

THE HNRNP F AND HNRNP H' ARE  
MAMMALIAN 3' MRNA PROCESSING FACTORS ACTING THROUGH GUANINE RICH  
SEQUENCES

by

Serkan Alkan

BS in Molecular Biology and Genetics, Bogazici University, 1999

Submitted to the Graduate Faculty of The School of Medicine,  
Graduate Program In Biochemistry and Molecular Genetics  
in partial fulfillment of the requirements for the degree of  
Doctor of Philosophy

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

SCHOOL OF MEDICINE

This dissertation was presented

by

Serkan Alkan

It was defended on

June 22, 2004

and approved by

Neal DeLuca, PhD, Department of Molecular Genetics and Biochemistry

Paula Grabowski, PhD, Department of Biological Sciences

Baskaran Rajasekaran, PhD, Department of Molecular Genetics and Biochemistry

Martin Schmidt, PhD, Department of Molecular Genetics and Biochemistry

Christine Milcarek, PhD, Department of Immunology  
Thesis Advisor/ Dissertation Director

# THE HNRNP F AND HNRNP H' ARE MAMMALIAN 3' MRNA PROCESSING FACTORS ACTING THROUGH GUANINE RICH SEQUENCES

Serkan Alkan, PhD

University of Pittsburgh, 2004

Heterogenous nuclear ribonucleoproteins (hnRNPs) are predominantly nuclear and have been shown to be involved in multiple cellular processes. There are more than 20 members of hnRNP protein family and they all bind RNA. Some hnRNPs like the ones in hnRNP H family was shown to have sequence specific RNA binding activity. hnRNP H potein family includes hnRNP H1, H2, H3 and F. This study mapped the location of hnRNP H/F binding site on the SV40 Late (SVL) pre-mRNA relative to the polyA signal. The binding site is characterized down to five consecutive Guanines. Disruption of hnRNP H/F binding site on the SVL pre-mRNA had a negative effect on polyadenylation *in vitro* and this effect is translated to downregulated gene expression in Gfp based *in vivo* assays. The fact that this effect was seen with constructs having single polyA sites implicates a pronounced effect on alternative polyadenylation. The second part of this study involves microarray experiments with A20 B-cell lymphoma and AxJ Plasmacytoma cells. More than 400 genes were found to be differentially expressed in AxJ cells. While this group includes many known B-cell stage specific genes, there are also numerous novel differentially expressed transcripts. The list includes transcription, splicing and polyadenylation factors, too. Elongation factor RNA polymerase II 2 (ELL2) is induced in AxJ cells as assayed by microarray and RT-PCR experiments. The overexpression of hnRNP F was shown revert the B-cell clock, in so as far as secretory to membrane Immunoglobulin ratio is concerned. In an attempt to identify the genes underlying this mechanism, microarray experiments were preformed with hnRNP F overexpressing AxJ plasmacytoma cells and

compared the gene expression profile with the one of mock transfected cells. There are about 30 differentially expressed transcripts upon forced expression of hnRNP F. Three B-cell stage specific genes were differentially expressed in hnRNP F overexpressing AxJ cells. ELL2 is among these genes. ELL2 is a known transcriptional elongation enhancer. A novel model regarding the secretory to membrane Immunoglobulin switch and involving ELL2 is proposed.

## **ACKNOWLEDGMENTS**

First and foremost, I would like to thank all the members of my family including Gral, Serdar and Sheyla Alkan for their sustained support for every aspect of my studentship. The stimulating working environment created by the Milcarek laboratory members was also greatly appreciated. I am grateful to all my committee members for their input and directions for my thesis research. I would like to thank Christine Milcarek for her professional supervision of my project and motivating me to attend meetings and seminars. Technical assistance and advice from members of Molecular Genetics& Biochemistry and Immunology Department has been invaluable throughout my studentship.

## TABLE OF CONTENTS

LIST OF TABLES .....	vii
LIST OF FIGURES .....	viii
ABBREVIATION LIST .....	ix
1. STATEMENT OF PROBLEM.....	1
2. INTRODUCTION: to the role of the hnRNP F/ H proteins in B-cell development .....	2
2.1. POSTTRANSLATIONAL MODIFICATION .....	7
2.2. TRANSCRIPTION .....	9
2.3. SPLICING.....	10
2.3.1. SILENCER .....	10
2.3.2. ENHANCER.....	12
2.4. POLYADENYLATION .....	13
3. COMMON METHODS.....	14
4. MANUSCRIPT ONE: REGULATION OF GENE EXPRESSION BY HNRNP H' AND F THROUGH FIVE CONSECUTIVE GUANINES IN 3'UTR .....	32
4.1. ABSTRACT.....	32
4.2. INTRODUCTION .....	33
4.3. RESULTS .....	36
4.3.1. Mapping the binding site of hnRNP F on SVL pre-mRNA.....	36
4.3.2. The mutations that significantly decrease hnRNP H'/F affinities decrease polyadenylation efficiency.....	43
4.3.3. Overexpression and downregulation of hnRNP F and H has opposite effects on gene expression.....	46
4.4. DISCUSSION.....	54
4.5. MATERIALS AND METHODS.....	58
5. MANUSCRIPT TWO: GLOBAL GENE EXPRESSION ANALYSIS REVEALS NOVEL B-CELL STAGE SPECIFIC GENES THAT ARE REGULATED BY HNRNP F.....	67
5.1. ABSTRACT.....	67
5.2. INTRODUCTION .....	68
5.3. RESULTS .....	71
5.4. CONCLUSIONS.....	81
5.5. METHODS .....	85
6. DISCUSSION.....	86
7. GENERAL DISCUSSION AND CONCLUSION.....	90
8. FUTURE DIRECTIONS .....	92
BIBLIOGRAPHY.....	96

## LIST OF TABLES

Table 1	Analysis of human hnRNP H protein family.....	4
Table 2	The sequence of the SVL based transcripts used in this study. ....	16
Table 3	Oligonucleotides used for <i>in vitro</i> transcription reactions. ....	18
Table 4	Oligonucleotides used for PCR, RT-PCR or cloning based experiments.....	23
Table 5	The sequence of wt and mutant GRS oligoribonucleotides.....	53
Table 6	Apparent dissociation constants (nM) of hnRNP F and H' proteins to wtSVL and SAA20 probes. ....	53
Table 7	Transcripts that are induced in AxJ Plasmacytoma cells.....	73
Table 8	Transcripts that are repressed in AxJ Plasmacytoma cells. ....	75
Table 9	Differentially expressed transcripts upon forced expression of hnRNP F in AxJ Plasmacytoma cells.....	78
Table 10	B-cell specific genes that are regulated by forced hnRNP F expression. ....	80

## LIST OF FIGURES

Figure 1	Domain structure of hnRNP H family proteins. Boxes stand for RNA Recognition Motif (RRM).....	4
Figure 2	Multiple alignment of experimentally proven hnRNP F/H binding RNAs by using Clustal W software.....	5
Figure 3	Multiple alignment of hnRNP H protein family by using BCM software.....	6
Figure 4	Short ribooligonucleotides were purified by using G-25 Superfine Sephadex columns.....	20
Figure 5	Plasmid maps that were used in this study.....	27
Figure 6	DNA sequencing chromatogram showing the PCR generated SAA18, SAA20 and SAA10 mutations located in pEGFP based plasmids. The wild type SVL sequence is given at the top.....	29
Figure 7	Different SVL pre-mRNA based probes that were used in this study.....	39
Figure 8	hnRNP F binds to Guanine rich sequence in SVL pre-mRNA specifically. ....	40
Figure 9	hnRNP F and H' bind to five consecutive Guanines in SVL pre-mRNA. ....	41
Figure 10	hnRNP H' binds to SVL pre-mRNA with a higher affinity compared to hnRNP F. ....	42
Figure 11	Diminishing the binding of hnRNP H' to SVL pre-mRNA decreases the polyadenylation efficiency <i>in vitro</i> . ....	44
Figure 12	Having no polyadenylation site reduces the GFP expression dramatically <i>in vivo</i> as assayed by fluorescent microscopy.....	49
Figure 13	Decreasing the binding of hnRNP H' to 3'UTR of gfp transcript reduces the GFP expression <i>in vivo</i> . ....	50
Figure 14	Forced expression of hnRNP F and H resulted in induction of GFP expression.....	51
Figure 15	Downregulation of hnRNP F and H by siRNAs diminished the GFP expression....	52
Figure 16	hnRNP F was purified by ion-exchange and metal affinity columns from E. coli extracts. BI, AI, SPT and FLT stand for Before Induction, After Induction, Supernatant and Flowthrough, respectively.....	63
Figure 17	hnRNP H was purified by glutathione-spearose and ion-exchange columns from E. coli extracts.....	64
Figure 18	Alternative processing of Ig $\gamma$ chain pre-mRNA result either in secretory or membrane specific IG.....	70
Figure 19	Venn diagram showing that three transcripts are differentially expressed both in Plasma cells and hnRNP F overexpressing AxJ cells.....	79
Figure 20	ELL2 is induced in AxJ cells compared to A20 cells as assayed by RT-PCR .....	80
Figure 21	Model: Ell2 Expression Results In More Processive RNA Polymerase II In Ig Secreting Cells So That It Skips the Membrane Specific 5' Splice Site .....	84

## ABBREVIATION LIST

A20	A20.2J mouse lymphoma cell line
ATP	adenosine 5'-triphosphate
APS	ammonium persulfate
AxJ	AxJ mouse plasmacytoma cell line
bp	base pair(s)
CBP	cap binding protein (complex)
cDNA	complementary DNA
CPSF	cleavage and polyadenylation specificity factor
CstF	cleavage-polyadenylation stimulation factor
CTD	carboxy-terminal domain of RNA polymerase II
DEAE	diethyl aminoethyl
DMEM	Dulbecco's modified Eagle's medium
DNase	deoxyribonuclease
DTT	dithiothreitol
EDTA	(Ethylenedinitrilo)tetraacetic acid
EMSA	ElectroMobility Shift Assay
fmol	femtomole(s)
<i>g</i>	gravity
<i>g</i>	gram(s)
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
<i>h</i>	hour(s)

HEPES	N-[2-Hydroxyethyl]piperazine-N'-[2-ethanesulphonic acid]
hnRNP	heterogeneous nuclear ribonucleoprotein
Ig	immunoglobulin
IVT	<i>In vitro</i> transcription
kD	kilodaltons
LB	Luria-Bertaini bacterial growth medium
M	molar
m <sup>7</sup> G	7-methyl guanine
mb	membrane-specific poly(A) site or membrane-specific mRNA
mg	milligram
ml	milliliter
min	minute(s)
mM	millimolar
mol	mole(s)
mRNA	messenger RNA
nt	nucleotide(s)
NTP	nucleoside 5'-triphosphate
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PBST	phosphate-buffered saline-TWEEN 20
PCR	polymerase chain reaction
PVA	polyvinyl alcohol
RNAP II	RNA polymerase II

PolyA	polyadenylation
RNase	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
sec	secretory-specific poly(A) site or secretory-specific mRNA; second(s)
SELEX	Systematic evolution of ligands by exponential enrichment
SVL	simian virus 40 late
TBE	Tris-borate-EDTA buffer
TE	Tris-EDTA buffer
TEMED	N,N,N',N'-tetramethylethylenediamine
UTR	Untranslated Region
tRNA	transfer RNA
μg	microgram
μl	microliter
μM	micromolar
UV	ultraviolet

## 1. STATEMENT OF PROBLEM

While the polyadenylation reaction requires multiple basal factors, it has been hypothesized that there are tissue specific modifiers of polyadenylation. Our lab (Veraldi et al. 2001) isolated polyadenylation (pA) complexes from cell lines representing early B-cells and plasma cells. The two proteins hnRNP F and H were found in a larger complex when the pA complex was reconstituted from early/memory B-cell extract. The location of the binding region of these two proteins relative to the pA signal was unmapped; the determination of their binding region on RNA might help explain the mechanism of action of these two proteins. The positive effect of hnRNP H on SV40 late pre-mRNA polyadenylation *in vitro* has been reported (Bagga et al. 1998). On the other hand, the hnRNP F was shown to have a negative effect on 3' cleavage reaction *in vitro* (Veraldi et al. 2001). The question of whether hnRNP H and F have opposite effects on polyadenylation or if they act together in the same direction remains to be answered. Even though they may act together in the same direction, their overall effect on polyadenylation may be gene specific. Furthermore, despite the extensive similarity of these two proteins, the possibility of differential RNA binding exists. Independent of whether they are able to bind different sequences or not, the optimal hnRNP H and F RNA binding sites have yet to be characterized.

While hnRNP H and F have been studied well *in vitro*, evidence is lacking that they do indeed regulate gene expression in the cell. The genomic scale analysis of the effect of these two proteins on gene expression will reveal whether they have a global role or act on a subset of specific genes

## 2. INTRODUCTION: to the role of the hnRNP F/ H proteins in B-cell development

This general introduction is intended to give the background about hnRNP proteins specifically the biological activities of hnRNP F and H proteins. The hnRNP F and H proteins were shown to play a specific role in B-cell development which will be elaborated in the next sections.

Heterogenous nuclear ribonucleoproteins (hnRNPs) display sequence specific RNA binding activity and associate predominantly with RNA pol II transcripts (Krecic and Swanson 1999). Most of them are located exclusively in the nucleus while others shuttle between cytoplasm and nucleus. RNA binding activities of these proteins are conferred by either RNA Recognition Motifs (RRMs) or RGG domains or both. The name RGG comes from the common Arginine-Glycine-Glycine repeat these domains have. RRM containing hnRNPs have multiple RRM, usually at least one RRM is responsible for sequence specific binding whereas the other nonspecific RRM increase the affinity of the protein to the specific sequence. RRM is an about 80 amino acid (aa) domain that has conserved submotifs octamer RNP-1 and hexamer RNP-2. Each RRM can fold into a globular domain independently and this domain itself is capable of binding to RNA. On the other hand, there are cases in which domains other than RRM like RGG and Glycine rich domains are required for optimal binding. Multiple RRM hnRNPs require a synergy among different RRM for optimal or high affinity sequence specific binding. A typical RRM has a  $\beta 1$ - $\alpha 1$ - $\beta 2$ - $\beta 3$ - $\alpha 2$ - $\beta 4$  structure with RNP-1 and RNP-2 being in the central anti-parallel  $\beta 3$  and  $\beta 1$  strands, respectively.  $\beta 3$  and  $\beta 1$  strands were shown to interact RNA directly by NMR (Gorlach et al. 1992).

Some hnRNPs like hnRNP K and hnRNP F were shown to bind single- as well as double stranded DNA (Lacroix et al. 2000) (Yoshida et al. 1999). Among the many proposed functions

for hnRNP proteins are mRNA trafficking, splicing, telomere length control, mRNA stability, transcription, polyadenylation.

Although the RGG domain is present in hnRNP proteins frequently, the primary sequence of this domain varies among different proteins. For example, in nucleolin RGGF is the minimal repeat quartet whereas hnRNP A has only GG repeats. In addition, the RGG domain is distinct from Glycine hinge domains that have multiple consecutive Glycine repeats and flexible structures (Birney et al. 1993).

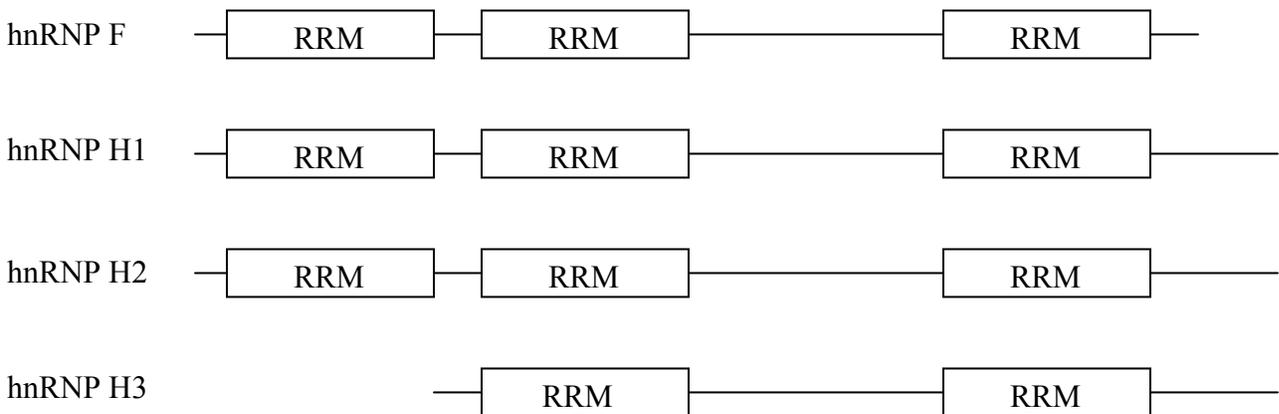
The hnRNP H protein family consists of hnRNP H1, H2 (H'), H3 (2H9) and hnRNP F (Table 1). The corresponding genes are *hnrp* H1, H2, H3 and *hnrp* F, respectively. Both hnRNP H1 and H2 (aka DSEF) are 449 aa proteins having three RRM. The domain structures of the hnRNP H protein family are illustrated in Figure 1. The 96% identity between these proteins may be due to a recent gene duplication event. On the other hand, hnRNP F is a 415 aa protein also having three RRM. Even though it shows a 72 and 70% identity to hnRNP H1 and H2, respectively, the corresponding RRM are more than 90% identical, strongly indicating that hnRNP F and hnRNP H1/H2 exhibit the same sequence selectivity. But hnRNP F lacks certain portions of the carboxy terminal domain present both in hnRNP H and H'. The 346 amino acid long hnRNP H3 is a relatively small variant member of the family that has been shown to have various alternatively spliced forms whose functions are not well characterized (Honore 2000). While all the members of hnRNP H family are Glycine rich (Table 1), a single amino acid Glycine makes up 24.3% of hnRNP H3.

