developed a method of removal of primer-adapters before each cycle of selection and followed by the repayment them back to allow amplification of bounded fraction. Finally, genomic SELEX as any other SELEX remains in-vitro technique which in certain conditions might not correspond to the in vivo situation and some of found complexes could not appear at all in living cells (Niranjanakumari, 2002). These lead to necessity of combination SELEX experiment with other in vivo techniques, for example, CLIP and CLIP-HITS technology (Ray, 2009).

1.5 Aim of the thesis

The main aim of this thesis was to identify genomic RNA targets of the *Arabidopsis thaliana* cyclophilin AtCyp59.

1.5.1 Specific aim: Establishment of the **Genomic SELEX** system for an affinity screen of **the AtCyp59 binding RNA partners.**

AtCyp59 belongs to a family of peptidyl/prolyl cis/trans isomerases. This family encodes 29 genes in *A. thaliana* genome (Romano, 2004). It is widely known that typically cyclophilins are single-domain small proteins which help other proteins in their folding and function. However, AtCyp59 consists of a catalytically active cis/trans isomerase domain, an RRM domain and a C-terminal positively charged domain. AtCyp59 is localized in the nucleus and implicated in nuclear RNA metabolism. This protein contains C-terminal positively charged domain enriched in Arginine-Serine (SR) dipeptides. Recently, it has been shown by deletion analysis that C-terminal domain of AtCyp59 is indispensible for interaction with SR proteins, an important family of splicing factors (Graveley, 2000; Gullerova, 2006). Interestingly, AtCyp59 binds to majority of SR proteins

through its C-terminal domain suggesting possible involvement of the protein during pre-mRNA splicing (Gullerova, 2006). As it is known from animal and plant studies, SR proteins are localized in the nuclear speckles, which serve as storage reservoirs (Lamond, 2003; Lorkovic 2004). However, AtCyp59 is not co-localized significantly with SR proteins. Rather, it is situated in very distinctive dots pattern near the periphery of speckles which are correlated with active transcription starts (Gullerova, 2006). Pull-down and immunoprecipitation analysis revealed that AtCyp59 physically interacts with the CTD of Pol II largest subunit (Gullerova, 2006). Moreover, phosphorylation of the CTD decreases upon transient over expression of the AtCyp59. It is well studied that pre-mRNA processing occurs co-transcriptionaly in which CTD domain of Pol II plays a very important role (Bentley, 2002). In general, CTD acts as a binding platform for various protein factors during transcription and at the same time recruits pre-mRNA processing proteins to the nascent transcripts from speckles (Proudfoot, 2002; Kornblitt, 2004). These data, together with AtCyp59 interaction with SR proteins and CTD of Pol II, propose function of AtCyp59 as a mediator between mRNA transcription and splicing.

Beside PPIase and C-terminal domains AtCyp59 contains an RNA recognition motif. It has been shown that this domain is highly conserved from yeast to human and more conserved in comparison with PPIase domain within diverged species. It binds artificial C- or G-rich oligomers of RNAs *in vitro*, (Gulerova, 2006). This data suggest an importance of RRM domain in AtCyp59 function. Since AtCyp59 interacts with CTD of Pol II and with SR proteins, AtCyp59 can be proposed to bring certain elements of splicing machinery to the transcribing pre-mRNA. Also, AtCyp59 possible interaction with pre-mRNA could help in splicing of this RNA by recruitment of other proteins, e.g. SR proteins.

In addition, absence of *in vivo* systems for studying AtCyp59 such as T-DNA mutant lines, over expressing lines or cell suspensions hinders to study function of this protein, especially in terms of its interaction with RNAs. Among *in vitro* methods available for searching for RNA targets, we have chosen the genomic SELEX, the method which utilizes genomic RNA library for the SELective EXponential enrichment of RNA ligands (Kim, 2003).

I planned to establish DNA library of *A. thaliana* suitable for further *in vitro* transcription to the RNA library. RNA library was planned to use in affinity selection with recombinant GST-tagged protein AtCyp59 on glutathione-coupled sepharose.

1.5.2 Specific aim: Validation and functional characterization of the found **RNA targets upon their interaction with AtCyp59** *in vitro* and *in vivo*.

One of AtCyp59 close homologs, *S. pombe* protein Rct1, is essential protein; Rct1 deletion mutant cell lines show growth and morphological defects already in heterozygous stage (Gullerova, 2007). Such an unusual for cyclophilin family proteins phenomenon might explain why there are no AtCyp59 mutant lines in *A. thaliana*. Next, partial deletion of Rct1 increases phosphorylation status of CTD in *S. pombe*. These lines of evidences indicate a great importance of highly conserved protein AtCyp59 and its homologues in transcription regulation (Gullerova, M., Barta, A., Lorkovic, Z.J., unpublished data).

Catalytically active PPIase domain of the protein changes conformation of a peptidyl bond followed by prolines. Since AtCyp59 interacts with SR proteins and CTD of Pol II, both of them could potentially be substrates for the PPIase domain of AtCyp59. As for the *S. pombe* homolog Rct1, it has been discovered that it interacts with kinase Cdk9 (Skrahina T., Lorcovic, Z.J., unpublished data). Cdk9 catalyses phosphorylation of serine 2 (S2) of the CTD of Pol II and by doing that promotes efficient transcription elongation (Peterlin, 2006). It has been shown that Cdk9 interacts with PPIase domain of Rct1 suggesting involvement of Rct1 in either elongation or termination stage of transcription. Furthermore, over expression of the Rct1 increases the occupancy of the Pol II mostly towards end of transcripts (Skrahina T., Lorkovic, Z.J., unpublished data). These data indicate that AtCyp59 might be also involved in the later stages of transcription where it may connect transcription and splicing through its multidomain structure. To further understand function of the AtCyp59 and, particularly, its RRM domain, bioinformatics analysis of the found RNA targets for AtCyp59 was planned. These data would reveal a common binding sequence for the protein as well as sequence localization in genome. Relevance of found RNA targets would be analyzed *in vitro* by an electromobility gel-shift assay (EMSA) and *in vivo* by an RNA immunoprecipitation (RIP). Overall this study would provide information whether and how RNA binding activity of the AtCyp59 links this protein to transcription and (or) splicing.

2. Materials and Methods.

2.1 Strains and cells handling

2.1.1 Arabidopsis cell lines and handling of cells

Genotypes of Arabidopsis thaliana lines used are listed in Table 2.1

General genetic methods, media and growth conditions were used as described previously (Manos, 1976) 2,4-Dichlorophenoxyacetic acid (Sigma) and isopentenyl adenosine (IPAR) (Sigma) were used at the final concentration 100 µg/ml.

Table 2.1. Genotypes of Arabidopsis Thaliana lines.

Line	Genotype	Reference
Wt	Col-O	ABRC stock
AtCyp59	pXVE::35S::Cyp59-HA	This study
RSZ33	pXVE::35S::RSZ33-HA	This study
YFP	pXVE::35S::YFP-HA	This study

2.1.2 *E.coli* strains.

Genotypes of E.coli strains used are listed in Table 2.2

Table 2.2. Genotypes of *E.coli* strains.

Strain	Genotype
XL-1-Blue	endA1 gyrA96(nal^R)thi-1 recA1 lac glnV44 F'[::Tn10 proAB$^+$lacI^{q}Δ(lacZ)M15] hsdR17($r_k^-m_k^+$)
BL21(DE3)	F$^-$ompT hsdSB ($r_B^-m_B^-$) gal dcm

2.2 Plasmid construction

2.2.1 GST-tagged plasmids

To obtain GST tagged RRM+Zn domain of AtCyp59 (residues 244 -322) its cDNA was amplified with BamHI-RMM fw and SalI-RRMZn rev primers from pGST-Cyp59 plasmid. The fragment was cut with BamHI and SalI and ligated into pGEX-4T-1 plasmid. GST tagged mutated RRM+Zn domain (*RRM+Zn) of AtCyp59 (changed residues: (286(Y) to D, 288(F) to D, 291(F) to D) was amplified in the same way but from the plasmid pDEDH-*RRM-Cyp59-HA. Obtained positive clones were verified by sequencing.

Construction of plasmids expressing pGST-Cyp59 and pDEDH-Cyp59-HA were described in (Gullerova, 2006). A plasmid, expressing mutated version of RRM domain of AtCyp59 was obtained using site-directed mutagenesis approach. Cyp59 gene with introduced mutations was amplified from pDEDH-Cyp59-HA plasmid using Cyp59_RRMmut3 fw and Cyp59_RRMmut3 rev primers. Then, circular original plasmid was digested by Dpn1 enzyme and linear new plasmid was ligated and transformed into XL-1 blue cells. Positive clones were verified by sequencing.

2.2.2 Arabidopsis binary vectors.

pMDC7 plasmid was obtained from ABRC stock center and contained minimal 35S CMV promoter under control of human estrogen inducible system.

To generate plasmid expressing HA-tagged full-length Arabidopsis Cyp59 protein, correspondent cDNA was amplified with following oligonucleotides: AtCyp59XhoI fw and AtCyp59RSpeI rev, which introduce HA, SpeI, XhoI sites and stop codon, sequentially. PCR product was cloned into XhoI/SpeI digested pMDC7, resulting in pMDC7-cyp59-HA plasmid.

To generate plasmid expressing HA-tagged Rsz33 protein, correspondent cDNA was amplified with following oligonucleotides: RSZ33fXhoI and RSZ33RHASpeI, which introduce HA, SpeI, XhoI sites and stop codon, sequentially. PCR product was cloned into XhoI/SpeI digested pMDC7, resulting in pMDC7-RSZ33-HA plasmid.

pMDC7-YFP plasmid encoded YFP yellow protein was obtained from ABRC stock center.

Oligonucleotides used for cloning are listed in Table 2.3

Table 2.3 Oligonucleotides for cloning

Name	**Restriction site**	**Sequence**	**In vector**
Atcyp59 RRM with Zn finger domains			
BamHI-RMM fw	BamHI	GACTAGGGATCCATGCCTGACAATGTGCTG	pGEX-4T-1
SalI-RRMZn rev	SalI	GACTAGGTCGACTCAACAGTCCTTGGCAATATG	pGEX-4T-1
Atcyp59 full-length and RSZ33			
AtCyp59XhoI fw	XhoI	TATACTCGAGATGTCAGTTCTTATTGTGACGGAG CCTT	pMDC7
AtCyp59RSpeI rev	SpeI	TATAACTAGTTCAAGCGTAATCTGGAACGTCATAT GGGTA	pMDC7
RSZ33fXhoI	XhoI	TATACTCGAGATGCCTCGCTATGATGATCGCTAT	pMDC7
RSZ33RHASpeI	SpeI	TATAACTAGTTCAAGCGTAATCTGGAACGTCATAT GGGTA	pMDC7
****RRM-Cyp59-HA***			
Cyp59RRMmut3 fw	[Phos]	GTTGTGCGATGCTGATATAGAGGATG	pDEDH
Cyp59 RRMmut3 rev	[Phos]	TGTCACCTGTCTTGAAATCCCGGAT	pDEDH

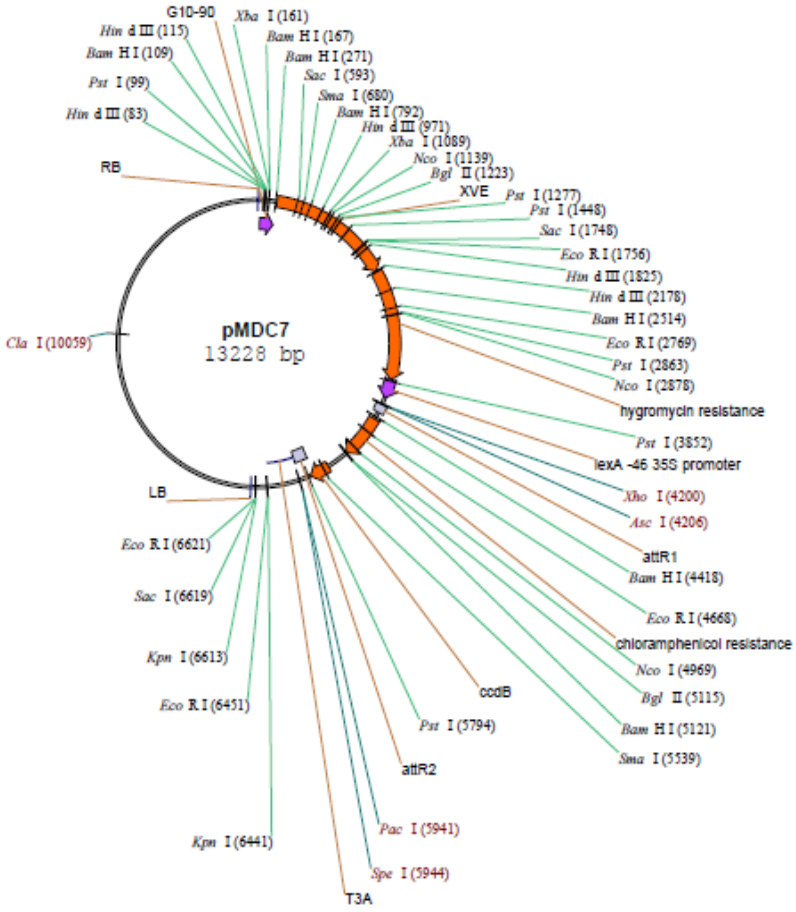

Fig. 2.1. Map of pMDC7 vector.

2.3 Arabidopsis cell-suspension generation

Each 500 µl of *wt* Col-O protoplasts cells were transformed with 150 µg of plasmid pMDC7-Cyp59-HA, pMDC7-RSZ33-HA or pMDC7-YFP via PEG inducible transformation as described in section 2.11.2. Transformed protoplasts were transferred to the 3cm cultivation plate, and incubated for 7 days in dark at 22 ^{o}C. Next, equal volume of B5-GM media supplemented with 100 µg/ml hygromycin in case of pMDC7-cyp59-HA and pMDC7-RSZ33-HA or 50 µg/ml kanamycin in case of pMDC7-YFP were added to recover the transformants. Cells were transferred into 100ml cultivation flask and incubated another 7 days in the dark at 22 ^{o}C and slow shaking at 60 rpm. Next, cells were transferred to the 300 ml cultivation flask, equal volume of B5-GM media with antibiotic were added and cells were incubated again in the dark at 22 ^{o}C at 160 rpm rotation. Starting for this point cells were diluted 1/5 times each week into fresh media with addition of selective reagent. To check presence of incorporated construct into genome

of *A. Thaliana* 1/5 of the cells were collected after 3 weeks time of incubation and dilution. Genomic DNA isolated and the presence of construct was checked with PCR using gene-specific primers listed in the table 2.3.

2.3.1 Induction of gene expression by estrogen in cell suspension. Proteosome inhibition assay.

To induce gene expression in cells carrying insertion of estrogen inducible cassette plasmid, 5 days old culture at a density of 4 mio cells/ml was induced by addition of 0.2, 2, 5, 100 µM final concentration of 17-β-estradiol human estradiol dissolved in DMSO. Then, cell suspension was incubated from 8 hours up to 5 days in the dark at 22 ^{o}C at 160 rpm rotation. Then, 500 µl aliquot of cells was spinned down, pellet was dissolved in 100 µl 2 X LB buffer, boiled for 5 min. 30 µl from it was loaded onto 10 % SDS-PAGE and analyzed by Western blot. 3 ml of induced culture were spinned down, grinded in liquid nitrogen and total RNA was isolated.

To perform proteosome inhibition experiment, 5 days old cell cultures (Cyp59, RSZ33, YFP) were induced by addition of 17-β-estradiol human estradiol at 100 µM final concentration and incubated for 24 hours at standard conditions. Then 26S proteosome inhibitor MG-115 (Sigma) was added to the culture at final concentration 100 µM dissolved in DMSO. Suspension was incubated further at standard conditions. After 1 hour, 2 hours and overnight, 1 ml of cells were collected and subjected to the western-blot analysis.

2.4 Overexpression and purification GST fusion proteins

The plasmids pGST, pGST-AtCyp59, pGST-Atcyp59,RRM+Zn, pGST-AtCyp59,*RRM+Zn were transformed into *E. coli* strain BL21(DE3). Overnight cultures grown at 37 ^{o}C supplemented with 100 µg/ml ampicilin were diluted 100 times and grown further at 37 ^{o}C till they reached an optical density of 0.8. Protein synthesis was initiated by addition of 1 mM final concentration of isopropyl-β-D-thiogalactopyranoside (IPTG). Next, cultures were incubated at 37 ^{o}C for 3 hrs. 500 ml of each culture was spinned down

at 4 ^{o}C 4000rpm for 15 min. Pellet was resuspended in 10 ml 1 X PBS buffer and sonicated (Bandelin HD 200 Sonoplus) on ice 4 times for 10 sec, at a power of 200 W, 50 cycles. Cell lysate was spinned down at 4000rpm for 10 min at 4 ^{o}C. The supernatant was mixed with 500 µl 4B glutathione sepharose beads (GEHealthcare) and incubated at 4 ^{o}C for 4 hours on slow rotation. The sepharose beads were washed three times with 15 ml of 1 X PBS buffer and the proteins were eluted 3 times with 500 µl of glutathione elution buffer incubated 10 min at room temperature. Elution buffer was exchange to the binding buffer (section 2.6) by overnight dialysis (1:20000). Finally, twenty microliters of each sample was resuspended in 60 µl of 2 X LB buffer, boiled for 5 min and 20 µl from it was loaded on 10% SDS-PAGE gel for analysis.

2.5 Arabidopsis library preparation

2.5.1 Preparation a plant material

Arabidopsis Thaliana Col-*O* seedlings were grown for 21 days in ø15 cm cultivation plates containing ½ MS Arabidopsis agar media. To prepare 2 plates, 30 mg of seeds were surface sterilized. Firstly 1.7 ml 70% ethanol with 0.1% Triton X-100 was added to the dry seed and seeds were incubated for 1 min at RT. Then ethanol was changed to 7.5% sodium hypochlorite with 0.1% Triton X-100. Seeds were further incubated for 7 min with slow rotation. Next, seeds were washed three times with ddH_2O for 10 min. Finally they were transferred to the ø15 cm cultivation plate. Plates with seeds were incubated for 2 - 3 days in dark at 4^{o}C for growth synchronization and then were moved to the growth chamber and were grown for 3 weeks at 22 ^{o}C 16 hrs of light and 8 hrs of dark.

2.5.2 Genomic DNA isolation

Three-week old seedlings were collected from the plate and homogenized in the liquid nitrogen. Genomic DNA was isolated from 2 gram of frozen tissue using a Plant DNeasy Mini kit (Qiagen) following the manufacturer's instruction. Finally, DNA was dissolved in 200 µl TE buffer and presence of RNA contamination was checked on 1%

agarose gel electrophoresis. DNA was stored for the short term at -20^{o}C and for the long time at -80^{o}C.

Afterwards, 30 µg of isolated genomic DNA was placed in a 13 ml round-bottom falcon tube and fragmented by ultrasound treatment using a Bandelin Sonoplus UW2070 device with a MS73 microtip. DNA was sonicated 8 times with 10 pulses for 10 seconds at 70% power. Sonication produced fragments from 100 bp to 4 kb in length, which were checked by agarose gel electrophoresis and compared with unshared control DNA. Then fragmented DNA was precipitated overnight in a presence of 1/10 volume 3 M NaOAc pH 5.4, and 3 volumes of absolute EtOH. The pellet was resuspended in 100 µl of TE buffer.

2.5.3 Primer labeling

In order to visualize the incorporation of the randomized primer-adaptors into genome, 2 µl of 10 µM primer (R_{ran} or F_{ran}) (see table 2.4) was added to 1 µl of 10 X PNK kinase buffer with 4 µl [γ^{32}P] ATP and 1 µl (10U) T4 polynucleotide kinase and 2 µl of water. The mixture was incubated at 37^{o}C for 1 hour. Then reaction was stopped by heating sample at 75^{o}C for 15 min and quickly chilled on ice. Unincorporated nucleotides were separated by a G-50 column from GEHealthcare following the manufacturer's instructions.

In the same time a DNA ladder was kinased in the same way e.g. 5µl of ϕx174/HinF I (250 ng) marker was added to a solution with 1.5 µl of [γ^{32}P] ATP, 2 µl of 10 X PNK kinase buffer, 1 µl 10U) T4 polynucleotide kinase and 10.5 µl of water. Solution was incubated at 37^{o}C for 10 min and then chilled on ice. One microliter for this mixture was mixed with 9 µl of formamide dye and used in 7M urea PAGE gel.

2.5.4 First and Second primer reaction annealing

The starting material was about 25 µg of fragmented and purified DNA as described above at a concentration of 1mM. The concentration of the primers was adjusted to allow annealing once every 40 nucleotides as described in (Lorenz, 2006). Firstly, 7µl of 255 µM unlabeled primer R_{ran} was added to the 72µl of DNA (25 µg). Mixture was divided in two parts to allow controlling the introduction of the forward (F_{ran})

and reverse primer (R_{ran}) separately. Next, one tube was supplemented with 2 µl of 2 µM radioactive R_{ran} primer and the other with 2 µl of water. Both tubes were incubated for 3 min at 93^{o}C, then placed on ice and further treated in parallel. After addition of 10 X Klenow buffer and deoxyribonucleotides to a final concentration of 1 mM, the reaction was started with 67 U of Klenow exo-minus enzyme (Fermentas) and incubated for 5 min on ice. After, the reaction was incubated for 25 min at room temperature and then 5 min at 50^{o}C. The reaction was inactivated by adding EDTA (final conc. 15 mM) and heating for 10 min at 75^{o}C. The reaction mixture was cleaned up from low molecular weight substances with YM-30 Millipore columns. At this point the efficiency of primer incorporation could be monitored in a denaturing 8% polyacrylamide gel.

The same protocol was applied for the forward primer (F_{ran}) reaction and where the radioactive F_{ran} primer was added to the non-radioactive sample.

Table 2.4. Oligonucleotides for library construction.

Name	**Sequence**
F_{ran}	AGGGGAATTCGGAGCGGGGCAGCNNNNNNNNN
R_{ran}	CGGGATCCTCGGGGCTGGGATGNNNNNNNNN

2.5.5 Preparative extraction of nucleic acids from PAGE

Two reaction samples from the previous step were combined, 80 µl of demonized formamide dye was added and resulted mixture was fractionated on a preparative 8% denaturing polyacrylamide (7M urea) gel for 2hrs at 100V. Labeled as described previously, size marker was added in the first line.

DNA (size 100-700 bp) was extracted from the gel using following steps. First, gel was divided into small pieces and frozen to -80^{o}C for 15 min. Then 500 µl of extraction buffer was added to each gel piece. The mixture was heated for 5 min at 95^{o}C and left overnight at 25^{o}C shaking (900 rpm). Alternatively shaking could be performed at 65^{o}C for 3 hours. Next, the gel mixture was filtrated though a 0.22 µm nitrocellulose filter (Millipore) and DNA was precipitated with 2 volume of EtOH for 3 hours at -80^{o}C.

2.5.6 *In vitro* RNA transcription

To introduce a T7 promoter sequence to the 5'end of the DNA library, the second pair of primers F_{clcf} and R_{clcr} (see table 2.5) were used. DNA library were amplified using these primers and phusion polymerase (Finnzyme) in 20 µl total of PCR reaction. The number of PCR cycles was kept 10 to avoid artificial byproducts. Typically, PCR was set uo as following: 10 s denaturation at 95^{o}C, 10 s annealing at 55^{o}C (-3^{o}C below the T_m of primer), 20 s elongation at 72^{o}C. PCR reaction was cleaned via phenol/chloroform extraction followed by PCR-clean up kit which has a low cut off to leave small DNA fragments in the library (e.g. Nucleospin extract II (Macherey-Nagel). Next, library was *in vitro* transcribed using a High Yield Transcription kit (Fermentas) Transcribed RNA was extracted with phenol/chloroform and precipitated with 2 volumes of EtOH overnight.

Table 2.5. Oligonucleotides for library amplification.

Name	**Sequence**
F_{clcf}	CCAAGTAATACGACTCACTATAGGGGAATTCGGAGCGGG
R_{clcr}	CGGGATCCTCGGGGCTG

2.5.7 RT-PCR

One-step RT-PCR kit (Qiagen) was used to reversibly transcribe RNA library and amplify resulted DNA in one tube. This kit was chosen because it contained two types of reverse transcriptases which allowed amplification of low and high abundant transcripts from the mixture. The number of PCR cycles during this reaction was kept to 7-9 to decrease formation of unspecific products. Concentration of the obtained DNA library was checked and if necessary, DNA was further amplified using one of the proof-reading polymerases. At this step, the DNA library can be store at -20^{o}C for at least 6 months or for a longer period at -80^{o}C.

2.5.8 Library quality control

The quality and comprehensiveness of the obtained DNA library was checked by PCR using 4 single copy gene primers (see table 2.6). PCR reaction was performed using standard protocol.

Table 2.6. Oligonucleotides for library quality control

Name	Gene number	Sequence	Product Size (bp)
RSZ32 Lp 595	AT3G53500	TTAGGGTTACTGCGTATTTGCACTCTC	344
RSZ32 Rp 939	AT3G53500	CCAGAGAGTTAAAACACGACCTCAGA	
Rsp31a Lp 3818	AT2G46610	TAGAGCAAGAGCTAGGAGTCCG	183
Rsp31a Rp 4096	AT2G46610	CCACATAGAGTGCAAAGCACATAC	
NUBQ	AT1G23410	GGTGCTAAGAAGAGGAAGAAT	245
CUBQ	AT1G23410	CTCCTTCTTTCTGGTAAACGT	
TPP_L	AT2G29630	TGAGTCAGAGTCAGCAATAAAGACA	200
TPP_R	AT2G29630	GCGGCAACAGTAGCTTCTTC	

Buffers:

½ MS Arabidopsis medium

2.2 g ½ GM (1A micro + ½ macro) (Duchefa)

1 ml MS vitamins (Sigma)

0.5 g MES buffer (Duchefa)

10 g Sucrose (Sigma)

7 g Agar (Duchefa)

Water to 1 L

pH 5.5 - 5.7 adjusted with 1 M KOH

TE buffer

10 mM Tris-HCl pH 8.0

1 mM EDTA pH 8.0

1 X PNK kinase buffer (New England Biolabs)

70 mM Tris-HCl

10 mM $MgCl_2$

5 mM Dithiothreitol

pH 7.6 at 25°C

Extraction buffer:

10 mM Tris-HCl pH 8.0

2 mM EDTA pH 8.0

0.3 M NaOAc pH 5.4

2.6 Affinity selection

2.6.1 Preparation of 4B Glutathione beads

In each cycle of Genomic SELEX a fresh aliquot of 40 µl 50% 4B glutathione sepharose (which is sufficient to bind up to 8 µg of a protein) was pre-washed 3 times with 400 µl binding buffer to remove residuals of ethanol and equilibrate sepharose. Next, beads were blocked with 0.5 ml (concentration 200µg/ml) tRNA (Sigma) dissolved in ddH_2O and incubated 30 min at 4^oC by slow rotation. Then, beads were washed three times with 400 µl of binding buffer.

2.6.2 Binding reaction and selection

To perform the protein-RNA binding reaction 10 µg of in-vitro transcribed RNA were dissolved in 100 µl binding buffer, heated for 5 min at 70^oC and then left for 10 min at 25^oC to refold an RNA pull. Next, the dialyzed to binding buffer protein of interest was added in a 3:1 molar excess of RNA over protein in the first 3 cycles. The stringency was increased and a ratio of 10:1 was used in later cycles. The mixture of RNA and protein was incubated for 30 min at 4^oC. Further, the blocked and washed GST-beads were added to

the binding reaction which was incubated for another 30 min at 4 ^{o}C with slow rotation. Then the beads were washed 3 times with 400 µl of binding buffer and then eluted twice with 100 µl of elution buffer according to the manufactures instructions (GEHealthcare). To release RNA from the RNA-protein complex, 400 µl FES buffer and 400 µl of Phenol pH 6.0 was added to the tube which was vigorously shaken for 10 min at 900 rpm. Then 200 µl of H_2O was added and the mixture was extracted with an equal volume of phenol/chloroform/isoamylalcohol: 25/24/1. Obtained RNA was precipitated overnight with 40 µg of glycogen, 1/10 volume of 3 M NaOAc pH 5.4, and 2 volumes of EtOH. The precipitate was dissolved in 20 µl water, cleaned from residuals of phenol with Megaclear RNA clean up kit. The yield of the selected RNA was measured by nanodrop. Selected RNA was subjected to the RT-PCR reaction following by in-vitro transcription followed by next round of an affinity selection.

Before the last cycle of selection the control – anti-GST selection with purified GST-tag protein was performed using 10:1 molar ratio GST over RNA library. Binding reaction was done as described above with only change of keeping unbound to the beads RNA fraction instead of beads fraction.

2.6.3 RT-PCR in the SELEX cycle.

To perform reverse transcription and subsequent amplification of the selected RNA, the one-step RT-PCR kit (Qiagen) and the library specific primers (see the table 2.5) were used following manufacturer's instructions. This kit was used because it contained two types of reverse transcriptases which allowed reverse transcription of low and high abundant transcripts from the mixture. The number of PCR cycles was kept to 7-9 to decrease formation of unspecific products (Zimmermann, 2010). Next, concentration of the received DNA library was checked and library was further amplified in 20 µl reaction using phusion polymerase (Finnzymes) to obtained required concentration of DNA library for the future *in vitro* transcription.

Buffers:

Binding buffer:

1 X PBS buffer

10 mM $MgCl_2$

PBS buffer:

0.135 M NaCl

27 mM KCl

8 mM Na_2HPO_4

2 mM NaH_2P0_4

FES buffer:

20 mM citric buffer pH 5.0

7 M urea

1mM EDTA pH 8.0

Elution buffer:

50 mM This-HCl, pH 8.0

10 mM reduced glutathione

2.7 Bioinformatics

For the small scale analysis CLC Main Workbench software from CLCBio were used to annotate, align and combine data. Large-scale analysis of 454 deep sequencing data was performed by Dr. Marek Zywicki in collaboration with group of Dr. Prof. Norbert Polacek in Innsbruck Medical University. For the motif discovery MEME suite software http://meme.sdsc.edu/meme4_4_0/cgi-bin/meme.cgi was used (Bailey, 1994).

2.8 In vitro RNA transcription

1. DNA template preparation for *in vitro* transcription:

Template for genomic RNA-targets not selected in Genomic SELEX: Partial sequence of gene of interest was amplified using phusion polymerase and following primer-pairs (see table 2.7) from wt cDNA of *A. thaliana* in 20 µl PCR reaction. Corresponded DNA fragment was excised from 2% agarose gel (TAE buffer) and purified using PCR extraction kit (Wizard SV Gel and PCR clean-up system from Promega). In the *in vitro* transcription 6 µl (100-200 ng) of obtained DNA template were used.

Template for 30nt RNAs containing binding motif variants: synthetic oligonucleotides containing selected with Genomic SELEX piece and complementary to T7 promoter sequence (see table 2.7) were dissolved in TES buffer to final concentration 50 µM. T7 promoter primer was dissolved in TES buffer similarly. 1 µl of oligonucleotide and 1 µl of T7 primer were mixed and 4 µl of ddH_2O were added. Reaction was heated up to 95^{o}C for 1 min and then cool down slowly to room temperature and used in the *in vitro* transcription.

2. In vitro transcription reaction:

To perform *in vitro* transcription 6µl of DNA template were added to the High-Yield In vitro transcription kit (Fermentas) and incubated at 37 ^{o}C for 4 hours followed by DNase I digestion for 20 min at 37 ^{o}C. Then, RNA was extracted by phenol/chloroform extraction and precipitated with 1/10 volume 3 M NaOAc, 1 µl (10 mg) glycogen (Roche) and 2.5 volume 96% ethanol overnight. Next, RNA was purified using RNA clean-up kit (DNA-Free RNA kit from ZYMO research) and finally dissolved in 20µl ddH_2O.