Both hnRNP F and H have been shown to bind polyG with an affinity so high that the interactions can persist at 2M NaCl. While there is not a general consensus for the optimal

hnRNP F/H binding site, it is clear from numerous studies that multiple runs of Guanines are necessary and sufficient for hnRNP F/H binding.

**Table 1** Analysis of human hnRNP H protein family

Analysis	hnRNP F	hnRNP H1	hnRNP H2	hnRNP H3
Length	415 aa	449 aa	449 aa	346 aa
Molecular Weight	45.7 kD	49.2 kD	49.3 kD	36.9 kD
1 microgram =	21.9 pMol	20.3 pMol	20.3 pMol	27.1 pMol
Isoelectric Point	5.4	5.9	5.9	6.4
Charge at pH 7	-14.1	-7.2	-7.2	-2.3
Chromosomal location	10q11	5	X	10q22
# RRM	3	3	3	2
# Glycine (%)	51 (12.3%)	66 (14.7%)	65 (14.5%)	84 (24.3%)



**Figure 1** Domain structure of hnRNP H family proteins. Boxes stand for RNA Recognition Motif (RRM).

A SELEX assay has been performed neither with hnRNP F nor with H yet but possible quarternary Guanine structures may complicate the outcome. While the quadruplex Guanine structure may be important, even decisive, for binding of the hnRNP H protein family, it still does not change the primary recognition sequence. This structure may be pronounced in short oligonucleotides compared to long RNA molecules. In order to determine the consensus sequence in naturally occurring hnRNP H/F binding RNA molecules, all the experimentally hnRNP F/H bound RNAs were collected and a multiple alignment was performed (Figure 2). While the consensus sequence includes consecutive Guanines, an Adenine residue downstream of the Guanine-rich stretch is also highly conserved.

<b>V1loopcons</b>	1	-AATAGTAGTAGCGTGGGGAAA--ATGATGATGGAGAAAGGAGAAATAAA	47
<b>Cftr-exon9-155G</b>	1	TATTAATTTCAAGATAGAAAGAGGACAGTTGTTGGGGGTTGCTGGATCCA	50
<b>RSV</b>	1	-----GAAUCG-ACAAAGGGGAGGAAGUGGGAGAAA--	30
<b>GRSwt</b>	1	-----GCGCGAGGUGUGGG-----	14
<b>Throidhorm</b>	1	-GGCGGCCAGAGGGUGUGCGGAG- CUGGUGGGGAGGAGCUUGGAGAGAAG	48
<b>C-SRC</b>	1	-----CUGAGGCUGGGGGCUCGUCUCUGCA	25
<b>Tat</b>	1	-----UUAUCCAUUUCAGAAUUGGGUGU---	23
<b>BETA-TRP</b>	1	-----AGUAAAUGUGGGACCUAGAGGAGGAGCUGAAAUUGU	38
<b>WA19tat</b>	1	-----GCCAGUUAUGAUAGGGACUUAGGGUGGGAAGCAUCCAGGA-	40
<b>844ins68_CBS</b>	1	-----ATCACTGGGGTGGATCATCCAGGTGGGGCTTTT-----	33
<b>Consensus</b>	1	-----G- G G GGG A	7
	48	AAACTGC-----	54
	51	CTGGAGCAGGCAAG-----	64
	30	-----	30
	14	-----	14
	49	GGACAAAGCUGGGGGCUGAGGGAGAACCCCA-----	80
	26	UGUGCUUCCU-----	35
	23	-----	23
	39	UACCAACAACUUGAAAUCUCUGGAAGCCCAAGCGGACAAGGU	80
	40	-----	40
	33	-----	33
	7	-----	7

**Figure 2 Multiple alignment of experimentally proven hnRNP F/H binding RNAs by using Clustal W software**

ClustalW is available at <http://www.ebi.ac.uk/clustalw/> (December 2004). Consensus sequence is given at the bottom. Sequences were taken from (Caputi and Zahler 2002), (Pagani et al. 2003), (Fogel and McNally 2000), this study, (Hastings et al. 2001), (Chou et al. 1999), (Jacquet et al. 2001), (Chen et al. 1999), (Caputi and Zahler 2001), (Romano et al. 2002), respectively.

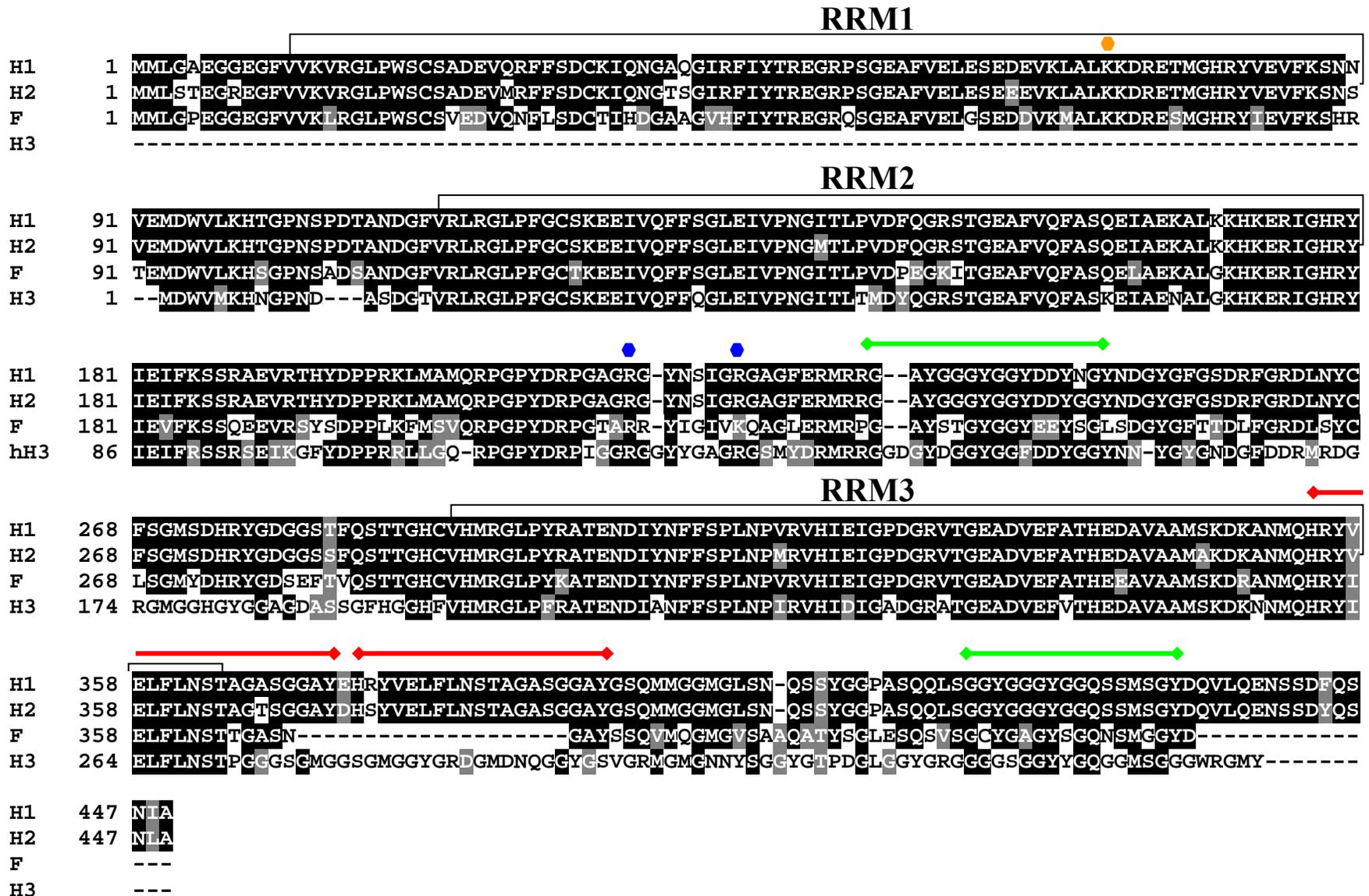


Figure 3 Multiple alignment of hnRNP H protein family by using BCM software.

RRMs are indicated with respect to hnRNP H1. Red and green lines stand for 19aa perfect and 16aa imperfect repeats on hnRNP H1, respectively. Blue hexagons indicate methylation sites whereas orange hexagons sumoylation sites

## 2.1. POSTTRANSLATIONAL MODIFICATION

The interaction of hnRNPs and RNA could be controlled at the level posttranslational modifications. Various hnRNPs were found to be phosphorylated, glycosylated, methylated and sumoylated. Despite the fact that the primary sequence of hnRNP F and H1 are 84% similar, their activities can be differentially modified posttranslationally.

The small ubiquitin-related modifier (SUMO) is an 11kD polypeptide that is attached to lysine residues in various proteins (Johnson 2004). SUMO is a member of the ubiquitin and ubiquitin-like superfamily. The SUMO consensus is a tetrapeptide of B-K-X-D/E where B stands for a hydrophobic amino acid, K is the lysine that is covalently linked to SUMO, X is any amino acid and D/E is an acidic residue. The hnRNP H1, H2 and F proteins have SUMO motif in their N-terminal region that is lacking in hnRNP H3 (Figure 3). The SUMO motif present in hnRNP H family is LKKD where the first amino terminal lysine is the residue SUMO is attached. hnRNP F and H1 were shown to be sumoylated experimentally (Mnaimneh et al. 2004). It is highly probable that the same modification also takes place with hnRNP H2. For hnRNP C and M, it has been proposed that Sumo modification takes place at the nuclear pore complexes (NPCs) (Vassileva and Matunis 2004) because the SUMO modification enzymes localize to NPCs. The suggested mechanism of action for SUMO modification in the case of hnRNP C and M is that the SUMO ligation decreases the affinity of hnRNP proteins to the mRNAs so that they can be transported through the nuclear pores. This mechanism maybe also valid for the hnRNP H family since the SUMO motif is present in the first RRM of hnRNP H1, H2 and F (Figure 3).

The majority of nuclear asymmetric dimethylarginine (DMA) residues are found in hnRNP proteins. From the sequence it is clear that hnRNP H1, H2 and H3 but not hnRNP F have potential asymmetric dimethylarginine residues (Figure 3). While the DMA consensus is given as

F/G-G-G-G-R-G-G-G/F (Kim et al. 1997), recent *in vitro* and *in vivo* studies unambiguously showed a second set of substrates that are loosely defined as having the R-X-R motif (Hyun et al. 2000). This motif, as an RGR triplet, is present twice in hnRNP H1, H2 and H3 but is not present in hnRNP F at all (Figure 3). Methylation was shown to modulate the protein-RNA and protein-protein interactions (McBride and Silver 2001). Dimethylation is covalent and the amino-alkyl bond is very stable and believed to be irreversible.

Initial studies found out that multiple hnRNPs have an uncommon modified amino acid: N<sup>G</sup>, N<sup>G</sup>-dimethylarginine. Arginine is the single methylated amino acid in hnRNP proteins and the major arginine methylated isoform is N<sup>G</sup>, N<sup>G</sup>-dimethylarginine (DMA) while N<sup>G</sup>-methylarginine is present in minute amounts. The hnRNPs contain about 65% DMA in the cell nucleus. When researchers (Honore et al. 1995) analyzed methylated hnRNPs in HeLa cells they found that hnRNP H but not hnRNP F was methylated as expected from the primary sequence (Figure 3).

The PolyA binding protein II (PABII) has been shown to be asymmetrically dimethylated at 16 different Arginines (Smith et al. 1999). Twelve of these modified Arginines are in the context of RXR where X is a small amino acid such Gly, Ala, Ser, Pro, Tyr. Peptide based studies also confirmed that the GRG sequence is the preferred substrate for methylation (Rawal et al. 1995).

While there is not experimental evidence that hnRNP H protein family is phosphorylated, they all have Serine and Tyrosine residues that score very high at NetPhos 2.0 software (<http://www.cbs.dtu.dk/services/NetPhos/>) and are potentially phosphorylated. Since there are so many potential phosphorylated amino acids, it is not possible to assess the differential effect, if any, of this modification to hnRNP H protein family members.

The main function of hnRNP proteins is to protect mRNA from degradation. As soon as the transcript is synthesized, the proteins cover the message body so that the pre-mRNA is not

sensitive to RNases. This binding can and does modulate half-life, splicing and polyadenylation of the transcript. The extent and the location of the hnRNP F/H binding dramatically alters the alternative splicing and polyadenylation pattern.

## 2.2. TRANSCRIPTION

The TATA-binding protein (TBP) binds to the TATA element to start the process of pre-initiation complex assembly at the RNAP II promoters; it remains at the promoter even after RNAP II starts to initiate transcript synthesis. It was found that hnRNP F associates with TBP (Yoshida et al. 1999) through immunoprecipitation. Interestingly, there are two isoforms of hnRNP F seen in 2D gels that have different isoelectric points and apparent molecular weights. One of the forms has a pI of 5.0 and molecular weight of about 50kD whereas the other one has a pI that is close to 6.0 and is about 10kD larger. The theoretical pI and MW of hnRNP F are 5.37 and 45.7 kD as calculated by Vector Nti, respectively (Table 1). So the second isoform is clearly a covalently modified form of hnRNP F and most possibly it is a SUMO modified form since the size of SUMO is roughly 11kD. Addition of SUMO affects the stability of hnRNP proteins {Vassileva, 2004 #1366}.

Although hnRNP F is expressed ubiquitously (Honore et al. 2004), it is present at high levels in thymus and spleen (Yoshida et al. 1999). Immunoprecipitation experiments showed a physical interaction between CTD of RNAP II and hnRNP F {Yoshida, 1999 #1233}. Studies hnRNP F was shown to bind to the SVL promoter specifically {Yoshida, 1999 #1043}. Recombinant hnRNP F RRM1 and RMM3 alone could bind to poly(rG) whereas an interaction could not be established with RRM2 alone. Interestingly Yoshida showed that the region between RMM2 and 3 is necessary and sufficient for polyG binding. This region consists of 38% charged amino acids

and may be an indication of nonspecific binding. The fact that hnRNP F binds to RNA and DNA by using nonoverlapping domains raises the possibility that it can bind to both at the very same time.

A yeast three hybrid screen was performed with untagged Cap Binding Protein 80 (CBP80) and GAL4-CBP20. The C-terminal two-thirds of hnRNP F was one of three positive clones. This interaction was verified by complementary *in vitro* binding assays and also extended to hnRNP H. Unlike hnRNP H1, hnRNP F preferentially binds to CBC-RNA complexes compared to naked RNA. This is most possibly due to the fact that the presence of CBC increases the overall affinity of hnRNP F to the RNA-CBC complex. The domain of hnRNP F responsible for CBP binding has not been determined yet. The binding of hnRNP F to the cap structure may further stabilize the transcript and prevent degradation. The minute amount of hnRNP F present in cytoplasm may also effect protein translation.

hnRNP F can bind to the TBP and RNAP II promoters, then load to the CTD of elongating RNAP II. It can either unload from CTD to the 5' cap of synthesized transcript and stabilize transcript or continue to travel with RNAP II.

## **2.3. SPLICING**

### **2.3.1. SILENCER**

The role of hnRNP H on splicing has been extensively studied with multiple different substrates. Given the complexity and multiplicity of the sequence elements affecting the single splice site choice, an auxiliary splicing factor may not act as a silencer or enhancer all the time. On the other hand, it is also clear from many studies, hnRNP H1 binds to silencers in most of the cases. As explained below, the mechanism of silencing differs in each case.

hnRNP H was shown to bind the Exonic Splicing Silencer (ESS) at the 5' end of the alternative exon 7 of rat  $\beta$ -tropomyosin gene (Chen et al. 1999). As the authors suggested, the silencing activity may be exerted by the disruption of SR protein association with exon 7 because a putative Exonic Splicing Enhancer (ESE) is located downstream of ESS.

In other studies, hnRNP H was shown to bind to another ESS at the 5' extremity of HIV-1 *tat* exon 2 and represses the upstream A3 splice site required for *tat* mRNA production (Jacquet et al. 2001). Silencing activity and binding of hnRNP H correlated in that study. It has been proposed that hnRNP H competes with U2AF<sup>35</sup> subunit for binding to the exon 5' extremity and exert its silencing effect by physically blocking the U2AF<sup>35</sup> binding site.

The 844ins68 is a frequent polymorphism of the Cystathionine  $\beta$ -Synthase (CBS) gene resulting from a 68 bp insertion. This insertion creates two identical 3' splice sites and dual closely spaced G-tracts that are just downstream to the proximal 3' ss (Romano et al. 2002). Binding of hnRNP H to Guanine runs inhibited the proximal 3' splice site and results in exclusive distal site utilization. Therefore, hnRNP H may either inhibit the recruitment of splicing factors to the proximal 3' ss or interact with other splicing factors to enhance the selection of the distal site. The fact that the block of G-rich sequences starts just 7 nt downstream of the proximal site favors the former hypothesis.

A single C toG substitution creates five consecutive Guanines in the Cystic Fibrosis Transmembrane Regulator (CFTR) exon 9, this in turn generates a new exonic splicing silencer and causes exon skipping (Pagani et al. 2003). The exon skipping activity was dominant to the adjacent enhancer and correlated with the binding of hnRNP H. A recent computational study aimed at identifying silencer elements that cause inhibition of pseudoexon splicing came up with three different motifs. One of three elements identified was U/GGU/AGGGG, which has been

shown to have silencing effect in transfection experiments (Sironi et al. 2004). The closest match to this sequence among the experimentally proven hnRNP H binding sites is the element GUUGGGGG, which was created with the above mentioned C-G mutation in CFTR exon 9 that resulted in reduced exon inclusion.

### **2.3.2. ENHANCER**

Alternatively processed exons are often characterized by suboptimal splice sites that are poorly recognized in the absence of enhancer sequences and trans-acting factors. HnRNP H protein family were shown to bind to several naturally occurring enhancer elements.

HnRNP H and F bind to purine rich intronic element called SE $\alpha$ 2 in the last intron of thyroid hormone receptor  $\alpha$ 2 (Tr $\alpha$ 2) message. This element consists of several consecutive Guanine residues. Binding of hnRNP F, hnRNP H and SF2/ASF to this enhancer results in the usage of upstream weak 5' splice site specific for Tr $\alpha$ 2 message (Hastings et al. 2001).

The inclusion of c-src N1 exon in cells of neuronal origin is accomplished by an intronic splicing enhancer downstream of the N1 5' splice site. The proteins that associate with this element include nPTB, hnRNP F, KSRP and hnRNP H. *In vitro* translated hnRNP F was shown to interact with recombinant hnRNP H at 100 mM NaCl (Chou et al. 1999) and therefore it is possible that hnRNP F and hnRNP H can exist as a heterodimer. HnRNP H is required for efficient splicing of c-src N1 exon (Markovtsov et al. 2000).

HnRNP H interacts also with an element just downstream of a point mutation of exon 6D in the HIV-1 Env gene. This interaction is required for the interaction of U1 snRNP with the enhancer. Binding of SC35 to the point mutant region may convert the hnRNP H-U1snRNP complex into a splicing enhancer. Assembly of components of U1 snRNP on Exon 6D requires hnRNP H binding (Caputi and Zahler 2002).

In most of these cases, hnRNP F was also shown to bind these elements albeit with a significantly lower affinity.

#### 2.4. POLYADENYLATION

The hnRNP H2 protein (aka DSEF) was shown to bind a downstream element of the SV40 Late (SVL) polyA region (Bagga et al. 1998). This element is G-rich and located 3' to the U-rich sequences to which CstF-64 binds. Several lines of evidence clearly showed that binding of hnRNP H2 to this element has a positive effect both on 3' cleavage and overall polyadenylation reactions. The effect of GRS has been shown to be position dependent; if it is placed more than 200 nt downstream of U-rich region, its effect would be abolished (Bagga et al. 1995).

The hnRNP F-GRS interaction has a negative effect on 3' cleavage reaction *in vitro*. Binding of hnRNP F but not hnRNP H2 inhibits the binding of CstF-64 both to the IgM and SVL pre-mRNA *in vitro* (Veraldi et al. 2001). Overexpression of hnRNP F in plasmacytoma cells resulted in slightly less secretory specific immunoglobulin (Veraldi et al. 2001) although membrane specific transcript abundance does not change. Depending on the Ig type, there are hnRNP F binding sites both downstream and upstream of the polyA signal. Even the negative effect of hnRNP F may be position dependent.

### 3. COMMON METHODS

#### RNA ELECTROMOBILITY SHIFT ASSAY PROTOCOL

Gel apparatus was set up with 1.5 mm spacers and comb at least an hour before starting the experiment. If the RNA probes are more than 50 nt, a 4% gel with 20:0.25 Acrylamide to bisacrylamide ratio was used. Native Gel Electrophoresis buffer contains 25mM Tris Base, 25 mM Boric acid, 1mM EDTA. 10 ml 20:0.25 Acrylamide solution, 40 ml Native Gel Electrophoresis buffer, 0.25g (0.5%) Agarose were mixed and this mix was heated to boiling and allowed to cool to 60<sup>0</sup>C. Then 375 ul 10% APS and 50 ul TEMED were added and the gel solution was immediately poured into the gel apparatus. The gel running buffer and polymerized gel were stored at least an hour at 4<sup>0</sup>C. If the RNA probes are less than 50 nt, then an 8% Native acrylamide (20:0.25) gel without agarose was used. For consistency, 8% gels were also boiled and cooled to 60<sup>0</sup>C and then APS and TEMED were added. This procedure diminished and usually eliminated RNA degradation during gel running.

A 25 tube mastermix was prepared each time. This mix contains: 7.5 ul purified Yeast tRNA (10mg/ml), 15 ul 100 mM DTT, 6 ul 100 mM ATP, 30 ul 500mM Creatine Phosphate, 15 ul 25 mM MgCl<sub>2</sub>, and 187.5 ul 10% PVA. 10 ul of this mix was added to each tube. Then 1 fmol (5 fmol if RNA probe is <50nt) radiolabeled RNA, protein and Buffer D is added to each tube. Every tube has 20 ul in total; volumes were equalized with Buffer D and proteins were always added last. Radiolabeled RNA were dissolved in 10mM TE whereas recombinant proteins were dissolved in Buffer D.

After mixing reaction components well, tube was spun briefly and incubated at 30<sup>0</sup>C for 20 minutes. Then tubes were put on ice and 2 ul of 6x DNA loading dye was added to each tube. Samples were loaded immediately to cold gel. The gels were run at 30 mAmp, 300V 15W for about 1-2 hours. Gel runs were always started with cold gel and cold buffer and were then run at

room temperature. It is very important to remember that the 14 nt GRS derived RNA probes run faster than Bromophenol blue. Then gels were dried on a gel dryer for 90 minutes and they were put either on the Phosphoimager or into autoradiography cassettes. Depending on the specific activity of the probes, Kodak X-omat film or phosphoimager screen was exposed 1-12 hours.

### **COMPETITION RNA EMSAs**

EMSAs with tritium labeled RNA competitors were done exactly as described above except that the reaction tubes contain one  $^{32}\text{P}$  labeled and one tritium labeled RNA probe.  $^{32}\text{P}$  labeled probes were added to tubes first and then tritium probes were added. As always, protein was added last. Autoradiography films or Phosphoimager screens were exposed for 1-12 hours. Tritium is used just to calculate concentration of competitors and tritium labeled probes are not visible on films or phosphoimager screen.

### **PLASMID BASED *IN VITRO* TRANSCRIPTION**

SV40 Late (SVL) pre-mRNA based transcripts were transcribed by using pSP65 plasmid having the SVL polyA region under the control of SP6 (A kind gift of Dr. Jeff Wilusz). The sequences of transcripts were given in Table 2. Templates for short SVL based transcripts were synthesized by PCR.

One microgram of template was used in a 20  $\mu\text{L}$  transcription reaction containing: 1x transcription buffer {40 mM Tris-HCl, pH 7.5 (SP6) or pH 7.9 (T7), 6 mM  $\text{MgCl}_2$ , 2 mM spermidine, 10 mM NaCl (SP6 only), 10 mM DTT (T7 only)} plus 12.5 mM DTT, 25  $\mu\text{M}$  rGTP, 500  $\mu\text{M}$  rCTP, rATP, rUTP each, , 20 units of RNasin<sup>TM</sup> ribonuclease inhibitor, 5 $\mu\text{L}$  of  $\alpha$ - $^{32}\text{P}$ -

**Table 2** The sequence of the SVL based transcripts used in this study.

<b>RNA Probe</b>	<b>Sequence</b>
247	GAAUACACGGAAUUCGAGCUCGCCCGGGGAUCCAGACAU GAUAAGAUACAUUGAUGAGUUUGGACAAACCACAACUAG AAUGCAGUGAAAAAAUGCUUUUUUGUGAAAUUUGUGA UGCUAUUGCUUUUUUGUAACCAUUUAAGCUGCAAUAA ACAAGUUAACAACAACAUAUGCAUUCAUUUUAUGUUUCA GGUUCAGGGGGAGGUGUGGGAGGUUUUUUAAAGCAAGUA AAACCUCUACAAAUGUGGUUAUGGCUGAUUA
248	GAAUACACGGAAUUCGAGCUCGCCCAACAACAACAUAUG CAUUCAUUUUAUGUUUCAGGUUCAGGGGGAGGUGUGGGA GGUUUUUUAAAGCAAGUAAAACCUCUACAAAUGUGGUUAU GGCUGAUUA
249	GAAUACACGGAAUUCGAGCUCGCCCGGGGAUCCAGACAU GAUAAGAUACAUUGAUGAGUUUGGACAAACCACAACUAG AAUGCAGUGAAAAAAUGCUUUUUUGUGAAAUUUGUGA UGCUAUUGCUUUUUUGUAACCAUUUAAGCUGCAAGAA ACAAGUUAACAACAACAUAUGCAUUCAUUUUAUGUUUCA GGUUCAGGGGGAGGUGUGGGAGGUUUUUUAAAGCAAGUA AAACCUCUACAAAUGUGGUUAUGGCUGAUUA
PCR-SVL- GEM (Probe D)	CAUACGAUUUAGGUGACACUAUAGAAUACACGGAAUUCG AGCUCGCCCGGGGAUCCAGACAUGAUAAAGAUACAUUGAU GAGUUUGGACAAACCACAACUAGAAUGCAGUGAAAAAA UGCUUUUUUUGUGAAAUUUGUGAUGCUAUUGCUUUUUU GUAACCAUUUAAGCUGCAAUAAACAAGUUAACAACAAC AAUUGCAUUCAUUUUAUGUUUCAGGUUCAGAAUACACGG AAUUCGAGCUCGGU
Probe F	CAUACGAUUUAGGUGACACUAUAGAAUACACGGAAUUCG AGCUCGCCCAACAACAACAUAUGCAUUCAUUUUAUGUUU CAGGUUCAGAAUACACGGAAUUCGAGCUCGGU
Probe E	GAAUACACGGAAUUCGAGCUCGCCCGGGGAUCCAGACAU GAUAAGAUACAUUGAUGAGUUUGGACAAACCACAACUAG AAUGCAGUGAAAAAAUGCUUUUUUGUGAAAUUUGUGA UGCUAUUGCUUUUUUGUAACCAUUUAAGCUGCAAUAA ACAAGUUAACAACAACAUAUGGGGAUCCUCUAGAGUCGAC CUGCAGGCAUGCAAGCUU
GRS	GGGGGAGGUGUGGG

GTP (3000 Ci/mmol; 10mCi/ml), and 50 units SP6 or T7 RNA polymerase. The reaction was incubated at 37<sup>0</sup>C for one hour. The DNA template was then digested by addition of 0.5 μL of 10 Units RNase-free DNase I with an incubation at 37°C for 15 minutes. The reaction was stopped by addition of 30 ul 0.1 M Tris and 50 ul 2xSDS buffer to a final volume of 100 ul. The radioactivity in 1 ul of the samples was determined with Scintillation counter. Then RNA transcripts were Phenol-Chloroform extracted and Ethanol precipitated and resuspended in 10 μL formamide loading dye (80% formamide, 1xTBE, 0.25% xyline cyanol, 0.25% bromophenol blue) and purified from a 5 % polyacrylamide-8M urea gel.

### ***IN VITRO* TRANSCRIPTION WITH DNA OLIGONUCLEOTIDE TEMPLATES**

All the transcription templates were ordered from DNA synthesis Center, Department of Biological Sciences Pittsburgh PA. Two oligos were used for each transcription: one only has the T7 promoter, the other having the antisense to the T7 promoter and antisense to wildtype or mutant GRS. This means for every oligo transcription., T7 promoter primer was always used, so only the second primer differed. The sequences of wild type and all 20 mutant GRS DNA oligo templates are given in Table 3. Mutations are also in antisense form but whenever they are transcribed, we get sense mutant GRS RNA Oligo.

A reaction mix containing 4 μg of each oligo was incubated at 95<sup>0</sup>C for 3 minutes in 0.025M EDTA and 10mM TE; then these samples were added to an *in vitro* transcription mix after being chilled for one minute on ice. The transcription reaction was performed as described below. We started with four ug of template because the rate limiting step in these reactions is transcriptional initiation.

A 14 nt scrambled RNA transcript SAA-SCR was created by using an identical strategy.

**Table 3****Oligonucleotides used for *in vitro* transcription reactions.**

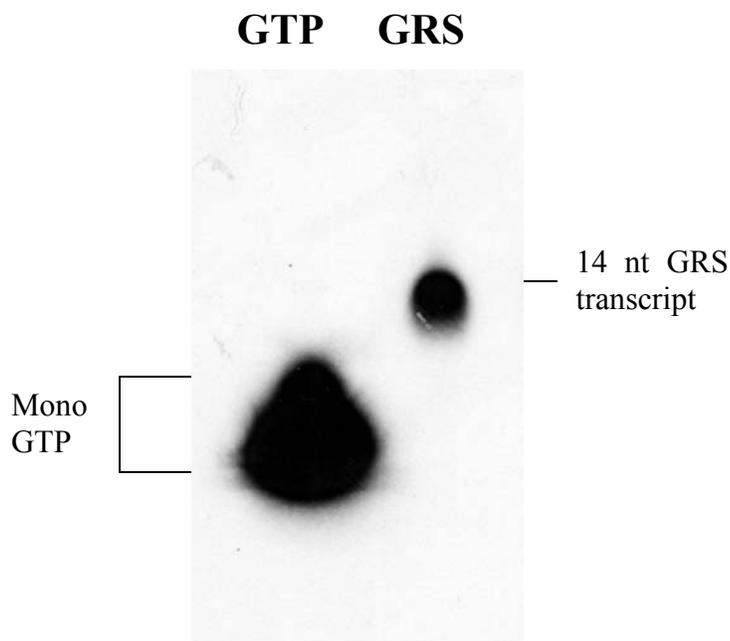
<b>Oligo Name</b>	<b>SEQUENCE</b>
wtGRS	CCCACACCTCCCCCTATAGTGAGTCGTATAT
SAA1	CCCACACCTCCCCCTTATAGTGAGTCGTATAT
SAA2	CCCACACCTCCCTCTATAGTGAGTCGTATAT
SAA3	CCCACACCTCCTCCTATAGTGAGTCGTATAT
SAA4	CCCACACCTCTCCCTATAGTGAGTCGTATAT
SAA5	CCCACACCTTCCCCCTATAGTGAGTCGTATAT
SAA6	CCCACACTTCCCCCTATAGTGAGTCGTATAT
SAA7	CCCACATCTCCCCCTATAGTGAGTCGTATAT
SAA8	CCCATACCTCCCCCTATAGTGAGTCGTATAT
SAA9	CCTACACCTCCCCCTATAGTGAGTCGTATAT
SAA10	CTCACACCTCCCCCTATAGTGAGTCGTATAT
SAA11	TCCACACCTCCCCCTATAGTGAGTCGTATAT
SAA12	CCCACACCTCCCTTTTATAGTGAGTCGTATAT
SAA13	CCCACACCTCCTTCTATAGTGAGTCGTATAT
SAA14	CCCACACCTCTTCCTATAGTGAGTCGTATAT
SAA15	CCCACACCTTTCCCTATAGTGAGTCGTATAT
SAA16	CCCACATTTCCCCCTATAGTGAGTCGTATAT
SAA17	CTTACACCTCCCCCTATAGTGAGTCGTATAT
SAA18	TTCACACCTCCCCCTATAGTGAGTCGTATAT
SAA19	CCCACACCTTCTCCTATAGTGAGTCGTATAT
SAA20	CCCACACCTTTTCCTATAGTGAGTCGTATAT
SAA-SCR	TCACGATACTGAGTTATAGTGAGTCGTATAT
T7prom	ATATACGACTCACTATA

## **PURIFICATION OF OLIGORIBONUCLEOTIDES**

After transcription with T7 RNAP, samples were subjected to DNase I digestion with 10 Units of DNase I at 37<sup>0</sup>C for 15 min. The Superfine G-25 RNA grade Sephadex media was purchased from Amersham Pharmacia and empty columns were purchased from Pharmacia. Columns were set up according to Pharmacia's instructions and they were stored in 20% Ethanol at 4<sup>0</sup>C. The G-25 Sephadex columns were initially spun for two minutes at 750g to get rid of the 20% Ethanol used as storage buffer. After this spin, samples were applied to columns, and they were spun another two minutes at 750g. Unincorporated nucleotides and truncated transcription products less than 10 nucleotides should be retained in Superfine G-25 Sephadex media whereas, at the same time, full length 14 nt transcripts will be in the flow through (Figure 4). By using this protocol, we were consistently able to get 2-15 pmol of pure RNA transcript.