Table 2.7. Oligonucleotides to prepare DNA template for *in vitro* transcription

Name	**Gene number**	**Sequence**	**Product size (bp)**
Oligonucleotides for PRC reaction			
SRP34a fw	AT3G49430	TAATACGACTCACTATAGGGTAATCATTATGAATGCAG GTTGAGC	132
SRP34a rev	AT3G49430	CACCACCTCCACCACCATAG	
RSZ22 fw	AT2G24590	TAATACGACTCACTATAGGGTAGAGCAGTCTCATAACC GTGGTG	132
RSZ22 rev	AT2G24590	ACTCATAGCACTTCAAATCAGAACC	
RSZ32 fw	AT3G53500	TAATACGACTCACTATAGGGAAGCTATTCCAGGTCACC AGTC	130
RSZ32 rev	AT3G53500	CCAAATTCACAACTGACCTGTAGC	
U2B fw	AT2G30260	TAATACGACTCACTATAGGGTAATGCAGCCTTCGTTCCA G	140
U2B rev	AT2G30260	ACTGTTCGAAGAGAAGCTGTAGCA	

Primers with anti-T7 promoter			
Rsz32	AT3G53500	GACTGTAACTACGGCTACGGCTTGGGCTCCTTCGGCGA CGAGGGGAGCGCCCTATAGTGAGTCGTATTAAATT	73
RNA_A	AT3G28430	GGTGGTGATGGTGACGGCAACAATGGCGGTGTCAGTG GCGATGGTGACTACAATGGCGGTGCCCCTATAGTGAGT CGTATTAAATT	86
RNA_B	At1g1571	TCGACACGGCCACGATCTAATCACTCACTCCCCTATAGT GAGTCGTATTAAATT	54
RNA_C	AT3G19430	CTGGCGACGATGGTGGCGGCGATGACAGTGGTGGTGA TGACGGTGGATACACTCCTCCCCTATAGTGAGTCGTATT AAATT	81
RNA_D	AT4G13340	GGAGGTGGAGGCGGTGGTGGTGGTGGCCCTATAGTGA GTCGTATTAAATT	50
RNA_E	at4centrom eric	GACGGTGATGACGTTGGTCGAGTGATGTCGCAGATGG AGCCCTATAGTGAGTCGTATTAAATT	63
RNA_F	AT2G40570	CTGACGTGGCTGCATATTGCTGAGGTGGCTCCCTATAG TGAGTCGTATTAAATT	54
RNA_G	CHrM	GCGGATGCATGTTTACTGTAAAAGTGGTTGTGTCTTAA CGGAATGATCTCAACTCGGCTACCCCTATAGTGAGTCG TATTAAATT	85
RNA_H	AT1G77850	AACGGCGGTGTTGCGGCGGCGGAGGAGAGGAGGAGC CCTATAGTGAGTCGTATTAAATT	59
RNA_K	AT3G49400	TGTTGGTCGAGGACAGTCTTCAAGTGATCGTCGTGGTG GCTACGGTGCCCTATAGTGAGTCGTATTAAATT	71
RNA_L	AT4g04350	CAATGAAGAAGTGGTGGATGGTGTTAGTGAGCGTGGT GGCCACCCCTATAGTGAGTCGTATTAAATT	67
RNA_M	AT2G07749	ACTGTCGCAATCAAACTCTCCGCCTGCCGGATCATCTCT ATCAACCGTCTCGGCCGCACCTTCTTCCGCATCTGTCTC AACTGCTCCCCTATAGTGAGTCGTATTAAATT	110
RNA_N	AT1G08700	CGGCGTGATGGCTCCCGTCTCCATCTGCATGTTCCTCGT CGTCCCCCTATAGTGAGTCGTATTAAATT	68

RNA_O	AT1G02065	GCGGCGGTTATGGTCAGCAAGTCGCTTACGGCAGCTCC GTTTCCCCTATAGTGAGTCGTATTAAATT	67
RNA_P	AT3G32377	CAGGAGCTGCGGTTGGAACTGCGACAGGAGCTGCGGG TGGAGGTGGCGTCGTTCCCTATAGTGAGTCGTATTAAA TT	77
RNA_R	ch3centrom eric	ATCGTGGTCTGCAGCACGCGCCTAACGGCGTGCCTCGG CATCAGCGTCCCTATAGTGAGTCGTATTAAATT	71
RNA_S	ch4other	GGTGGTGGCCGCGGCGGTGGTAGTGTCCGACGGTGGT GACGGCCGACCCTATAGTGAGTCGTATTAAATT	70
RNA_T	AT5G35057	GTCGTTGGCCGAGCTGGTGGTGGCATCGTTGGTCGAG CTGGTGGTGCGCCCTATAGTGAGTCGTATTAAATT	72
T7 promoter		AATTTAATACGACTCACTATAGG	23

TES buffer:

10 mM Tris-HCl pH 8.0
1 mM EDTA
0.1 M NaCl

2.9 Electro mobility gel shift-assay (EMSA)

2.9.1 Binding reaction

To perform a binding reaction obtained RNA oligonucleotide (see table 2.8) or in-vitro transcribed RNA piece (see table 2.9) was mixed with binding buffer, loading buffer and appropriately diluted protein (Atcyp59,RRM+Zn, AtCyp59,*RRM+Zn, AtCyp59-full-length) (for example see following table 2.10). The mixture (20µl final volume) was slightly vortexed for 30 sec and left at RT (or in case of full-length protein – on ice) for 15 min and then loaded to the acryl amide native gel (6% or for the short sequences 10%).

Table 2.10. Example of protein and RNA titration for the binding reaction

Probe	RNA [µM]	5X Loading Buffer [µl]	10X Binding Buffer [µl]	Protein [nM]
1	0,1	4	2	0
2	0,1	4	2	7
3	0,1	4	2	14
4	0,1	4	2	35
5	0,1	4	2	70
6	0,1	4	2	140
7	0,1	4	2	280
8	0,1	4	2	350
9	0,1	4	2	560
10	0,1	4	2	700

2.9.2 Native gel electrophoresis and detection

Probes were run in pre-caste native PAGE 6% or 10% (Invitrogen) in 0.5 x TBE buffer at RT on 90 V for 30-40 min. Then gel was transferred to the 50ml RNA detection solution and incubated in dark for 20 min slowly shaking. Stained gel was rinsed few times with water to remove excess of the dye. Next, gel was scanned in the Thyphoon 900 imager (GEHealthcare) at 526 nm emission filter excited with green laser (532 nm) at 800 ppt resolution. Pictures were processed in Adobe Photoshop. Kd was calculated as described by (Ryder, 2008).

Table 2.8 Synthetic RNA oligonucleotides for EMSA

Name	Gene number	Sequence	Length (nt)
7nt binding motif variants			
RNA_A sh		GUUGCCG	7
RNA_B sh		GUGGCCG	7
RNA_C sh		GUCGCCA	7

RNA_D sh		GCCGCCA	7
RNA_E sh		GGUGCCG	7
RNA_F sh		GCAGCCA	7
RNA_G sh		GUAGCCG	7
RNA_H sh		GCCGCCG	7
RNA_K sh		GUAGCCA	7
RNA_L sh		GUGGCCA	7
RNA_M sh		GCGGCCG	7
RNA_N sh		GGAGCCA	7
RNA_O sh		GCUGCCG	7
RNA_P sh		GACGCCA	7
RNA_R sh		GAUGCCG	7
RNA_S sh		GCGGCCA	7
RNA_T sh		GAUGCCA	7
RNA sequences selected by genomic SELEX containing binding motif variants			
RNA_A	AT3G28430	GCACCGCCAUUGUAGUCACCAUCGCCACUGACACCGCCAUUGUUGCCGUCACCAUCACCACC	62
RNA_B	AT1G1571	GAGUGAGUGAUUAGAUCGUGGCCGUGUCGA	30
RNA_C	AT3G19430	GAGGAGUGUAUCCACCGUCAUCACCACCACUGUCAUCGCCGCCACCAUCGUCGCCAG	57
RNA_D	AT4G13340	CCACCACCACCACCGCCGCCACCUCC	26
RNA_E	ch4centromeric	CUCCAUCUGCGACAUCACUCGACCAACGUCAUCACCGUC	39
RNA_F	AT2G40570	AGCCACCUCAGCAAUAUGCAGCCACGUCAG	30
RNA_G	CHrM	GUAGCCGAGUUGAGAUCAUUCCGUUAAGACACAACCACUUUUACAGUAAACAUGCAUCCGC	61
RNA_H	AT1G77850	CUCCUCCUCUCCUCCGCCGCCGCAACACCGCCGUU	35
RNA_K	AT3G49400	CACCGUAGCCACCACGACGAUCACUUGAAGACUGUCCUCGACCAACA	47
RNA_L	AT4G04350	GUGGCCACCACGCUCACUAACACCAUCCACCACUUCUUCAUUG	43

RNA_M	AT2G07749	GAGCAGUUGAGACAGAUGCGGAAGAAGGUGCGGCCG AGACGGUUGAUAGAGAUGAUCCGGCAGGCGGAGAGU UUGAUUGCGACAGU	86
RNA_N	AT1G08700	GGACGACGAGGAACAUGCAGAUGGAGACGGGAGCCAU CACGCCG	44
RNA_O	AT1G08700	GAAACGGAGCUGCCGUAAGCGACUUGCUGACCAUAAC CGCCGC	43
RNA_P	AT3G32377	AACGACGCCACCUCCACCCGCAGCUCCUGUCGCAGUUC CAACCGCAGCUCCUG	53
RNA_R	ch3centrome ric	ACGCUGAUGCCGAGGCACGCCGUUAGGCGCGUGCUGC AGACCACGAU	47
RNA_S	ch4other	UCGGCCGUCACCACCGUCGGACACUACCACCGCCGCGG CCACCACC	46
RNA_T	AT5G35057	CGCACCACCAGCUCGACCAACGAUGCCACCACCAGCUC GGCCAACGAC	48
Rsz32	AT3G53500	CGCUCCCCUCGUCGCCGAAGGAGCCCAAGCCGUAGCCG UAGUUACAGUC	50

Table 2.9. Genomic RNA targets containing binding pattern and not selected by Genomic SELEX

Name	Gene number	Sequence	Length
SRP34a	AT3G49430	GUUUCUAAUCAUUAUGAAUGCAGGUUGAGCUUGCACA UGGUGGUCGAGGACAGUCUUCAAGUGAUCGUCGUGGU GGCUACGGUGGUGGUGGCAGCGGCUAUGGUGGUGGA GGUGGUGGUG GUGGAUCAGCUCGGU	132
RSZ22	AT2G24590	GUAGAGCAGUCUCAUAACCGUGGUGGUGGUGGAGGUC GUGGUGGUGGUCGUGGAGGAGGUGAUGGUGGUCGUG GACGUGGUGGUUCUGAUUUGAAGUGCUAUGAGUGUG GUG	132
RSZ32	AT3G53500	GGCAGGGUGGAAGCUAUUCCAGGUCACCAGUCAAAUCC CGCUCCCCUCGUCGCCGAAGGAGCCCAAGCCGUAGCCGU	130

		AGUUACAGUCGAGGUCGCAGCUACAGGUCAGUUGUGA AUUUGG	
U2B	AT2G30260	UUAAUGCAGCCUUCGUUCCAGCCGAGCGGGCAAGAAAC AAUGCCACCAAACAACAUACUCUUCAUUCAGAAUCUCCC ACACGAGACAACAAGCAUGAUGCUACAGCUUCUCUUCG AACAGU	140

Buffers:

10 x TBE buffer:

890 mM Boric acid

20 mM EDTA

890 mM Tris-base

H_20 to 1 liter

1 X Binding buffer:

10 mM HEPES-KOH, pH 7.9

10 mM $MgCl_2$

50 mM KCl

1 mM DTT

0.025% Nonnidet P-40

Supplemented with protease inhibitor cocktail (Roche) and RNAse inhibitor (Promega) upon usage.

Loading buffer:

50% glycerol

0.01% Bromphenol blue

0.01% Xylene Cyanol

RNA detection buffer:

SYBR Green II (Invitrogen) 1:10000 dilution in 0.5 X TBE buffer

Protein detection buffer:

SYPRO Red (Invitrogen) 1:5000 dilution in 7.5% Acetic acid.

2.10 Protoplasts isolation

Five-days old (4-5 mio cells/ml) *wt* Col-0 *A. thaliana cell* suspension (30 ml) grown at 22 ^{o}C in the dark with 160 rpm shaking in B5-GM media were collected to the falcon tube and spinned down at 25 ^{o}C 1000 rpm for 5 min. Cell pellet was resuspended in 25 ml enzyme solution. B5-0.34 GM media was added to obtain 50 ml volume, and mixture was transferred to the 15 cm cultivation plate for 1 -1.5 hrs in the dark with slow 60rpm shaking. Cells were checked under microscope every 30 min till the stage when more than 50% cell population were looked like single-cell colonies of round-shape. After this point cells were transferred to the falcon tube and spinned down at 25 ^{o}C 1000 rpm for 5 min. Pellet was resuspended in 40 ml B5-0.28MS media and cells again spinned down at 25 ^{o}C 1000 rpm for 5 min. At this stage protoplast cells should float over the solution. Such cells were transferred to the 13 ml round-bottom falcon tube, washed again with B5-0.28MS media and spinned down for 7 min 800 rpm. Last step was repeated one more time to obtained homogeneous population of protoplast. (This work was performed together with Mag. Andrij Belokurov)

Buffers:

B5-GM media:

4.5 g/ Murahigo & Skug (macro & micro + B5 vitamin) l (Duchefa)

100µl/l 2,4-Dichlorophenoxyacetic acid (auxin stock - 10mg/ml)

30 g/l Sucrose (Sigma)

pH 5,7 adjusted with 1 M KOH

B5-0.28MS media

4.5 g/l B5 vitamins powder (Duchefa)

96 g/l Sucrose (sigma)

pH 5,5 adjusted with 1M KOH

Enzyme solution:

1% Cellulase (Duchefa)

0.2% Macerozyme (Duchefa)

In B5-0.34GM media

Filter sterilized, stored -20 ^{o}C

2.11 PEG inducible protoplasts transformation

2.11.1 Plasmid DNA preparation

In this work following plasmid were used: pDEDH-Cyp59-HA (plasmid encodes full-length Atcyp59 protein fused with HA-epitope at C-terminus. (Gullerova, 2006), pDEDH-Cyp59*RRM-HA (plasmid, derived from pDEDH-Cyp59-HA with 3 point mutations in the most conserved amino acids in RRM domain (described above)), pGREEN-MPK6-HA (plasmid encodes Map kinase 6 protein fused with HA epitope. (Kind gift of Dr. Prof. Irute Meskeine)), pDEDH-GFP (encodes green fluorescence protein. (Gullerova, 2006)). Each construct was transformed onto *E.coli* XL-1 blue competent cells and plasmid DNA was isolated using Maxi DNA preparation (Qiagen). Final concentration was adjusted to the 500 ng/µl.

2.11.2 Protoplasts transformation

One hundred µl of protoplasts (2 mio cells/ml) in tube were carefully mixed with 30 µl plasmid DNA (concentration 0.5 µg/µl) to obtain homogenous solution. 300 µl PEG-600 pH 9.0 were added to the mixture and obtained solution was left at RT for 15 min. Next, 1.5 ml 0.275M $Ca(NO_3)_2$ was added to the tube, and resulted solution was mixed and spinned down at RT 1000rpm for 5 min without break. Further, supernatant was discarded and pellet resuspended in 500 µl B5-0.34GM media. Transformed protoplasts were left in the dark at 22 ^{o}C for 24 hours. (This work was performed together with Mag. Andrij Belokurov)

Buffers:

<u>B5-0.34GM media</u>

4.5 g/l B5 vitamins powder (Duchefa)

30.5 g Glucose (Serva)

30.5g Mannitol (Serva)

1 mg/l 2,4D (auxin-hormone to block cell-differentiation) (10 mg/ml)

pH 5.5 adjusted with 1 M KOH

PEG 6000 solution:

60 g/l PEG 6000 (sigma)

16.4 g/l Manitol (serva)

4.7g/l $Ca(NO_3)_2*4H_2O$

pH 9.0 adjusted with 1M KOH

0.275M $Ca(NO_3)_2$

64.94 g $Ca(NO_3)_2 *4H_2O$/l

2.12 SDS-PAGE and Western blotting

Protein extract were prepared from ~400.000 pelleted protoplast cells by adding 60 µl of 2 x LB buffer, then was boiled for 5 min. 25 µl of extract were loaded to 10% SDS-PAGE and run at 150 V for 1.5 h. After, gel was transferred onto PVDF membrane (Millipore). Western blotting was performed using standard protocol. Primary antibody was Anti-HA, rat, monoclonal (Roche) diluted 1:5000 and secondary antibody - rabbit anti-rat Immunoglobulin G (Sigma) diluted 1:10000. Secondary antibody was conjugated with horseradish peroxidase therefore western blot was developed using cheminiluminescence kit (GE Healthcare) and exposed to Kodak Biomax MR film.

Buffers:

2 X LB Buffer

4% SDS

20% glycerol

10% β-mercaptoethanol

0.004% bromphenol blue

0.125 M Tris-HCl, pH 6.8

2.13 Total RNA isolation

Transformed protoplasts cells (24 hrs after transformation) (~600.000 cells) were spinned down at 25 ^{o}C for 5 min, 4000 rpm (table centrifuge). Pellet was resuspended in 450 µl RTL buffer (Qiagen). Total RNA was isolated and treated with DNase I following

manufacturer's instruction (Qiagen Plant RNeasy mini kit). Purified RNA were dissolved in 30 µl of water and stored at -80 ^{o}C.

2.14 RNA immunoprecipitation (RIP)

2.14.1 Magnetic-beads conjugation

Dry magnetic dynalbeads (3 mg = 3*10^8 beads) were dissolved in 60 µl of 0.1 M sodium phosphate buffer, vortexed for 30 seconds and slowly rotated for another 10 min till complete dissolution. Beads were put on magnet to remove supernatant and buffer was exchanged to the mixture of 60µl antibody (monoclonal mouse anti-HA, concentration 1mg/ml (sigma)), 60µl of 0.3M sodium-phosphate buffer and 60µl 3 M $(NH_4)_2SO_4$. Mixture was vortexed again for 30 sec. Reaction tube was parafilmed to prevent liquid evaporation and incubated at 30 ^{o}C for 18 hrs with slow rotation. Afterwards, supernatant was removed and beads were quickly washed with 120 µl 100 mM Glycine - HCl pH 2.5. Buffer was exchanged to the 120 µl 10 mM Tris-HCl pH 8.8. Second fast wash were performed with 120 µl of 100 mM triethylamine (Make fresh every time by adding 168 µl of stock solution to 11.156 ml of ddH_2O) and subsequently exchanged to 400 µl 1 x PBS. Beads were incubated for 5 minutes on rocker or shaker. Beads were further washed 3 times with PBS buffer, then once with PBS+0.5% Triton-X-100 for 5 min and once with PBS+0.5% Triton-X-100 for 15 minutes. Finally, beads were resuspended in total 200 µl of 1 x PBS+0.02% NaN_3 and stored at 4^{o}C.

2.14.2 Immunoprecipitation

Three mio. transformed protoplasts cells (24hrs after transformation) were spinned down for 5 min at 25 ^{o}C, 4000 rpm on table centrifuge. Pellet was resuspended in 450 µl lysis buffer and incubated for 20 min on ice. Afterward cells were sonicated (Bandelin HD 200 Sonoplus) 4 times for 10 sec on ice, at a power of 200 W, 10 cycles to promote nuclear membrane breakage. Then, cells were spinned down and supernatant diluted 10 times with binding buffer to decrease SDS concentration to 0.1%. At this stage

20 µl pre-washed in binding buffer magnetic dynalbeads, coupled with anti-HA antibodies were added to the supernatant and incubated at 4 ^{o}C for 1 h with slow rotation. After one hour solution was transferred to the eppendorf and moved on magnetic holder. The supernatant with unbound material was removed. Magnetic beads were washed there times with 1ml binding buffer and then three times with 1 ml washing buffer. Finally, to remove RNA-protein complex from the beads and digest proteins, 400 µl washing buffer was added to the beads supplemented with 100 µg proteinase K (sigma) and 5 µl 10% SDS. The mixture was incubated for 30 min at 55 ^{o}C followed by RNA extraction with phenol/chloroform. Extracted RNA were precipitated with 1/10 volume NaOAc, 10 µl (100mg) glycogen (Roche) and 2.5 volume 96% ethanol overnight. Next day, precipitate was spinned down at 4^{o}C for 15 min at 14000rpm then washed with 500 µl 80% ethanol to remove residual amount of salts and dissolved in 80 µl RNase – free water. Next, DNase I treatment and RNA clean-up were performed according to manufacturer's instructions (Qiagen RNeasy Plant Mini kit). Purified RNA was dissolved in 30 µl H_2O and stored at -80 ^{o}C.

Buffers:

<u>0.1M sodium-phosphate buffer pH 7.4</u>

2.62 g $NaH_2PO_4xH_2O$

14.42 g $Na_2HPO_4xH_2O$

Adjust to 1L with water and if necessary pH.

<u>1 x PBS pH 7.4</u>

0.26 g $NaH_2PO_4xH_2O$

1.44 g $Na_2HPO_4xH_2O$

8.78 g NaCl

Adjust to 1L.

<u>100mM Glycine –HCl pH 2.5</u>

0.375g/50 ml glycine. Adjust pH with HCl.

<u>Lysis buffer</u>

50 mM HEPES-KOH, pH 7.9

2.5 mM $MgCl_2$

1 mM EDTA

1% SDS

Supplemented with RNAse inhibitor(Promega).

Binding buffer

10 mM HEPES-KOH, pH 7.9

2 mM $MgCl_2$

50 mM KCl

1 mM DTT

0.025% Nonnidet P-40

Supplemented with protease inhibitor cocktail(Roche) and RNAse inhibitor(Promega) upon usage.

Washing buffer:

100 mM HEPES-KOH, pH=7.9

2.15 Semi-quantitative RT-PCR

Each RNA sample from immunoprecipitation (10 µl) or 1 µg of total isolated RNA were reversibly transcribed with M-MLV reverse transcriptase (Promega) and 15-mer oligo-dT in 20 µl total for 15 min at 42 ^{o}C as recommended by Promega. Then sample was diluted with 80 µl of H_2O and 2 µl from it was used for PCR reaction. 20 µl PCR was performed with Phsuion polymerase (Finnzymes) supplemented with 0.1 µl 100 µM each primer-pairs for the target genes (see table 2.11) and all other PCR components as recommended by Finnzymes.

Table 2.11. Oligonucleoties for RT-PCR for targets genes.

Name	Gene number	Sequence	Product size (bp)
Domino_rt_F	AT5G62440	GCGACCCCGAAAGCCGAGAC	452
Domino_rt_R	AT5G62440	TGTCCATTGCCGTTAGCTCCAGG	
SRp34a_rt_F	AT3G49430	TGGATGGCTGTCGCTTGAGGG	620
SRp34a_rt_R	AT3G49430	TCTCGACATTGCCCTGGGGG	
Gar1_rt_F	AT3G03920	CACCAATGAGAGGCGGCGGG	579
Gar1_rt_R	AT3G03920	AGCTTCCACGAGAGCCACCG	
RSZ22a_rt_F	AT2G24590	TGGGTTGCTAGAAGACCTCCTGGT	702
RSZ22a_rt_R	AT2G24590	ACGGCAGATACAACTATGGCT	
HAP5B_rt_F	AT1G56170	GGGTTTCGCGAGATCTCACTCTCA	704
HAP5B_rt_R	AT1G56170	TGGTTCCCAGCAGAGCAGAGC	
RSZ32 lp595	AT3G53500	TTAGGGTTACTGCGTATTTGCACTCTC	640
RSZ32 rp2390	AT3G53500	TGTAGCTGCGACCTCGACTGTAAC	
SR45_rt_F	AT1G16610	CGGGCTCTCCTATCCGCCGT	469
SR45_rt_R	AT1G16610	GGAGGTGGTGGTGGCGGTGA	
RS41_rt_F	AT5G52040	GGTCGCACAGGACGCAGACT	1066
RS41_rt_R	AT5G52040	ACGACAAGCGATTTCGAATGGAGTCA	
AT3G28430_rt_F (RNA A)	AT3G28430	CCCACTTATGCAGTGAAGATACTCCG	1062
AT3G28430_rt_R (RNA A)	AT3G28430	TGCAGCCGGCTAAAGGTGCC	
AT4G13340_rt_F (RNA D)	AT4G13340	TCTCCAGCACCAACTCCAGTTTATTG	420
AT4G13340_rt_R (RNA D)	AT4G13340	GCGGTGGAGGCGGAGAGCTA	
tRNA_fw (RNA F)	AT2G40570	AGCGTTTTCCGGATAGCATGTCG	736
tRNA_rev (RNA F)	AT2G40570	TGGTCGAGCATTCACCGCGT	
ARF17_fw (RNA H)	AT1G77850	CTCGGATCACATGGTTTCAAGGCA	1160
ARF17_rev (RNA H)	AT1G77850	GCCAGACTCTGCAGGACCGC	

WD-40_fw (RNA K)	AT3G49400	TCTCCCATGGTGGAACTTACTGCT	650
WD-40_rev (RNA K)	AT3G49400	GCAGCCTTCTGCGACCTGGC	
AT2G07749_fw (RNA M)	AT2G07749	TCCTCGGGGGCAGACCACAC	750
AT2G07749_rev (RNA M)	AT2G07749	CTCGGTGGGATGGGTCGGGT	
SQM_fw (RNA O)	AT1G02065	GCGGATCTGAGCCACGCGAA	1170
SQM_rev (RNA O)	AT1G02065	TCGGGATCCCCCACACCACA	

3. Results

3.1 Selection of the RNA targets for AtCyp59 with Genomic SELEX

3.1.1 Genomic DNA library of *Arabidopsis Thaliana*

Arabidopsis thaliana cyclophilin Cyp59 (AtCyp59) has been shown to be localized in the nucleus (Weighardt, 1999). Moreover, unlike majority of the *A. thaliana* cyclophilins AtCyp59 possesses multidomain structure containing catalytically active domain, RNA recognition motif and C-terminal charged domain. In last decade it has been demonstrated that RRM domain of AtCyp59 could potentially bind artificial RNA and has preferences to the CG-rich sequences (Gullerova, 2006). In order to indentify whether AtCyp59 is capable to bind real messenger RNA and if this binding has influence on transcription and splicing we decided to set up a Genomic SELEX experiment. Choice of method was based on absence of any over expressing protein system (i.e., cell suspension, stable plant line) as well as lack of T-DNA insertion mutation lines.

Genomic SELEX consists of two principal steps: library development and affinity selection. During library construction the specific adapter sequences (Figure 3.1) are incorporated on the both sides of the genomic DNA library fragments using Klenow extension reaction for further utilization of *in vitro* transcription followed by amplification. To ensure ligation of primer-adaptors to the maximum number of available places in the *A. thaliana* genome, genomic DNA was fragmented.

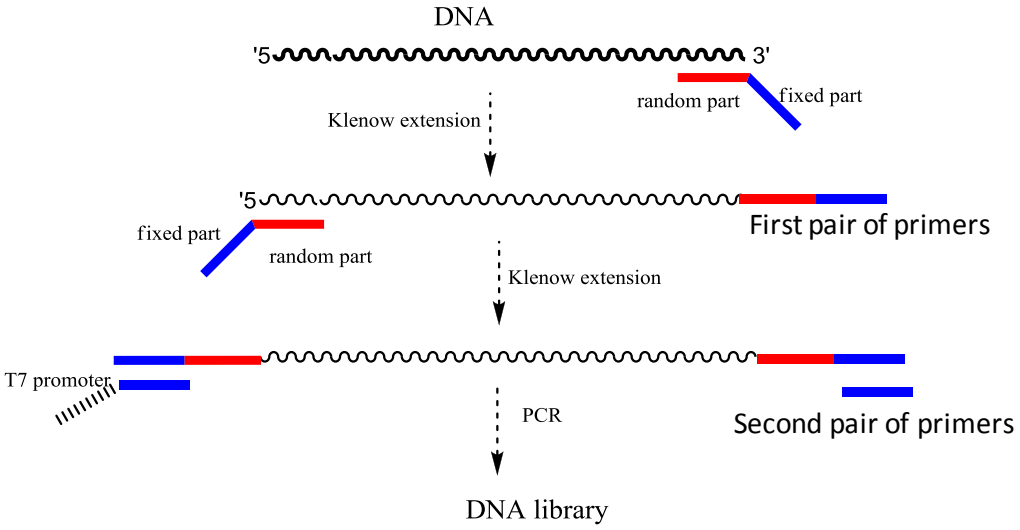

Figure 3.1: Scheme of Arabidopsis DNA library preparation. Examples of the two-step Klenow extension reaction with single stranded DNA fragment (in curls). Primers-adapters are in red-blue, where red – is a randomized sequence, blue – fixed sequence. The second pair of primers is complementary to the fixed part of the first pair and in dashes – depicted T7 promoter sequence on the 5'-end of one adaptor.

At the beginning of library development genomic DNA was isolated from approximately 2g of adult (3-weeks old) *A. thaliana* wild-type Columbia leafs grown *in vitro* on MS media. Obtained DNA (30µg) was fragmented by ultra-sound treatment (Figure 3.2 A) to increase availability of genomic pieces for further ligation of adaptors (Lorenz, 2006).

Primers and DNA-sized marker had been labeled on the 5'-end with γ-[^{32}P]-ATP to visualize incorporation efficiency before introduction of primer-adaptors to the fragmented genomic DNA of *A. thaliana*. Primer-adaptors R_{ran} and L_{ran} (table 2.4 Materials and methods) were designed to contain randomized part which should suit well to increase their incorporation to every available piece of DNA; and the fixed-part which was screened to be not present in the *A. thaliana* genome to prevent enrichment of false-positive sequences. Adaptors were annealed to the fragmented genomic DNA using Klenow extension reaction in the two-step manner. R_{ran} and L_{ran} primers were introduced separately to control incorporation efficiency. Aliquots of each reaction as well as labeled primers were resolved on PAGE followed by autoradiography (Figure 3.2 B). Yield of single Klenow extension reaction as it is well known does not exceed 10% (Singer, 1997). Finally

only around 1% for initial fragmented genomic DNA got incorporated with adaptors. That explains necessity for substantial amount of start material.