## ***IN VITRO* TRANSCRIPTION WITH <sup>3</sup>H LABELED UTP**

This type of transcription reaction is very similar to the ones described above. This is written in a pretty chatty style. Since we have need a large quantity of tritium labeled competitors and the labeling percentage is not crucial, we increased the total volume of the reaction by 2.5 fold to 50 ul. The amount each individual reagent was increased by 2.5 fold except that each rNTP amount was increased 10 fold. A total of  $2.17 \times 10^{-10}$  mol of <sup>3</sup>H-UTP was used for each 50 ul reaction.



**Figure 4** Short ribooligonucleotides were purified by using G-25 Superfine Sephadex columns.

After *in vitro* transcription, transcription products were purified by using G-25 Sephadex columns as described in Methods section. A denaturing 12% (19:1) PAGE gel was run with purified wt GRS transcript (lane 2) and mono  $\alpha$ - $^{32}$ P-GTP (lane 1).

#### **CALCULATION OF INCORPORATION PERCENTAGE**

After *in vitro* transcription (IVT) reaction, 1/100 of the total volume is counted radioactivity was determined at the high energy channel 3 on a Beckmann Scintillation counter and the total  $^{32}$ P is calculated from this measurement. Either a denaturing Acrylamide-Urea gel was run with the IVT samples or if the transcript size was less than 20 nucleotides, purification was done by using G-25 Superfine Sephadex columns. In the former case, the corresponding bands were cut out and RNA is purified from the gel piece. The final Ethanol precipitated material was counted at high energy channel 3 and the total final counts were calculated from this measurement. This final total was divided by the initial total; the outcome is the Incorporation percentage. The incorporation percentage is multiplied with the total number of labeled and unlabeled GTP or UTP (Depending on the usage of  $\alpha$ - $^{32}$ P-GTP or  $\alpha$ - $^{32}$ P-UTP). The result was also divided by the

number of Guanines or Adenines in the transcript in order to get the total number of moles of transcript.

### **RNA-PROTEIN UV CROSSLINKING ASSAY**

UV crosslinking reaction mixture contains 10-20 fmol of  $\alpha$ -<sup>32</sup>P-GTP labeled, *in vitro* transcribed RNA, 40  $\mu$ g/ml Phenol-Chloroform purified yeast tRNA, 1 mM ATP and 0.7 mM MgCl<sub>2</sub> and recombinant proteins. Each sample was brought up to a final volume of 25  $\mu$ l with Buffer D. Once again, recombinant proteins were added last in the UV crosslinking experiments. Reactions were mixed and incubated at 30°C for 10 minutes. Then samples were transferred to a 96-well microtiter plate on ice and irradiated with ultraviolet light for 12 minutes at 3000  $\mu$ watts/cm<sup>2</sup> (1.8 J/cm<sup>2</sup>) in a UVStratalinker 1800 (Stratagene). Energy emitted by UV light causes covalent bond formation between RNA and proteins in close proximity. Unbound and unprotected RNA were digested with RNaseA (final concentration 50  $\mu$ g/ml) at 37°C for 15 minutes and samples are run on a 8% (19:1) SDS-PAGE gels. Then the gel was dried for one hour on a gel drier and the dried gel was exposed for 2-4 days and analyzed on the Storm 860 PhosphoImager (Molecular Dynamics) with the Imagequant software (Amersham Biosciences). Mobility of the proteins will shift to some extent since each protein is also crosslinked to short RNAs.

### **PCR MUTATION STRATEGY**

The pSP65-SVL plasmid has about 250 nt SVL polyA region under the control of SP6 promoter. By having a 5' PCR primer with T7 promoter overhang and a 3' primer with mutant GRS overhang, we were able to get PCR products having aT7 promoter and a mutant GRS. All the PCR primers we used to create GRS mutants in the context of 225 nt SVL were given below. As

a common 5' primer we use the one labeled 248-5'CC. The T7 promoter is red shaded in this primer. Since we initially wanted to put the SAA18 and SAA20 mutations in the context of 225 nt SVL, we just designed two 3' primers for these mutants and another one for the wild type. 248A-3' has SAA20 whereas 248B-3' has SAA18 mutations in them in the antisense form. The oligo 248wt-3' is designed to amplify the wild type SVL. The wt SVL we amplified with 248-5'CC and 248wt-3' is identical to probe A. The sequences of these oligonucleotides are given in Table 4.

There is one complication in this PCR however. The mt 3' primers 248A-3' and 248B-3' are 46 nt long and the former has three, while the latter has two mismatches to the pSP65-SVL plasmid. This means we may get a certain percentage of annealing of mt 3' primers to the wild type SVL sequence present on the pSP65-SVL. We solved this issue by digesting pSP65-SVL with restriction enzyme M<sub>n</sub>II. One of the cutting sites of M<sub>n</sub>II is located just one nucleotide 5' of the GRS and except at this site it does not cut SVL region. So we digested pSP65-SVL with M<sub>n</sub>II and used the digested plasmid as template. These PCR products will be used in *in vitro* 3' cleavage assays.

### **CLONING SVL MUTANTS TO pEGFP-C1 PLASMID**

Originally, the pEGFP-C1 plasmid (Clontech) was designed so that a second open reading frame can be cloned to the C-terminal multicloning site (MCS) in frame of gfp Orf. In this way, the protein of interest will be GFP tagged. Clontech has put the SV40 early polyadenylation signal just after the MCS. We found another use for this vector. Without cloning any Orf in to the MCS, we can take out the default SV40 early polyA site from the plasmid and put in our mutant or wt SVL polyA regions.

**Table 4 Oligonucleotides used for PCR, RT-PCR or cloning based experiments.**

<b>Oligo Name</b>	<b>SEQUENCE</b>
ELL2-3'	CTTCTGGTCCAAGCAGAG
ELL2-5'	AGGAGTTAAAGAAGGGTGC
248-5'CC	TAATACGACTCACTATAGAATACACGGAATTCGAGCT
248A-3'	AAAAAACCTCCCACACCTTTTCTGAACTGAAACATAAAATGAAT
248B-3'	AAAAAACCTCCCACACCTCCCTTTGAACTGAAACATAAAATGAAT
248wt-3'	AAAAAACCTCCCACACCTCCCCCTGAACTGAAACATAAAATGAAT
248-5'CC	TAATACGACTCACTATAGAATACACGGAATTCGAGCT
247A- 3' CLONING	TACCGACGCGTAAAAAACCTCCCACACCTTTTCTGAACTGAAACATAAAAT GAAT
247B- 3' CLONING	TACCGACGCGTAAAAAACCTCCCACACCTCCCTTTGAACTGAAACATAAAAT GAAT
247WT- 3' CLONING	TACCGACGCGTAAAAAACCTCCCACACCTCCCCCTGAACTGAAACATAAAAT GAAT
Eco1	AATTCGATCGGATATCAGTACTA
Mlu1	GCTAGCCTATAGTCATGATGCGC

There is a unique EcoRI site in the MCS of pEGFP-C1 (Clontech). About 300 bp downstream of MCS, there is unique MluI site. By performing double digestion, we can take out the plasmid's polyA sequences without any damage to the backbone of the vector. The final task is to create PCR amplified mt SVL having EcoRI and MluI sites which will be used as insert. The previously used common 5' primer 248-5'CC has an EcoRI site on it. We designed three other 3' primers; we just added MluI sites to the 5' of the previously used 3' primers. Please note that there are also another five nucleotides 5' of the MluI sites. They are just there to enhance MluI

digestion. 247A-3'Cloning, 247B-3'cloning have SAA20 and SAA18 mutations on them in the antisense form, respectively. The sequences of these oligonucleotides were given in Table 4.

Each PCR was performed with M<sub>n</sub>II digested pSP65-SVL plasmid and common 5' primer 248-5'CC and one of the 3' cloning primers. The final PCR products will 252 nt (224 nt SVL plus overhangs) and EcoRI-MluI double digested PCR product size will be 219 nt which is 33 nt shorter than the PCR product. The digestion can be observed on a 4% Agarose gel. The engineered pEGFP plasmid having the SAA18 and SAA20 mutations were named pEGFP-SAA18 and pEGFP-SAA20, respectively.

To remove the default polyA site of the pEGFP-C1 vector, we made use of two oligonucleotides: EcoI and MluI (Table 4).

Annealing of these two complementary oligos with 5' EcoRI and 3' MluI overhangs creates a perfect insert for the pEGFP vector. pEGFP-nopolyA was created by first digesting out the EcoRI-MluI fragment from the pEGFP-C1 and then cloning the annealed oligos into the EcoRI and MluI sites of pEGFP-C1. Oligos were mixed and denatured at 95<sup>0</sup>C for 1 minute in 0.025M EDTA and 10mM TE. They are allowed to anneal at RT for seven minutes. Then the ligation reaction performed with annealed oligos. Annealed EcoI-MluI oligos have also unique PvuI, EcoRV and ScaI restriction sites that were not present in pEGFP-C1 backbone. These unique sites were very useful to identify positive clones.

### **CLONING STRATEGY: CLONING THROUGH SELECTIVE DIGESTION**

The rate limiting factor for most of the cloning experiments is the quantity of the insert. This problem usually is a result of inefficient gel purification of the digested fragments. Although the

ligations can be done in Agarose gel pieces, this reaction also suffers from inefficiency and inconsistency.

In this study, essentially all the cloning was done without gel purification of the digested fragments. This technique makes use of the fact that the DH5alpha bacterial cells are transfected with circular plasmids about 1000 times more efficiently than any linearized plasmid (Conley and Saunders 1984) unlike human 293T cells. After ligation was performed, unwanted ligated products were linearized with a final selective restriction enzyme digestion that cuts the unwanted circular plasmids but not the ones of interest (Spear 2000) . The final digestion was either performed for 16 hr at RT or for 4 hr at 37<sup>0</sup>C to ensure a complete digestion. This technique dramatically reduced the false positive clones compared to conventional cloning methods. The Fast-Link ligase (Epicentre) was used for all the ligation reactions. The reactions, 15 ul total volume, were set up with: 1.5 ul 10 mM ATP, 1.5 ul 10x Ligation Buffer (330 mM Tris-Acetate pH 7.8, 660 mM Potassium Acetate, 100 mM Magnesium Acetate, 5 mM DTT), 1 ul Fast-Link Ligase (2U/ul), about 500 ng DNA and DEPC H<sub>2</sub>O. The ligation was performed at RT for 30 min. The Insert:vector ratio was 1:5 and 1:15. Ligation reactions were heat inactivated at 65<sup>0</sup>C for 15 minutes prior to transformations.

## **SELECTIVE DIGESTION AND TRANSFORMATION**

The enzyme PstI cuts between EcoRI and MluI sites of parental vector pEGFP-C1 whereas it does not cut any other pEGFP plasmids: i.e. pEGFP-SVLwt, pEGFP-SAA18, pEGFP-SAA20, or pEGFP-nopolyA (Figure 5). The PstI digestion was performed in a 15 ul reaction containing 1.5 ul 10x PstI Buffer, 0.5 ul PstI (10U/ul), 10 ul DEPC H<sub>2</sub>O and 3 ul heat inactivated ligation mix (about 100 ng DNA). Although we were aware that PstI is in excess in this reaction, we

performed it either for 4 hr at 37<sup>0</sup>C or 16 hr at RT to ensure that the digestion went to completion. The PstI enzyme was inactivated at 65<sup>0</sup>C for 20 min after the digestion.

Invitrogen's Max Efficiency DH5alpha cells 50 ul were transferred from 1.5 ml tubes to 15 ml conical ones. One µl from the final digestion (about 6.7 ng DNA) was gently added to these cells without pipetting up and down. This mix was incubated for 30 min on ice. Then cells were heat shocked for 45 sec at 42<sup>0</sup>C without swirling. After another 2 min incubation on ice, 250 ul of room temperature, rich SOC media was added to each tube. Cells were shaken for one hour at 37<sup>0</sup>C to allow npt2 gene expression. 200 ul of these cells were plated on Kanamycin plates (50ug/ul Kanamycin). In most of the cases, the cloning worked so well that, a 1:100 dilution of bacterial culture was necessary to isolate individual colonies.

This strategy can be applied for most of the cloning experiments. It is applied for our cloning experiments with enormous success with a very low false positive percentage.

From 6-10 clones were picked from Kanamycin plates and 3 ml of YT plus Kanamycin media in 15 ml tubes were inoculated with these clones. These cultures were grown for 16 hr at 37<sup>0</sup>C. Plasmid minipreps (Qia-gen) were prepared from 1.5 ml of saturated cultures. The identity of the clones were determined by diagnostic digestions. An EcoRI-MluI double digestion was performed for each cloned plasmid; the cloned fragment is 216 bp in each case whereas the parent plasmid yields 283 bp fragment. Once we made sure that the plasmid had the appropriate size insert, it is sent for DNA sequencing.

Approximately 150 ml YT plus Kanamycin media was inoculated with 10 ul of the bacterial culture having the desired plasmid and this culture was allowed to grow at 37<sup>0</sup>C for 16 hr. Approximately 100 ml of the saturated culture was used for maxi-prep plasmid purification with

Qia-gen®'s maxi-prep kit. Maxi-prep procedure usually resulted in about 500 µg of plasmid DNA.

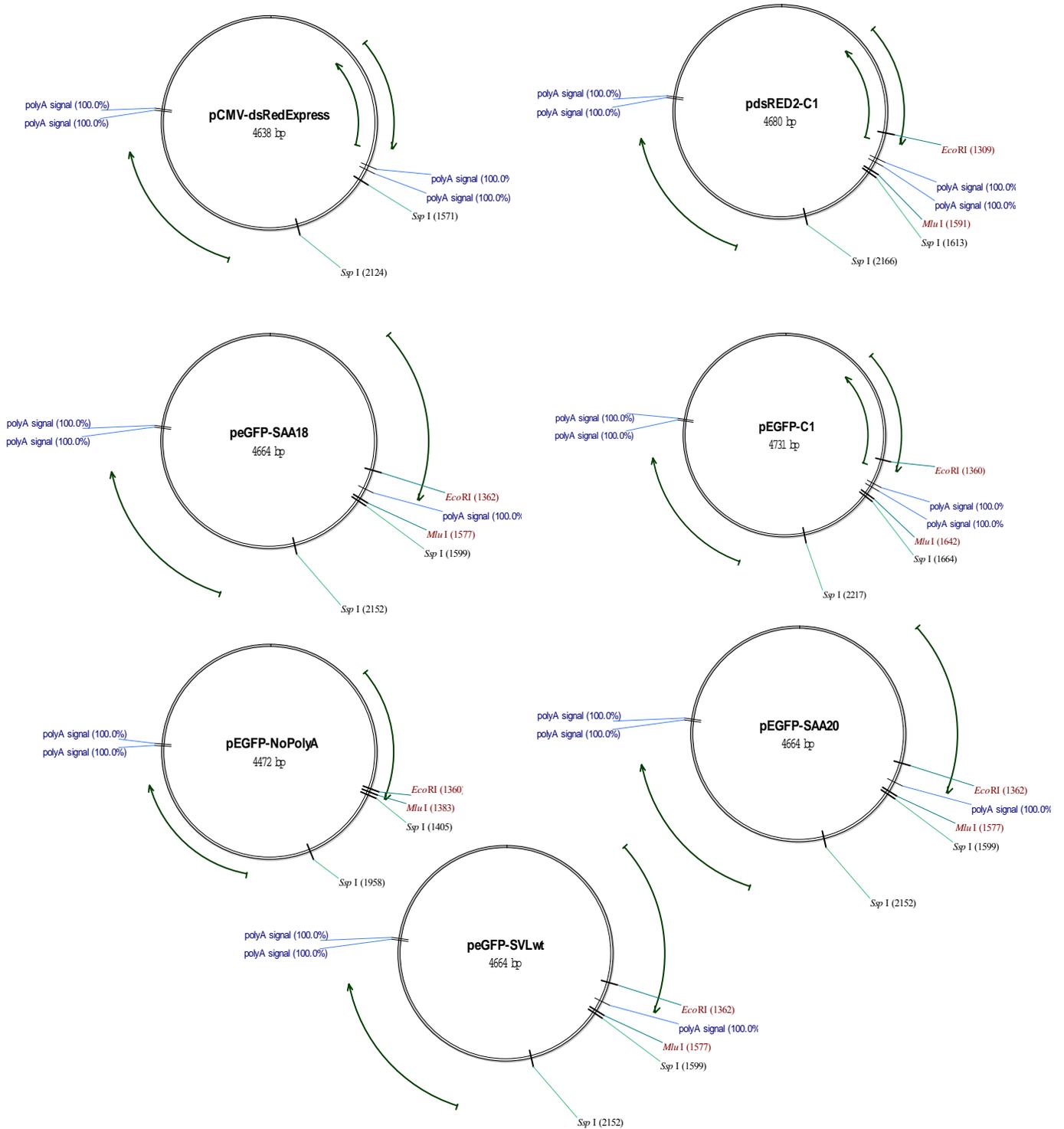


Figure 5 Plasmid maps that were used in this study.

## **DNA SEQUENCING**

All the DNA sequencing reactions were performed by University of Pittsburgh Proteomics and Genomics Laboratory. The sequencing files were analyzed by Chromas 2.0 software available at <http://www.technelysium.com.au/chromas.html> (December 2004). Chromatograms corresponding to four of the cloned plasmids were shown in Figure 6.

## **PHOSPHORYLATION**

Phosphorylation reactions were performed with HeLa, A20 and AxJ extracts as explained in (Hirose et al. 1999)

## **ANTIBODIES**

Polyclonal rabbit anti-peptide antibodies R-7263 and R-7264 were used to detect hnRNP F and hnRNP H1/H2, respectively. Horseradishperoxidase-conjugated secondary antibodies (sheep anti-mouse IgG and goat anti-rabbit IgG) were purchased either from Sigma or Boehringer Mannheim. Murine anti-RNAP II antibody is a kind gift of Dr. Baskaran Rajasekaran. Murine anti-GFP and anti-neomycin antibodies were purchased from Clontech and Upstate Cell Signaling solutions.

## **INSTRUMENTS**

All UV absorbances for quantification of nucleic acid concentrations and visible absorbances for protein quantification were performed on a Bio-Rad Smartspec 3000 model spectrophotometer. Scintillation counting was performed in a Beckman Instruments model LS5000 TD counter. Almost all the PCR reactions were performed with Bio-Rad Icyler gradient Thermal cycler.

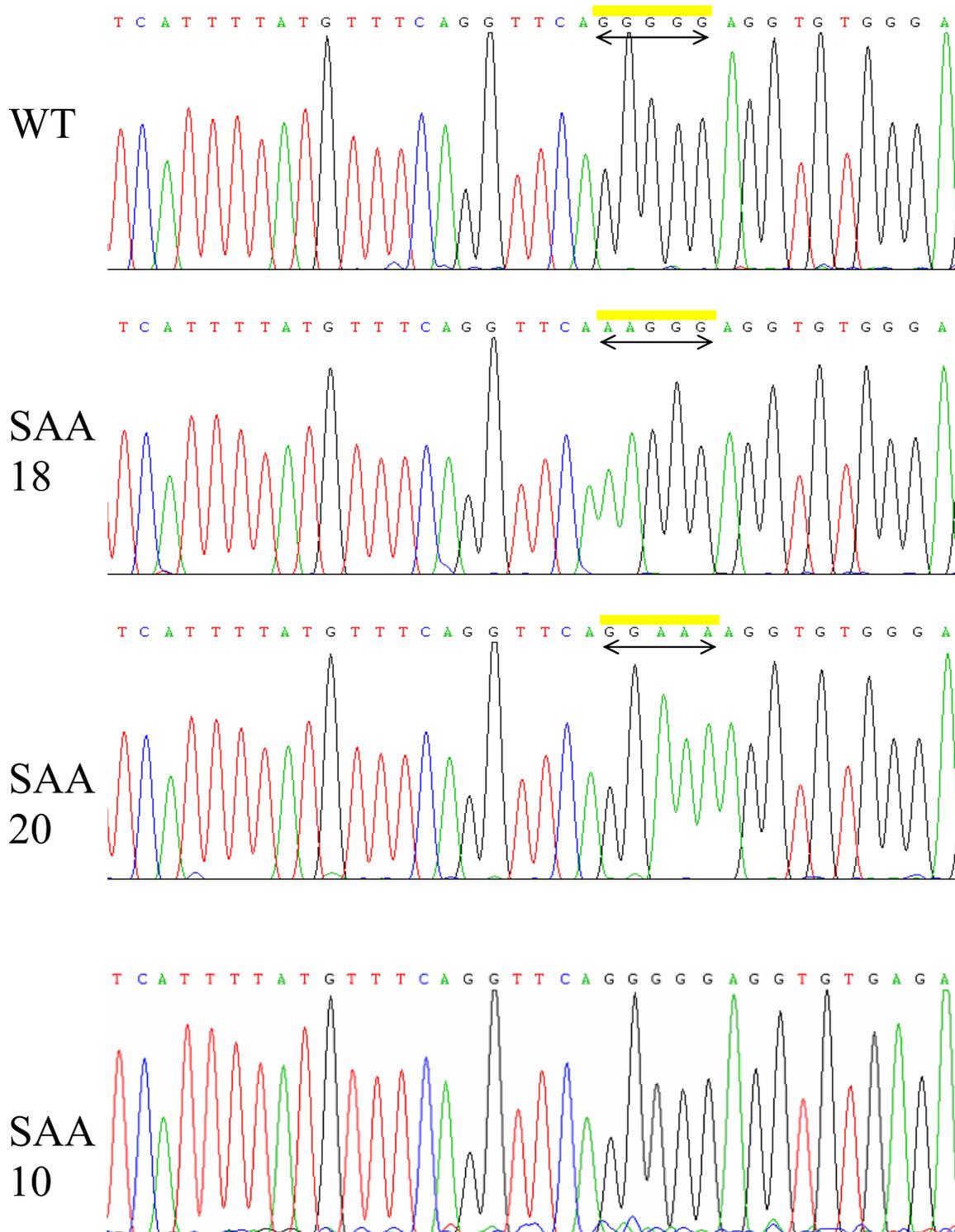


Figure 6 DNA sequencing chromatogram showing the PCR generated SAA18, SAA20 and SAA10 mutations located in pEGFP based plasmids. The wild type SVL sequence is given at the top.

## **BIOINFORMATICS**

Various DNA, RNA and protein analysis were performed by Vector Nti 7.0 software. All the cloning experiments, plasmid map creation, PCR Oligonucleotide design, Blast searches, Agarose-Acrylamide gel run simulations, analysis of restriction enzyme cuts were done by using the same Vector Nti software. Most of the protein, DNA and RNA alignments were visualized by BioEdit software available at <http://www.mgb.pitt.edu/moleculartoolbox.htm>.

## **MAMMALIAN CELL CULTURE**

A20 B-lymphoma, HeLa, 293T cells were grown as explained in American Type Culture Collection website ([www.atcc.org](http://www.atcc.org)). AxJ cells were grown as explained in (Veraldi et al. 2001).

## **TRANSIENT TRANSFECTION**

Human embryonic kidney 293T cells were transfected with linearized plasmids by using lipotransfection method. Plasmids were linearized with Ssp1 resulting in blunt ended fragments. The digestion was performed for 4hr at 37<sup>0</sup>C. This digestion disrupts only the fl origin in all vectors. 1.5-3.0x 10<sup>5</sup> 293T cells were plated on 6 well plates. 2 ug equimolar plasmid, serum free DMEM and Fugene 6 transfection agent (Invitrogen) were combined in 100 ul volume. This mix was incubated 15 min at RT and added to the wells having 2 ml serum positive DMEM. After incubating 18-41 hours, media (DMEM, 10% FCS, 20mM Hepes) were taken out and cells were washed with 1xPBS and trypsinized with 0.5 ml 1xTrypsin. After adding 2ml of serum positive DMEM to block trypsin, cell were spun down in 5 ml Falcon tubes. Cell pellets were

sequentially washed with 1xPBS and then with 1xFACS buffer (5% FCS, 1xPBS, 0.01% NaAzide). Final cell pellets were fixed in 1xParaformaldehyde.

## **FLOW CYTOMETRY ANALYSIS**

All the flow cytometry experiments shown in this study were performed with Coulter Epics XL flow cytometer operating with a single 488nm laser present in UPCI Hillman Cancer Centre, Pittsburgh. Both EGFP and DsRed proteins were excited with 488nm laser and emissions were recorded in FL1 (green) and FL3 (red) channel. EGFP is optimized for flow cytometry to be used with the 488 nm laser that is present in almost all flow cytometers. Since we used the optimal excitation wavelength of EGFP, there is a certain overlap to FL3 as in all EGFP flow experiments. Therefore, a modest compensation was done by subtracting 3-10% of the FL1 channel from FL3 depending on the cell line. A total of  $5-20 \times 10^3$  events were recorded for each sample. The in depth analysis was done with the software WinMdi available at <http://facs.scripps.edu/software.html> (December 2004).

**4. MANUSCRIPT ONE: REGULATION OF GENE EXPRESSION BY HNRNP H' AND F THROUGH FIVE CONSECUTIVE GUANINES IN 3'UTR**

**4.1. ABSTRACT**

The two proteins hnRNP F and H are remarkably similar in sequence yet differ in expression pattern in a variety of tissues. Here, using purified proteins, the dissociation constant of hnRNP H to the SVL pre-mRNA was shown to be at least three fold less than that of hnRNP F when measured with the filter binding assay. An extensive mutational analysis in the 14 nt G-rich sequence (GRS) downstream of the SVL polyA site was performed and the binding of both hnRNP F and H was significantly diminished by the disruption of the five Guanines and enhanced by the presence of a downstream Adenine at position 13. Interruption of this binding site by mutation had a mild negative effect on *in vitro* polyadenylation assays and a pronounced effect on protein expression using the Green fluorescent protein (GFP) based reporter in mammalian cells. The same effect was verified by A13 mutation, with enhanced binding, resulting in increased GFP expression. Downregulation of hnRNP F/H levels by RNAi diminished the GFP level. In addition, forced expression of hnRNP F and H resulted in elevated gene expression. These effects were sequence specific to the G5 block. This study shows that both hnRNP F and hnRNP H are able to bind to the same sequence in the 3' UTRs but with different affinities and that the relative levels of hnRNP F and H as well as the target RNA sequence can influence gene expression.

## 4.2. INTRODUCTION

Heterogenous nuclear ribonucleoproteins (hnRNPs) display sequence specific RNA binding activity and associate predominantly with RNA pol II transcripts (Krecic and Swanson 1999). Most of them are located exclusively in the nucleus while others shuttle between cytoplasm and nucleus. RNA binding activities of these proteins are conferred by either RNA Recognition Motifs (RRMs) or Arginine, Glycine, Glycine –rich (RGG) domains or both. The RRM containing hnRNPs typically have multiple RRM; usually at least one RRM is responsible for sequence specific binding whereas the other RRM increase the affinity of the protein to the specific sequence. Among the many proposed functions for hnRNP proteins are mRNA trafficking, splicing, telomere length control, mRNA stability, transcription, and polyadenylation (Krecic and Swanson 1999) .

The hnRNP H protein family consists of hnRNP H1, H2 (DSEF or aka H'), H3 (2H9) and hnRNP F. The corresponding genes are *hnrp* H1, H2, H3 and *hnrp* F, respectively. hnRNP H1 and H2 are 449 aa proteins having three RRM (Figure 1). The 96% identity between H1 and H2 may ensure identical physiological function; because of the identity we will use the “H” nomenclature for both throughout. In contrast, hnRNP F is a 415 aa protein having three RRM and it lacks the very carboxy terminal part present in hnRNP H. While it shows a 72 and 70% identity to hnRNP H1 and H2, overall, the corresponding RRM are more than 90% identical. The 346 amino acid long hnRNP H3 is a member of the family that has been shown to have various alternatively spliced forms whose functions are not well characterized (Honore 2000). The role of hnRNP H/F on splicing has been extensively studied with multiple different substrates. It appears that in most of the cases hnRNP H/F bind to splicing silencers (Grabowski 2004) (Sironi et al. 2004).

The hnRNP F protein was implicated in the polyA site choice within the immunoglobulin heavy chain (Veraldi et al. 2001) since over-expression of it in a plasma cell resulted in decrease in the usage of secretory polyA site that is normally be chosen over the membrane one. Purified hnRNP F inhibited binding of CstF-64 to both the secretory and membrane specific polyA sites. The precise binding site of hnRNP F was not mapped. Using an *in vitro* purification method of polyadenylation complexes (Veraldi et al. 2001) a larger complex devoid of polyadenylation factors was observed in A20 B-cell lymphoma containing both hnRNP F and hnRNP H proteins; this complex was absent in AxJ Plasmacytoma cells where hnRNP F expression is slightly lower. The 14 nt Guanine Rich Sequence (GRS) region in SV40 late (SVL) pre-mRNA was shown to bind hnRNP H2 protein, also known as DSEF-1 (Bagga et al. 1995). Binding of recombinant hnRNP H to GRS was shown to activate 3' processing *in vitro*. Diverse pre-mRNAs including IVA2 and Igmu can be influenced by hnRNP H (Arhin et al. 2002) so the question of sequence specificity remains. While several GRS mutants were made and analyzed *in vitro* (Bagga et al. 1995), a conclusive mutation analysis of GRS region on SVL pre-mRNA had not been performed to determine the minimal, optimal binding site. In one study the sequence GGGA was implicated as the binding site of hnRNP H family proteins (Caputi and Zahler 2001) and a run of Guanines in another study (Buratti et al. 2004). When 9 out of 14 nucleotides of the SVL GRS were deleted the observed *in vitro* 3' processing efficiency dropped about 2.5 fold (Cooke et al. 1999). In addition, the amount of specific polyadenylation complexes formed on the same mutant dropped significantly (Cooke and Alwine 2002) supporting the idea that binding of hnRNP H or F to the GRS increases the rate of complex formation, hence overall 3' processing efficiency.

To better understand the function and the effect of these two hnRNP proteins F and H on polyadenylation and hence gene expression, we wanted to characterize their binding region relative to the polyA signal and relative affinities. In this study, we have shown that hnRNP F and H bind a five Guanine stretch downstream of polyA site on SVL pre-mRNA. The consensus binding site of hnRNP F and H includes a downstream Adenine and upstream Guanine. Diminishing the binding of hnRNP F and H by mutating the RNA inhibited polyadenylation *in vitro* and gene expression *in vivo*. Increasing and decreasing the levels of hnRNP F and H by forced protein expression or RNAi in mammalian cells elevated or decreased gene expression, respectively, in a sequence dependent matter. Multiple lines of evidence indicate that hnRNP F and H proteins are important modulators of gene expression.

### 4.3. RESULTS

#### 4.3.1. Mapping the binding site of hnRNP F on SVL pre-mRNA.

Purified hnRNP H protein was shown to bind to a G-rich sequence (GRS) of 14 nts downstream of the cleavage site on SVL pre-mRNA (Bagga et al. 1995). While the three RBDs of hnRNP F show extensive homology to hnRNP H, the exact hnRNP F binding site on the SVL region had not been determined. In order to pinpoint the binding region of hnRNP F on SVL, we made use of 7 different SVL pre-mRNA based constructs (Figure 7). The wild type SVL RNA probe is 224 nt long having 3' cleavage site and both upstream and downstream sequences. Electromobility Shift Assays (EMSAs) were performed with each of these RNA samples and histidine tagged recombinant hnRNP F. Since hnRNP F was shown to bind Cap binding complex, all the *in vitro* transcription reactions were performed without cap analog. Despite the fact that the transcripts were uncapped, RNA stability was maintained.

We found that hnRNP F protein was able to bind three of the RNA samples: wild type SVL, SVL having a polyA signal mutation AAGAAA, and a portion of the SVL having only the downstream G-rich signal (Figure 8A and 8B). This first data set points toward hnRNP F binding somewhere in the region downstream of the cleavage site.

Substitution of the downstream cleavage region with vector pGem sequence in probe C, abolished the binding of hnRNP F (Figure 8B). Replacement of the 14 nt GRS with pGem sequences in probe D and F, also eliminated the hnRNP F binding (Figure 8A and 8B). These data strongly suggest that hnRNP F binds to the G-rich downstream region. Before doing any mutational analysis, competition gelshifts were performed to determine if the interaction observed between hnRNP F and GRS is a result of specific RNA-protein binding (Figure 8C and

8D). The binding was able to be competed by a specific RNA (GRS+) but not by a non-specific RNA, demonstrating the specificity of binding.

There are 11 Guanines in the 14 nt G-rich region downstream of GU-rich section making it a good candidate for the hnRNP F binding site. In order to determine if hnRNP F indeed binds to the GRS, we directly transcribed the 14 nucleotide GRS (Table I) by oligonucleotide based *in vitro* transcription (Milligan JF 1987). The GRS region binds to hnRNP F with reasonable affinity compared to wild type SVL RNA probe (Figure 8B).

After narrowing the binding region of hnRNP F to 14 nt, we wanted to determine the specific nucleotides important for both hnRNP F and H binding. For this purpose, a serial single G→A mutation study was performed to detect which G residues are essential for the interaction of hnRNP F and hnRNP H. There are 11 G's in GRS and each was mutated to A individually for a total of 11 single mutants. The exact sequence of each mutant is given in Table I.

Gel shift data for single substitution mutants are summarized in Figure 4A. The most potent mutants are SAA3 and SAA5 which have mutation at the 3<sup>rd</sup> and 5<sup>th</sup> positions, respectively. These mutants show about 15-30 percent of wt GRS binding to hnRNP F and H. SAA10 and 11 bind to hnRNP F and H more efficiently than wt GRS. The effect of each mutation is similar on both hnRNP F and H binding although SAA3 and SAA9 have a slightly more pronounced effect on hnRNP F than on H. In summary, EMSA results indicate that the third and fifth G's are the most important ones for both hnRNP F and hnRNP H binding (Figure 9A). The SAA3 mutation disrupts the longest G-stretch in the GRS exactly in the middle.