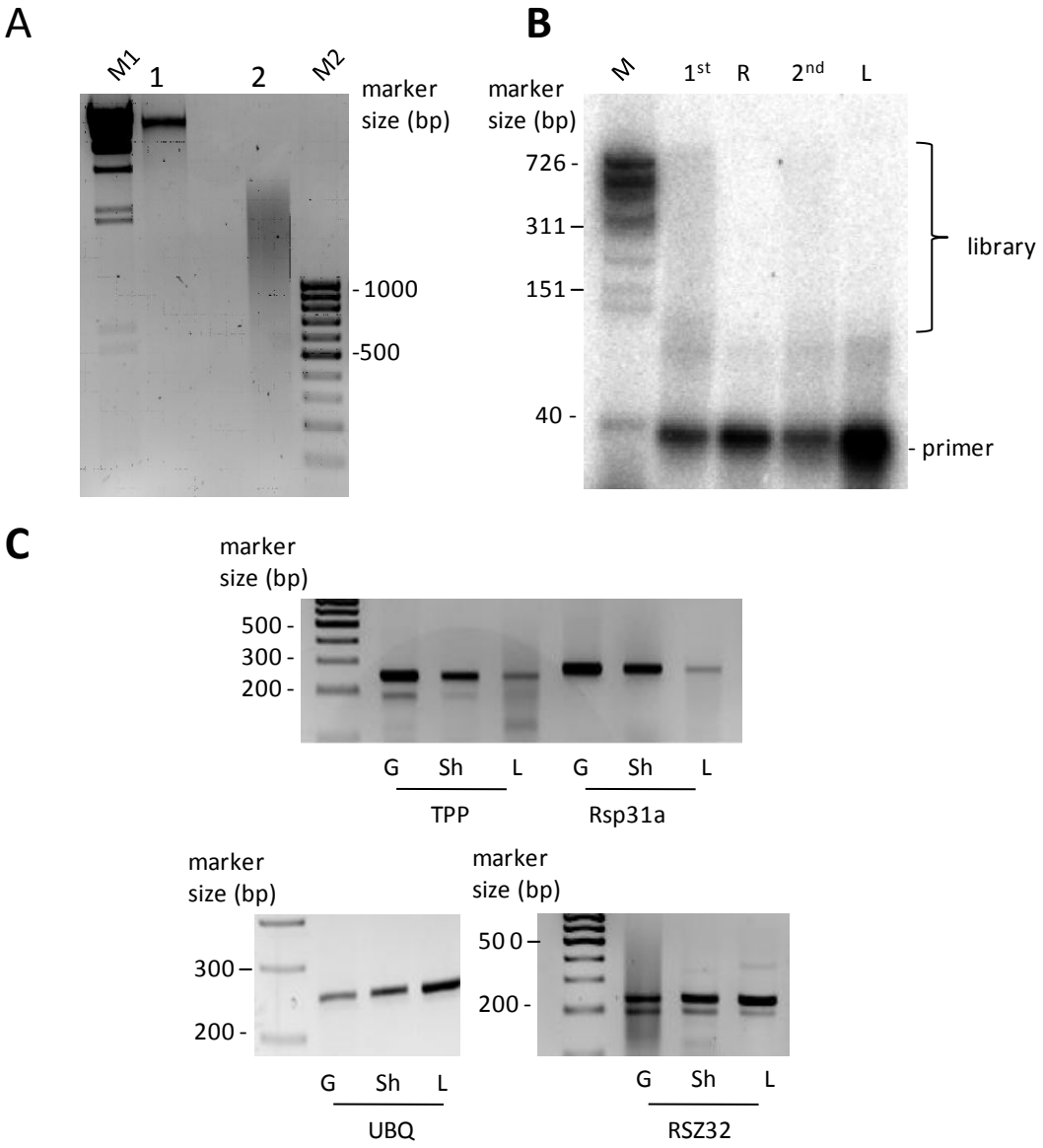

Figure 3.2: DNA library preparation and verification. **A** – 1.2 % agarose gel electrophoresis, M1– λ_{indIII} marker; 1- isolated genomic DNA from Arabidopsis thaliana; 2 – fragmented DNA; M2 – 100 bp DNA sized marker (Fermentas); **B** – Autoradiogram of PAGE- gel electrophoresis, on side (M) – DNA size marker, 1st- fragmented DNA with incorporated R primer, R – labeled primer-adaptor, 2nd – second step of annealing reaction of fragmented DNA with both R and L primers ligated, L – P^{32}- labeled primer before reaction. **C** – Gel-electrophoresis of the PCR amplification with genome-specific primers. Here is on the side DNA size marker, G – PCR on genomic DNA before manipulations, Sh- fragmented genomic DNA, L – DNA-library with

incorporated primers on the both sides. Sequences of gene specific primers are in the table 2.6 (Materials and Methods).

After two-step Klenow reaction DNA library was resolved on a preparative PAGE followed by autoradiography. DNA fragments of size from 100 bp to 700 bp were excised and eluted. Then obtained DNA library was further amplified by PCR to introduce the second pair of adapters (table 2.5 Materials and Methods) as well as T7 adaptor sequence for *in vitro* transcription reaction. During this step only those sequences which had gained randomized adaptors were amplified because second pair of primers is complementary to the fixed-part of the randomized primers. Next, DNA library was *in vitro* transcribed using T7 polymerase followed by reverse transcription and amplification utilizing only second pair of primers. These steps ensure that only those fragments which contain T7 promoter sequence on the 5'-end and the adapter sequence on the 3'-end were left in the system. Finally, resulted "clean" DNA library was screened by amplification with couple of gene-specific primers to check its diversity. DNA library was compared with initial genomic DNA of Arabidopsis thaliana and fragmented DNA (figure 3.2. C). PCR analysis has shown the presence of four selected genes in all libraries. Based on that observation obtained DNA library was further used for affinity selection with AtCyp59 and also subjected to the 454 deep sequencing to ensure its representation.

3.1.2 Selection of RNA targets of AtCyp59.

AtCyp59 has multidomain organization where each domain is responsible for different action; for example, C-terminal charged domain interacts with SR proteins and PPIase domain binds to the CTD of Polymerase II (Gullerova, 2006). Since we were interested in the functional contribution of RNA recognition domain of the AtCyp59, we decided to clone this domain in frame with N-terminal GST- tag as shown in figure 3.3 A. RRM domain of the AtCyp59 was over expressed in *E. coli* and purified using glutathione Sepharose-4B. Figure 3.3B shows SDS-PAGE of isolated recombinant GST-tagged AtCyp59, RRM+Zn. Purified protein was dialyzed in SELEX binding buffer and used in affinity selection cycle (as described in Material and Methods).

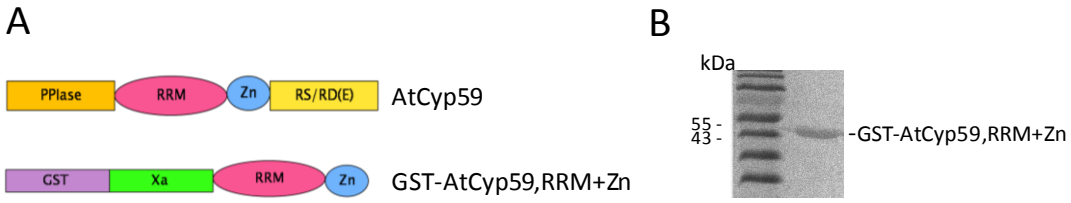

Figure 3.3: Purification of the RRM-Zn domain of AtCyp59. A – Schematic representation of the domain organization of full-length AtCyp59 and GST-tagged construct of RRM+Zn motif of AtCyp59. PPIase – peptidyl/prolyl cis-trans isomerase domain; RRM – RNA recognition motif; Zn – zink –finger motif class CCHC, RS/RD(E) – domain enriched in Argenine, Serine, Aspartate, Glutmate; GST- Glutathione S-transferase; Xa - Factor Xa protease site. **B –** Coomassie blue-stained gel of purified recombinant GST tagged RRM+Zn domain of cyclophilin AtCyp59. Molecular mass standard in kilodaltons are indicated on the side.

The second principal step in Genomic SELEX procedure is the selection of RNA targets through multiple rounds of affinity selection between protein and RNA pool, as it is schematically depicted in figure 3.4. In general, during each cycle of selection the DNA library is transcribed using T7 polymerase into an RNA library. Then, the RNA pool is incubated with a protein in appropriate molecular ratio followed by separation of RNA-protein complex from unbound RNAs on glutathione sepharose. Resulted complex is eluted from sepharose and then an RNA fraction is extracted for further reverse transcription followed by amplification to ensure that same amount of material is employed at each cycle. Thus, going to the next round of selection, along with increasing molecular ratio of RNA over the protein we enhance stringency of selection leaving in the system only high affinity RNA binders for AtCyp59 protein.

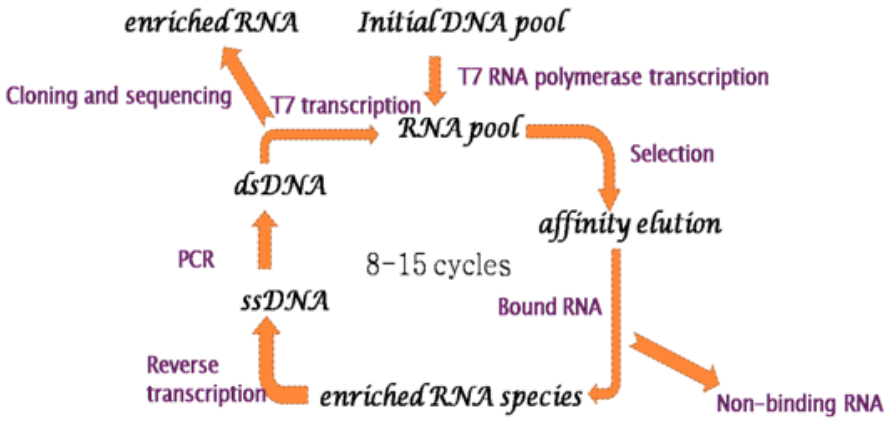

Figure 3.4: Schematic representation of SELEX process. Cyclic enrichment of RNA pool through subsequent rounds of selection with AtCyp59, RRM+Zn protein on glutathione sepharose.

Crucial step in the affinity selection part of the Genomic SELEX is separation of RNA-protein complex from unbound RNA. Since we utilized GST-tagged fusion version of the RRM domain with Zn finger of the protein AtCyp59, we decided to use 4B Glutathione sepharose for that purpose. Gluthathione sepharose is a reversible system therefore RNA-protein complex could be easily eluted by excess of reduced glutathione. Unspecific RNA binding was prevented by pre-incubation of glutathione sepharose with excess of tRNA.

Principe of Genomic SELEX is to subsequently enrich RNA binder pool, which is capable for binding to the protein throughout cyclic process. During the first three rounds of selection we utilized mild binding conditions with molar ratio of RNA pool to the protein 3:1 to allow possibility for weak RNA-binders to be present in the system. After we had seen substantial enrichment (figure 3.5), we increased molar ratio of RNA to the protein up to 10:1 and continued selection. This step ensures that only specific RNA targets are present in the selected pool. After the 9^{th} step when RNA recovery level had reached 2% out of possible 10% (due to 10:1 molar ratio) we performed additional control step of selection. Instead of AtCyp59 protein we used recombinant GST protein and collected unbound RNA fraction to discard all unspecific RNA artifacts. After 11 cycles we collected a selected RNA pool and performed 454 deep sequencing followed by bioinformatics analysis. Prior to the 454 sequencing we had done preliminary analysis of the selected RNA library employing reverse transcription followed by TOPO T/A

(Promega) cloning system. Sanger DNA sequencing of obtained clones revealed that 253 out of 300 clone sequences contained 40-50 nt fragments of *A. thaliana* genome. Based on that observation we decided to perform 454 deep sequencing of selected library.

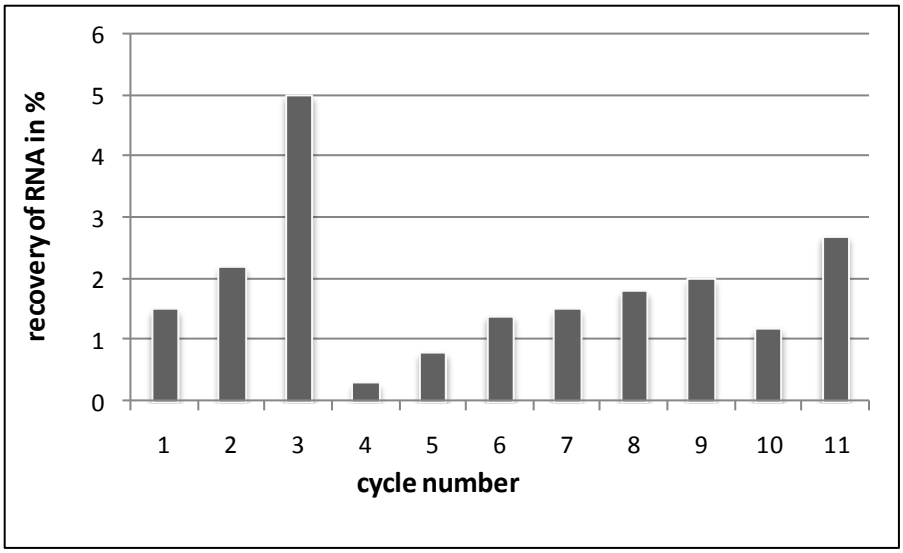

Figure 3.5: Enrichment of the RNA sequences from *Arabidopsis thaliana* genome that binds cyclophilin AtCyp59. In a vertical axle calculated RNA recovery as percentage of collected RNAs that were bound to the protein and start RNA material in each cycle. During 1-3 cycle molecular ration of RNA to the protein were 3:1, cycle 4-9, 11 – ratio 10:1, cycle 10 – anti GST selection.

3.2 Bioinformatics analysis of the sequenced libraries.

3.2.1 Analysis of the Initial DNA library.

Prior to selection with AtCyp59, the initial DNA library had been checked by PCR amplification with few gene-specific primers. As it is shown in figure 3.2 C all selected genes were present in constructed library. Nevertheless, the question we asked is that if this library resembles genome of *A. thaliana* in existence of genomic features. To answer this question, initial DNA library was subjected to the deep sequencing using 454 technology. About 20000 reads (2% of the entire genome) were obtained and aligned

back to the genome of *A. thaliana* after primer-adaptor sequences had been discarded. As shown in figure 3.6 initial DNA library contains all genomic elements such as exons, introns, pseudogenes, transposones, as original genome does. However, analysis revealed that there is a decrease of abundance of intergenic regions and gene intronic sequences in the initial DNA library. From the other hand there is an enrichment of gene exons sequences and rRNA sequences in the initial library in contrast to the genome. This observation could be explained by the differences in availability of diverse genomic element for primer-adaptor incorporation, i.e. status of the chromatin condensation can influence on availability (Johnson et al., 2002). However, distribution of genomic elements across the initial library resembles genome of *A. thaliana* and distinctions in abundance are not crucial. We concluded that initial DNA library is representative enough to perform an affinity selection with the protein of interest.

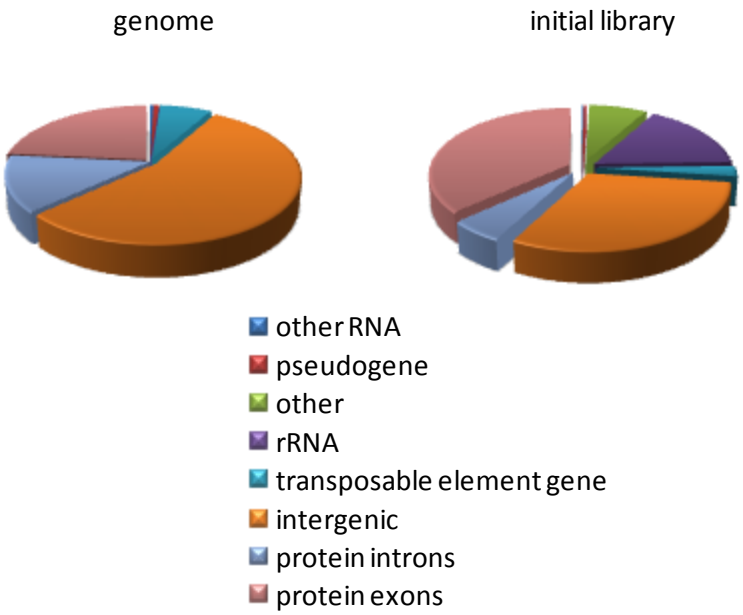

Figure 3.6: Comparison between sequences composition in the genome and initial DNA library before selection. Genomic elements are indicated in colors and calculated as percentage from total number of annotated sequences in genome or obtained sequences after 454 sequencing.

Initial DNA library showed a slightly different sequences distribution in comparison to the genome of *A. thaliana*. To understand whether this library biased towards excess of any particular nucleotides variations we compare di-nucleotides distribution between

the initial DNA library and genome of the *Arabidopsis thaliana*. As shown in the figure 3.7 there is almost two-fold overrepresentation of CG-rich sequences in the library in contrast to the genome and a small underrepresentation of the AT-rich sequences. These divergences would be considered in further analysis.

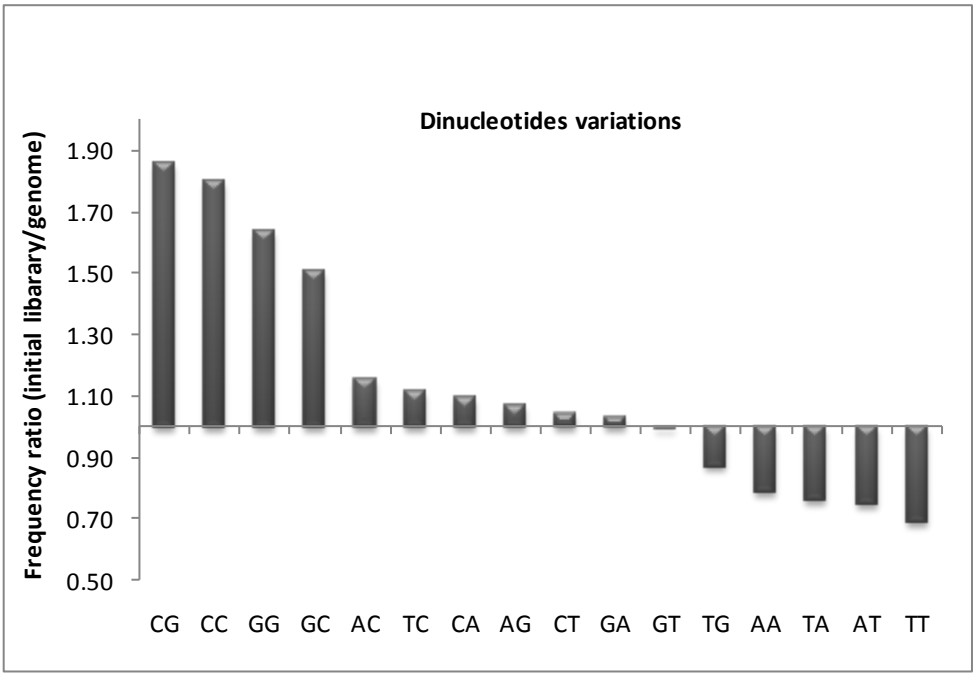

Figure 3.7: Dinucleotides bias distribution in the initial library. Frequency ratio is calculated as difference between abundance of di-nucleotides in the library to the genome. When number is >1 then this variation is overrepresented in the library and if <1 – underrepresented.

3.2.2 Sequence composition of the selected library is different from the genome and from the initial DNA library.

After 11 rounds of the affinity selection we sequenced resulted SELEX library using 454 deep sequencing technologies. Obtained 20.000 sequencing reads were aligned back to the Arabidopsis genome and compared with the initial library and genome of *A. thaliana* in abundance of the featured in figure 3.8 genomic elements. SELEX library drastically differs from the initial DNA library as well as from genome. There is a significant enhancement of the sequences annotated as proteins exons and decrease of

intergenic sequences and protein introns. Also, it is remarkable that an unusual expand in rRNA sequences in the initial library diminished after selection with AtCyp59. From the other side, transposable elements seem to be more present in the SELEX library compare to either the initial library or the genome. These observations led us to the following conclusions. First, SELEX library varies from both genome and the initial DNA library and is enriched only in certain types of genomic elements that show the evidence of the directed selection. Second, increase in the exonic sequences could serve as an indicator of mRNA selection by RNA binding protein which might support our hypothesis that AtCyp59 is involved in pre-mRNA splicing or transcription. Finally, we saw an enrichment of sequences correspondent to transposable elements (turquoise on figure 3.8) which could due to a favorable amplification of the repetitive sequences (Economou et. al 1990).

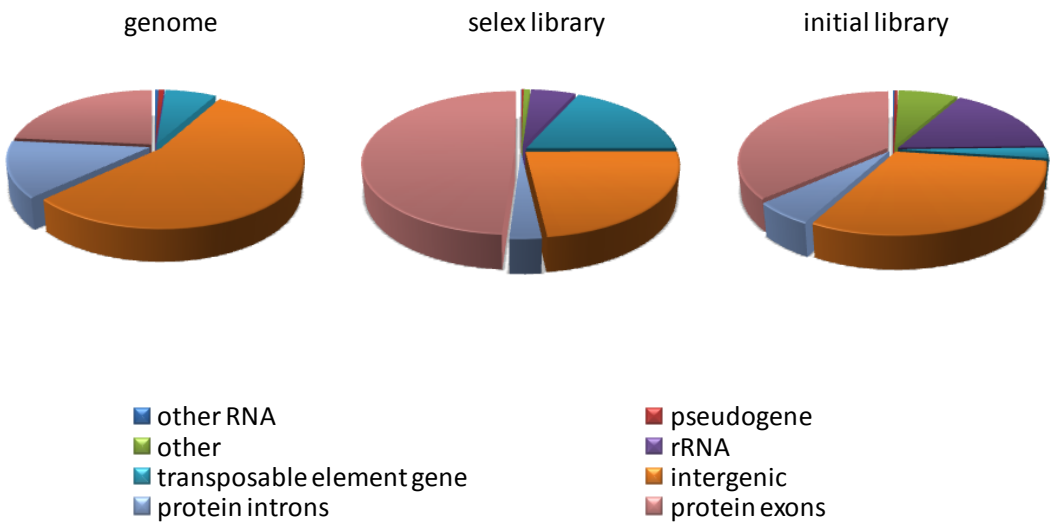

Figure 3.8: Abundance of various genomic elements in the libraries and genome of *A.Thaliana*. Genomic elements are indicated in colors and calculated as percentage from total number of annotated sequences in genome or total number of obtained sequences.

3.2.3 Analysis of the SELEX library

To find out the quality of obtained reads in the SELEX library we performed statistical analysis of the sequences. At the beginning, all 23.000 reads were treated to discard primer-adaptors sequences which were used to develop library and further during the selection procedure. As shown in the table 3.1 over 70% of the reads contained accurate adaptors from the both ends and hold long enough sequence in between primers (more than 18 nt). Next question we asked, how many reads from that trimmed sequences could be aligned back to the genome of *Arabidopsis thaliana*. As an analysis revealed over 90% of this reads were found in the genome of *Arabidopsis thaliana*. To determine repetitiveness of the sequencing reads we calculated how frequently they could be found in the genome. Almost all reads were mapped less than 5 times in the genome (table 3.1, figure 3.11) suggesting that sequencing reads belong to non-repetitive sequences in the genome of *Arabidopsis thaliana*.

Table 3.1 General results of the 454 sequencing.

Total number of reads	23186
Number of trimmed sequences from both ends (longer than 18 nt)	17654 (71%)
Number of reads mapped to the genome	15786 (89%)
Number of reads mapped to five or less loci in the genome	15690 (88%)

Next general question we were interested in was about length distribution of the sequence reads we obtained after 454 deep sequencing. There were few assumptions we made concerning possible length diversity. First, during the library development we selected sequences in range 100-700 bp after introducing randomized primer-adaptors. Second, during the selection procedure, where several round of PCR amplification were involved, there is a tendency to better amplify sequences of the small size. Finally, at the time when we sequenced SELEX library 454 technologies allowed to sequence reads in

length up to 200nt. Giving these assumptions, after removal of the primer-adaptor sequences we expected to find out reads in our SELEX library in the range of from 20 to 150 nt in length as revealed to be correct (shown in the figure 3.9).

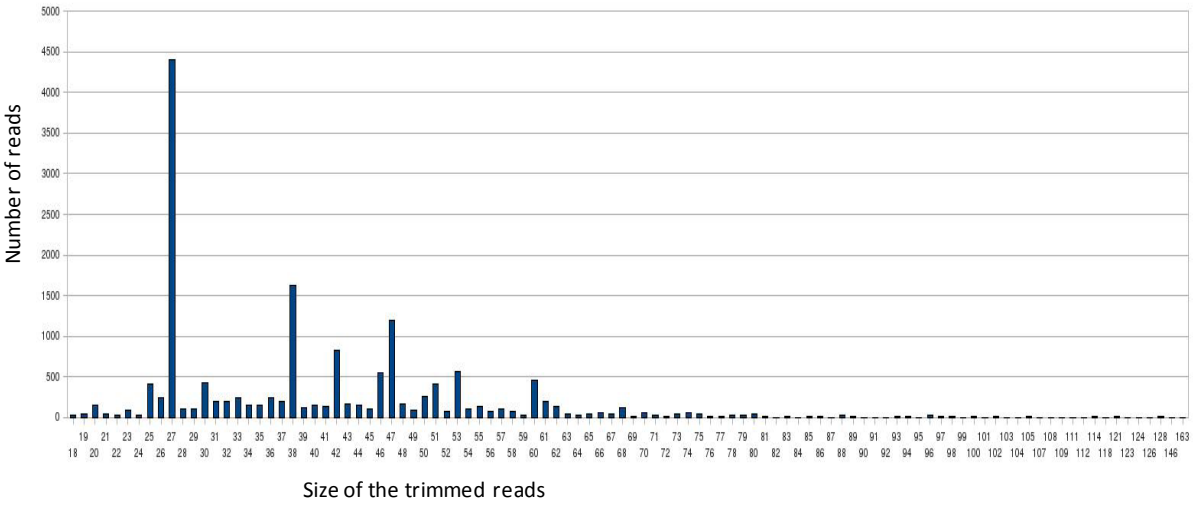

Figure 3.9: Size distribution of the sequenced reads from selected library with AtCyp59. Diagram displays dependence of number of sequenced reads obtained from their length.

As shown in the figure 3.9, over 25% of the obtained reads were 27nt in length after primer-adaptor sequences had been removed. This striking pre-determination by size led us to the question whether these sequences were distinguished from each other or belonged to particular class or gene. Further analysis of this cluster revealed that 70% of these sequences were mapped to two genes in the genome: YCF 2.2 from the chloroplast genome and AT1G15710 gene from nuclear genome. These two genes indeed represented 7% and 10.6% from all sequenced reads in the SELEX library respectively. To understand what made these sequences so abundant in the selected pool, we performed a minimal free energy prediction which is shown in the figure 3.10. It has been shown from previous structural studies of RRM-containing proteins (Allain, 2000) that they interact with single-stranded RNA and create stacking bounds between aromatic amino acids from the protein and heterocyclic residues from RNA molecules (Maris, 2005). This knowledge led to the assumption that potential binding site for the AtCyp59 could

correspond to the bulge structure in RNA. From the predicted structures exhibited in the figure 3.10 the most abundant sequences contain such potential structural elements.

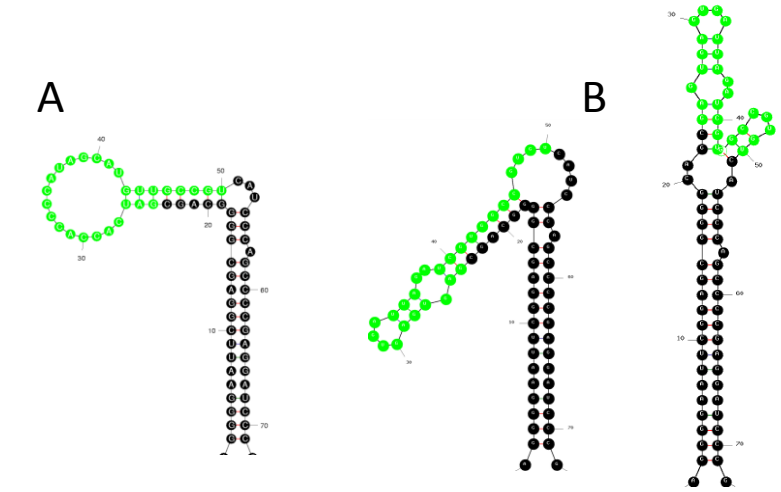

Figure 3.10: Minimum free energy structure prediction for the two most abundant reads. Local RNA structural prediction based on minimal thermodynamically Gibbs energy, in green shown mapped to the genome sequence and in black – primer sequences. The whole piece represents the sequence which was involved in the rounds of an affinity selection. A – YCF 2.2 annotated gene from the chloroplast genome to which belonged 7% of sequenced reads, B – AT1G1570 gene from nuclear genome to which belonged 10.6% of sequenced reads.

As mentioned above, over 90% of the reads mapped back to the genome were found in less than 5 loci in the *A. thaliana* genes. We further investigated how many reads could be mapped once, twice or more to the genome. As shown in the figure 3.11, more than 50% from sequenced reads could be found only once or twice in the Arabidopsis genome suggesting that they belong to the non-repetitive and gene-related elements in the genome. Double mapping also could potentially indicate sequences belonging to the homologous genes which emerged after several rounds of genome duplication.

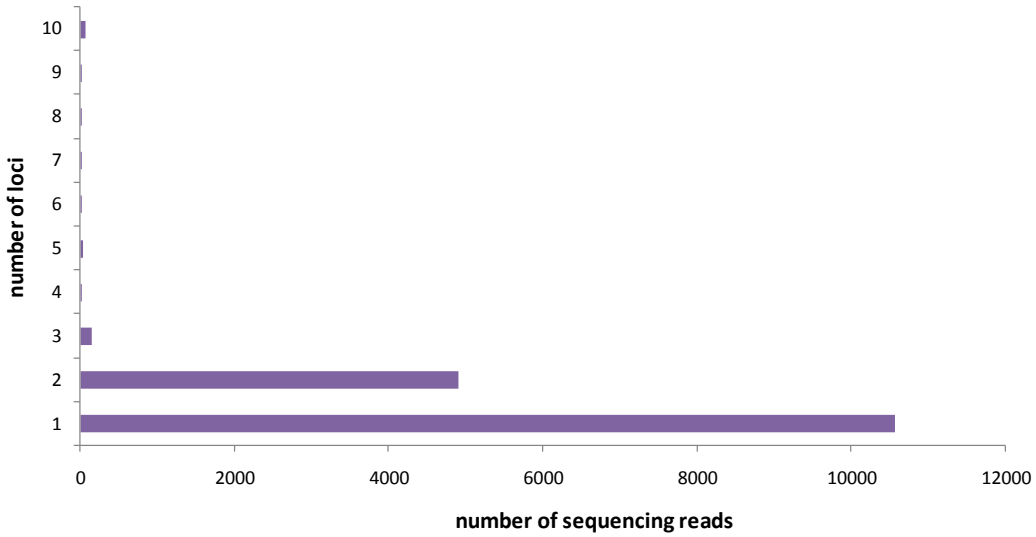

Figure 3.11 Sequencing reads distribution by mapping to the loci in the *Arabidopsis thaliana* genome. From over 20.000 sequenced reads almost 50% are showing single presence to particular location in the genome. Some could be found twice or more times suggesting possible affiliation to the duplicated genes.

Based on the preliminary analysis of the sequenced RNA pool selected with cyclophilin AtCyp59, several conclusions could be made. First, majority of obtained reads contained expected primer sequences at both ends which could be adequately assigned to the 5'- or 3'- end. Second, reads were predictably distributed by their length and could be aligned back to the genome. Finally, overall sequence distribution in the SELEX library tremendously varied from initial library and from genome. Sequence element distributions suggested a force driven selection towards restricted number of sequences showing high affinity to the RRM+Zn domain of the AtCyp59.

3.2.4 Orientation of genes found in the SELEX library.

Statistics examination of the library after selection revealed that substantial number of reads belongs to the protein coding genes, particularly to their exonic parts. To further investigate an abundance of the protein coding genes in overall sequences pool we decided to compare representation of sequences which belong to the protein coding

genes over other genomic elements such as intergenic regions and transposable elements as shown in the figure 3.12 A. Also after sequences had been mapped to the genome and assembled by genomic loci we compared abundance of contigs corresponded to the protein coding genes over other sorts of sequences in the number of genes as shown in the figure 3.12 B. Obtained results demonstrated that sequences of the protein coding genes are the main element in selected with AtCyp59 library.