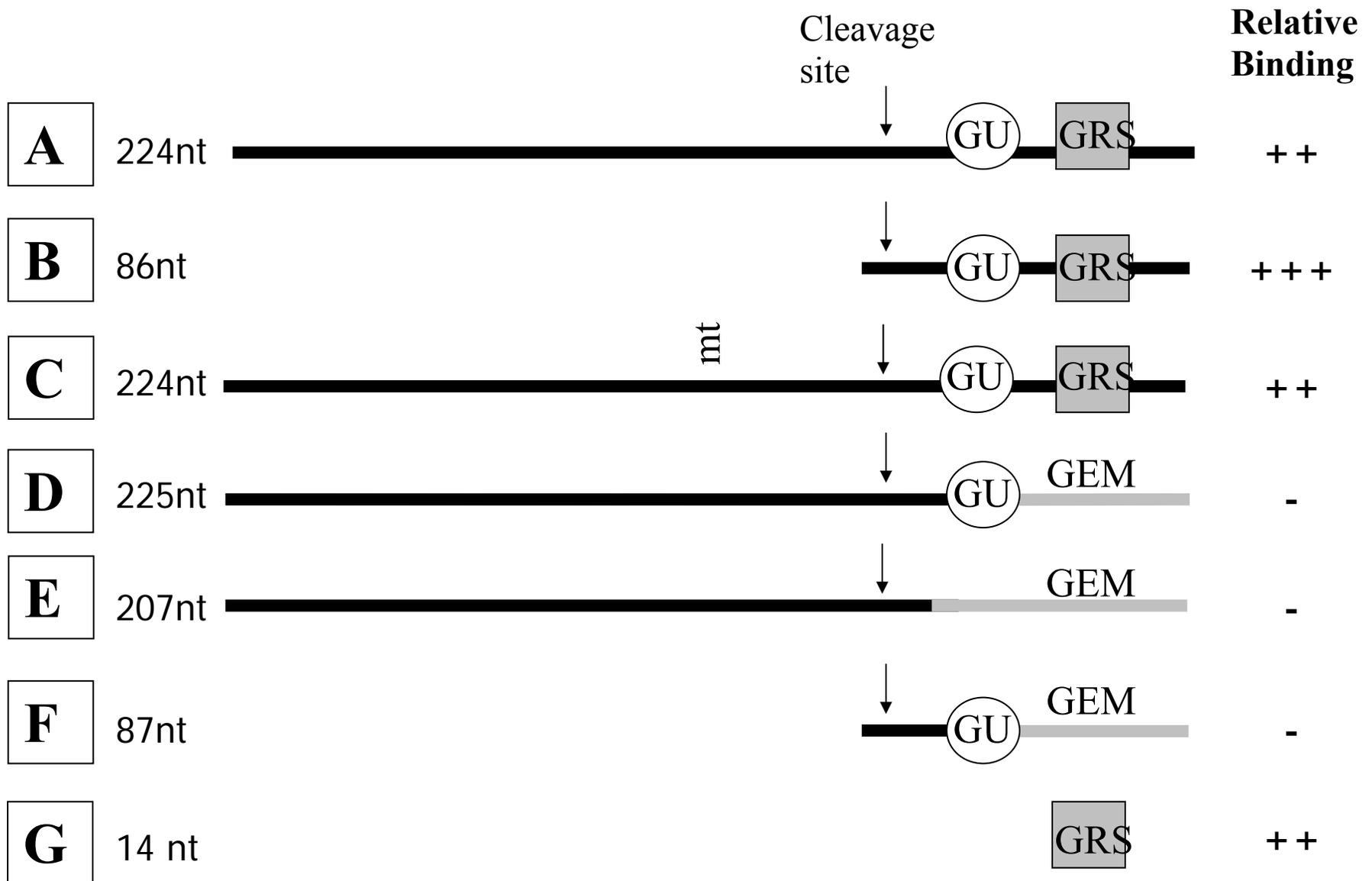
Then we made every possible two consecutive G→A mutations in GRS (Table 5) and performed EMSAs with them (Figure 9B). Mutants SAA16 and 17 are our most effective mutants in this case. SAA 16 has mutations at the 3<sup>rd</sup> and 4<sup>th</sup> positions whereas SAA17 is mutant

at the 2<sup>nd</sup> and 3<sup>rd</sup> nucleotides (Table I). After this set of experiments, it was clear that the first five Guanines are decisive for both hnRNP F and H binding.

We made two other mutants: one having mutations at the 3<sup>rd</sup> and 5<sup>th</sup> positions (SAA 19) the other having triple G→A substitutions between the third and fifth residues (SAA 20) (Table 5). The latter completely knocked out both hnRNP F and H binding whereas the former showed some residual binding of hnRNP H as assayed by EMSA (Figure 9B).

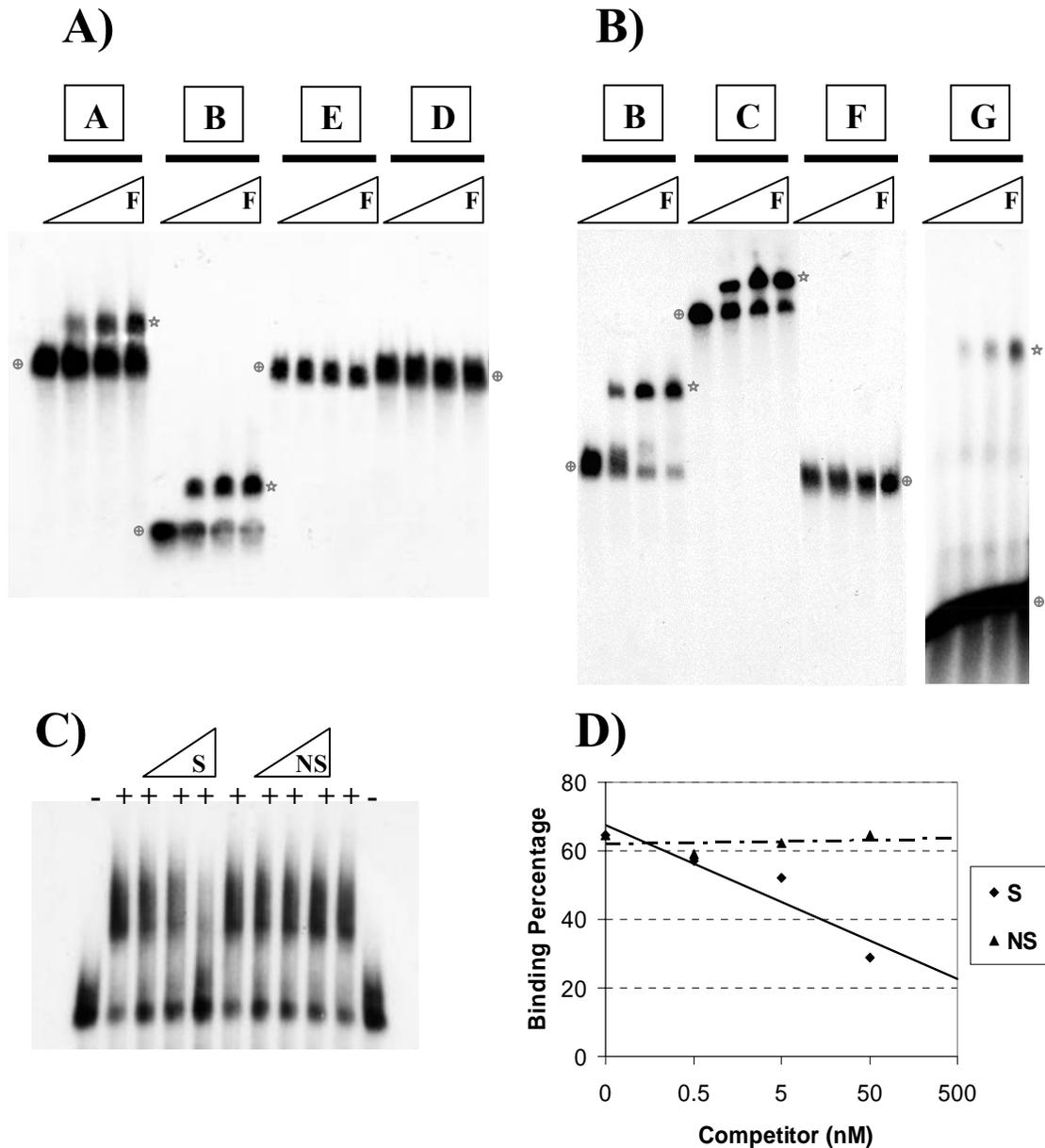
Mutants SAA 10 and 11 bound to both hnRNP F and H better than the original GRS by a factor of about two fold. While our mutation analysis showed that a Guanine stretch is critical for binding, it may be possible that a downstream Adenine is required for optimal binding. The non-specific nature of quarternary Guanine structures discouraged us to pursue a possible SELEX experiments. However, all the experimental binding sites of hnRNP F and H present in the literature were collected and a multiple alignment was performed with them (Figure 2). The consensus binding sequence shown in Figure 2 includes a Guanine stretch, an Adenine located 8-9 nucleotides downstream of this stretch, and an upstream Guanine. This parallels to our experimental results. The locations of SAA10 and 11 mutations coincide with the conserved Adenine when the GRS is aligned to the consensus sequence.

The apparent dissociation constants (K<sub>d</sub>s) of hnRNP F and H to SVL pre-mRNA were determined precisely by using filter binding assay (Liu et al. 2002). Either the 86 nt wild type probe B (Figure 7) or the one having the SAA20 mutation inserted were used for the binding assays. Proteins were extensively purified as described in Materials & Methods. We found that hnRNP H binds to wtSVL with an affinity (37 nM) that is more than three fold higher than that of hnRNP F (121 nM, Figure 10 and Table II). This may ensure that physiologically when both are



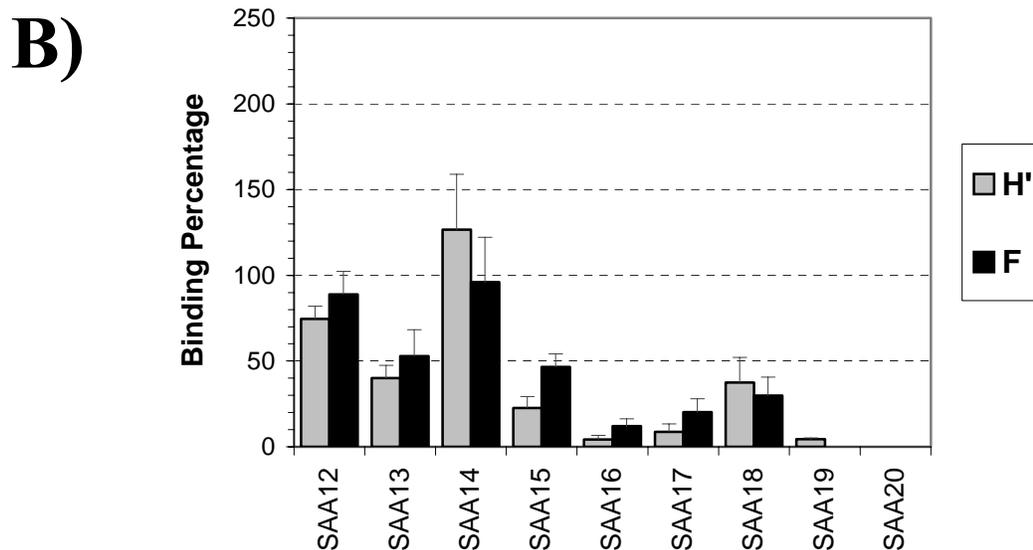
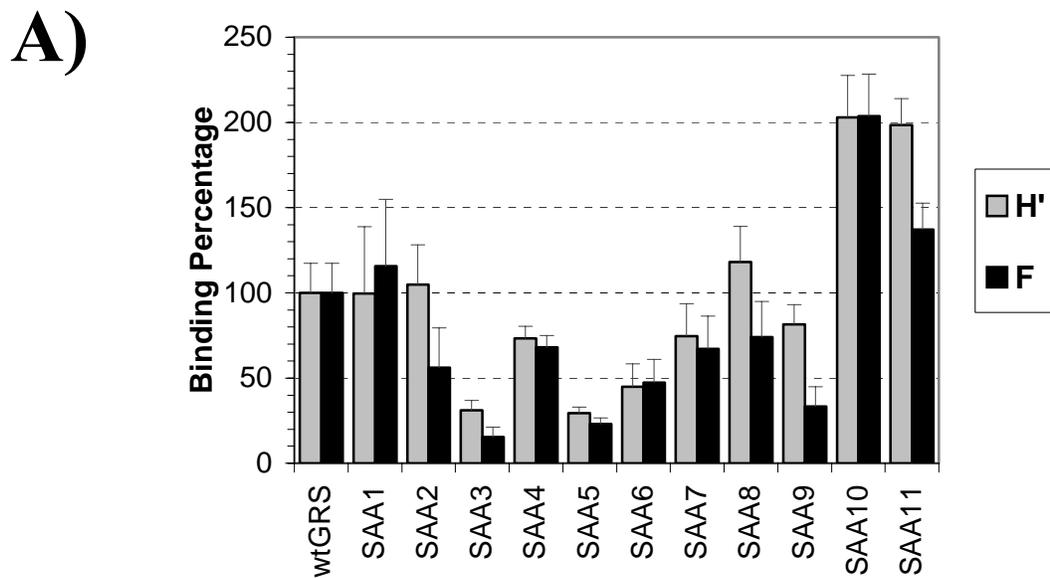
**Figure 7** Different SVL pre-mRNA based probes that were used in this study.

The length of each RNA is indicated on the left whereas the relative binding affinity to hnRNP F is given on the right side. Probe A is wild type SVL probe. GEM stands for vector pGem sequences.



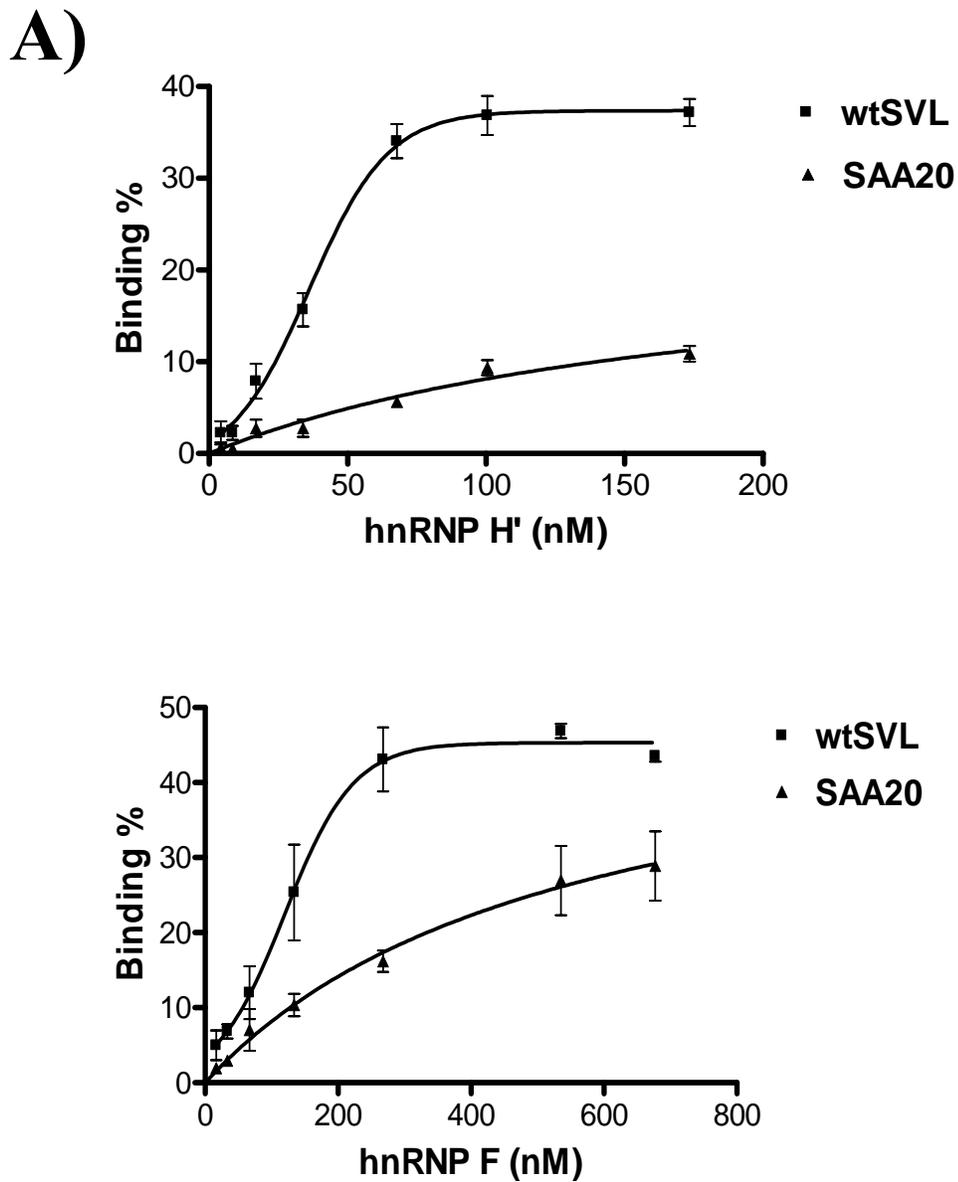
**Figure 8** hnRNP F binds to Guanine rich sequence in SVL pre-mRNA specifically.

A) Electromobility shift assays (EMSA) were performed with 50 pM  $^{32}$ P-GTP labeled SVL pre-mRNAs in 20  $\mu$ l reactions. 0, 55, 110, 220 nM of recombinant his-hnRNP F were used to shift each RNA probe. Samples were run on a 4% (20:0.25) native Acrylamide gel having 0.5% Agarose. Stars indicate the positions of hnRNP F-RNA complexes. Crosshairs show the positions of unshifted RNA probes. B) EMSAs were performed as described in A). Samples are run on an 8% (20:0.25) native Acrylamide gel. 250 pM probe G was used last four lanes C) EMSAs were performed with 50 pM probe B and either 0.5, 5 or 50 nM specific (S) or nonspecific competitors (NS). + and - signs stand for the presence and absence of 165 nM his-hnRNP F, respectively. D) Quantification of lanes 3-5 and 7-9 in C) is shown. Quantification was done by using Scion Image software available at [http://www.scioncorp.com/frames/fr\\_download\\_now.htm](http://www.scioncorp.com/frames/fr_download_now.htm) (December 2004).



**Figure 9** hnRNP F and H' bind to five consecutive Guanines in SVL pre-mRNA.

A) Relative binding of wt and mutant GRS oligoribonucleotides to hnRNP F and H' proteins. EMSAs were done with 250 pM of each probe three times with 330 nM his-hnRNP F and 210 nM GST-hnRNP H'. Binding percentage of each mutant relative to wtGRS is shown in the graph. Error bars represent standard error. B) Relative binding of mutant GRS oligoribonucleotides to hnRNP F and H' proteins. EMSAs were done as described in A). Binding percentage of each mutant relative to wtGRS is shown in the graph. Error bars represent standard error.



**Figure 10** **hnRNP H' binds to SVL pre-mRNA with a higher affinity compared to hnRNP F.**

A) Binding curves of hnRNP H' to wt probe B and probe B having SAA20 mutation as determined by filter binding assay. Each reaction was performed three times and error bars represent standard error. The curves are fitted to the experimental data points by using Prism 4 software version 4.01 (Graphpad software Inc). B) Binding curves of hnRNP F to very same RNA molecules were obtained as described in part A).

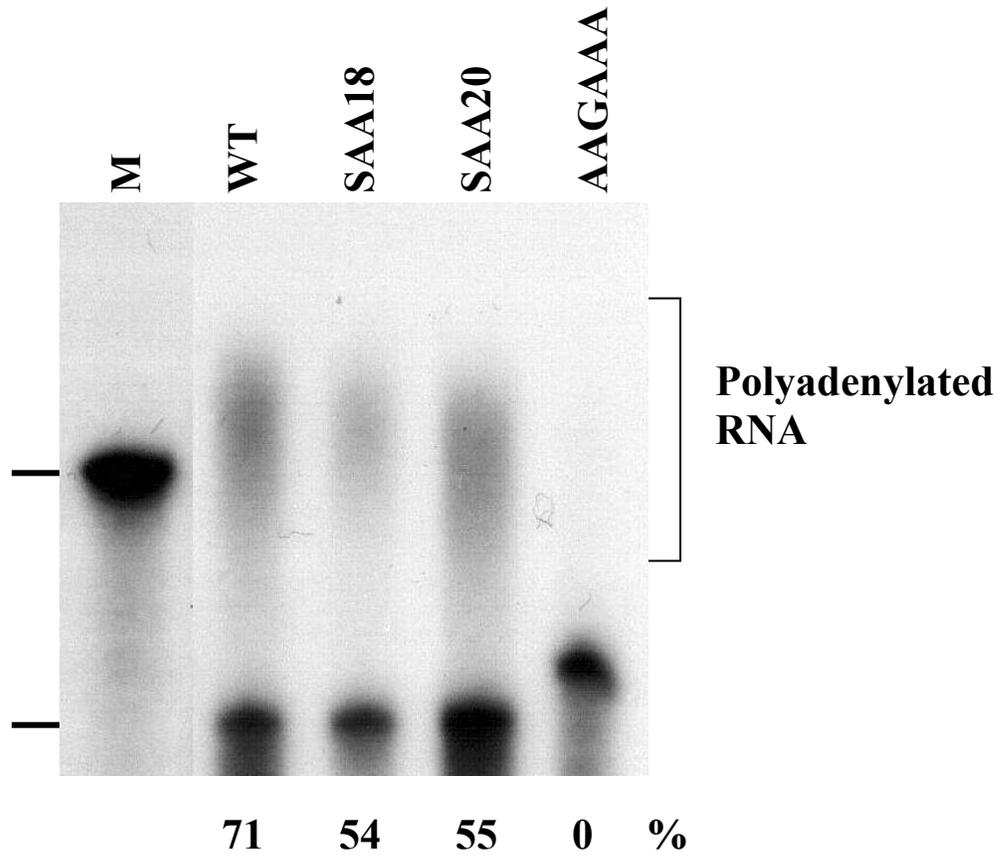
present in equal amounts and homogenously located, the binding sites are preferentially occupied by hnRNP H.

The SAA20 mutation decreased the affinity of both proteins about 4-5 fold (Figure 10 and Table II). The only difference between the 86 nt wt and SAA20 probe is three consecutive G→A mutations. The filter binding data showed that hnRNP F and H have parallel sequence selectivity.

#### **4.3.2. The mutations that significantly decrease hnRNP H'/F affinities decrease polyadenylation efficiency**

To determine if decreasing hnRNP H/F binding has an effect *in vitro* on the strength of polyA site, SAA18 and SAA20 mutations (Table I) were put in the context of 224 nt SVL pre-mRNA (Probe A in Figure 7) by PCR. Then *in vitro* polyadenylation assays were performed multiple times with these templates, a representative assay is shown in Figure 11. Wild-type SVL probe was most efficient at polyadenylation. Decreasing the affinity of the hnRNP H and F on SVL pre-mRNA inhibited the overall polyadenylation percentage, indicating that the loss of the G5 tract had a negative effect on pA addition. Since hnRNP F/H depletion and add-back type of *in vitro* polyadenylation experiments have already been performed by (Bagga et al. 1998) (Veraldi et al. 2001), we want to extend our findings to the cell culture level. The 224 nt wt and mutant SAA20 and SAA10 SVL polyA regions were cloned into the 3' Untranslated Region (UTR) of mammalian GFP expression plasmid pEGFP-C1 (Clontech). Human embryonic kidney 293T cells were transfected with these constructs. Whenever GFP pre-mRNA is efficiently cleaved and polyadenylated, we expect to observe higher expression and more intense GFP fluorescence. Before testing the mutated GRS sequences in the pEGFP plasmids, we wanted to

make sure that the polyA site had a measurable effect in our gene expression assay. This is important because



**Figure 11** Diminishing the binding of hnRNP H' to SVL pre-mRNA decreases the polyadenylation efficiency *in vitro*.

Polyadenylation assays performed *in vitro* with internally labeled wtSVL and SVL having SAA18 and SAA20 mutations. Lane M has a RNA marker that is 225 nt longer than unpolyadenylated SVL substrate. Last lane contains the reaction performed with AAGAAA polyA signal mutant. All the reaction products were run on a 5% Denaturing gel.

we made use of the pEGFP parent plasmid which has a very strong CMV immediate early promoter in 293T cells and the overall expression of GFP might have been so high that a single strong or weak polyA site might not have made a significant difference in expression.

To construct the baseline vector, the default polyA site of the *gfp* open reading frame was removed from the pEGFP-C1 plasmid. Furthermore, the 224 nt wtSVL polyA site was synthesized by PCR and the default polyA site of pEGFP-C1 was replaced by this PCR product (See Methods section for detail). The former plasmid was named pEGFP-nopolyA whereas the latter as pEGFP-wtSVL. Both plasmids were linearized before transfection and do not have any downstream, non-consensus polyA signals.

We transfected 293T cells by lipid mediated transfection. After 18-41 hours of incubation, cells were fixed for quantitative flow cytometry analysis. Numerous replications of this experiment showed a very reproducible 9-10 fold induction with the wtSVL samples versus the no polyA construct, as quantified by flow cytometry (Figure 12 and 13A). This motivated us to proceed with cloning the mutants in the GFP reporter.

Mutations SAA20 and SAA10 were inserted into the context of the 224 nt SVL construct by PCR, then they were cloned into the pEGFP vector. Final plasmids contain a single polyA site having either wtSVL or SVL with the SAA20 or SAA10 mutation. All the plasmids were verified by DNA sequencing; other than the mutation, they are identical.

Quantitative flow cytometry results with the 293T cell line showed a decrease in GFP mean intensity of pEGFP-SAA20 and an increase with the SAA10 transfected cells (Figure 13A)

relative to wild type SVL, respectively. This result correlated with our *in vitro* binding assays. Whenever there is more hnRNP H/F binding (SAA 10), more GFP is expressed and when there is less binding (SAA20) the opposite is true.

The effect of the SAA20 mutation was verified by using a western blot for the GFP protein in the A20 and 293T transfected cells. The A20 cell line is a B lymphoma where hnRNP F was relatively high (Veraldi et al. 2001). Transfection of A20 cells showed about 2-3 fold repression of GFP expression in the case of pEGFP-SAA20 (Figure 13B) and 293T cells showed a slight decrease in GFP. Neomycin Phosphotransferase (npt2) gene is present in all pEGFP plasmids; the Npt2 levels serve as an internal transfection control and are comparable among different samples as determined by western blotting (Figure 13B). Therefore we concluded that weakening the H/F binding site decreases the gene expression.

The expression difference between wtSVL and SAA20 mutant version of GFP plasmids is the greatest in the case of the A20 cell line (Figure 13B). We speculated that this might be due to an altered amount of hnRNP H and hnRNP F in A20 compared to the 293T cell line. We determined the relative levels of the two proteins in these cell lines by western blot (Figure 13C). Significantly more hnRNP H as compared to hnRNP F is found in the 293T cells whereas the A20 cell line has comparable hnRNP H and hnRNP F levels. In the former case, hnRNP H may still be present at high enough levels to show some binding to our SAA20 mutant whereas we believe that A20 cell line exhibits less H' binding to the mutant.

#### **4.3.3. Overexpression and downregulation of hnRNP F and H has opposite effects on gene expression**

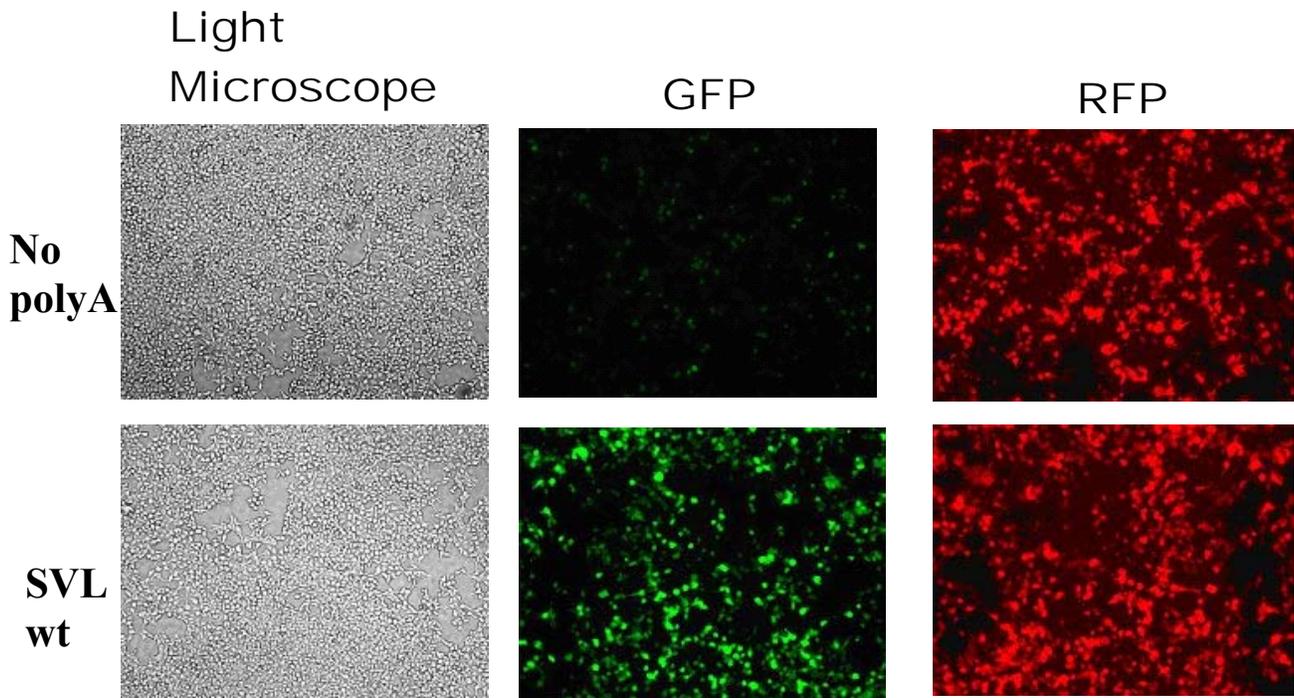
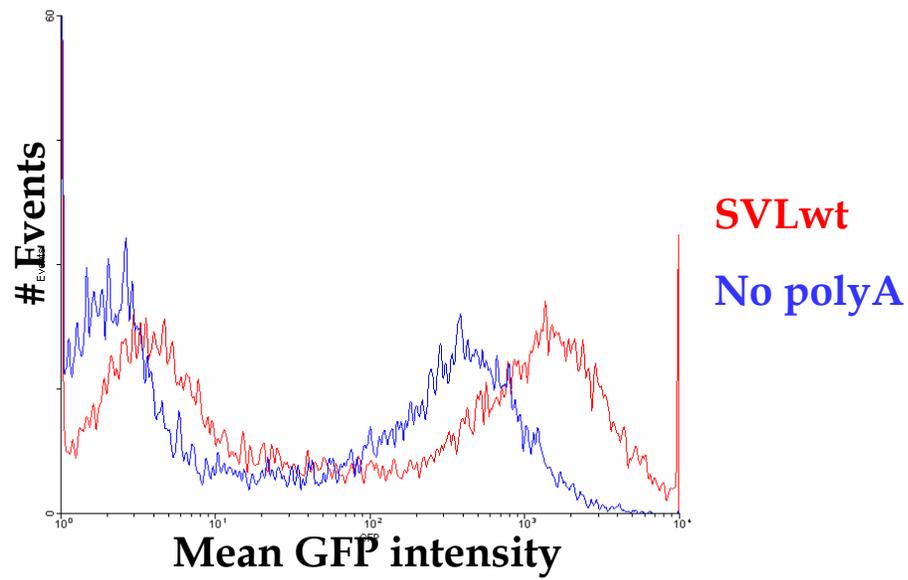
If our hypothesis is correct, then the modulation of the levels of hnRNP F and H would result in differences in gene expression on a *gfp* gene with a single polyA signal. We first

overexpressed hnRNP F and hnRNP H separately in 293T cells (Figure 14A); the overexpression levels are modest, about a 20-30% increase over the endogenous protein levels. The hnRNP F and H overexpressing cells were cotransfected with pEGFP-SVLwt plasmid. Flow cytometric quantification of GFP levels shows an about 2 and 3 fold induction by hnRNP F and H forced expression, respectively (Figure 14B). The same experiment was repeated by replacing the pEGFP-SVLwt with pEGFP-SAA20. Increasing the levels of hnRNP F and H does not effect the gfp expression with SAA20 plasmid, since it binds the proteins very poorly and the extra proteins expressed by cells is not great. This shows that the effects of hnRNP F and H are sequence dependent.

The complementary experiment would be the downregulation of these two proteins by RNAi. However the extensive similarity between hnRNP F and H poses a problem: RNAi for hnRNP F could also downregulate hnRNP H, and vice versa. In addition, the hnRNP H siRNA should not only target hnRNP H1 but also hnRNP H2 at the same time to be effective. Having tried various hnRNP F and hnRNP H specific siRNA duplexes, we were finally able to get the two different siRNA duplexes we needed (See Methods). The hnRNP F specific siRNA only downregulated hnRNP F but not hnRNP H and vice versa (Figure 15A). The hnRNP H specific siRNA was designed to a region where H1 and H2 messages are identical but differ from hnRNP F (See the methods section).