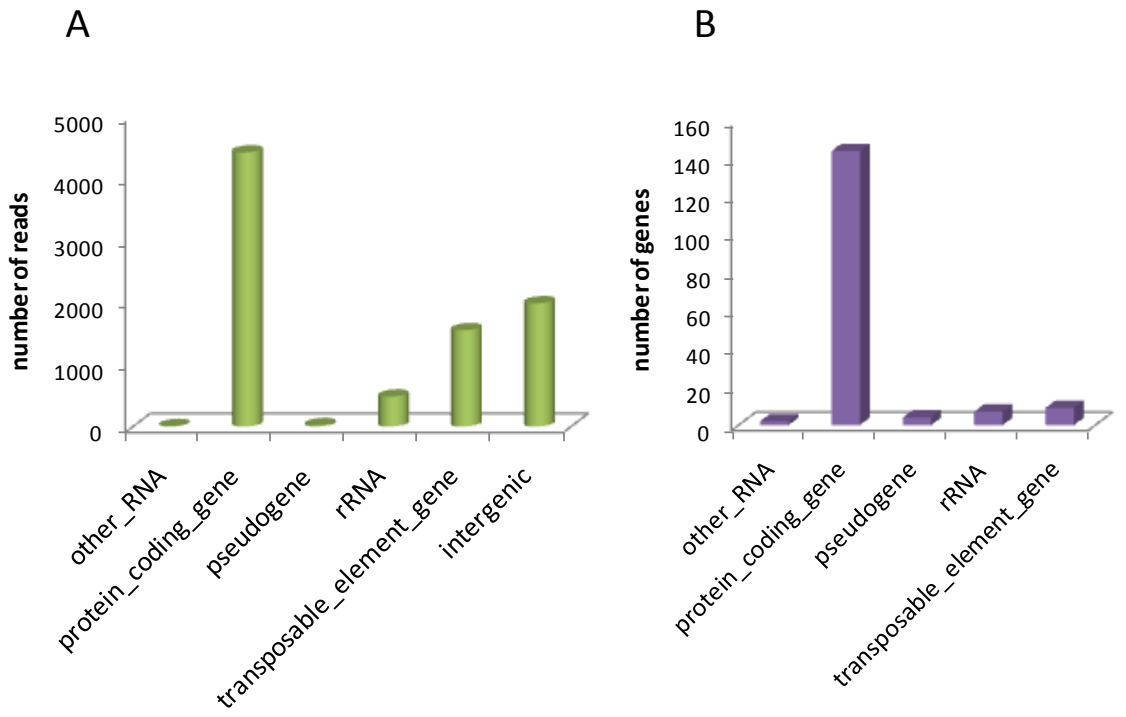

Figure 3.12 Mapped to the genome sequence elements composition. A – Representation of different genomic elements which had been revealed after selection with AtCyp59 in dependence on number of reads**. B –** Representation of different aligned contigs had been generated after selection with AtCyp59 in dependence on number of genes

As discussed previously, protein coding gene sequences in the SELEX library represented the most prominent batch of data suggesting that the affinity driven selection took place and main targets of RRM+Zn domain of the AtCyp59 lied in the mRNA sequences. We wanted to know whether all of the protein coding gene sequences had arisen from sense annotated gene strand. As exhibited in the figure 3.13 A those sequences were equally distributed in sense and antisense orientation toward annotated gene. This observation could be explained by the essence of genomic SELEX method in

which no preferences had been set up for the abundance of sequences towards transcribed genes because initial sequences were appeared from genomic DNA.

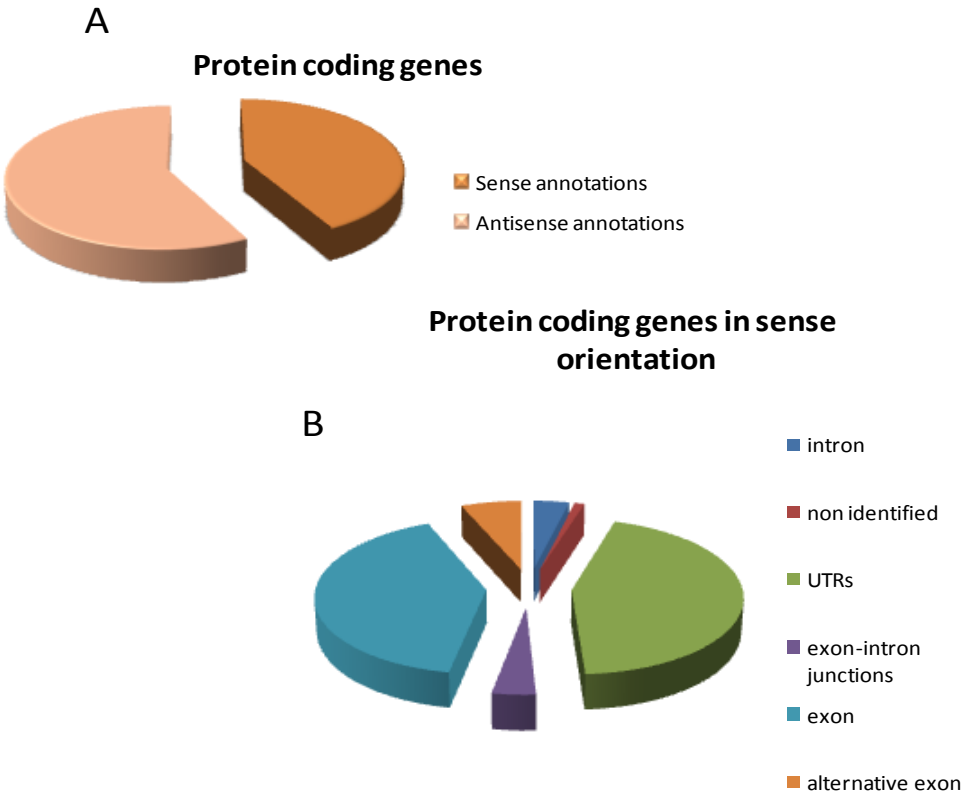

Figure 3.13 Genome annotations of the protein coding genes. A – Strand distribution according to the transcribed gene. **B** – Localization of the reads within transcribed gene in the sense orientation.

Next, we decided to look deeper into the pre-mRNA structure of the transcribed genes and aligned sequences corresponded to these genes into their structure. Average length of reads in selected pool accounted for 40-45nt whereas typical size distribution of the pre-mRNA genes in the *A. thaliana* ranged in >1000bp. As displayed in the figure 3.13 B correspondent sequenced reads were distributed predominantly over exons or untranslated regions within a pre-mRNA structure of transcribed genes.

Combining all revealed data together we concluded that potential RNA targets of the RRM+Zn domain of AtCyp59 were localized in the sequences belonging to the protein coding genes. These sequences were equally distributed in sense and antisense

orientation towards annotated genes in the *A. thaliana* genome. Sense oriented portions of sequences were localized in the exonic sequences or UTRs in the pre-mRNA structure.

3.2.5 Analysis of AtCyp59 binding motif.

General purpose of any Genomic SELEX experiment is an identification of the common binding motif within obtained sequencing pool for the protein of interest. To address this question we aligned back sequencing reads to the genome and sorted revealed data by the genomic loci and similarity (when sequences were appeared from duplicated genes) into contigs (clusters). Then, we arranged these clusters by number of reads presented in each cluster. This number represented depth of sequencing of each cluster-locus. To determine the common binding motif we took into consideration the first 10 most abundant clusters and subjected them to alignment with MEME suite program (Bailey, 1994). Obtained pattern shown in the figure 3.14 represents the predicted binding motif for the RRM+Zn domain of AtCyp59. This pattern is GC-rich which supported previous findings (Gullerova, 2006). It is well known that composition of the exons in the pre-mRNA in the *A. thaliana* is bias towards GC content (Carels, 2000). These findings further indicated that RNA binding partners of RRM+Zn domain of AtCyp59 lie within exonic regions of mRNAs.

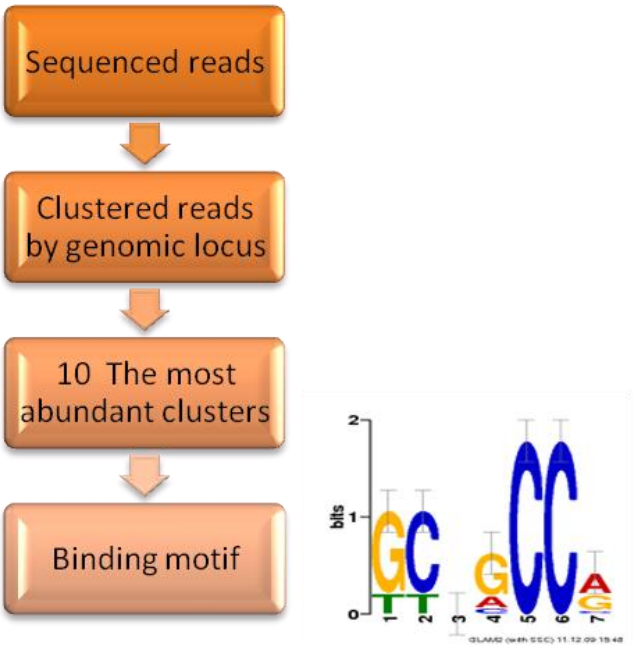

Figure 3.14 Scheme of the bioinformatics identification of the binding motif in the selected pool. Binding motif prediction was based on sequences alignment of the 10 most abundant clusters in analyzing SELEX library after several rounds of affinity selection with AtCyp59 using MEME suite program tool.

Predicted binding motif was identified using the first 10 most abundant sequencing clusters. To determine whether this motif represents whole SELEX library we calculated percentage of contigs or reads contained binding motif in the selected library. Results shown in the table 3.2 demonstrated that predicted binding pattern is found in 50% of the sequence reads which suggests possibility of existence of other potential binding motifs. Next, we asked whether this binding sequence was significantly meaningful. To answer this question we compared presence of the binding motif in the initial library to its abundance in the SELEX library. As displayed in the table 3.2 number of reads containing motif was significantly higher in the selected library in contrast to the initial pool. Presence of binding motif in the initial library could be explained by previous findings that initial library was predominantly biased towards GC-rich sequences. Summarizing these observations we could conclude that sequences containing predicted binding motif were gradually selected during Genomic SELEX experiment.

Table 3.2 Binding motif distribution within the libraries shown in number of reads and contigs.

	Absolute number of motifs	Normalized number of motifs/1000bp	Number of sequences with/without motif
SELEX library			
assembled unique contigs	61	2.62	54/413 (11.56%)
reads	9421	13	9261/8313 (52.9%)
Initial library			
assembled unique contigs	152	0.94	138/938 (12.83%)
reads	967	1.07	916/10266 (8.19%)

To identify how frequently binding motif could be found within sequence reads we aligned predicted pattern to the contigs of similar sequences selected during SELEX. Results exhibited in the figure 3.15 demonstrated that majority of sequences in the SELEX library possessed the binding motif only once or maximum three times. Thus binding pattern could serve as the signature of the particular sequence. And moreover, that RRM+Zn domain of the AtCyp59 interacts with potential RNA targets in sequence-specific manner.

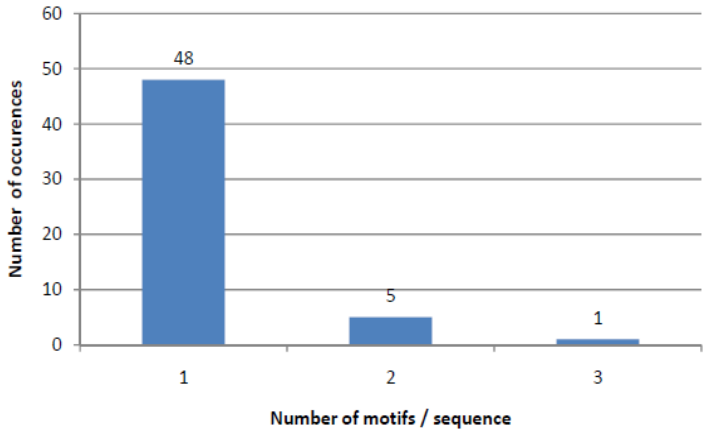

Figure 3.15 Distribution of the binding motif per sequence in the selected with AtCyp59 library. Diagram indicates frequency of occurrence of binding motif within unique sequenced reads (contigs).

Predicted binding pattern could be divided into several unique motifs because of possible nucleotide variations in certain positions within binding motif. To narrow down binding motif variations to the more defined pattern or subclass we differentiated each binding motif variant by its abundance in the SELEX library. Resulted diagram displayed in the figure 3.16 revealed that determined binding pattern could be re-written as GCWGCCG.

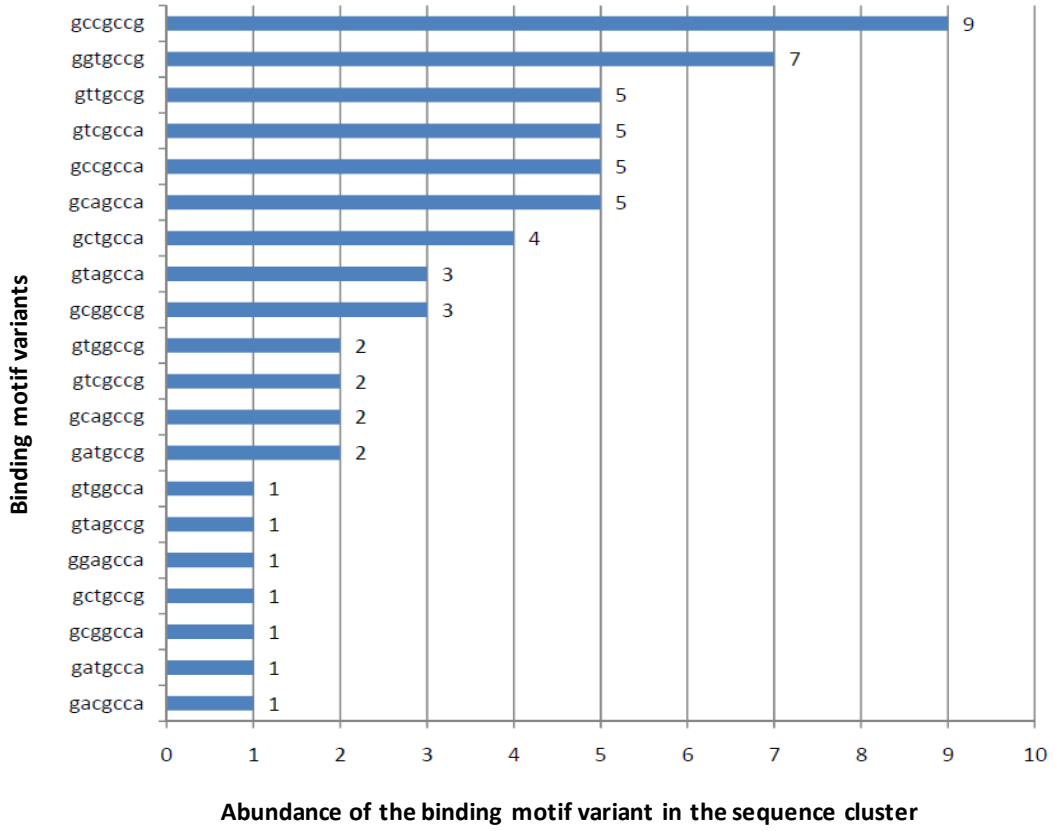

Figure 3.16 Frequency of occurrence of the binding motif variants in the SELEX library. Each bar us represented the number of unique assembled contigs each containing binding motif variant in the library selected with the RRM+Zn domain of AtCyp59.

3.2.6 Distribution of AtCyp59 binding motif in the genome.

Determined binding motif as shown above occurred in 50% of the sequenced reads from the library selected with AtCyp59 and was GC-rich. However, based on library development procedure, essence of the Genomic SELEX experiment, and number of reads

obtained after 454 deep sequencing we assumed that those reads could not cover the genome of the *A. thaliana*. Therefore, we decided to look back to the genome and check how abundant in the genome found binding motif is and whether this pattern keeps its localization within exonic sequences in the mRNA structure. To answer this question we aligned predicted motif over whole genome of *A. thaliana*. Results shown in the table 3.3 displayed that over 70% of mRNAs in the genome of *A. thaliana* contained binding motif from which over 20% contained the motif within exonic sequences and only 3% -within introns. Figure 3.17 exhibited the normalized per 1000 bp abundance of the binding motif in the mRNA structure. Depicted data suggest predominant localization of the binding pattern within exonic sequence on mRNA. These striking findings suggested that AtCyp59 could potentially interact with every mRNA in the genome and act as a general regulator. This tendency could circumstantially explain the absence of any T-DNA insertion mutant lines of AtCyp59 gene, because such mutation might disturb multiple processes in the living cell.

Table 3.3 Binding motif distribution in the *Arabidopsis thaliana* genome.

	Absolute number of motifs	Normalized number of motifs/1000bp	Number of sequences with/without motif
full mRNAs	41429	0.68	27720/11920 (69.93%)
exons	52197	0.86	44176/154044 (22.29%)
introns	5555	0.21	5349/143231 (3.37%)

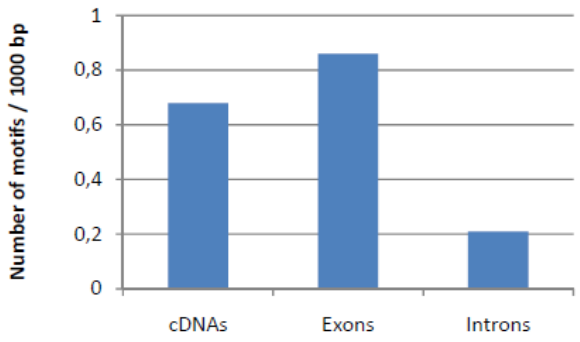

Figure 3.17 Distribution of the AtCyp59 binding motif in the *Arabidopsis thaliana* genome on messenger RNA per thousand base-pairs. Each bar represented the normalized per

1000 bp number of cDNAs, exons or introns containing binding motif in the genome of *Arabidopsis thaliana*.

To identify how often predicted binding motif could be found within each mRNA in the *A. thaliana* genome, we aligned this motif to the mRNA sequences in the genome. Results exhibited in the figure 3.18 displayed the similar trend of binding motif distribution as previously shown on figure 3.15. Majority of genes in the genome contain binding motif once or maximum three times on the gene-body which further proves sequence specificity of interaction between RRM+Zn domain of the AtCyp59 and its potential RNA targets. This fact supports that this motif might serve as mRNA signature in particular process or at particular stage of mRNA life.

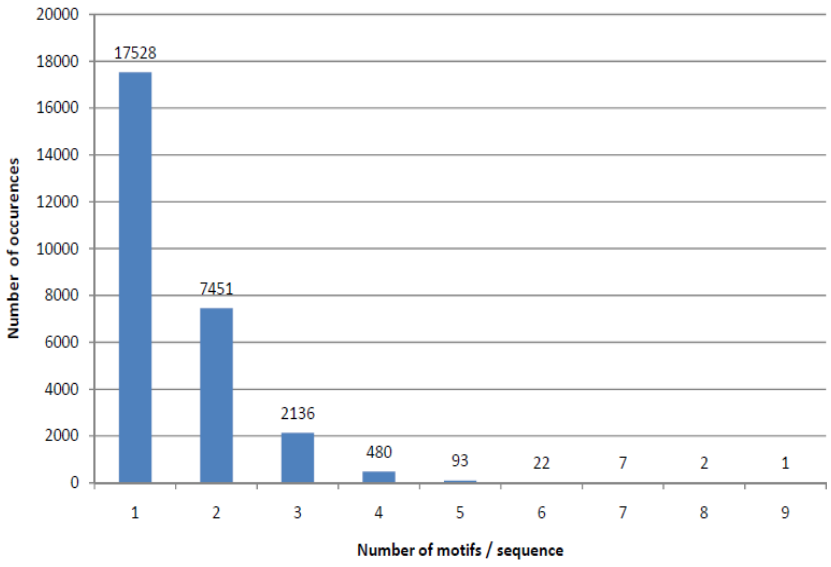

Figure 3.18 Distribution of the binding motif per sequence in the *Arabidopsis thaliana* genome. Each bar on the diagram indicated frequency of occurrence of binding motif within gene sequence in the genome of the *Arabidopsis thaliana*.

3.2.7 AtCyp59 binding motif localization near the future stop codon on the mRNAs in the genome.

As discovered above predicted binding motif occurred once per sequence in the majority of mRNAs in the *A. thaliana* genome. Also it is localized within exonic regions in

the pre-mRNA stricture. To determine whether binding motif had tendency to be present towards beginning, middle or end of the pre-mRNA gene we quantified number of binding motif occurred normalized to the certain position on the pre-mRNA. Results displayed in the figure 3.19 did not show any significant preferences of distribution of binding motif on overall position on pre-mRNA. However, when we correlated abundance of translation termination sites at normalized position on pre-mRNA to the number of binding motif we found considerable enrichment of occurrence of predicted motif towards the end of the pre-mRNAs.

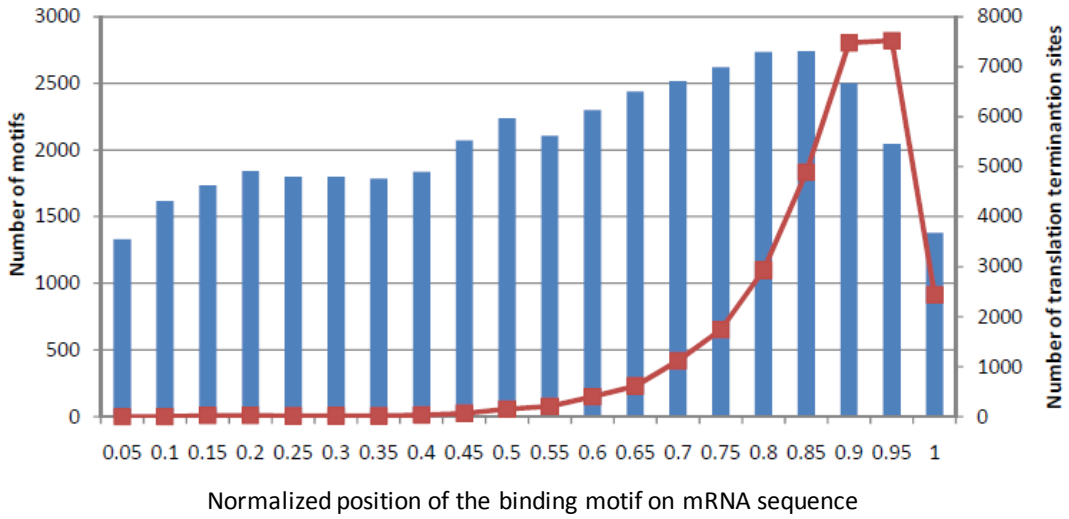

Figure 3.19 Binding motif distribution across mRNA sequence in the genome. On the X axis plotted normalized by one length of the pre-mRNA genes. On the right side – absolute number of the binding motif and on the left – absolute number of the translation termination sites. Each blue bar represented number of binding motifs which could be found within for example 0.05 from the beginning of the each mRNA gene. Red dots line represented number of translation termination sites which could be found within for example 0.05 from the beginning of the each mRNA gene.

To further investigate this phenomenon we plotted number of found binding motifs to the normalized position of the translation termination site. Results displayed in the figure 3.20 showed significantly high possibility to find binding motif within gene-body with increasing possibility to localize binding motif closer to the translation termination site and almost complete absence of predicted motif after translation stop codon.

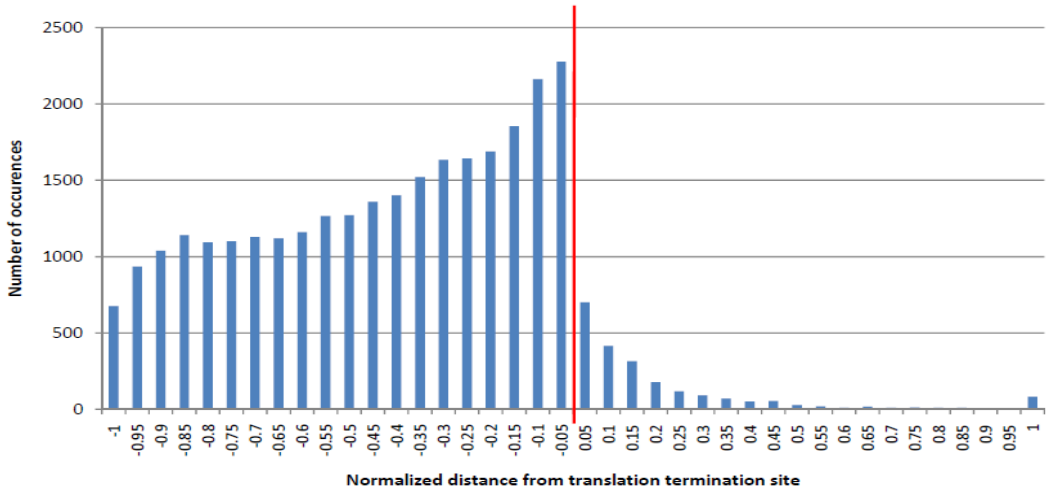

Figure 3.20 Position of the binding motif normalized to the translation stop in the *Arabidopsis thaliana* genome. On the X axis plotted normalized by one distance from the translations termination site on the mRNA genes. Each blue bar represented number of binding motifs which could be found at particular distance from the translational termination site on the mRNA gene.

Finally, combining latest bioinformatics observation we concluded that predicted binding motif occurred in over 70% pre-mRNA genes in *A. thaliana* genome which could potentially indicate function of the AtCyp59 as the general factor in the transcription and/or splicing. Genome-wide data supported localization of the binding motif within exonic sequences on gene-body structure which additionally sustained previous studies (Gullerova, 2006). It was also supported by correlation between high GC-conent of the binding motif and exonic sequences in the genome. Lastly, high possibility to find the binding motif near the translation termination site on the pre-mRNA gene structure raised assumption that binding of the AtCyp59 to mRNA could contribute to the 3'-end processing of the pre-mRNA. AtCyp59 interaction with mRNA might potentially serve as indicator of the end of the transcribing gene either for the Polymerase II or the spliceosome machinery.

3.3 *In vitro* binding studies of found RNA targets to the AtCyp59.

To investigate and verify binding affinity of AtCyp59 to selected targets we used recombinant full-length, RRM+ZN and mutated versions of GST-tagged AtCyp59 protein. Mutations in the RRM domain were designed to disturb three aromatic conserved amino acids which have been shown to be indispensable for RNA recognition (Y286D, F288D and F291D). Schematic pictures of the constructed vectors are shown in the figure 3.21 A. Proteins were expressed and purified using glutathione beads (coomassie blue stained gel is depicted on the figure 3.21 B). Before utilization of these proteins in the EMSA analysis they all were dialyzed into appropriate binding buffer.

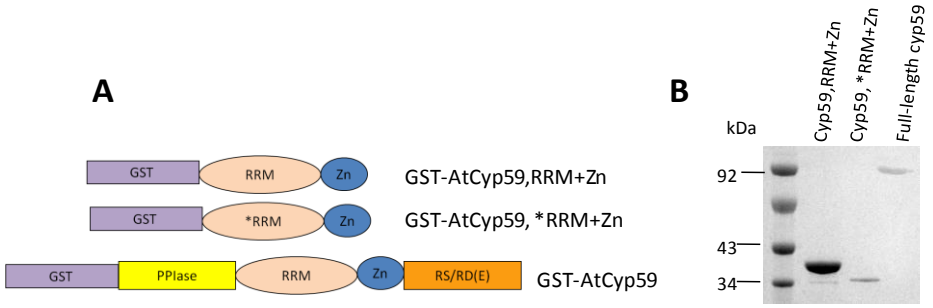

Figure 3.21 Protein expression panel. A – Schematic representation of variants of the AtCyp59 used in the in vitro studies. **B –** Commasie gel staining of the GST-tagged purified proteins. Molecular weight marker is shown on the left side**.**

3.3.1 RRM-Zn domain of the Atcyp59 binding studies to 7nt variants of the binding motif.

Intensive bioinformatics studies on SELEX library revealed a common binding motif which could be written as G[NC]NGCCW. This binding consensus as shown above is significantly abundant in the library selected with RRM+Zn domain of the AtCyp59 and presumably assembles with protein-coding genes. In the genome of *A. thaliana* this motif represents over 70% of the mRNA genes and is localized within exons. Moreover, as it

was discovered, presence of this motif increases towards translation termination site but not within UTRs. Also, as it was mentioned above, this binding consensus could be divided into 16 single motifs each associated with specific contig in the SELEX library. To verify binding affinity of the RRM+Zn domain of AtCyp59 to the binding motif we checked each motif variant using electro mobility gel-shift assay (EMSA). For that purpose we took 7nt synthesized RNA representing each binding motif variants and incubated it with variable concentration of the recombinant protein of AtCyp59 (RRM+Zn domain) dialyzed in the binding buffer. Then we separated resulted binding reactions on the native 10% polyacrylamide gel in 0.5X TBE buffer. As shown in the figure 3.22 different motif variants displayed quite diverse affinity to the RRM+Zn domain of the AtCyp59. For instance, exhibit 3.22A and 3.22D shows the best binding affinity to the protein with approximated constant of dissociation (K_d) of as little as 40nM, whereas other RNAs showed moderate affinity to the protein (figure 3.22 B) or even weak affinity (3.22 C).

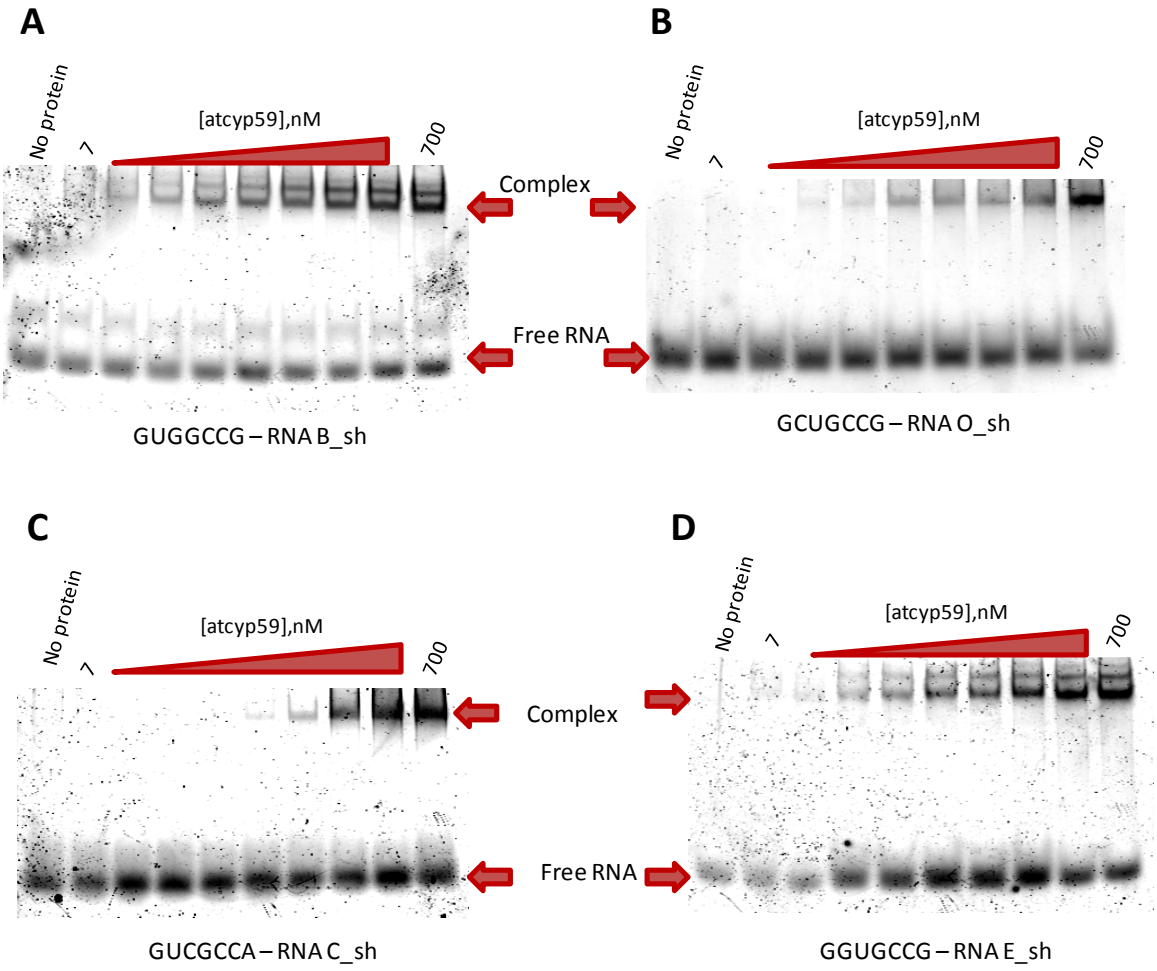

Figure 3.22 Gel Shift assay with RRM+Zn domain of AtCyp59 and 7nt variants of the binding motif. SYBR GREEN II stain of the 10% native polyacrylamide gel electrophoresis of the 7nt RNA derived from 16 possible binding consensus variants incubated with increasing concentration of the RRM+Zn domain of the AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. Free RNA and RNA-protein complex formation is indicated with arrow. **A,D** – examples of the high affinity binding between RNA and the protein with approximated K_d = 40nM; **B** - example of the moderate affinity binding between RNA and the protein with approximated K_d = 105nM; **C** - example of the low affinity binding between RNA and the protein with approximated K_d = 300nM.

Results of the *in vitro* binding experiments with all possible 7nt RNA variants representing binding consensus are summarized in the table 3.4. K_D's vary from 40 to 300nM range suggesting high order of sequence specificity of the RRM domain of protein to the RNA sequence. As it could be seen from the first three RNA candidates in the table 3.4 RRM domain of the AtCyp59 recognized the seed sequence of GNNGCCW with higher affinity if

the last nucleotide is G and the second is G or U, which leads to the revised sequence of the consensus GG(U)NGCCG. Interestingly, we did not find strong correlation of the abundance of these variants in the SELEX library with its binding efficiency to the protein, however some of the most abundant motif subclasses appeared on the top of the table 3.4.

Table 3.4 Range of dissociation constants of 7nt variants of the binding motif, based on gel shift assay.

Name	Sequence	Abundance in the selected contigs	K_d [nM]
RNA E_sh	GGUGCCG	7	40±10
RNA B_sh	GUGGCCG	2	40±10
RNA G_sh	GUAGCCG	1	40±10
RNA F_sh	GCAGCCA	5	105±25
RNA O_sh	GCUGCCG	1	105±25
RNA T_sh	GAUGCCA	1	105±25
RNA P_sh	GACGCCA	1	105±25
RNA H_sh	GCCGCCG	9	200±35
RNA A_sh	GUUGCCG	5	200±35
RNA K_sh	GUAGCCA	3	200±35
RNA M_sh	GCGGCCG	3	200±35
RNA C_sh	GUCGCCA	5	300±40
RNA D_sh	GCCGCCA	5	300±40
RNA R_sh	GAUGCCG	2	300±40
RNA L_sh	GUGGCCA	1	300±40
RNA N_sh	GGAGCCA	1	300±40
RNA S_sh	GCGGCCA	1	300±40

3.3.2 Leveling effect of full-length protein on weak and strong binders of the binding motif variants.

Next, we were interested in whether other domains of AtCyp59 of the protein could contribute to the state of interaction between RNA and protein.

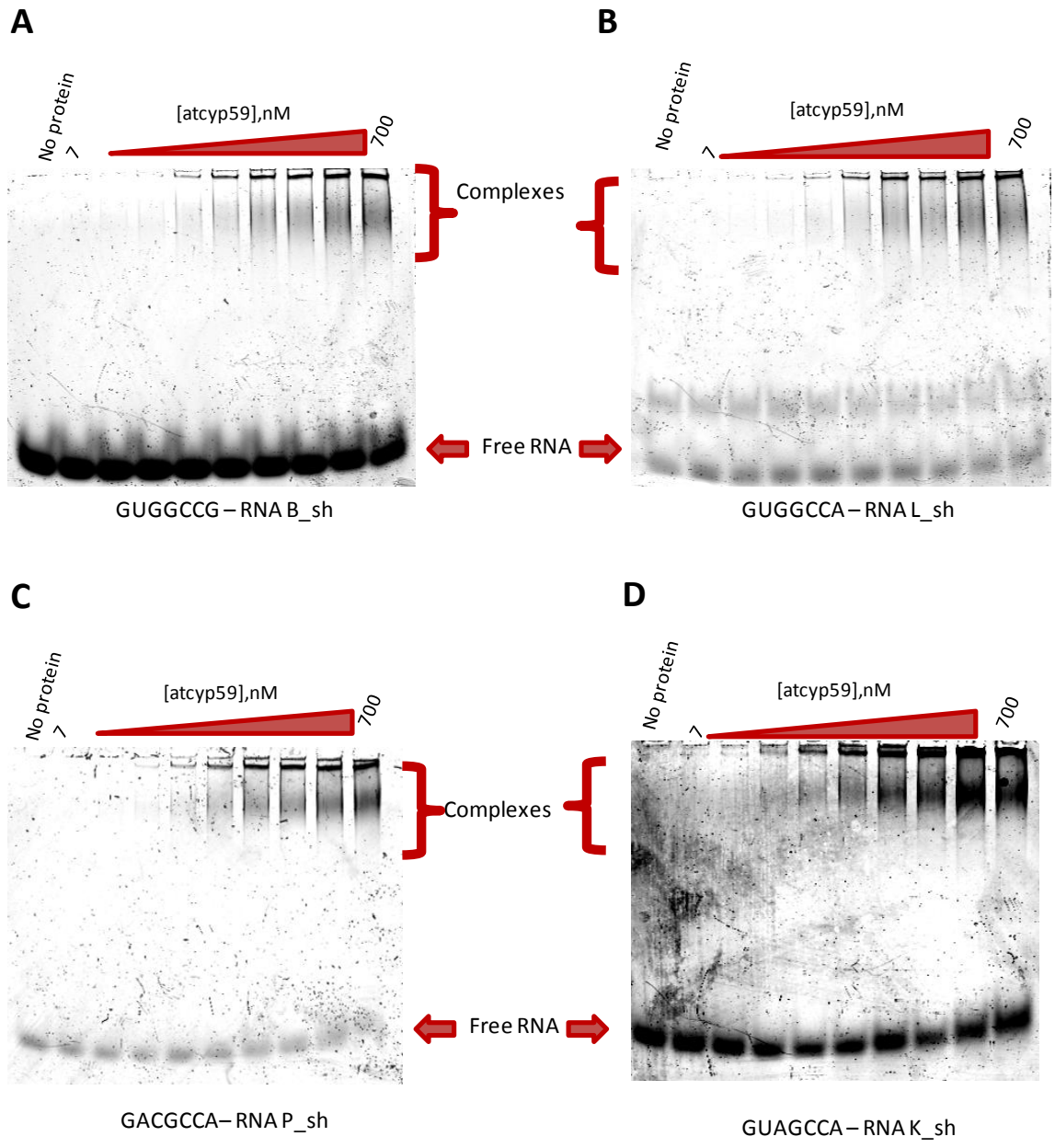

Figure 3.23 Gel shift assay of the binding reaction between full-length protein AtCyp59 and 7 nt RNA binding motif variants. SYBR GREEN II stain of the 10% native polyarylamide gel electrophoresis of the 7nt RNA derived from 16 possible binding consensus variants incubated with increasing concentration of the full-length AtCyp59. Concentration of the protein is shown

above in nM range. First line in every gel is the line where protein was not added. Free RNA and RNA-protein complex formation is indicated with arrow. **A-D** – example of the equal affinity binding between RNA and the protein with approximated K_D = 120nM.

For that reason, we performed the same screen of 8 from 16 (two from the each group showing different binding specificity) of 7nt RNA variants of the binding consensus and employing full-length AtCyp59. Again each RNA variant was incubated with increasing concentration of the protein and then resolved on the 10% native polyacrylamide gel. Examples of the analysis are shown on the figure 3.23 (A-D). As displayed in this exhibit, surprisingly all tested RNA variants showed similar binding affinity to the full-length protein which could be approximated using this method to the 120nM. Summarized results are presented in the table 3.5. As mentioned above, AtCyp59 consists of three major parts: PPIase catalytically active domain, RRM – RNA-binding domain and C-terminal positively charged domain. Both N- and C-terminal domains are responsible for the protein-protein interactions and had been found not to participate directly in the RNA binding (Gullerova, 2006). However, these experiments suggested that initial recognition of the RNA sequences carried out by the RRM+Zn domain of the AtCyp59 might be accompanied by C- and N-terminal domains. Also this recognition is highly sequence-specific and RRM+Zn domain requires GC-rich sequences for the efficient binding. Moreover, alterations in binding affinity between short and full-length protein clearly indicated that N- and C-terminal domains of the AtCyp59 participated in binding and tuned it down to middle binding affinity (table 3.4). One of the proposed hypotheses could be that other domains of the cyclophilin AtCyp59 are required for the recognition of the local pre-mRNA structure. Fact that very strong binders started to exhibit moderate affinity could potentially indicate that in the *in vivo* situation binding between cyclophilin and RNA might be flexible and reversible, probably during short period of time. So-called “leveling” of binding within binding consensus indicated that majority of mRNA in genome could bind AtCyp59 as it was shown by bioinformatics analysis and supported the theory that cycplophlin could play role of general regulatory factor in *A. thaliana*.

Table 3.5 Comparative dissociation constants between RRM+Zn and full-length protein atCyp59.

Name	Sequence	Abundance in the selected contigs	K_d [nM]
RRM+Zn domain of the AtCyp59			
RNA B_sh	GUGGCCG	2	40±10
RNA E_sh	GGUGCCG	7	40±10
RNA T_sh	GAUGCCA	1	105±25
RNA P_sh	GACGCCA	1	105±25
RNA K_sh	GUAGCCA	3	200±35
RNA M_sh	GCGGCCG	3	200±35
RNA L_sh	GUGGCCA	1	300±40
RNA S_sh	GCGGCCA	1	300±40
Full-length AtCyp59			
RNA B_sh	GUGGCCG	2	120±25
RNA E_sh	GGUGCCG	7	120±25
RNA T_sh	GAUGCCA	1	120±25
RNA P_sh	GACGCCA	1	120±25
RNA K_sh	GUAGCCA	3	120±25
RNA M_sh	GCGGCCG	3	120±25
RNA L_sh	GUGGCCA	1	120±25
RNA S_sh	GCGGCCA	1	120±25

3.3.3 Decrease in binding affinity of RRM+Zn domain to the RNA binding motif upon mutations in conservative aromatic amino acids

Previous results showed that RRM+Zn domain of the AtCyp59 recognizes target RNA in the sequence specific manner. EMSAs using 7nt-long RNAs compromising consensus binding sequences identified in the Genomic SELEX experiment revealed different affinities with which RRM+Zn domain and full-length AtCyp59 bind to their

target RNAs. To outline the contribution of RRM domain into RNA-binding function of AtCyp59, we mutated three conserved residues which are important for RNA binding function of RRM (Wang, 2008).

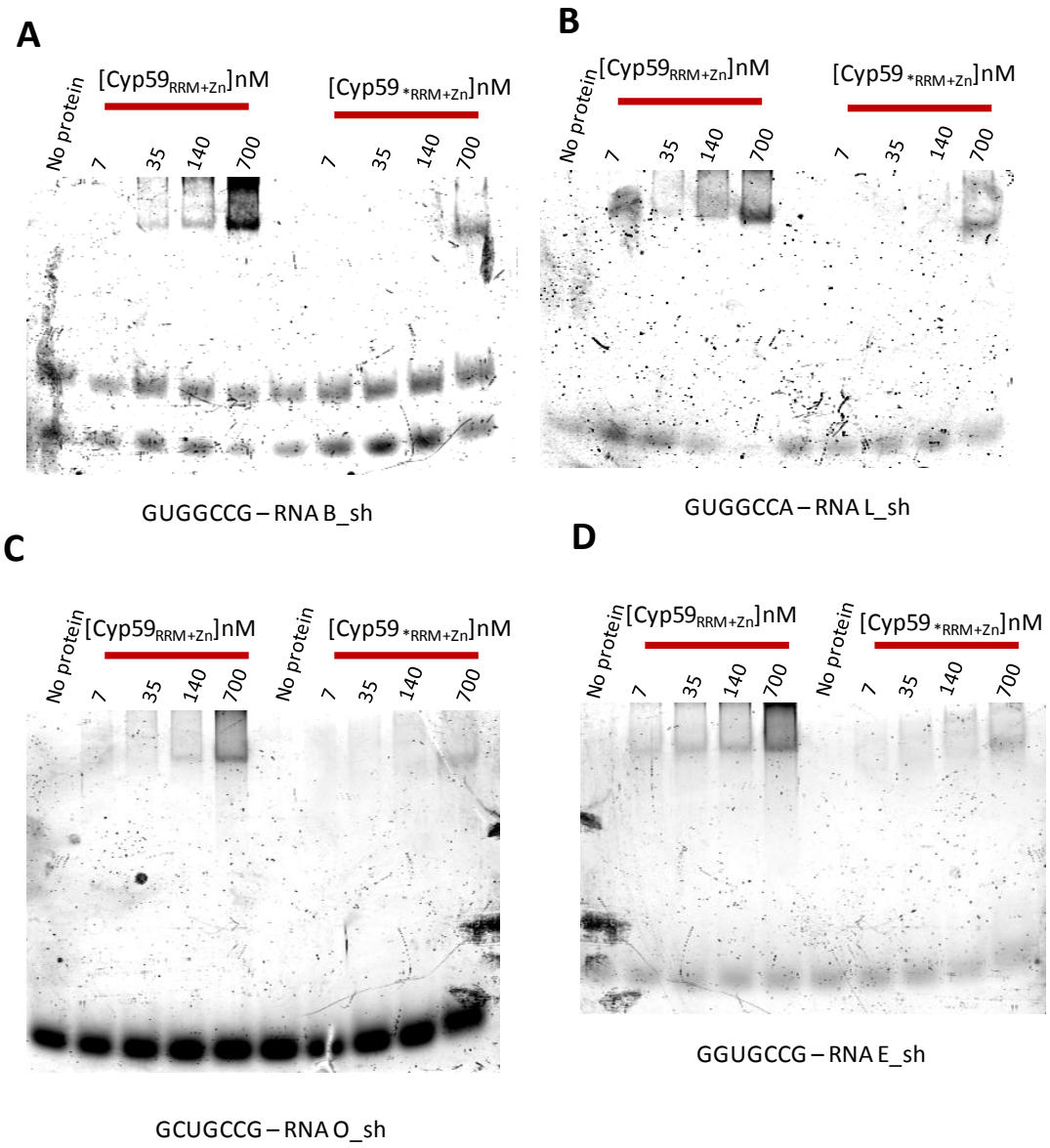

Figure 3.24 Gel shift assay of the binding reaction between RRM+Zn domain and mutated *RRM+Zn domain of AtCyp59 and 7 nt RNA binding motif variants. SYBR GREEN II stain of the 10% native polyacrylamide gel electrophoresis of the 7nt RNA derived from 16 possible binding consensus variants incubated with increasing concentration of the RRM+Zn domain of the AtCyp59 or *RRM+Zn domain of AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. RNA sequence used in the assay as well as short name is indicated below each gel.

Recombinant mutant version of RRM+Zn domain (*RRM+Zn) was purified and dialyzed into binding buffer. Using the same EMSA assay we compared binding efficiency of normal and mutated protein to the 7nt RNA variants representing binding consensus pattern. As shown in the figure 3.24 A-D, disturbance of three conserved aromatic residues in the RRM domain of the AtCyp59 significantly decreased binding to the target RNAs. In this assay we tested one RNA variant from each four groups which showed differential affinity to the RRM+Zn domain of AtCyp59. Further details with approximated from this experiment K_D's are shown in the table 3.6.

Table 3.6 Dissociation constants of mutated RRM-Zn domain of protein atCyp59 to the 7 of variants of the binding motif.

Name	Sequence	Abundance in the selected contigs	K_d [nM]
*RRM+Zn domain of the AtCyp59			
RNA B_sh	GUGGCCG	2	>700
RNA L_sh	GUGGCCA	1	>700
RNA O_sh	GCUGCCG	2	>700
RNA E_sh	GGUGCCG	7	>700

As displayed in the table 3.6 all tested 7nt RNAs interacted with mutated protein in unspecific manner suggesting that mutations of essential for RNA-binding activity of RRM domain amino acids were crucial for sequence-specific recognition of RNAs and required for formation of the RNA-protein complexes.

3.3.4 AtCyp59 binds to the RNA sequence specifically.

To further investigate requirements for an interaction between protein and its RNA targets, and test the specificity of RNA recognition of AtCyp59, we performed gel-shift assay using RNA sequence (scramle) which was not neither predicted as binding

motif nor found within sequenced pool after Genomic SELEX experiment. In this binding reaction we also used wild type and mutated version of the RRM+Zn domain of AtCyp59.

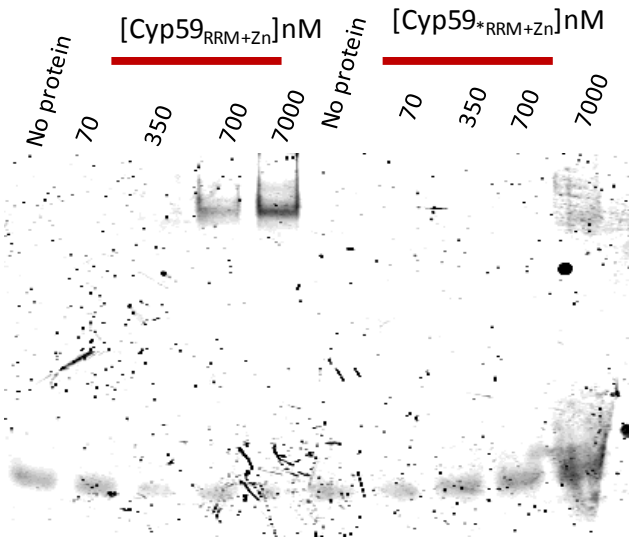

Figure 3.25 Gel shift assay. SYBR GREEN II stain of the 10% native polyacrylamide gel electrophoresis of the 7nt RNA unspecific RNA incubated with increasing concentration of the RRM+Zn domain of the AtCyp59 or *RRM+Zn domain of AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. RNA sequence used in the assay as well as short name is indicated below the gel.

As shown in the figure 3.25 scrambled RNA interacted with both proteins extremely unspecific with approximated K_d was lower than 7μM (table 3.7). WT RRM domain of the AtCyp59 had affinity to the non-specific RNA in the same range as if RRM domain was mutated. And mutated version of protein showed even lower affinity to the scrambled RNA.

Table 3.7 Dissociation constants of non-specific 7nt RNA-oligo to the RRM+Zn and *RRM+Zn domain of the AtCyp59.

Name	Sequence	K_d [μM]
RRM+Zn domain of the AtCyp59		
RNA SC	UAAUUUU	0.7±0.1
*RRM+Zn domain of the AtCyp59		
RNA SC	UAAUUUU	>7

Discovered by genomic SELEX common binding RNA consensus sequence was verified to bind to AtCyp59 *in vitro* using 7nt RNA fragments. Site-directed mutagenesis studies revealed that RRM+Zn domain is responsible for the initial RNA sequence recognition upon binding to particular RNA piece because mutations in the essential aromatic aminoacids in the RRM domain resulted in significant decrease in binding affinity and abolished binding of RNA to the protein. It is noteworthy that mutations in the RNA motif also decreased binding of both WT and mutated versions of AtCyp59 to that RNA sequence suggesting AtCyp59 interacts with RNA in the sequence-specific manner.

3.3.5 Differential binding of RRM+Zn domain of AtCyp59 to mRNAs containing binding sequence variants.

As it was previously described (figure 3.22), diverse common binding motif variants of 7nt-long RNAs have differential affinity to the RRM+Zn domain of the ATCyp59. However, during the Genomic SELEX experiment these motif variants were encased into longer RNA pieces involved in several round of selection with RRM+Zn domain of AtCyp59. To investigate possible contribution of the local mRNA structure in binding to the protein, we performed EMSA with RNA fragments as they were sequenced

and used in the selection process (each of these fragments contains one particular motif variant per sequence).

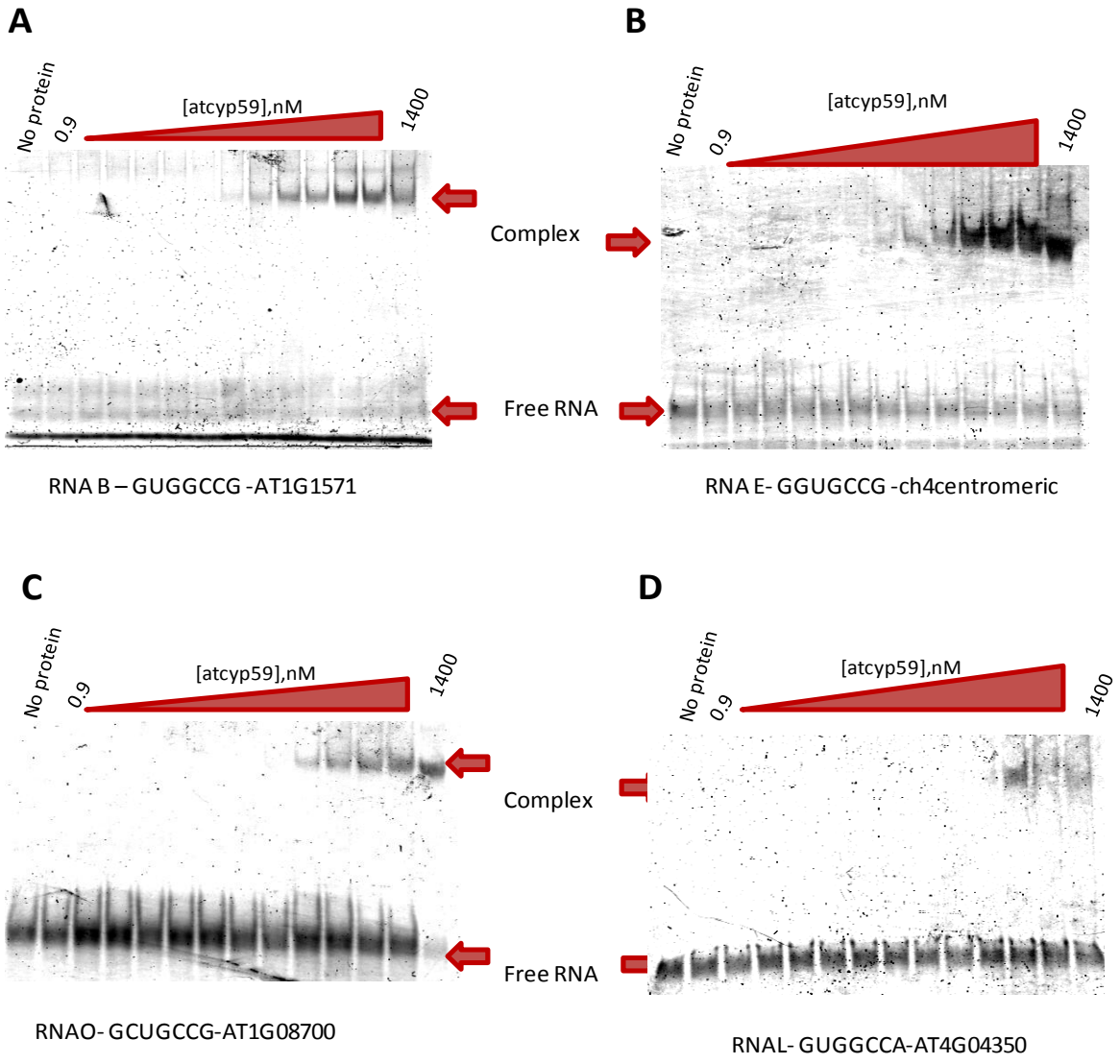

Figure 3.26 Gel shift assay. SYBR GREEN II stain of the 8% native polyacrilamide gel electrophoresis of the approx. 30nt RNA containing one of the 16 possible binding consensus variants incubated with increasing concentration of the RRM+Zn domain of the AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. Free RNA and RNA-protein complex formation is indicated with arrow. **A,B** – example of the high affinity binding between RNA and the protein with approximated K_d = 180nM; **C** - example of the moderate affinity binding between RNA and the protein with approximated K_d = 390nM; **D** - example of the low affinity binding between RNA and the protein

with approximated K_d = 800nM.The binding reaction is between RRM+Zn domain of AtCyp59 and selected pieces of RNA containing the variants of binding motif.

Examples of *in vitro* interaction between 30nt RNA fragments and *WT* RRM+Zn domain of AtCyp59 (figure 3.26) indicated that, similarly to the previous data, RRM+Zn domain of the AtCyp59 differentially binds to 30nt-long RNAs as to 7nt-long RNAs.

Table 3.8 Dissociation constants of 30nt RNA selected and sequenced with 454 deep sequencing, containing binding motif to the RRM+Zn domain of the AtCyp59.

Name	Sequence of motif	Gene name	K_d [nM]
RNA E	GGUGCCG	ch4centromeric	180±50
RNA B	GUGGCCG	AT1G1571	180±50
RNA G	GUAGCCG	CHrM	180±50
RNA O	GCUGCCG	AT1G08700	390±70
RNA F	GCAGCCA	AT2G40570	390±70
RNA T	GAUGCCA	AT5G35057	220±60
RNA P	GACGCCA	AT3G32377	220±60
RNA H	GCCGCCG	AT1G77850	600±75
RNA A	GUUGCCG	AT3G28430	600±75
RNA K	GUAGCCA	AT3G49400	600±75
RNA M	GCGGCCG	AT2G07749	600±75
RNA D	GCCGCCA	AT4G13340	800±80
RNA C	GUCGCCA	AT3G19430	800±80
RNA L	GUGGCCA	AT4G04350	800±80
RNA N	GGAGCCA	AT1G08700	>1400
RNA R	GAUGCCG	ch3centromeric	>1400

We observed several groups or subclasses of the 30nt-long RNA-variants that exhibited high (figure 3.26 A, B), middle (figure 3.26 C), low (figure 3.26 D) or almost no affinity to the protein. Interestingly, this graduation was completely similar to the one we

observed previously with 7-nt long RNA variants. This finding suggested that principal RNA sequence recognition undertaken by RRM+Zn domain of the AtCyp 59 does not depend on RNA length. Summarized analysis of the all 16 variants of the common binding sequence is shown in the table 3.8. However, overall affinity of RRM+Zn domain of AtCyp59 to the 30nt-long RNA fragments was decreased compare to the 7nt-long RNA pieces. It seems that local RNA structure might influence interaction between RNA and the protein and alter overall binding efficiency. However, general tendency in differentiating between variable nucleotide preferences remained the same as it was observed earlier.

3.3.6 Full-length protein promotes binding to the long RNA targets

As it is described above, longer RNA fragments showed lower affinity to the RRM+Zn domain of AtCyp59 *in vitro*. To determine whether full-length protein AtCyp59 with N- and C-terminus domains contributes to the binding to longer fragments of the RNA, we performed EMSA analysis with two from each four subgroups of differentiated affinity RNA sequences (Table 3.8). Results shown in the figure 3.27 demonstrated that binding affinity of the several RNAs had been changed and improved upon interaction with full-length protein. Generally, approximated K_D were significantly higher upon binding to full version of the AtCyp59 for all tested variants of the RNA fragment containing one of the variant of the common binding motif. Most prominent improvement showed RNAs which before exhibited very low binding to the shorter version of the protein. Only slight increase in the binding efficiency was detected for the top three RNA fragments which previously showed highest affinity in 7nt-long form. Summarized results are presented in the table 3.9.

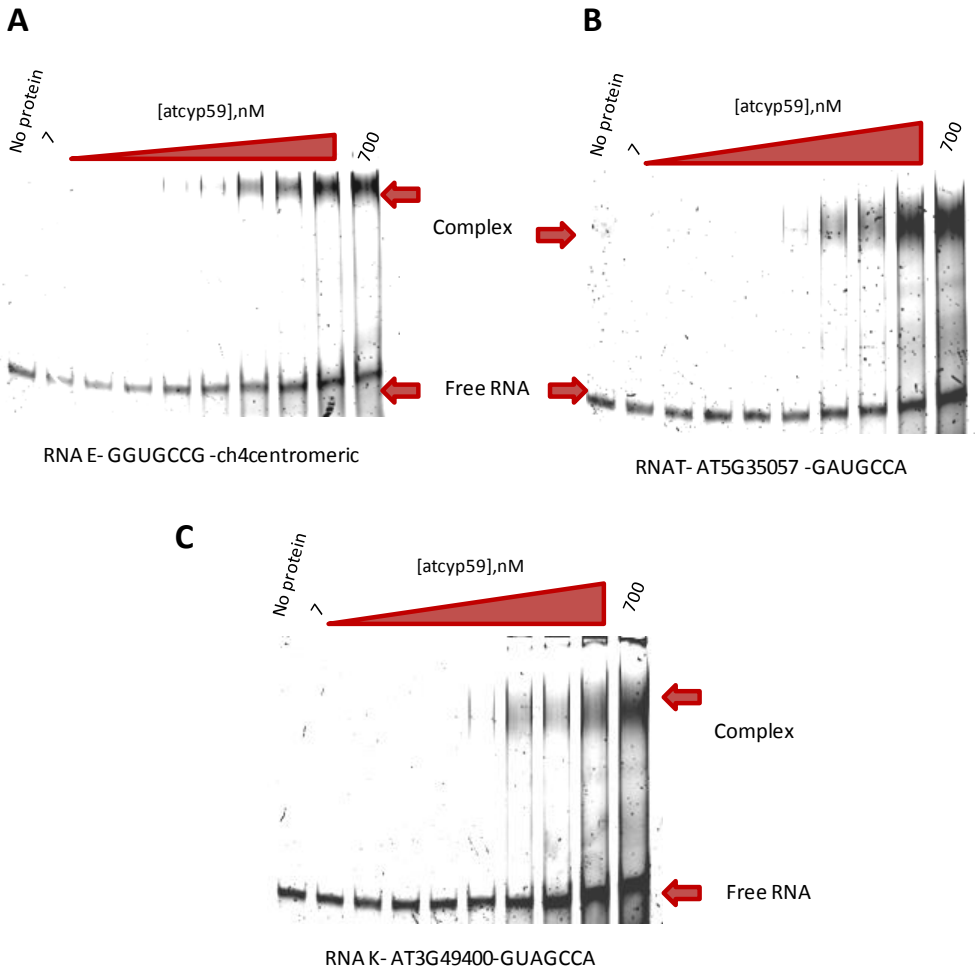

Figure 3.27 Gel shift assay. SYBR GREEN II stain of the 8% native polyacrilamide gel electrophoresis of the approx. 30nt RNA containing one of the 16 possible binding consensus variants incubated with increasing concentration of full-length version AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. Free RNA and RNA-protein complex formation is indicated with arrow. **A-D** – example of the high affinity binding between RNA and the protein with approximated K_d = 120-230nM.

As displayed in the table 3.9 dissociation constants of interaction between full-length protein and long RNA fragments are varied between 120-230nM. This results excelled from one obtained with 7 nt RNA sequences which showed complete equivalency. Obtained data suggested that local RNA structure interferes with RNA sequence recognition upon protein binding and full-length protein helps to grade with obstacles by contributing into binding efficiency. However, other AtCyp59 domains could

not fully equalize binding to long RNA fragment which might suggest that *in vivo* other protein factors should be taken into consideration.

Table 3.9 Dissociation constants of 30nt RNA selected and sequenced with 454 deep sequencing, containing binding motif to the RRM+Zn domain of the AtCyp59.

Name	Sequence of motif	Gene name	K_d [nM]
RNA E	GGUGCCG	ch4centromeric	110±50
RNA B	GUGGCCG	AT1G1571	110±50
RNA O	GCUGCCG	AT1G08700	140±70
RNA T	GAUGCCA	AT5G35057	110±60
RNA K	GUAGCCA	AT3G49400	180±75
RNA M	GCGGCCG	AT2G07749	180±75
RNA D	GCCGCCA	AT4G13340	230±80
RNA R	GAUGCCG	ch3centromeric	230±80

3.3.7 Mutations in the RRM domain of AtCyp59 decrease binding efficiency to the target mRNAs

As it was previously discussed (figure 3.24), mutations in the RRM domain of the AtCyp59 significantly decreased its binding to the RNA sequences and in some cases completely prevented interaction. To indentify whether this trend is still present using longer RNA fragments, we decided to perform EMSA analysis using few examples from different sub-classes of 30 nt-long RNAs. We compared binding efficiency between WT and mutated RRM domain of the AtCyp59. Selected RNA fragments were incubated with increasing concentration of the one of the protein and then resolved in the 8% native polyacrylamide gel. Results are shown in the figure 3.28.

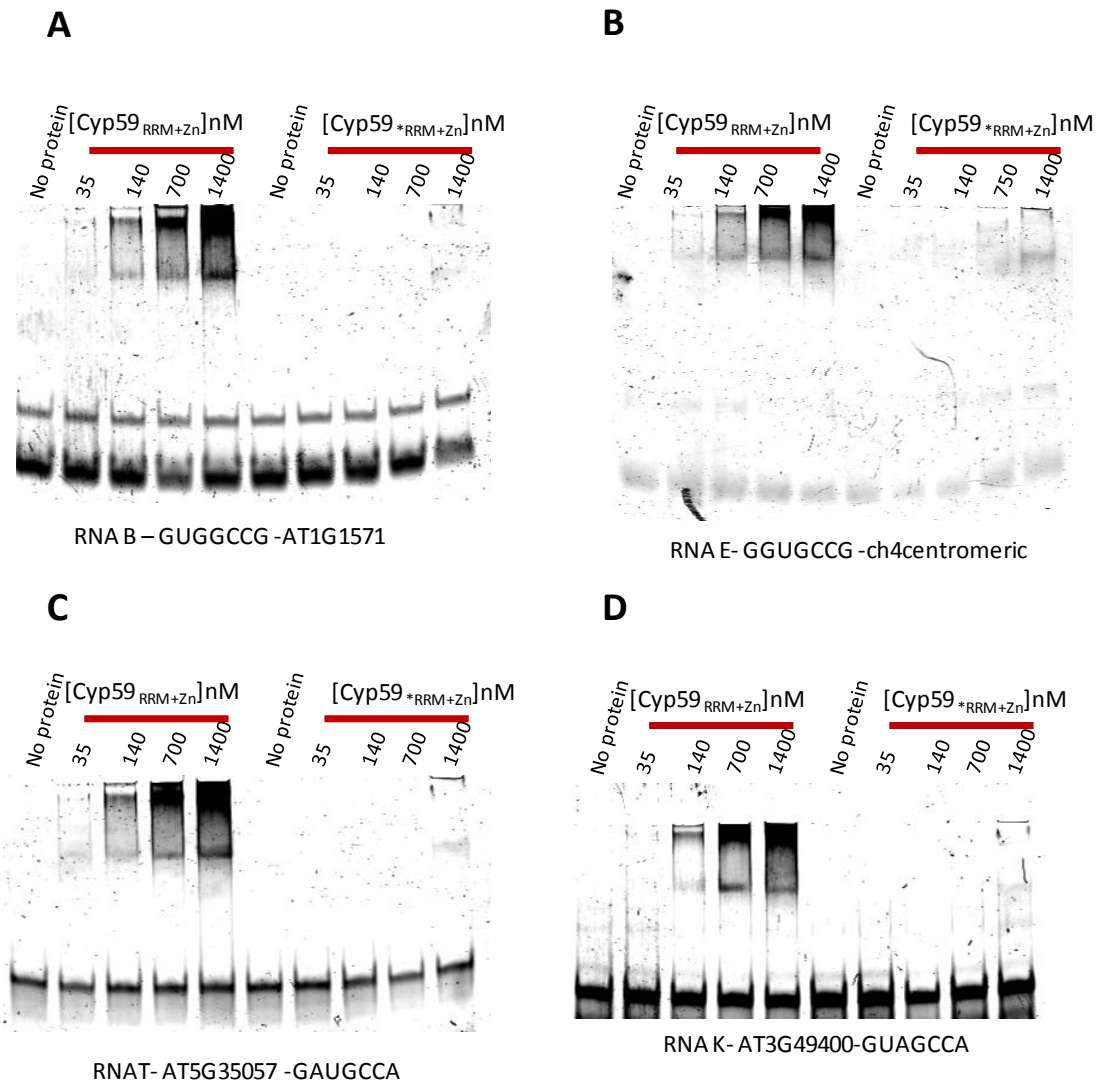

Figure 3.28 Gel shift assay. SYBR GREEN II stain of the 8% native polyacrylamide gel electrophoresis of the 30nt RNA containing one of the binding motif derived from 16 possible binding consensus variants incubated with increasing concentration of the RRM+Zn domain of the AtCyp59 or *RRM+Zn domain of AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. RNA sequence used in the assay as well as short name is indicated below each gel.

Tested 30nt RNAs interact with mutated protein in the unspecific manner (figure 3.28). This suggests that these mutations were crucial for sequence-specific recognition of RNAs and were required for the formation of RNA-protein complexes.

3.3.8 Genomic RNA-targets which were not selected also bind to the AtCyp59.

Bioinformatics analysis revealed that over 70% of mRNA in the genome of *A. thaliana* contain common binding consensus which was then verified by *in vitro* studies. This finding indicated that each of these mRNA could potentially bind to the RRM+Zn domain of the AtCyp59. To determine whether some of the non-sequenced and non-selected with Genomic SELEX experiment RNAs containing one of the variant of the binding motif could interact with RRM+Zn domain of the AtCyp59, we decided to repeat EMSA analysis using particular mRNAs from genome. For that purpose we chose four mRNA from genome which contained three different variants of the binding motif localized in their last exon near the future translation stop codon. Also these mRNAs are known to be spliced and some of them could be spliced alternatively. To performed gel-shift assay we took approx 150nt sequence around binding motif on their sequences. Resulted RNA fragments were further incubated with increasing concentration of wild-type or mutated version of the recombinant RRM+Zn domain of AtCyp59 and then resolved on 6% native polyacrylamide gel. In parallel to wild-type version, we also used mutated version of the protein to checked whether every of the selected RNAs interact specifically with RRM+Zn domain (figure 3.29).

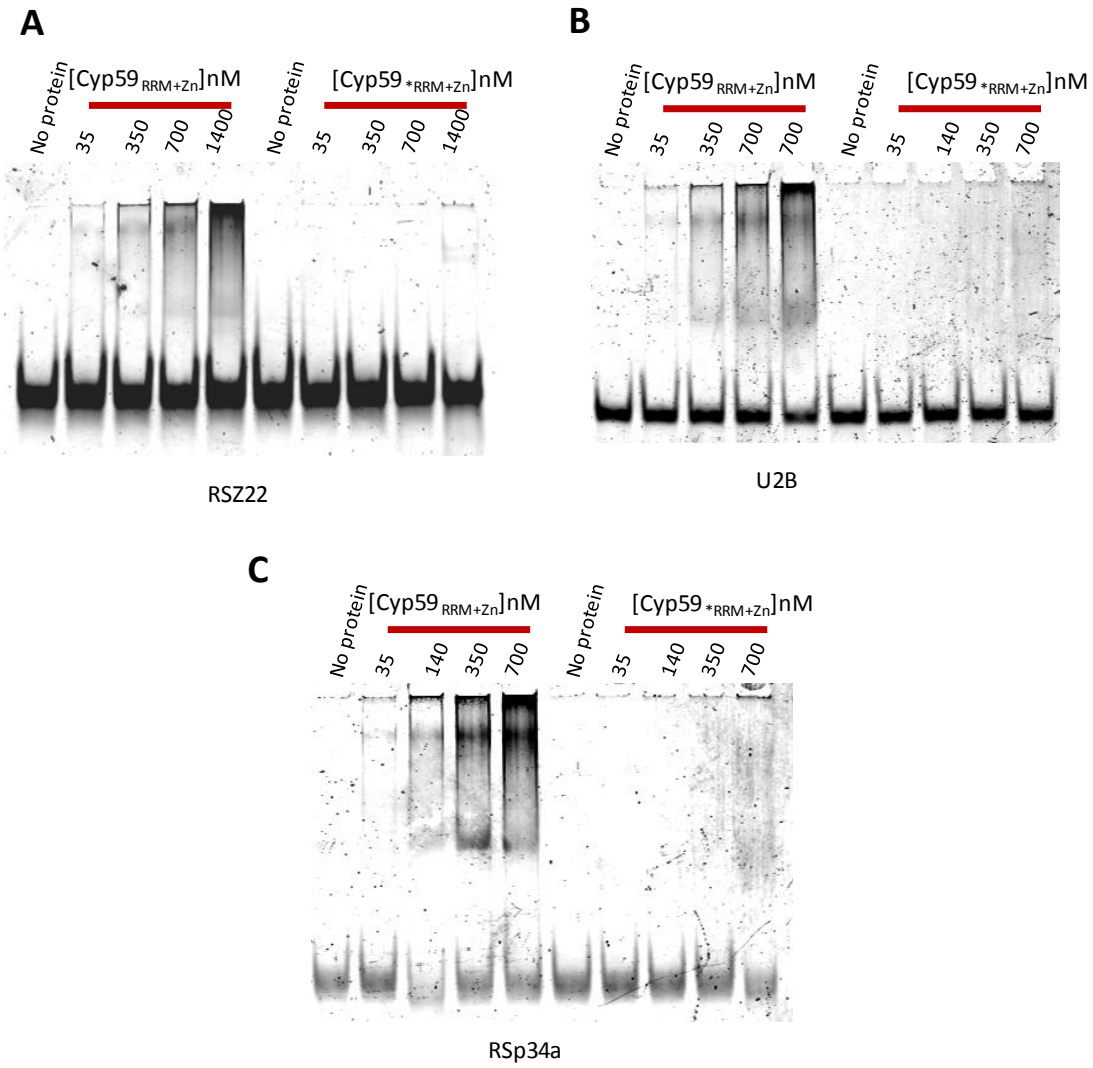

Figure 3.29 Gel shift assay SYBR GREEN II stain of the 6% native polyacrylamide gel electrophoresis of the 100-200nt RNA containing one of the binding motif close to the translational stop codon incubated with increasing concentration of the RRM+Zn domain of the AtCyp59 or *RRM+Zn domain of AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. RNA sequence used in the assay as well as short name is indicated below each gel.

Chosen RNA fragments from genome which contained one of the binding sequences near the future translation stop codon and was not selected with genomic SELEX showed moderate binding affinity to the RRM+Zn domain of the AtCyp59 (fig. 3.29). Furthermore, this interaction seemed to be sequence specific because mutations in the RRM domain completely abolished complex formation between RNA and the protein.

These results proved the data obtained with bioinformatics analysis and also suggested that AtCyp59 could potentially act as general regulatory factor.

One of the four selected RNAs was also tested using EMSA with full-length version of the AtCyp59. Very prominent complex formation between RNA and full-length protein was observed in this case (figure 3.30).

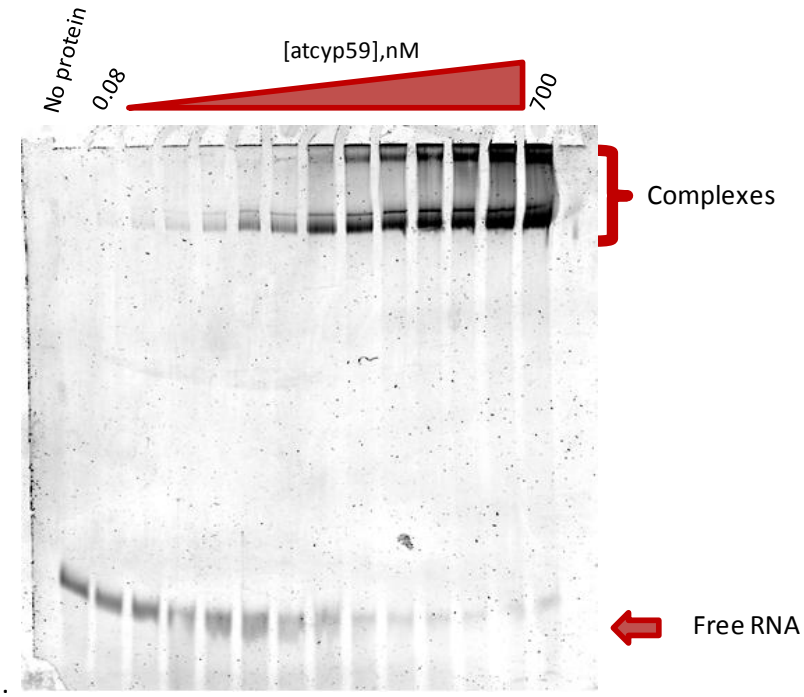

Figure 3.30 Gel shift assay. SYBR GREEN II stain of the 6% native polyacrylamide gel electrophoresis of the 100-200nt RNA (RSZ32) containing one of the binding motif close to the translational stop codon incubated with increasing concentration of the full-length version of the AtCyp59. Concentration of the protein is shown above in nM range. First line in every gel is the line where protein was not added. RNA sequence used in the assay as well as short name is indicated below each gel.

Data presented in this section indicated that predicted binding consensus sequence could be proved using EMSA. Binding its RNA targets of different length, RRM+Zn domain recognized them in sequence-specific manner. Mutations in either RRM domain or RNA target significantly interfere with formation of ribonucleoprotein complex. Finally, other domains of the AtCyp59 positively contributed to complex

formation especially when long RNA fragments were used. It suggests implication of other domains in dealing with local mRNA structure.

3.4 RNA immunoprecipitation in the protoplast overexpressing HA tagged AtCyp59.

3.4.1 Mutations in the RRM domain of the AtCyp59 decrease level of the protein overexpression.

In vitro studies described in the previous section revealed that RRM+Zn domain of AtCyp59 is responsible for binding to RNA fragments containing predicted binding consensus sequence. To determine whether this binding occurs *in vivo* and also whether over expression of the AtCyp59 has an effect on expression of target mRNAs, we decided to over express full-length protein in the *A. thaliana* protoplasts system. We cloned C-terminally HA-tagged full-length versions of AtCyp59 with wild-type and mutated RRM domain in the vector pDEDH under 35SCMv promoter. Also, we used pGREEN vector containing MAP kinase 6 gene and pDEDH fused with GFP but without HA-tag sequence (figure 3.31 A).

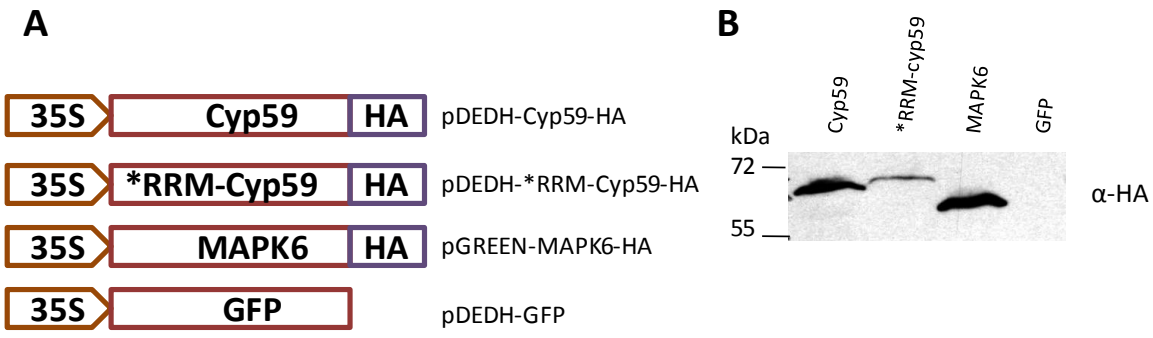

Figure 3.31 Protoplasts expression. A – Schematic representation of the constructs used in the PEG-inducible transformation experiments. 35S – 35S CM virus promoter sequence, HA – hemagglutinine antigene, *RRM – RRM domain in which three essential aminoacids were mutated, GFP – green fluorescents protein, MAPK – MAP kinase six. B-Western blot assay of cells

overexpressiing depicted constructs were performed using anti-HA antibody. Molecular weight marker in kilodaltons is displayed on the left side

As shown in the figure 3.31 B, mutated version of the AtCyp59 exhibited lower level of expression suggesting either protein instability or implication of unknown cell regulators preventing formation of abnormal protein. From the other hand, wild-type protein showed significant level of expression comparable with level of MAPK6. GFP construct in this case was used to indirectly check transformation efficiency which was approximately 20-30%. Since protein level of expression of target protein was different we decided to check whether level of mRNA of the Cyp59 was also affected upon protoplast transformation. For that purpose, second portion of the previously transformed protoplast were used for total RNA isolation. Resulted RNA was utilized in semi-quantitative RT-PCR with primers to exogeneous constructs in case of Cyp59 and *RRM-cyp59 and endogeneous Cyp59 in case of MAPK6 and GFP.

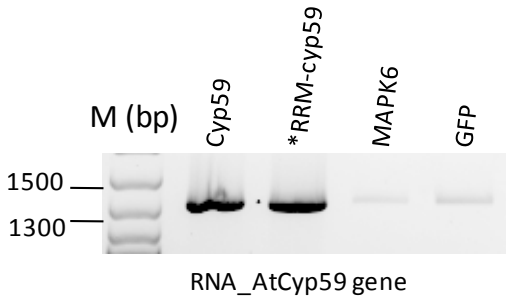

Figure 3.32 Protoplasts RNA expression. Semi-quantitative RT-PCR with primers to the AtCyp59 gene. Molecular weight markeris shown in the left side in bp.

As displayed in the figure 3.32 on mRNA level (exo - and endogenous), both AtCyp59 variants was equally expressed. Compare to the other constructs used where only endogenous level of cyp59 mRNA was detected in both cases wild-type and mutated RRM of AtCyp59 was similarly over expressed. These results suggested that level of protein expression is controlled on post-translational level and formation of the abnormal version of the AtCyp59 is unfavorable process.

3.4.2 Level of expression of target genes doesn't change significantly upon expression of wild-type or mutated AtCyp59 protein.

To identify whether upon AtCyp59 (WT or mutant version) over expression level of the target mRNAs are changed, we performed semi-quantitative RT-PCR analysis using primer pairs amplifying mRNAs of the target genes. We decided to test mRNAs which contain one of the binding motifs (figure 3.33).

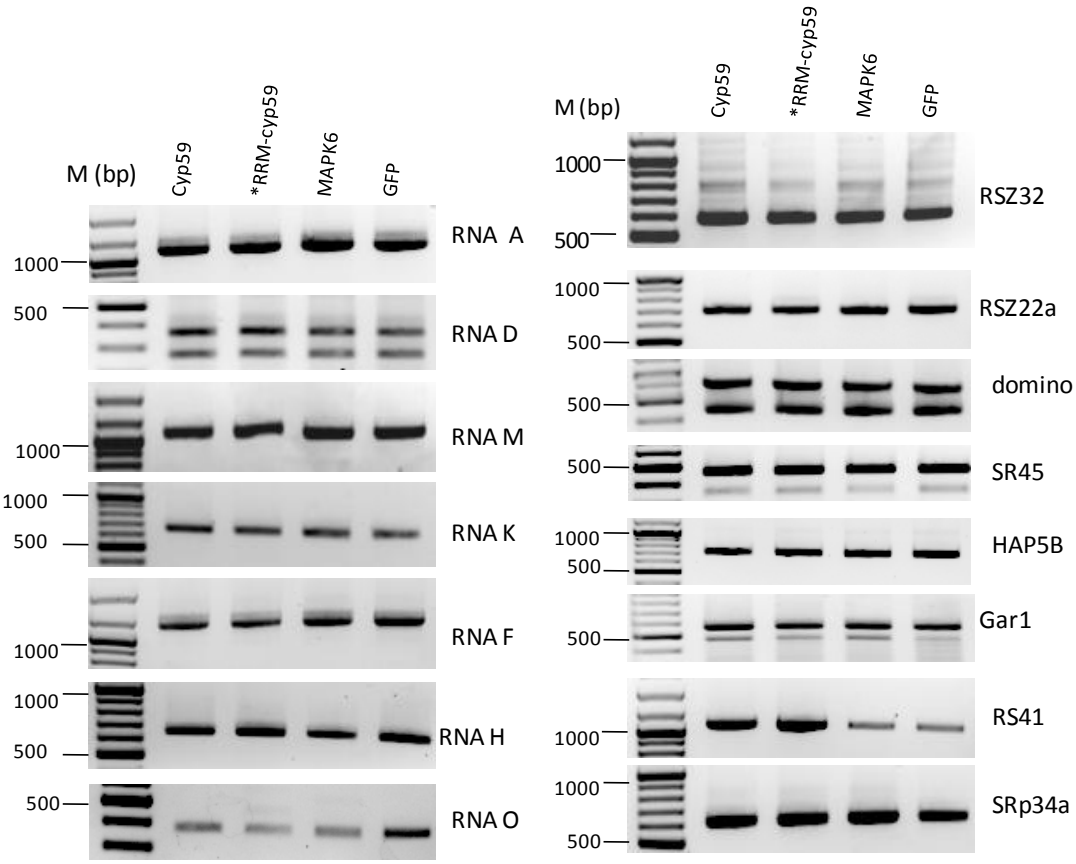

Figure 3.33 Protoplasts mRNA expression. Semi-quantitative RT-PCR with primers to the genes indicated on the right side. Molecular weight markers are shown on the left side in bp.

None of the tested mRNA genes showed significant alterations in the expression level (figure 3.33). This observation might be explained by the fact that only 20-30% of cells were actually transformed with construct of interest and that RT-PCR method is not sensitive enough to detect small changes. Second explanation might be that alterations in

the expression of the target genes are minimal and could not be detected by RT-PCR method. The third, these alterations might be happening during particular stage of plant development.

3.4.3 mRNAs containing binding motif co - immunoprecipitate only wild-type protein.

To test whether target mRNAs interact with wild-type protein AtCyp59 *in vivo*, we precipitated RNA-protein complex protoplasts cells expressing WT AtCyp59 (mutant version served as negative control) using magnetic beads coupled with anti-HA antibody. RNA fraction was purified from precipitated complexes and used for the semi-quantitative RT-PCR with gene-specific primers (figure 3.34).

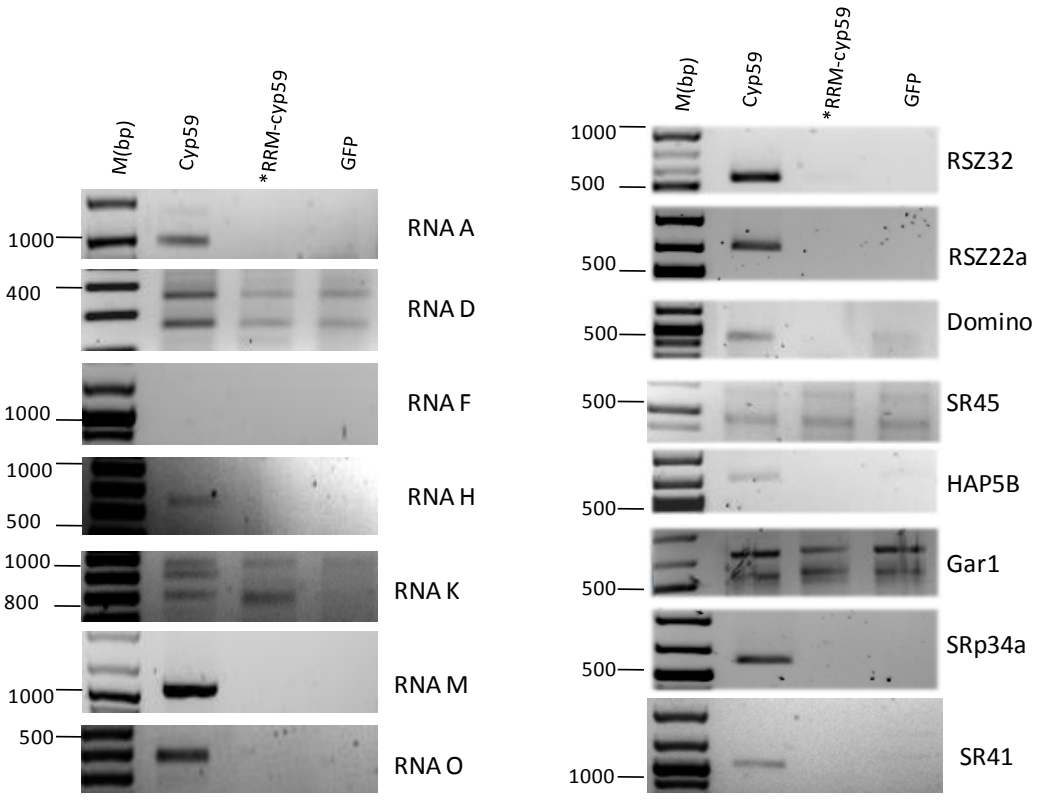

Figure 3.34 Protoplasts RNA immunopresipitation. Semi-quantitative RT-PCR with primers to the genes indicated on the right side. Molecular weight marker is shown in the left side in bp.

As displayed in the figure 3.34, only wild-type cyclophilin protein was able to form complex with target genes tested *in vivo*. Mutated version of the protein as well as controlled protein such as GFP did not from a complex with RNA, which suggests absence of non-specific binding of RNA to the magnetic beads. Interestingly, some target mRNAs especially those which were sequenced in genomic SELEX experiment were precipitated by none of the proteins which might suggest either their implication in particular developmental or growth stage or their overall low presence of this mRNAs in the protoplast system. Also in some cases we saw some background of control immunoprecipitations which might indicate uncomplete washing of the beads.

Generally, from the data obtained we concluded that RRM+Zn domain of the AtCyp59 binds sequence specifically RNA containing common binding pattern GG(U)NGCCG *in vitro* and *in vivo*. Also, mutations in RRM domain of the protein prevent complex formation *in vivo* and significantly decrease binding *in vitro*. In addition, formation of mutated version of AtCyp59 is unfavorable in protoplast expression system on the protein level or on mRNA level.

3.5 Expression of the HA-tagged AtCyp59 protein suppressed by unknown mechanism in the estrogen-inducible cell-suspension system.

3.5.1 Estrogen–inducible expression cassette is integrated in the *Arabidopsis thaliana* genome.

As discussed above, AtCyp59 could be over expressed in the protoplast system. When we checked mRNA level of the selected target genes we didn't find significant up- or down-regulation of these genes upon expression of wild-type or mutated version of the protein. However, almost every tested target gene containing binding motif was shown to immuno-precipitate with wild-type version of the AtCyp59 but not with mutated version or GFP. Nevertheless, RNA level after immuno-precipitation was very

low and sometimes was difficult to detect (figure 3.34). Also protoplast transformation efficiency in general is very low (20-30%) which brings high level of heterogeneity into the system. To overcome these obstacles, we decided to develop stable cell-line suspension over expressing AtCyp59. We decided to use inducible system since we knew from previous experiments that constant expression of AtCyp59 under 35S promoter could not be accomplish. We cloned C-terminally HA-tagged AtCyp59, RSZ33 (another AS factor (Lopato, 2002)) and YFP gene (figure 3.35 A) to the pMDC7 vector, where 35SCMv promoter are kept under control of the LexA operator which became active upon addition of the human estrogen (Chang, 2005).

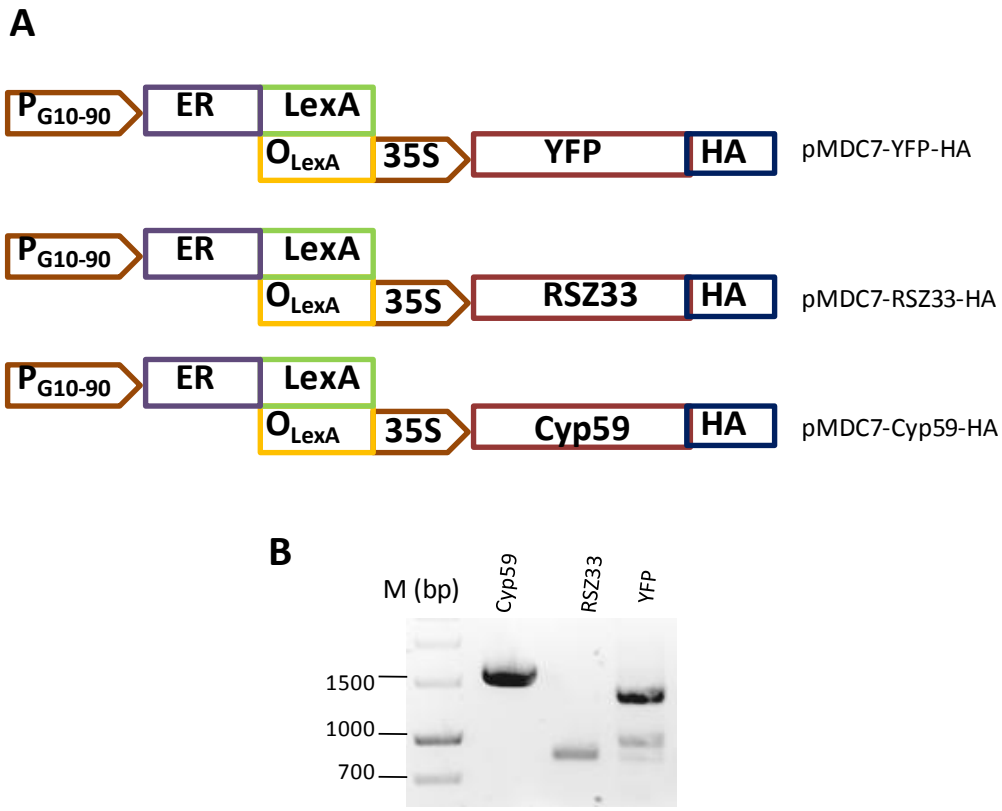

Figure 3.35 Estrogen-inducible Arabidopsis thaliana system. A- Schematic representation of the constructs**. B-** PCR amplification of the isolated genomic DNA **to** check incorporation of the constructs**.**

Designed vectors carrying genes of interest were transformed into isolated protoplasts and left for 24 hours to regenerate. Transgenic cell-suspensions were cultivated for several weeks of growing on antibiotics selective media. To check construct

DNA integration into genome of *A. thaliana* we isolated genomic DNA from each construct and performed PCR amplification using specific primers. As shown in the figure 3.35 B, targeted genes were successfully incorporated into genome.

3.5.2 Protein expression is suppressed upon estrogen induction but level of mRNA produced remains to be sufficient

To carry out over expression experiment we took 5 days-old cell culture (after every week routine culture dilution) and induced it by 5µM of 17-β-estradiol. Then 1ml aliquots were collected at different time points and analyzed by Western blot analysis to indentify start of protein over expression.

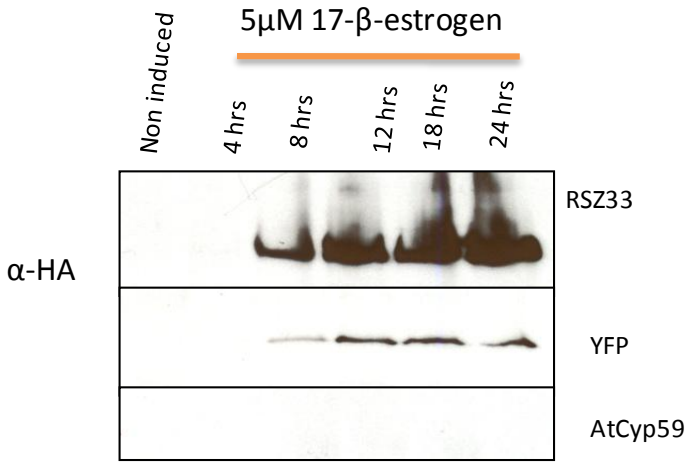

Figure 3.36 Estrogen-inducible Arabidopsis thaliana system. Western blot analysis of induced with 17 β-estradiol cultures expressing AtCyp59, YFP or RSZ33 genes with HA-tag on the C-terminus. Analysis performed with anti-HA antibodies.

HA-tagged protein AtCyp59 showed zeroed level of protein expression whereas proteins RSZ33 and YFP began to be over expressed between 4 and 8 hours after induction reaching saturated level of protein expression in 24 hours upon induction. Increasing 17 β-estradiol final concentrations to 50 and 100 µM did not result in any detectable level of AtCyp59 over expression. We hypothesized that protein undergoes

degradation by a putative cellular mechanism either on transcriptional or translational level. To check whether mRNA level of exo- and endogenous AtCyp59 were affected upon induction of the protein expression we isolated total mRNA from the cells collected at the time points shown in the figure 3.36. Then, we isolated total RNA from these cells and performed reverse transcription using oligo-dT primer. Semi-quantitative RT-PCR of exo- or endogenous levels of the AtCyp59 are presented in the figure 3.37.

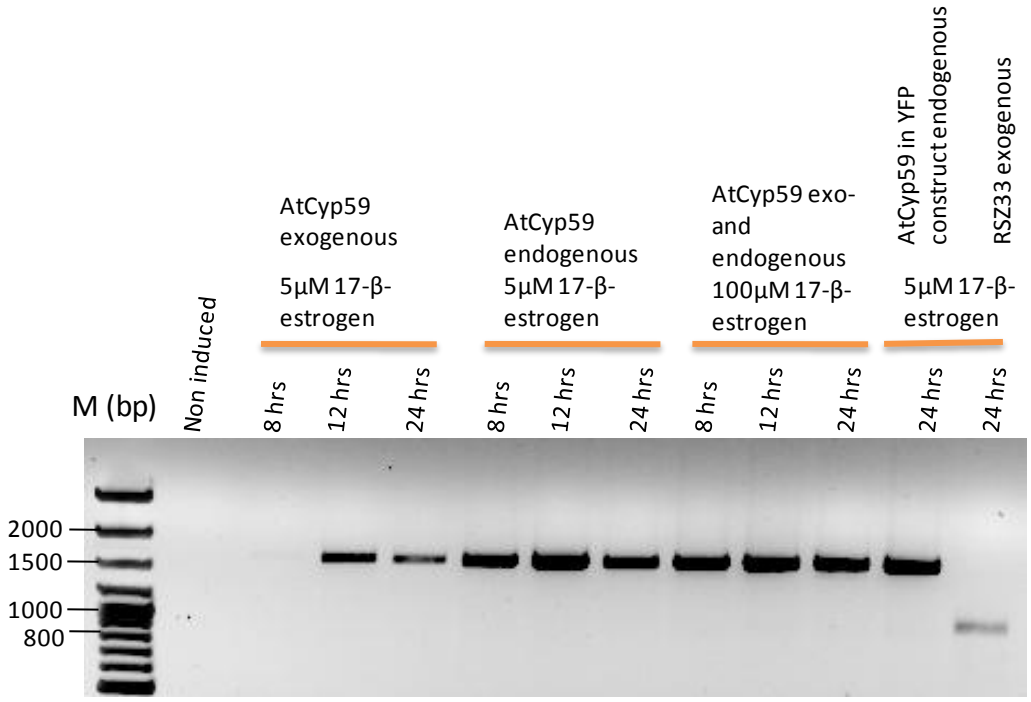

Figure 3.37 RNA expression analysis of the cell suspension culture induced with 17 β-human estradiol. PCR amplification of the AtCyp59 gene using primers to exo- and endogenous mRNAs. Cells were induced to the final 5 and 100µM estrogen concentration and collected after 8, 12, or 24 hours.

As shown in the figure 3.37 the elevated levels of the exogenous mRNA encoding HA-tagged AtCyp59 are produced in 12 hours upon induction with 5µM estradiol and remain the same till 24 hours check-point. From the other hand, endogenous level of mRNA coding for AtCyp59 remains the same level for all time checked regardless increasing concentration of inducer. The same level of endogenous mRNA of gene AtCyp59 was detected in the cells produced HA-tagged YFP in 24 hours upon treatment. Finally, we checked level of exogenous mRNA encoding HA-tagged RSZ33 in the cells

producing RSZ33 protein upon induction with 5µM estradiol. In this case, level of exogenous RSZ33 mRNA was lower than level of exogenous AtCyp59 mRNA. Summarizing these observations, exogenous HA-tagged version of AtCyp59 gene was over expressed on transcriptional level but we didn't detect any produced HA-tagged protein AtCyp59, meaning that either the mRNA might not be translated or the protein might undergo degradation after translation. To check whether we could prevent protein degradation by blocking 26S proteasome degradation pathway, after estradiol induction we incubated cells with 26S proteasome inhibitor MG132 for 1, 4, and 24 hours; however we didn't detect any traces of the protein. Summarizing, we concluded that additional production above wild-type level of the AtCyp59 is non favorable process and tightly regulated.

4. Discussion

Cyclophilins belong to a protein super family of immunophilins (Galat, 2003). These proteins change peptide bond preceding proline from *trans*- to *cis*- conformation (Marks, 1996). It is well known that imid bond could exist in two distinct conformations, rotation of which is accompanied with relatively high energy barrier and requires a catalyst (Pahlke, 2005). Cyclophilins have a catalytically active domain – peptidyl-proline *cis/trans* isomerase (PPIase) - which helps to level down energy barrier between two conformations and to accelerate the rotation of the peptidyl-prolyl bond. Changes in the peptidyl-prolyl bond are indispensable for overall protein function and folding as they have been discovered to influence on many processes in the living cell (He, 2004). *A. thaliana* genome encodes a maximum known so far number of cyclophilins – 29 genes (He, 2004). Majority of these proteins are very small in size, with single-domain organization and localized predominantly in the cytoplasm (Peltier, 2002). However, there are three genes of cyclophilin's family in *A.Thaliana* which showed nuclear localization. These are *A. Thaliana* cyclophilins AtCyp59, AtCyp64 and AtCyp93 (Romano 2004). Among them, AtCyp59 is one of the most interesting proteins as it has a multidomain organization which contains catalytically active PPIase domain on the N-terminus followed by RRM domain and Argenine-Serine rich domain (SR) on the C-terminus (Weighardt, 1999).

It has been previously shown (Gullerova 2006,) that C-terminal domain of AtCyp59 is responsible for protein-protein interaction. It binds to majority of another SR proteins but this binding seems to be unspecific (Gullerova, 2006). In general, SR proteins are argenine-serine rich protein containing also one or two RNA recognition motifs (Wu, 1993). SR proteins serve as global positive cis-acting regulators of constitutive and/or alternative splicing (Robberson, 1990). They promote splicing on the proximal splicing site by binding to specific elements on the pre-mRNA stricture – exonic splicing enhancers (ESE) (Martinez-Contreras, 2007). In general, splicing is performed by huge multicomponent RNA-protein complex – spliceosome. During intron excision spliceosome undergoes several structural rearrangements in pre-mRNA, snRNP structure and many changes in protein-protein interactions (Burge, 1998). It is still no completely clear how all of these changes are regulated. Typically the most flexible and reversible regulation is via

rounds of phosphorylation/dephosphorylation. It is well established that activity of majority of SR proteins are controlled by their phosphorylation status (Mermoud, 1994). Moreover, such regulation of phosphor-status of SR proteins is mediated by several kinases and phosphatases which could directly interact with spliceosome and influence on splicing (Misteli, 1999). Taking this into account, AtCyp59 interaction with SR proteins might bring a new level of regulation into pre-mRNA processing. Also AtCyp59 binds to the SR proteins by its C-terminal RS/RD region. From the other side, SR proteins have many prolines residues in between serines-argenines repeats which conformation might be important for availability of serines to be phosphorylated. Interestingly, it has been shown recently, that many cyclophilins in human found to be associated with spliceosome (Rappsilber, 2002). It has been found that human cyclophilins interact with many components of splicing machinery such as tri-snRNP complex, and SR proteins as well as with pre-mRNA (Pushkarsky, 2001; Horowitz, 2002). Thus, AtCyp59 might potentially mediate pre-mRNA processing on protein-protein level by interaction with SR proteins.

Although AtCyp59 is localized in the nucleus, it does not resemble speckles pattern of SR proteins (Gullerova, 2006). AtCyp59 shows specific dots pattern within the nucleus which are associated with actively transcribed genes (Gullerova, 2006). Further analysis has showed that AtCyp59 interacts *in vitro* and *in vivo* with C-terminal domain (CTD) of the Polymerase II (Gullerova, 2006). CTD is the largest of Pol II domains and consists of several Tyr-Ser-Pro-Thr-Ser-Pro-Ser heptapeptide repeats (up to 52 repeats in humans) (Corden, 1990). Pol II is responsible for transcription of all mRNAs in the cell. During the transcription two serines in CTD of Pol II got dynamically phosphorylated in processes regulated by specific kinase for each serine (Buratowski, 2003). Since conformation of preceding prolines can influence on availability of serine residues for phosphorylation, interaction between AtCyp59 and CTD of Pol II could play role in phosphorylation status of serines on CTD (Schiene, 2000). Apparently, close homolog of AtCyp59 in *S. pombe*, Rct1, interacts with non-catalytic doamin of Cdk9 kinase (Skrahina, unpublished), kinase which is responsible for phosphorylation of the Ser2 on the CTD (Peterlin, 2006). This phosphorylation mark is appeared when Pol II switches from initiation stage to the elongation and Ser 2 stays phosphorylated throughout elongation cycle of transcription (Egloff, 2008). As it has been shown, PPIase domain of Rct1 plays an important role in

interaction with non-kinase part of Cdk9 (Skrahina, unpublished). This data suggest that Rct1 might change the proline bond preceding Ser2 to make it accessible for the Cdk9 kinase action on CTD of Pol II. As for the AtCyp59, it has been shown that upon over expression of the protein in cell culture, level of CTD phosphorylation dramatically decreases which leads to cell death (Gullerova, 2006). From the other hand overespression of Rct1 in yeast led to morphological cell defects. Also Rct1 is essential gene because deletion of this gene is lethal (Gullerova, 2007). It is worth to mention that the level of CTD phosphorylation in both overexpression and knock-out of Rct1 decreses (Skrahina, unpublished). These data may propose that expression of AtCyp59 seems to be tightly regulated and its PPIase domain might be involved in transcription regulation during the elongation phase.

As it is mentioned above, AtCyp59 contains RRM domain which could potentially bind RNAs. As it was discovered, it showed a high level of conservation even between distinctive species and even higher conservation then its PPIase domain, suggesting an additional impact of this domain into overall function of the protein (Gullerova, 2006). It has been shown that RRM domain could bind synthetic poly(rC) and poly(rG) (Gullerova, 2006). However, the absence of *in vivo* model system where AtCyp59 would be over expressed or knock-down makes it difficult to study protein function *in vivo.* One possible solution for this problem would be the usage of antibodies raised against endogenous AtCyp59 or its parts for immunoprecipitation-based techniques, i.e. protein- or RNA-immunoprecipitation from different tissues on different developmental stages. Having no anti-AtCyp59 antibodies, we have decided to focus on available *in vitro* methods to study the role of RRM domain of AtCyp59.

Among *in vitro* methods, Genomic SELEX appeared to be powerful screening technique which allows employing of recombinant protein of interest for selecting of specific RNA aptamers sequences with genomic origin (Gold, 1997). Genomic SELEX consists of two principal steps: genomic library development (Figure 3.1) and affinity selection (Figure 3.4). Crucial point in the whole procedure is quality of DNA library before selection, its representation and nucleotide distribution (Singer, 1997). It remains important to analyze library before selection, and availability of next generation sequencing technologies gives a great tool for it. Our results showed that initial DNA library contained same genomic elements as the *A. Thaliana* genome (Figure 3.6). The

number of sequences we obtained could potentially cover only 2% of genome, however distribution of the genomic elements in the initial library were only slightly differ than in the genome (Figure 3.6). Thus, number of protein exons and rRNA genes were higher whereas number of intergenic regions was lower in the initial DNA library. Consistent with this, there were increase in CG content and decrease in AT content in di-nucleotide distribution in the initial library compare to *A. Thaliana* genome (Figure 3.7). Interestingly, our nucleotide distribution is different from one observed by group of Schroeder (Zimmermann, 2010) when they performed "neutral" (no-protein) genomic SELEX experiment employing *E. coli* genomic library. This fact suggests that nucleotide bias we observed was not due to amplification forces on library and might lie in library delepoment ingeneral or primer design. Also, differences between *A. Thaliana* genome and initial library were considered to be insignificant and library was qualified for further selection with AtCyp59. However, for better understanding causes of initial DNA library differences compare to genome which could be due to the differences in genome chromatin condensation (Johnson, 2002), incomplete randomness of primers, preferences of adaptors to particular genome regions, further analysis together with deeper sequencing coverage should be performed.

Analysis of enriched SELEX library after selection with AtCyp59 revealed drastically increase in sequences derived from protein exons (Figure 3.8), also further decrease in rRNA and intergenic region compare to initial library and from genome of *A. Thaliana*. Our quality control of obtained after SELEX library showed that more than 80% of reads we sequenced contained right 5'- and 3'- adaptors from the both sides. The sequences in between adaptors were longer than 18 nt and 90% from those reads could be allocated back to the genome of *A. Thaliana* (Table 3.1). We have found that over 50% of obtained sequences derived from protein coding genes (Figure 3.12 A, B). Sequencing reads were equally distributed over sense and antisense strand towards an annotated gene (Figure 3.13 A). From those sequences, which were in sense orientation and came from protein coding genes, were identified that they were localized in exonic or UTR's parts of the pre-mRNA (Figure 3.13 B). Thus, AtCyp59 binds to the RNA targets which are derived from protein coding genes, particularly exonic parts.

Next, we were able to indentify common RNA binding consensus sequence which seems to be recognized by AtCyp59 (Figure 3.14). Previous structural studies on proteins,

containing RRM domain had revealed that RNA recognition motif usually consists of 4 β-sheets and 2 α-helices which divide RRM motif into two parts (Allain, 2000). Structural analysis has revealed that RNA molecule interacts with RRM domain in between these parts making stacking with aromatic rings of the amino acids (Maris, 2005) Also, it is known that the size of typical RNA binding pocket in RRM domain allows adoption of 5-15 ribonucleotides (Wang 2008). In this RNA-binding pocket three conserved aromatic amino acids make a stacking interaction with RNA bases (Wang, 2008). The motif, which we identified, consists of 7nt-long RNA consensus enriched with G and C nucleotides and GCCR core sequence (Table 3.2). This is in consistent with our previous data which showed that AtCyp59 interacts with poly-C and poly-G RNA oligonucleotides (Gullerova, 2006). The binding consensus we have indentified consists of several single motif variations where each variation was sequenced different number of times (Figure 3.16). We checked *in vitro* binding for all motif variants with AtCyp59 (Figure 3.22). Interestingly, binding motif variants have shown variable binding affinity to the RRM+Zn domain of AtCyp59 where binding consensus could be narrowed down to G(GU)GCCG (Table 3.4). It is well known that dissociation constant varies a lot depending on nature of RNA-protein complex, its stability and proposed function. For instance, binding of U1 snRNA to its protein is very tight with K_D's in range of few nM (Hall, 1992). From the other side, recently K_D's of RNA-protein complex between the human cyclophilin containing RRM domain and its RNA targets have been identified. This K_D was in range of few µM suggesting that this complex is very flexible (Hom, 2010). Our K_D approximation for the binding of 7nt RNA targets to RRM+Zn domain of AtCyp59 are lying in 100's nM range. These data suggest moderate complex formation. Interestingly, binding of the full-length AtCyp59 to the same 7nt binding motif variations showed equal affinity to all binding motif variants (Figure 3.23, Table 3.5) suggesting involvement of other domains of AtCyp59 in this interaction. Also, we have noticed that full-length AtCyp59 could not differentiate between binding motif variants.

We assume that AtCyp59 interacts with its RNA targets sequence specifically since mutations in aromatic amino acids in RRM domain of AtCyp59 which are known to be involved in RNA recognition prevent binding of RRM+Zn domain of AtCyp59 to its 7nt binding motif variants (Figure 3.24). Also, mutations in core binding pattern of 7nt RNA drastically decrease its affinity to RRM+Zn domain of AtCyp59 (Figure 3.25).

Sequences involved in the genomic selection and obtained after 454 deep sequencing were longer than 7nt with average 30-40 nt (Figure 3.9). We investigated binding affinity of these longer RNA pieces to the RRM+Zn domain of AtCyp59 in vitro (Figure 3.26). We have found that RRM+Zn domain of AtCyp59 interacts with RNA sequences containing variants of binding motif, but its affinity is lower and ranges in K_D >300nM (Table 3.8). However, specificity of RRM+Zn domain of AtCyp59 to particular binding motif variant stays in the same range as for 7nt RNA. We concluded that presence of flanking RNA sequences slows down its sequence-specific recognition by protein. Interestingly, full-length AtCyp59 equalizes and improves its binding affinity to the longer RNA pieces containing binding motif variants (Figure 3.27). Again, as with 7nt RNAs full-length protein does not differentiate between RNAs containing different binding motif variants and binds to them with approximately similar K_D around 150-200 nM (Table 3.9). Here we concluded that other domains (PPIase, C-terminus region) of AtCyp59 facilitate RNA recognition by RRM+Zn domain, probably due to dealing with possible local RNA structures. Also, it has been shown that AtCyp59 might have RNA chaperon activity (Gullerova, unpublished), when it can modify RNA structure (Rajkowitsch, 2007). Thus, full-length AtCyp59 positively regulates binding to its 30nt RNA targets containing common binding consensus and might also change local RNA structure to help RRM+Zn domain recognizing RNA sequence. Finally, we confirmed that RRM+Zn domain of AtCyp59 interacts with 30nt RNAs containing binding motif variants, since mutation in RRM domain of AtCyp59 decrease protein affinity to the RNA (Figure 3.28). We also confirmed interaction between AtCyp59 and its RNA targets *in vivo*, using transient expression of the AtCyp59 in protoplasts cells (Figure 3.34). We have found that AtCyp59 does not affect expression of its targets upon over expression in protoplasts (Figure 3.33). It may suggest that increase in AtCyp59 protein production does not change overall mRNA expression during the 24 hours after protoplast transformation. We also noticed that not all RNA targets, which were identified and proved *in vitro*, were immunoprecipitated along with AtCyp59 *in vivo*, suggesting that interaction of the protein with its RNA might be flexible and happen at particular time point, for example at different stages of plant development (Figure 3.34). Interestingly, we have noticed that mutations in aromatic amino acids in RRM domain of AtCyp59 negatively affected the over expression of mutant protein in protoplast cells (Figure 3.31 B). However, exogenous

mRNA levels of both mutated and *wt* proteins remained on the same level (Figure 3.32). We speculate that mutations in the RRM domain of AtCyp59 might influence protein stability therefore overall protein level decreases. We also have found that mutated version of protein does not immunoprecipitate its RNA targets further showing sequence specificity of RRM+Zn domain of AtCyp59 (Figure 3.34).

Genomic SELEX suits well for common binding RNA motif identification for the particular protein, but due to the procedure could not give a full scope of possible targets in the genome. Thus, we performed a genome screen to indentify binding consensus distribution in *A. thaliana* genome. We have found that binding motif could be localized in over 70% of mRNAs in the genome (Table 3.3). It is overrepresented among exons on the pre-mRNAs (Figure 3.17) and typically could be found ones or maximum three times on the same mRNA (Figure 3.18). Thus, presence of the binding motif on mRNA might serve as mRNA signature for AtCy59 recognition. Also, we have shown that full-length AtCyp59 binds equally to RNAs containing different variant of the binding motif (Figure 3.27). Therefore, presence of this motif in more than 50% of the mRNA genes might indicate that AtCyp59 acts as universal factor in transcription or splicing. We have found out that position of the RRM+Zn of AtCyp59 binding motif correlates with position of future translation termination site on pre-mRNA structure (Figure 3.19, 3.20). We tested binding affinity *in vitro* (Figure 3.29, 3.30) and *in vivo* (Figure 3.34) of few mRNAs from genome which we didn't find during genomic SELEX and which contained binding motif close to the future stop codon. We have found that RRM+Zn domain of AtCyp59 interacts with those mRNA pieces *in vitro* and majority of those mRNA were immunoprecipitated with AtCyp59 in protoplast (Figure 3.34). We hypothesized that binding of AtCyp59 to the mRNA in close proximity to the future stop codon might have influence on proper transcription termination and 3'-end mRNA processing. Notably, it has been discovered that cyclophilins are involved in 3'-end mRNA processing (Noble, 2005). For example, protein responsible for the polyadenylation cleavage PcfI could only interact with CTD phosphorylated on Ser2 and when proline preceding Ser2 is in *trans* conformation (Licatalosi, 2002). We have also found out that PPIase activity of AtCyp59 slows down upon binding to the polyA mRNA fraction or specific RNA found to be AtCyp59 target in genomic SELEX (Bannikova O., Skrahina T., Barta A. unpublished). Thus, we speculate that

binding of AtCyp59 to the pre-mRNA might serve as a signal to the Pol II of the mRNA close end and by that also influence phosphorylation status of the CTD of Pol II.

In order to study function of AtCyp59 *in vivo*, we tried to establish estrogen inducible AtCyp59 cell suspension system (Figure 3.35 A). Although we have indentified that construct was integrated to the *A.thaliana* genome (Figure 3.35 B) we were unable to induce protein expression (Figure 3.36). Strikingly, other constructs in the same system we established showed elevated level of protein expression (Figure 3.36). Interestingly, we also didn't notice an increase in mRNA level of AtCyp59 upon estrogen induction (Figure 3.37) compare to endogenous AtCyp59 mRNA level. Thus, we assume that expression of AtCyp59 is under strong control and might be regulated on mRNA or protein level. It has been known that expression of AtCyp59 is tightly regulated as it drastically influences on level of CTD of Pol II phosphorylation which might explain tight regulation of AtCyp59 expression and absence of its stable over expression lines and T-DNA mutant lines (Gullerova, 2006). We assume that tight regulation of the AtCyp59 expression might be associated with its mRNA binding which influences on PPIase activity of AtCyp59 (Skrahina, unpublished results) thus altering phosphorylation status of CTD of Pol II. Thus, further studies on transcription efficiency of target genes upon expression of AtCyp59 might shed the light on AtCyp59 regulatory function.

Summing up obtained data, genomic SELEX allowed us to identify the binding RNA consensus motif for highly conserved multidomain protein AtCyp59. This motif is present in more than 50% of *A. thaliana* mRNAs and localized in last exons close to future translation termination sites. The binding of AtCyp59 to its RNA targets was confirmed *in vitro* by EMSA and *in vivo* by RIP from protoplasts transiently expressing HA-tagged protein. Analysis of K_D of complexes formed between either full-length AtCyp59 or its RRM+Zn domain with RNA target sequences of different length has shown that the other domains of AtCyp59 might facilitate RNA binding by RRM+Zn domain. It is noteworthy that those domains are C-terminal RS/RD domain which is responsible for protein-protein interactions with AS-associated SR proteins and N-terminal PPIase domain which activity is inhibited upon RNA binding by full-length protein. Since PPIase activity of AtCyp59 may influence on phosphorylation status of Ser2 in CTD repeats of Pol II and this activity is negatively regulated by binding of AtCyp59 to its RNA targets which are indeed localized close to their translation termination sites, we believe that AtCyp59 represents a global

regulator of transcription and plays an important role on interplay between transcription and following splicing of pre-mRNAs.

5. Reference

Ahn SH, Kim M, Buratowski S (2004) Phosphorylation of serine 2 within the RNA polymerase II C-terminal domain couples transcription and 3' end processing. *Mol Cell* **13:** 67-76

Ali GS, Palusa SG, Golovkin M, Prasad J, Manley JL, Reddy AS (2007) Regulation of plant developmental processes by a novel splicing factor. *PLoS ONE* **2:** e471

Allain FH, Bouvet P, Dieckmann T, Feigon J (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin. *EMBO J* **19:** 6870-6881

Allison LA, Moyle M, Shales M, Ingles CJ (1985) Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell* **42:** 599-610

Ast G (2004) How did alternative splicing evolve? *Nat Rev Genet* **5:** 773-782

Auboeuf D, Dowhan DH, Kang YK, Larkin K, Lee JW, Berget SM, O'Malley BW (2004) Differential recruitment of nuclear receptor coactivators may determine alternative RNA splice site choice in target genes. *Proc Natl Acad Sci U S A* **101:** 2270-2274

Auweter SD, Fasan R, Reymond L, Underwood JG, Black DL, Pitsch S, Allain FH (2006) Molecular basis of RNA recognition by the human alternative splicing factor Fox-1. *EMBO J* **25:** 163-173

Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2:** 28-36

Barbazuk WB, Fu Y, McGinnis KM (2008) Genome-wide analyses of alternative splicing in plants: opportunities and challenges. *Genome Res* **18:** 1381-1392

Barrass JD, Beggs JD (2003) Splicing goes global. *Trends Genet* **19:** 295-298

Barta A, Kalyna M, Lorkovic ZJ (2008) Plant SR proteins and their functions. *Curr Top Microbiol Immunol* **326:** 83-102

Baskaran R, Dahmus ME, Wang JY (1993) Tyrosine phosphorylation of mammalian RNA polymerase II carboxyl-terminal domain. *Proc Natl Acad Sci U S A* **90:** 11167-11171

Bauren G, Belikov S, Wieslander L (1998) Transcriptional termination in the Balbiani ring 1 gene is closely coupled to 3'-end formation and excision of the 3'-terminal intron. *Genes Dev* **12:** 2759-2769

Bauren G, Wieslander L (1994) Splicing of Balbiani ring 1 gene pre-mRNA occurs simultaneously with transcription. *Cell* **76:** 183-192

Bedard KM, Daijogo S, Semler BL (2007) A nucleo-cytoplasmic SR protein functions in viral IRES-mediated translation initiation. *EMBO J* **26:** 459-467

Bentley DL (2005) Rules of engagement: co-transcriptional recruitment of pre-mRNA processing factors. *Curr Opin Cell Biol* **17:** 251-256

Berardini TZ, Bollman K, Sun H, Poethig RS (2001) Regulation of vegetative phase change in Arabidopsis thaliana by cyclophilin 40. *Science* **291:** 2405-2407

Beyer AL, Osheim YN (1991) Visualization of RNA transcription and processing. *Semin Cell Biol* **2:** 131-140

Bienroth S, Keller W, Wahle E (1993) Assembly of a processive messenger RNA polyadenylation complex. *EMBO J* **12:** 585-594

Bird G, Zorio DA, Bentley DL (2004) RNA polymerase II carboxy-terminal domain phosphorylation is required for cotranscriptional pre-mRNA splicing and 3'-end formation. *Mol Cell Biol* **24:** 8963-8969

Birney E, Kumar S, Krainer AR (1993) Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* **21:** 5803-5816

Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* **72:** 291-336

Boggs RT, Gregor P, Idriss S, Belote JM, McKeown M (1987) Regulation of sexual differentiation in D. melanogaster via alternative splicing of RNA from the transformer gene. *Cell* **50:** 739-747

Bourgeois CF, Popielarz M, Hildwein G, Stevenin J (1999) Identification of a bidirectional splicing enhancer: differential involvement of SR proteins in 5' or 3' splice site activation. *Mol Cell Biol* **19:** 7347-7356

Bourquin JP, Stagljar I, Meier P, Moosmann P, Silke J, Baechi T, Georgiev O, Schaffner W (1997) A serine/arginine-rich nuclear matrix cyclophilin interacts with the C-terminal domain of RNA polymerase II. *Nucleic Acids Res* **25:** 2055-2061

Boutz PL, Stoilov P, Li Q, Lin CH, Chawla G, Ostrow K, Shiue L, Ares M, Jr., Black DL (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. *Genes Dev* **21:** 1636-1652

Brow DA (2002) Allosteric cascade of spliceosome activation. *Annu Rev Genet* **36:** 333-360

Brown JW, Birmingham A, Griffiths PE, Jossinet F, Kachouri-Lafond R, Knight R, Lang BF, Leontis N, Steger G, Stombaugh J, Westhof E (2009) The RNA structure alignment ontology. *RNA* **15:** 1623-1631

Brown JW, Simpson CG, Thow G, Clark GP, Jennings SN, Medina-Escobar N, Haupt S, Chapman SC, Oparka KJ (2002) Splicing signals and factors in plant intron removal. *Biochem Soc Trans* **30:** 146-149

Brown KM, Gilmartin GM (2003) A mechanism for the regulation of pre-mRNA 3' processing by human cleavage factor Im. *Mol Cell* **12:** 1467-1476

Burge CB, Padgett RA, Sharp PA (1998) Evolutionary fates and origins of U12-type introns. *Mol Cell* **2:** 773-785

Bunka DH, Mantle BJ, Morten IJ, Tennent GA, Radford SE, Stockley PG (2007) Production and characterization of RNA aptamers specific for amyloid fibril epitopes. *J Biol Chem* **282:** 34500-34509

Buratowski S (2003) The CTD code. *Nat Struct Biol* **10:** 679-680

Buratowski S (2008) Transcription. Gene expression--where to start? *Science* **322:** 1804-1805

Caceres JF, Misteli T, Screaton GR, Spector DL, Krainer AR (1997) Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *J Cell Biol* **138:** 225-238

Carels N, Bernardi G (2000) Two classes of genes in plants. *Genetics* **154:** 1819-1825

Carte J, Wang R, Li H, Terns RM, Terns MP (2008) Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev* **22:** 3489-3496

Cech TR (2009) Crawling out of the RNA world. *Cell* **136:** 599-602

Chen Y, Carlini DB, Baines JF, Parsch J, Braverman JM, Tanda S, Stephan W (1999) RNA secondary structure and compensatory evolution. *Genes Genet Syst* **74:** 271-286

Colwill K, Pawson T, Andrews B, Prasad J, Manley JL, Bell JC, Duncan PI (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J* **15:** 265-275

Corden JL (1990) Tails of RNA polymerase II. *Trends Biochem Sci* **15:** 383-387

Corden JL (2007) Transcription. Seven ups the code. *Science* **318:** 1735-1736

Corey RB, Pauling L (1953) Fundamental dimensions of polypeptide chains. *Proc R Soc Lond B Biol Sci* **141:** 10-20

Cramer P, Caceres JF, Cazalla D, Kadener S, Muro AF, Baralle FE, Kornblihtt AR (1999) Coupling of transcription with alternative splicing: RNA Pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell* **4:** 251-258

Crawford JB, Patton JG (2006) Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol Cell Biol* **26:** 8791-8802

Croft MT, Moulin M, Webb ME, Smith AG (2007) Thiamine biosynthesis in algae is regulated by riboswitches. *Proc Natl Acad Sci U S A* **104:** 20770-20775

Damgaard CK, Tange TO, Kjems J (2002) hnRNP A1 controls HIV-1 mRNA splicing through cooperative binding to intron and exon splicing silencers in the context of a conserved secondary structure. *RNA* **8:** 1401-1415

Das R, Yu J, Zhang Z, Gygi MP, Krainer AR, Gygi SP, Reed R (2007) SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol Cell* **26:** 867-881

de la Fuente van Bentem S, Anrather D, Roitinger E, Djamei A, Hufnagl T, Barta A, Csaszar E, Dohnal I, Lecourieux D, Hirt H (2006) Phosphoproteomics reveals extensive in vivo phosphorylation of Arabidopsis proteins involved in RNA metabolism. *Nucleic Acids Res* **34:** 3267-3278

de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR (2003) A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12:** 525-532

de la Mata M, Kornblihtt AR (2006) RNA polymerase II C-terminal domain mediates regulation of alternative splicing by SRp20. *Nat Struct Mol Biol* **13:** 973-980

Deshler JO, Rossi JJ (1991) Unexpected point mutations activate cryptic 3' splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev* **5:** 1252-1263
Djordjevic M (2007) SELEX experiments: new prospects, applications and data analysis in inferring regulatory pathways. *Biomol Eng* **24:** 179-189

Dornan J, Taylor P, Walkinshaw MD (2003) Structures of immunophilins and their ligand complexes. *Curr Top Med Chem* **3:** 1392-1409

Dye MJ, Gromak N, Proudfoot NJ (2006) Exon tethering in transcription by RNA polymerase II. *Mol Cell* **21:** 849-859

Economou EP, Bergen AW, Warren AC, Antonarakis SE (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. *Proc Natl Acad Sci U S A* **87:** 2951-2954

Egloff S, Murphy S (2008) Cracking the RNA polymerase II CTD code. *Trends Genet* **24:** 280-288

Fanghanel J, Fischer G (2004) Insights into the catalytic mechanism of peptidyl prolyl cis/trans isomerases. *Front Biosci* **9:** 3453-3478

Feilner T, Hultschig C, Lee J, Meyer S, Immink RG, Koenig A, Possling A, Seitz H, Beveridge A, Scheel D, Cahill DJ, Lehrach H, Kreutzberger J, Kersten B (2005) High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. *Mol Cell Proteomics* **4:** 1558-1568

Feng Y, Chen M, Manley JL (2008) Phosphorylation switches the general splicing repressor SRp38 to a sequence-specific activator. *Nat Struct Mol Biol* **15:** 1040-1048

Fischer G, Bang H, Mech C (1984) [Determination of enzymatic catalysis for the cis-trans-isomerization of peptide binding in proline-containing peptides]. *Biomed Biochim Acta* **43:** 1101-1111

Fu XD (1995) The superfamily of arginine/serine-rich splicing factors. *RNA* **1:** 663-680

Fulgosi H, Vener AV, Altschmied L, Herrmann RG, Andersson B (1998) A novel multi-functional chloroplast protein: identification of a 40 kDa immunophilin-like protein located in the thylakoid lumen. *EMBO J* **17:** 1577-1587

Galat A (2003) Peptidylprolyl cis/trans isomerases (immunophilins): biological diversity--targets--functions. *Curr Top Med Chem* **3:** 1315-1347

Gaur RK, Beigelman L, Haeberli P, Maniatis T (2000) Role of adenine functional groups in the recognition of the 3'-splice-site AG during the second step of pre-mRNA splicing. *Proc Natl Acad Sci U S A* **97:** 115-120

Gerber HP, Hagmann M, Seipel K, Georgiev O, West MA, Litingtung Y, Schaffner W, Corden JL (1995) RNA polymerase II C-terminal domain required for enhancer-driven transcription. *Nature* **374:** 660-662

Gold L, Brown D, He Y, Shtatland T, Singer BS, Wu Y (1997) From oligonucleotide shapes to genomic SELEX: novel biological regulatory loops. *Proc Natl Acad Sci U S A* **94:** 59-64

Gold L, Singer B, He YY, Brody E (1997) SELEX and the evolution of genomes. *Curr Opin Genet Dev* **7:** 848-855

Golovkin M, Reddy AS (1998) The plant U1 small nuclear ribonucleoprotein particle 70K protein interacts with two novel serine/arginine-rich proteins. *Plant Cell* **10:** 1637-1648

Gomes NP, Bjerke G, Llorente B, Szostek SA, Emerson BM, Espinosa JM (2006) Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program. *Genes Dev* **20:** 601-612

Goodall GJ, Filipowicz W (1989) The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* **58:** 473-483

Gopinath SC (2007) Methods developed for SELEX. *Anal Bioanal Chem* **387:** 171-182

Gornemann J, Kotovic KM, Hujer K, Neugebauer KM (2005) Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell* **19:** 53-63

Graveley BR (2005) Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123:** 65-73

Grishin NV (2001) KH domain: one motif, two folds. *Nucleic Acids Res* **29:** 638-643

Gullerova M, Barta A, Lorkovic ZJ (2006) AtCyp59 is a multidomain cyclophilin from Arabidopsis thaliana that interacts with SR proteins and the C-terminal domain of the RNA polymerase II. *RNA* **12:** 631-643

Gullerova M, Barta A, Lorkovic ZJ (2007) Rct1, a nuclear RNA recognition motif-containing cyclophilin, regulates phosphorylation of the RNA polymerase II C-terminal domain. *Mol Cell Biol* **27:** 3601-3611

Hall KB, Stump WT (1992) Interaction of N-terminal domain of U1A protein with an RNA stem/loop. *Nucleic Acids Res* **20:** 4283-4290

Hamilton GS, Steiner JP (1998) Immunophilins: beyond immunosuppression. *J Med Chem* **41:** 5119-5143

Handschumacher RE, Harding MW, Rice J, Drugge RJ, Speicher DW (1984) Cyclophilin: a specific cytosolic binding protein for cyclosporin A. *Science* **226:** 544-547

Hartmuth K, Barta A (1986) In vitro processing of a plant pre-mRNA in a HeLa cell nuclear extract. *Nucleic Acids Res* **14:** 7513-7528

He Z, Li L, Luan S (2004) Immunophilins and parvulins. Superfamily of peptidyl prolyl isomerases in Arabidopsis. *Plant Physiol* **134:** 1248-1267

Hirose T, Sugita M, Sugiura M (1993) cDNA structure, expression and nucleic acid-binding properties of three RNA-binding proteins in tobacco: occurrence of tissue-specific alternative splicing. *Nucleic Acids Res* **21:** 3981-3987

Hirose Y, Tacke R, Manley JL (1999) Phosphorylated RNA polymerase II stimulates pre-mRNA splicing. *Genes Dev* **13:** 1234-1239

Hom RA, Chang PY, Roy S, Musselman CA, Glass KC, Selezneva AI, Gozani O, Ismagilov RF, Cleary ML, Kutateladze TG (2010) Molecular mechanism of MLL PHD3 and RNA recognition by the Cyp33 RRM domain. *J Mol Biol* **400:** 145-154

Honig A, Auboeuf D, Parker MM, O'Malley BW, Berget SM (2002) Regulation of alternative splicing by the ATP-dependent DEAD-box RNA helicase p72. *Mol Cell Biol* **22:** 5698-5707

Horowitz DS, Lee EJ, Mabon SA, Misteli T (2002) A cyclophilin functions in pre-mRNA splicing. *EMBO J* **21:** 470-480

Huang Y, Steitz JA (2001) Splicing factors SRp20 and 9G8 promote the nucleocytoplasmic export of mRNA. *Mol Cell* **7:** 899-905

Huang Y, Steitz JA (2005) SRprises along a messenger's journey. *Mol Cell* **17:** 613-615

Hunt AG (2008) Messenger RNA 3' end formation in plants. *Curr Top Microbiol Immunol* **326:** 151-177

Hutchison S, LeBel C, Blanchette M, Chabot B (2002) Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem* **277:** 29745-29752

Isshiki M, Tsumoto A, Shimamoto K (2006) The serine/arginine-rich protein family in rice plays important roles in constitutive and alternative splicing of pre-mRNA. *Plant Cell* **18:** 146-158

James TD, Cashel M, Hinton DM (2010) A mutation within the {beta} subunit of Escherichia coli RNA polymerase impairs transcription from bacteriophage T4 middle promoters. *J Bacteriol*

Johnson L, Cao X, Jacobsen S (2002) Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr Biol* **12:** 1360-1367

Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* **12:** 5-14

Kallen J, Sedrani R, Zenke G, Wagner J (2005) Structure of human cyclophilin A in complex with the novel immunosuppressant sanglifehrin A at 1.6 A resolution. *J Biol Chem* **280:** 21965-21971

Kalyna M, Barta A (2004) A plethora of plant serine/arginine-rich proteins: redundancy or evolution of novel gene functions? *Biochem Soc Trans* **32:** 561-564

Kalyna M, Lopato S, Barta A (2003) Ectopic expression of atRSZ33 reveals its function in splicing and causes pleiotropic changes in development. *Mol Biol Cell* **14:** 3565-3577

Kalyna M, Lopato S, Voronin V, Barta A (2006) Evolutionary conservation and regulation of particular alternative splicing events in plant SR proteins. *Nucleic Acids Res* **34:** 4395-4405

Kameoka S, Duque P, Konarska MM (2004) p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *EMBO J* **23:** 1782-1791

Kieffer LJ, Seng TW, Li W, Osterman DG, Handschumacher RE, Bayney RM (1993) Cyclophilin-40, a protein with homology to the P59 component of the steroid receptor complex. Cloning of the cDNA and further characterization. *J Biol Chem* **268:** 12303-12310

Kim S, Shi H, Lee DK, Lis JT (2003) Specific SR protein-dependent splicing substrates identified through genomic SELEX. *Nucleic Acids Res* **31:** 1955-1961

Kim YK, Kim VN (2007) Processing of intronic microRNAs. *EMBO J* **26:** 775-783

Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J (2009) Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat Genet* **41:** 376-381

Komarnitsky P, Cho EJ, Buratowski S (2000) Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev* **14:** 2452-2460

Kops O, Zhou XZ, Lu KP (2002) Pin1 modulates the dephosphorylation of the RNA polymerase II C-terminal domain by yeast Fcp1. *FEBS Lett* **513:** 305-311

Kornblihtt AR (2006) Chromatin, transcript elongation and alternative splicing. *Nat Struct Mol Biol* **13:** 5-7

Kotlajich MV, Crabb TL, Hertel KJ (2009) Spliceosome assembly pathways for different types of alternative splicing converge during commitment to splice site pairing in the A complex. *Mol Cell Biol* **29:** 1072-1082

Kramer A (1996) The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem* **65:** 367-409

Krecic AM, Swanson MS (1999) hnRNP complexes: composition, structure, and function. *Curr Opin Cell Biol* **11:** 363-371

Lacadie SA, Rosbash M (2005) Cotranscriptional spliceosome assembly dynamics and the role of U1 snRNA:5'ss base pairing in yeast. *Mol Cell* **19:** 65-75

Lacadie SA, Tardiff DF, Kadener S, Rosbash M (2006) In vivo commitment to yeast cotranscriptional splicing is sensitive to transcription elongation mutants. *Genes Dev* **20:** 2055-2066

Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcarcel J (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* **109:** 285-296

Lamond AI, Spector DL (2003) Nuclear speckles: a model for nuclear organelles. *Nat Rev Mol Cell Biol* **4:** 605-612

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T,

Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ (2001) Initial sequencing and analysis of the human genome. *Nature* **409:** 860-921

Li Q, Lee JA, Black DL (2007) Neuronal regulation of alternative pre-mRNA splicing. *Nat Rev Neurosci* **8:** 819-831

Lin S, Coutinho-Mansfield G, Wang D, Pandit S, Fu XD (2008) The splicing factor SC35 has an active role in transcriptional elongation. *Nat Struct Mol Biol* **15:** 819-826

Lin S, Fu XD (2007) SR proteins and related factors in alternative splicing. *Adv Exp Med Biol* **623:** 107-122

Long JC, Caceres JF (2009) The SR protein family of splicing factors: master regulators of gene expression. *Biochem J* **417:** 15-27

Loomis RJ, Naoe Y, Parker JB, Savic V, Bozovsky MR, Macfarlan T, Manley JL, Chakravarti D (2009) Chromatin binding of SRp20 and ASF/SF2 and dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation. *Mol Cell* **33:** 450-461

Lopato S, Forstner C, Kalyna M, Hilscher J, Langhammer U, Indrapichate K, Lorkovic ZJ, Barta A (2002) Network of interactions of a novel plant-specific Arg/Ser-rich protein, atRSZ33, with atSC35-like splicing factors. *J Biol Chem* **277:** 39989-39998

Lopato S, Gattoni R, Fabini G, Stevenin J, Barta A (1999) A novel family of plant splicing factors with a Zn knuckle motif: examination of RNA binding and splicing activities. *Plant Mol Biol* **39:** 761-773

Lorenz C, Gesell T, Zimmermann B, Schoeberl U, Bilusic I, Rajkowitsch L, Waldsich C, von Haeseler A, Schroeder R (2010) Genomic SELEX for Hfq-binding RNAs identifies genomic aptamers predominantly in antisense transcripts. *Nucleic Acids Res* **38:** 3794-3808

Lorenz C, von Pelchrzim F, Schroeder R (2006) Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat Protoc* **1:** 2204-2212

Lorkovic ZJ, Hilscher J, Barta A (2008) Co-localisation studies of Arabidopsis SR splicing factors reveal different types of speckles in plant cell nuclei. *Exp Cell Res* **314:** 3175-3186

Lorkovic ZJ, Wieczorek Kirk DA, Lambermon MH, Filipowicz W (2000) Pre-mRNA splicing in higher plants. *Trends Plant Sci* **5:** 160-167

Lu KP, Zhou XZ (2007) The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease. *Nat Rev Mol Cell Biol* **8:** 904-916

Lu KP, Finn G, Lee TH, Nicholson LK (2007) Prolyl cis-trans isomerization as a molecular timer. *Nat Chem Biol* **3:** 619-629

Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP (1993) Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* **73:** 1067-1078

Manos PJ, Goldthwaite J (1976) An improved cytokinin bioassay using cultured soybean hypocotyl sections. *Plant Physiol* **57:** 894-897

Maris C, Dominguez C, Allain FH (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* **272:** 2118-2131

Marivet J, Frendo P, Burkard G (1995) DNA sequence analysis of a cyclophilin gene from maize: developmental expression and regulation by salicylic acid. *Mol Gen Genet* **247:** 222-228

Marks AR (1996) Cellular functions of immunophilins. *Physiol Rev* **76:** 631-649

Martinez-Contreras R, Cloutier P, Shkreta L, Fisette JF, Revil T, Chabot B (2007) hnRNP proteins and splicing control. *Adv Exp Med Biol* **623:** 123-147

Matlin AJ, Clark F, Smith CW (2005) Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* **6:** 386-398

Mauger DM, Lin C, Garcia-Blanco MA (2008) hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. *Mol Cell Biol* **28:** 5403-5419

McCracken S, Rosonina E, Fong N, Sikes M, Beyer A, O'Hare K, Shuman S, Bentley D (1998) Role of RNA polymerase II carboxy-terminal domain in coordinating transcription with RNA processing. *Cold Spring Harb Symp Quant Biol* **63:** 301-309

Meinhart A, Kamenski T, Hoeppner S, Baumli S, Cramer P (2005) A structural perspective of CTD function. *Genes Dev* **19:** 1401-1415

Melamud E, Moult J (2009) Stochastic noise in splicing machinery. *Nucleic Acids Res* **37:** 4873-4886

Mermoud JE, Cohen PT, Lamond AI (1994) Regulation of mammalian spliceosome assembly by a protein phosphorylation mechanism. *EMBO J* **13:** 5679-5688

Mi H, Kops O, Zimmermann E, Jaschke A, Tropschug M (1996) A nuclear RNA-binding cyclophilin in human T cells. *FEBS Lett* **398:** 201-205

Michlewski G, Sanford JR, Caceres JF (2008) The splicing factor SF2/ASF regulates translation initiation by enhancing phosphorylation of 4E-BP1. *Mol Cell* **30:** 179-189

Misteli T, Caceres JF, Spector DL (1997) The dynamics of a pre-mRNA splicing factor in living cells. *Nature* **387:** 523-527

Moore MJ, Proudfoot NJ (2009) Pre-mRNA processing reaches back to transcription and ahead to translation. *Cell* **136:** 688-700

Moore MJ, Silver PA (2008) Global analysis of mRNA splicing. *RNA* **14:** 197-203

Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ (2008) Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* **15:** 902-909

Muto Y, Pomeranz Krummel D, Oubridge C, Hernandez H, Robinson CV, Neuhaus D, Nagai K (2004) The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol* **341:** 185-198

Ngo JC, Chakrabarti S, Ding JH, Velazquez-Dones A, Nolen B, Aubol BE, Adams JA, Fu XD, Ghosh G (2005) Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol Cell* **20:** 77-89

Ngo JC, Giang K, Chakrabarti S, Ma CT, Huynh N, Hagopian JC, Dorrestein PC, Fu XD, Adams JA, Ghosh G (2008) A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. *Mol Cell* **29:** 563-576

Niranjanakumari S, Lasda E, Brazas R, Garcia-Blanco MA (2002) Reversible cross-linking combined with immunoprecipitation to study RNA-protein interactions in vivo. *Methods* **26:** 182-190

Noble CG, Hollingworth D, Martin SR, Ennis-Adeniran V, Smerdon SJ, Kelly G, Taylor IA, Ramos A (2005) Key features of the interaction between Pcf11 CID and RNA polymerase II CTD. *Nat Struct Mol Biol* **12:** 144-151

Oubridge C, Ito N, Evans PR, Teo CH, Nagai K (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372:** 432-438

Pahlke D, Freund C, Leitner D, Labudde D (2005) Statistically significant dependence of the Xaa-Pro peptide bond conformation on secondary structure and amino acid sequence. *BMC Struct Biol* **5:** 8

Peltier JB, Emanuelsson O, Kalume DE, Ytterberg J, Friso G, Rudella A, Liberles DA, Soderberg L, Roepstorff P, von Heijne G, van Wijk KJ (2002) Central functions of the lumenal and peripheral thylakoid proteome of Arabidopsis determined by experimentation and genome-wide prediction. *Plant Cell* **14:** 211-236

Peterlin BM, Price DH (2006) Controlling the elongation phase of transcription with P-TEFb. *Mol Cell* **23:** 297-305

Phatnani HP, Greenleaf AL (2006) Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev* **20:** 2922-2936

Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* **108:** 501-512

Pushkarsky T, Zybarth G, Dubrovsky L, Yurchenko V, Tang H, Guo H, Toole B, Sherry B, Bukrinsky M (2001) CD147 facilitates HIV-1 infection by interacting with virus-associated cyclophilin A. *Proc Natl Acad Sci U S A* **98:** 6360-6365

Query CC, Moore MJ, Sharp PA (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev* **8:** 587-597

Rajkowitsch L, Schroeder R (2007) Dissecting RNA chaperone activity. *RNA* **13:** 2053-2060

Ranganathan R, Lu KP, Hunter T, Noel JP (1997) Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent. *Cell* **89:** 875-886

Rappsilber J, Ryder U, Lamond AI, Mann M (2002) Large-scale proteomic analysis of the human spliceosome. *Genome Res* **12:** 1231-1245

Ratajczak T, Carrello A, Mark PJ, Warner BJ, Simpson RJ, Moritz RL, House AK (1993) The cyclophilin component of the unactivated estrogen receptor contains a tetratricopeptide repeat domain and shares identity with p59 (FKBP59). *J Biol Chem* **268:** 13187-13192

Ray D, Kazan H, Chan ET, Pena Castillo L, Chaudhry S, Talukder S, Blencowe BJ, Morris Q, Hughes TR (2009) Rapid and systematic analysis of the RNA recognition specificities of RNA-binding proteins. *Nat Biotechnol* **27:** 667-670

Reddy AS (2007) Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu Rev Plant Biol* **58:** 267-294

Reed R (2000) Mechanisms of fidelity in pre-mRNA splicing. *Curr Opin Cell Biol* **12:** 340-345

Reyes-Reyes M, Hampsey M (2007) Role for the Ssu72 C-terminal domain phosphatase in RNA polymerase II transcription elongation. *Mol Cell Biol* **27:** 926-936

Ritchie DB, Schellenberg MJ, MacMillan AM (2009) Spliceosome structure: piece by piece. *Biochim Biophys Acta* **1789:** 624-633

Robberson BL, Cote GJ, Berget SM (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10:** 84-94

Romano PG, Horton P, Gray JE (2004) The Arabidopsis cyclophilin gene family. *Plant Physiol* **134:** 1268-1282

Rossi F, Labourier E, Forne T, Divita G, Derancourt J, Riou JF, Antoine E, Cathala G, Brunel C, Tazi J (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature* **381:** 80-82

Sachs AB, Sarnow P, Hentze MW (1997) Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* **89:** 831-838

Sanford JR, Gray NK, Beckmann K, Caceres JF (2004) A novel role for shuttling SR proteins in mRNA translation. *Genes Dev* **18:** 755-768

Sapra AK, Anko ML, Grishina I, Lorenz M, Pabis M, Poser I, Rollins J, Weiland EM, Neugebauer KM (2009) SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Mol Cell* **34:** 179-190

Sato H, Hosoda N, Maquat LE (2008) Efficiency of the pioneer round of translation affects the cellular site of nonsense-mediated mRNA decay. *Mol Cell* **29:** 255-262

Sauliere J, Sureau A, Expert-Bezancon A, Marie J (2006) The polypyrimidine tract binding protein (PTB) represses splicing of exon 6B from the beta-tropomyosin pre-mRNA by directly interfering with the binding of the U2AF65 subunit. *Mol Cell Biol* **26:** 8755-8769

Saunders A, Core LJ, Lis JT (2006) Breaking barriers to transcription elongation. *Nat Rev Mol Cell Biol* **7:** 557-567

Schiene C, Fischer G (2000) Enzymes that catalyse the restructuring of proteins. *Curr Opin Struct Biol* **10:** 40-45

Schneider D, Gold L, Platt T (1993) Selective enrichment of RNA species for tight binding to Escherichia coli rho factor. *FASEB J* **7:** 201-207

Scholz C, Schindler T, Dolinski K, Heitman J, Schmid FX (1997) Cyclophilin active site mutants have native prolyl isomerase activity with a protein substrate. *FEBS Lett* **414:** 69-73

Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M (2003) Structural basis for the molecular recognition between human splicing factors U2AF65 and SF1/mBBP. *Mol Cell* **11:** 965-976

Sharma S, Kohlstaedt LA, Damianov A, Rio DC, Black DL (2008) Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. *Nat Struct Mol Biol* **15:** 183-191

Shen H, Kan JL, Green MR (2004) Arginine-serine-rich domains bound at splicing enhancers contact the branchpoint to promote prespliceosome assembly. *Mol Cell* **13:** 367-376

Shimada T, Fujita N, Maeda M, Ishihama A (2005) Systematic search for the Cra-binding promoters using genomic SELEX system. *Genes Cells* **10:** 907-918

Shin C, Feng Y, Manley JL (2004) Dephosphorylated SRp38 acts as a splicing repressor in response to heat shock. *Nature* **427:** 553-558

Shuman S (2001) Structure, mechanism, and evolution of the mRNA capping apparatus. *Prog Nucleic Acid Res Mol Biol* **66:** 1-40

Simpson CG, Fuller J, Maronova M, Kalyna M, Davidson D, McNicol J, Barta A, Brown JW (2008) Monitoring changes in alternative precursor messenger RNA splicing in multiple gene transcripts. *Plant J* **53:** 1035-1048

Simpson CG, Manthri S, Raczynska KD, Kalyna M, Lewandowska D, Kusenda B, Maronova M, Szweykowska-Kulinska Z, Jarmolowski A, Barta A, Brown JW (2010) Regulation of plant gene expression by alternative splicing. *Biochem Soc Trans* **38:** 667-671

Singer BS, Shtatland T, Brown D, Gold L (1997) Libraries for genomic SELEX. *Nucleic Acids Res* **25:** 781-786

Singh NN, Singh RN, Androphy EJ (2007) Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res* **35:** 371-389

Spellman R, Smith CW (2006) Novel modes of splicing repression by PTB. *Trends Biochem Sci* **31:** 73-76

Stamnes MA, Shieh BH, Chuman L, Harris GL, Zuker CS (1991) The cyclophilin homolog ninaA is a tissue-specific integral membrane protein required for the proper synthesis of a subset of Drosophila rhodopsins. *Cell* **65:** 219-227

Sterner DA, Carlo T, Berget SM (1996) Architectural limits on split genes. *Proc Natl Acad Sci U S A* **93:** 15081-15085

Stewart DE, Sarkar A, Wampler JE (1990) Occurrence and role of cis peptide bonds in protein structures. *J Mol Biol* **214:** 253-260

Stiller JW, Cook MS (2004) Functional unit of the RNA polymerase II C-terminal domain lies within heptapeptide pairs. *Eukaryot Cell* **3:** 735-740

Takagaki Y, Ryner LC, Manley JL (1989) Four factors are required for 3'-end cleavage of pre-mRNAs. *Genes Dev* **3:** 1711-1724

Tazi J, Kornstadt U, Rossi F, Jeanteur P, Cathala G, Brunel C, Luhrmann R (1993) Thiophosphorylation of U1-70K protein inhibits pre-mRNA splicing. *Nature* **363:** 283-286

Tuerk C, Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249:** 505-510

Venkataraman K, Brown KM, Gilmartin GM (2005) Analysis of a noncanonical poly(A) site reveals a tripartite mechanism for vertebrate poly(A) site recognition. *Genes Dev* **19:** 1315-1327

Verdecia MA, Bowman ME, Lu KP, Hunter T, Noel JP (2000) Structural basis for phosphoserine-proline recognition by group IV WW domains. *Nat Struct Biol* **7:** 639-643

Vinciguerra P, Stutz F (2004) mRNA export: an assembly line from genes to nuclear pores. *Curr Opin Cell Biol* **16:** 285-292

Wachtel C, Manley JL (2009) Splicing of mRNA precursors: the role of RNAs and proteins in catalysis. *Mol Biosyst* **5:** 311-316

Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK, Breaker RR (2007) Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* **19:** 3437-3450

Wahl MC, Will CL, Luhrmann R (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell* **136:** 701-718

Wahle E, Ruegsegger U (1999) 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* **23:** 277-295

Waldmeier PC, Zimmermann K, Qian T, Tintelnot-Blomley M, Lemasters JJ (2003) Cyclophilin D as a drug target. *Curr Med Chem* **10:** 1485-1506

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470-476

Wang HY, Lin W, Dyck JA, Yeakley JM, Songyang Z, Cantley LC, Fu XD (1998) SRPK2: a differentially expressed SR protein-specific kinase involved in mediating the interaction and localization of pre-mRNA splicing factors in mammalian cells. *J Cell Biol* **140:** 737-750

Wang Y, Han R, Zhang W, Yuan Y, Zhang X, Long Y, Mi H (2008) Human CyP33 binds specifically to mRNA and binding stimulates PPIase activity of hCyP33. *FEBS Lett* **582:** 835-839

Warf MB, Berglund JA (2007) MBNL binds similar RNA structures in the CUG repeats of myotonic dystrophy and its pre-mRNA substrate cardiac troponin T. *RNA* **13:** 2238-2251

Warf MB, Diegel JV, von Hippel PH, Berglund JA (2009) The protein factors MBNL1 and U2AF65 bind alternative RNA structures to regulate splicing. *Proc Natl Acad Sci U S A* **106:** 9203-9208

Warf MB, Berglund JA (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem Sci* **35:** 169-178

Weighardt F, Cobianchi F, Cartegni L, Chiodi I, Villa A, Riva S, Biamonti G (1999) A novel hnRNP protein (HAP/SAF-B) enters a subset of hnRNP complexes and relocates in nuclear granules in response to heat shock. *J Cell Sci* **112 (Pt 10):** 1465-1476

Wen JD, Gray CW, Gray DM (2001) SELEX selection of high-affinity oligonucleotides for bacteriophage Ff gene 5 protein. *Biochemistry* **40:** 9300-9310

Wen JD, Gray DM (2004) Selection of genomic sequences that bind tightly to Ff gene 5 protein: primer-free genomic SELEX. *Nucleic Acids Res* **32:** e182

West S, Gromak N, Proudfoot NJ (2004) Human 5' --> 3' exonuclease Xrn2 promotes transcription termination at co-transcriptional cleavage sites. *Nature* **432:** 522-525

Wickens M, Anderson P, Jackson RJ (1997) Life and death in the cytoplasm: messages from the 3' end. *Curr Opin Genet Dev* **7:** 220-232

Wilcox CB, Rossettini A, Hanes SD (2004) Genetic interactions with C-terminal domain (CTD) kinases and the CTD of RNA Pol II suggest a role for ESS1 in transcription initiation and elongation in Saccharomyces cerevisiae. *Genetics* **167:** 93-105

Wu JY, Maniatis T (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75:** 1061-1070

Wu S, Romfo CM, Nilsen TW, Green MR (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature* **402:** 832-835

Xiao SH, Manley JL (1997) Phosphorylation of the ASF/SF2 RS domain affects both protein-protein and protein-RNA interactions and is necessary for splicing. *Genes Dev* **11:** 334-344

Xu Q, Modrek B, Lee C (2002) Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res* **30:** 3754-3766

Xu YX, Manley JL (2004) Pinning down transcription: regulation of RNA polymerase II activity during the cell cycle. *Cell Cycle* **3:** 432-435

Zhang J, Corden JL (1991) Identification of phosphorylation sites in the repetitive carboxyl-terminal domain of the mouse RNA polymerase II largest subunit. *J Biol Chem* **266:** 2290-2296

Zhang L, Doudna JA (2002) Structural insights into group II intron catalysis and branch-site selection. *Science* **295:** 2084-2088

Zhang Z, Krainer AR (2004) Involvement of SR proteins in mRNA surveillance. *Mol Cell* **16:** 597-607

Zhao H, Xing D, Li QQ (2009) Unique features of plant cleavage and polyadenylation specificity factor revealed by proteomic studies. *Plant Physiol* **151:** 1546-1556

Zhou XZ, Kops O, Werner A, Lu PJ, Shen M, Stoller G, Kullertz G, Stark M, Fischer G, Lu KP (2000) Pin1-dependent prolyl isomerization regulates dephosphorylation of Cdc25C and tau proteins. *Mol Cell* **6:** 873-883

Zhou Z, Licklider LJ, Gygi SP, Reed R (2002) Comprehensive proteomic analysis of the human spliceosome. *Nature* **419:** 182-185

Zhu J, Krainer AR (2000) Pre-mRNA splicing in the absence of an SR protein RS domain. *Genes Dev* **14:** 3166-3178

Zimmermann B, Bilusic I, Lorenz C, Schroeder R (2010) Genomic SELEX: A discovery tool for genomic aptamers. *Methods*

Zimmermann B, Gesell T, Chen D, Lorenz C, Schroeder R (2010) Monitoring genomic sequences during SELEX using high-throughput sequencing: neutral SELEX. *PLoS ONE* **5:** e9169

Zorio DA, Blumenthal T (1999) U2AF35 is encoded by an essential gene clustered in an operon with RRM/cyclophilin in Caenorhabditis elegans. *RNA* **5:** 487-494

6. Curriculum Vitae

OLGA O. BANNIKOVA

PERSONAL INFORMATION:
Schelleingasse 14-16/3/11
1040, Vienna, Austria
Phone: +4368110650412
E-mail: olga.bannikova@univie.ac.at

Date and place of birth:
18.02.1985
Chelyabinsk, Russian Federation

CURRENT STATUS

PhD international Fellow, ViennaBioCenter (VBC), University of Vienna, Vienna, Austria

Work Experience:

2007-present **Medical University**, Vienna, Austria

Doctoral Student Researcher, Department of Medical Biochemistry, Laboratory of Dr.Andrea Barta

- Developed screening protocol for identification RNA targets for proteins of interest. Results revealed insight into mechanism of transcription regulation near end of transcript in *Arabidopsis Thaliana;*
- Mentored two graduate students on experimental design, data analysis and thesis evaluation;
- Result led to contribution in the book "RNA splicing: The complete guide", by Wiley-Blackwell, *in pres;*

Education:

2007-present **University of Vienna**, Vienna, Austria

Doctoral student in Molecular Biology with future degree (*Dr. rer.nat*)

Successfully conducting project "*Finding RNA targets for cyclophilin atCyp59 by Genomic SELEX in Arabidopsis Thaliana",* which uncovered essential role of protein in connecting pre-mRNA processing and transcription

2002-2007 **Lomonossov Moscow State University**, Moscow, Russian Federation

Diploma degree (*Dipl*.) in Chemistry, Specialization Bioorganic Chemistry

Diploma thesis: *"Synthesis of 2'- pyrenyl modified oligonucleotides for Single Nucleotide Polymorphism determination in gene 23S RNA Helicobacter Pylori".* Project led to development of new screening method of single mutation in bacterial genome for laboratory use.

- Degree with distinctions, TOP10 students GPA (4.85/5.0);
- Had 2 semesters of economics and passed exam with distinction;

1997-2002 **Secondary school N8**, Kogalym, Russian Federation

- Intensive courses in Mathematics, Physics and English;
- Leaving certificate with honors;
- Excellent marks (5.0/5.0) cumulative GPA;

Professional skills:

Leadership:

- Practical course of Genetics for undergraduate students (>20 people), VBC, 2009. Design and set up experiments to teach students molecular biological techniques used in gene expression analysis;

- European Research Night, Vienna, 2009. Explain to public mechanism of alternative splicing in terms of game “How to combine words to create a sentence;

Communication and presentation:

- progress reports every 3-4 months;
- presentations at international conferences;
- teamwork and collaborations with scientific groups worldwide;
- working experience in international multilanguage institution;

Project management:

- coordination of scientific, financial and administrative aspects of projects;
- proven ability to work under pressure and meet deadlines;
- multitasking developed due to combination of successful academic work and extracurricular activities;

Analytical skills:

- analysis and interpretation of complex experimental data;
- problem solving with a combination of personal experience and expertise from others;

Other skills and achievements:

Languages:

- Russian: native
- English: Fluent, working languages at VBC, school with English language bias;
- German: intermediate (Mittelstufe B1 level);

Scholarships:

- VBC International PhD Program Fellowship, 2007-present;
- Second price in International student scientific conference “Lomonossov”, Moscow, 2007;
- Price in regional school competition in Chemistry, 2000.

Courses and seminars:

- Successful presentations in English, VBC, 2009;
- Sharpen your communication skills, VBC,2010;
- Series of seminars on Consulting in Life Sciences, Vienna, 2009-2010;
- IESE MBA Summer School, Barcelona, Spain, 2010
- BioBusiness Summer School, Amsterdam, Netherland, 2010

Other interests:

- Studying in internet-based Photography-school, 2010-present;
- Playing piano. Finished music school in class piano with honors, Siberia, 2000.
- Playing badminton.
- Counted cross-stitching
- Jigsaw puzzles