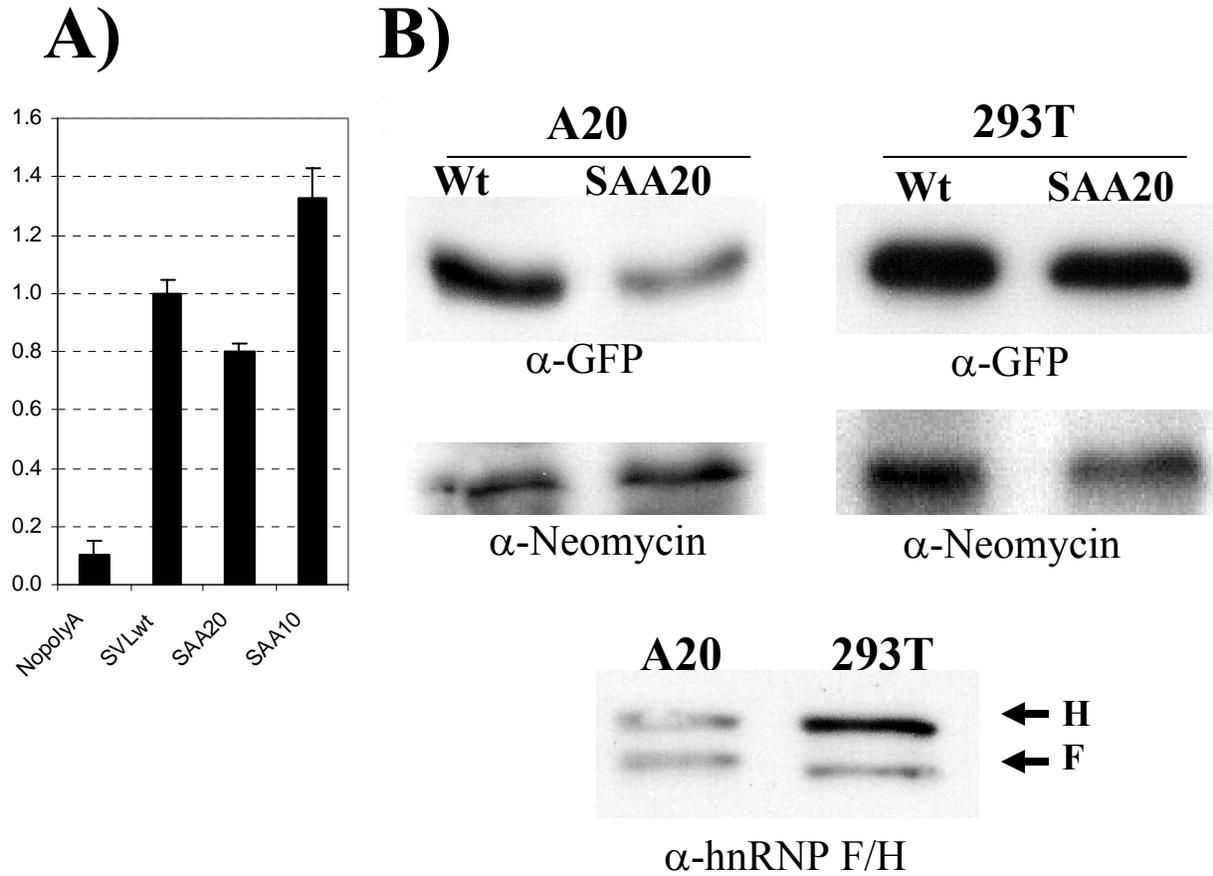
Downregulation of hnRNP H and F both resulted in slightly decreased gfp expression with pEGFP-SVLwt plasmid as assayed by flow cytometry (Figure 15B). Interestingly, whenever the pEGFP-SAA20 plasmid was transfected, the effect of downregulation by the siRNA was much larger (Figure 15B). A potential reason for this experimental outcome is that not only is the GRS mutant (=SAA20), i.e. it binds with a 4 to 5-fold reduced affinity to hnRNP

F and H, but also that the hnRNP F and H levels are down by 4-5 fold. This experimental situation is opposite to that shown in Figure 14B where the RNA (SVLwt) binds to hnRNP F and H with a higher affinity and there is more hnRNP F or H in the cells.



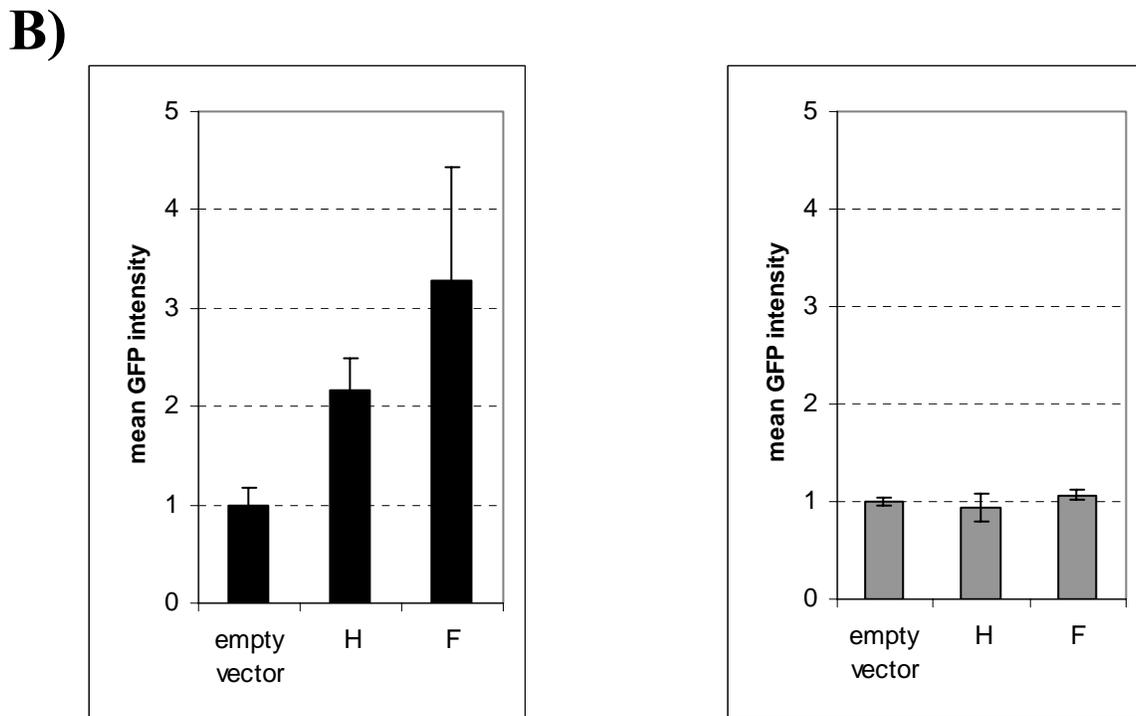
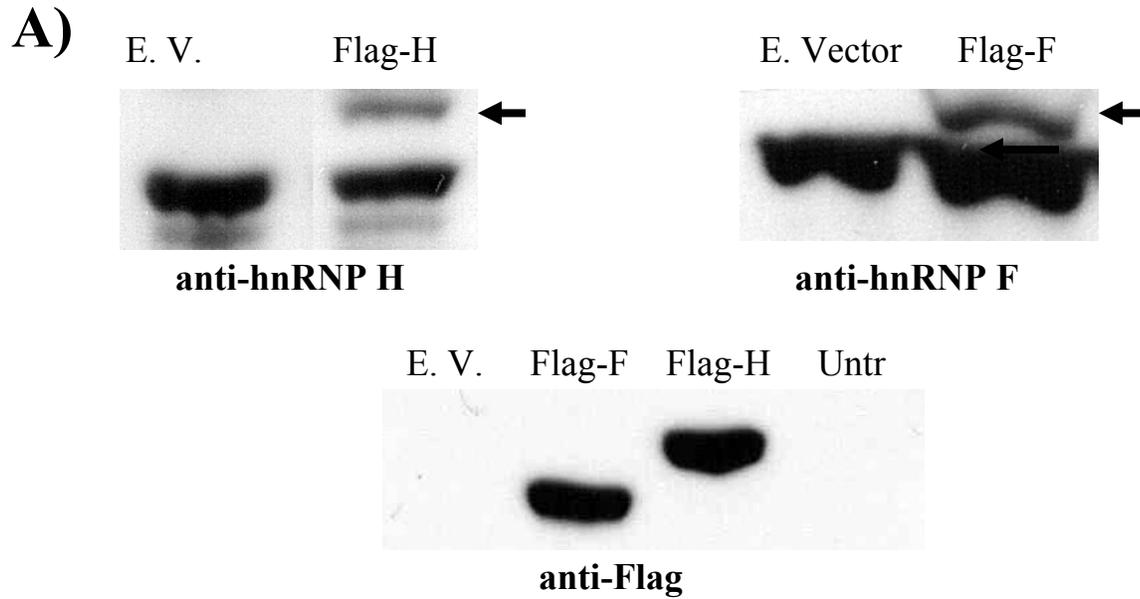
**Figure 12** Having no polyadenylation site reduces the GFP expression dramatically *in vivo* as assayed by fluorescent microscopy.

Fluorescent and light micrographs of pEGFP-nopolyA, pEGFP-SVLwt transfected 293T cells. The pictures of live monolayer 293T cells in the first column were taken under bright light with a Nikon T-100 microscope. GFP and RFP (DsredExpress) proteins were detected as described in Materials & Methods.



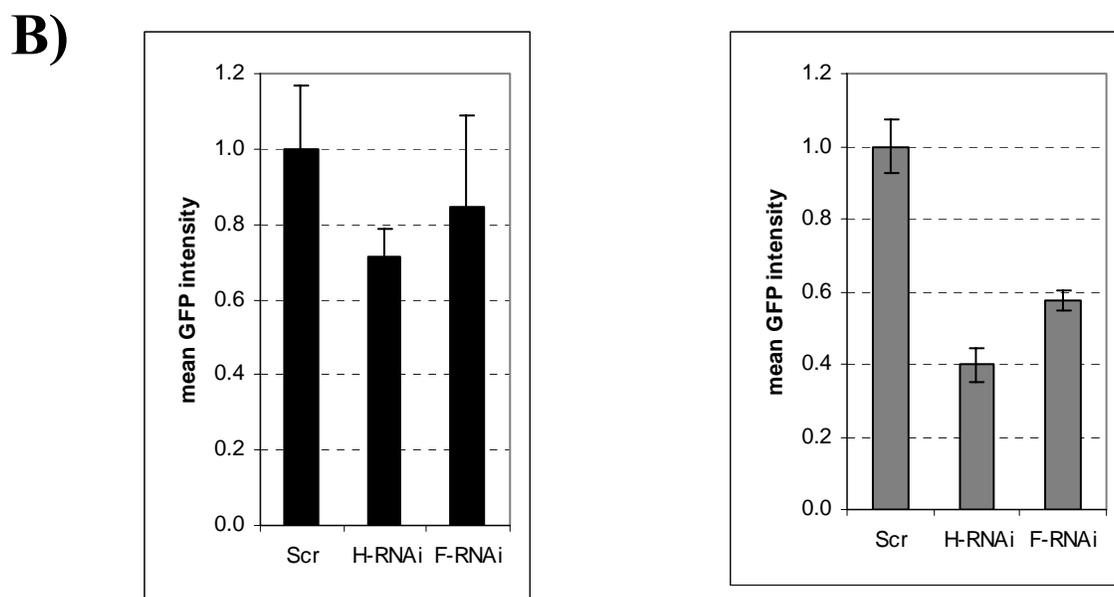
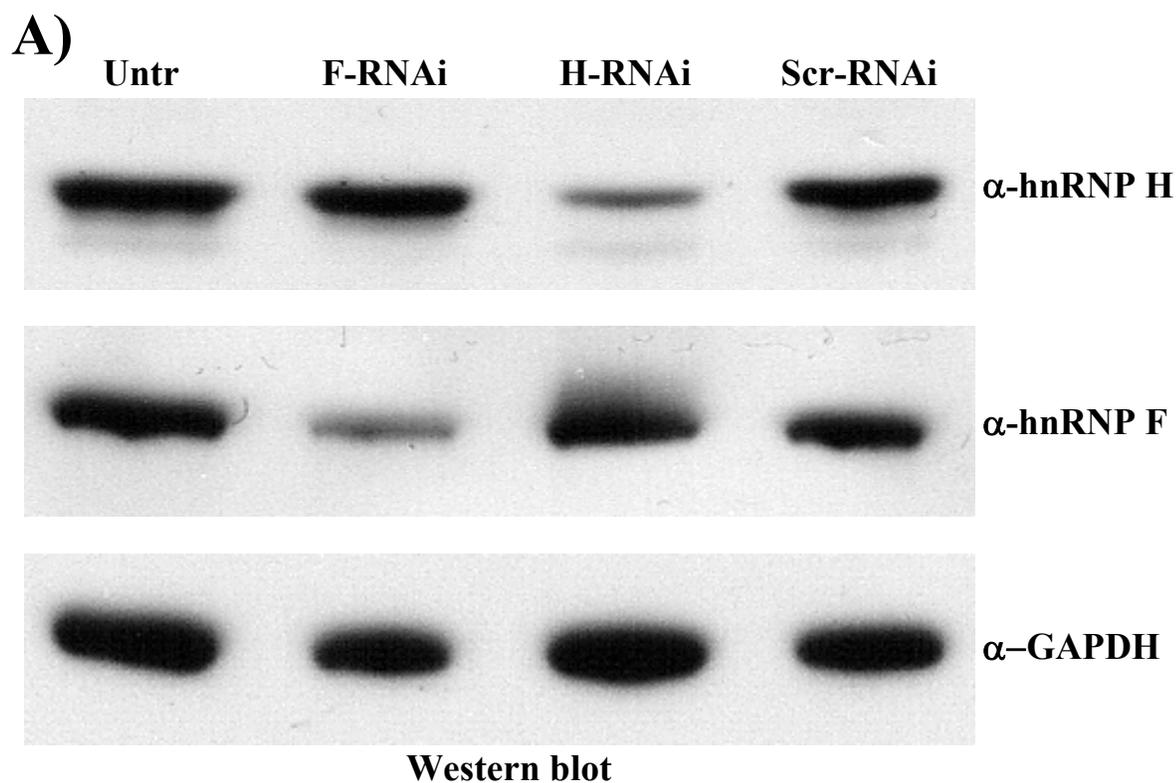
**Figure 13**      **Decreasing the binding of hnRNP H' to 3'UTR of gfp transcript reduces the GFP expression *in vivo*.**

- A) Mean Gfp fluorescence intensities corresponding to pEGFP-nopolyA, pEGFP-SVLwt, pEGFP-SAA20 transfected 293T cells were quantified by flow cytometry. Each transfection was performed three times. Error bars represent the standard error. The same results were verified by western blot shown on the right panel. Middle panel shows the quantification of this western blot. B) GFP levels in pEGFP-SVLwt, pEGFP-SAA20 transfected A20 cells were determined by western blots with anti-GFP antibody. Anti-Neomycin antibody is used for loading control. Quantification of this western blot is given on the left panel. C) Relative hnRNP H'/H and hnRNP F levels in A20 and 293T nuclear extracts were detected with a western blot assay using with hnRNP H and hnRNP F antibodies.



**Figure 14      Forced expression of hnRNP F and H resulted in induction of GFP expression**

A) Flag tagged hnRNP F and H were overexpressed in 293T cells as explained in Methods section. Western blots performed with anti-hnRNP F, anti-hnRNP H and anti-Flag antibodies confirmed the overexpression. Untr stands for untransfected cells whereas empty vector samples are the cells transfected with the backbone plasmid alone. B) Quantitation of GFP levels in hnRNP F and H overexpressing cells by flow cytometry. The left panel shows the data corresponding 293T cells cotransfected with pEGFP-SVLwt whereas the right panel with pEGFP-SAA20.



**Figure 15** Downregulation of hnRNP F and H by siRNAs diminished the GFP expression.

**A)** Western blots performed with anti-hnRNP H1/H2, anti-hnRNP F and anti-GAPDH antibodies. Untr stands for untransfected cells and Scr-RNAi is a negative control siRNA that does not match any message. **B)** Quantitation of GFP levels by flow cytometry. The left panel shows the data corresponding 293T cells cotransfected with pEGFP-SVLwt whereas the right panel with pEGFP-SAA20.

**Table 5** The sequence of wt and mutant GRS oligoribonucleotides

<b>GRSwt</b>	GGG GGA GGU GUG GG
<b>SAA1</b>	<b>AGG</b> GGA GGU GUG GG
<b>SAA2</b>	G <b>AG</b> GGA GGU GUG GG
<b>SAA3</b>	G <b>GA</b> GGA GGU GUG GG
<b>SAA4</b>	GGG <b>AGA</b> GGU GUG GG
<b>SAA5</b>	GGG <b>GAA</b> GGU GUG GG
<b>SAA6</b>	GGG GGA <b>AGU</b> GUG GG
<b>SAA7</b>	GGG GGA <b>GAU</b> GUG GG
<b>SAA8</b>	GGG GGA GGU <b>AUG</b> GG
<b>SAA9</b>	GGG GGA GGU <b>GUA</b> GG
<b>SAA10</b>	GGG GGA GGU GUG <b>AG</b>
<b>SAA11</b>	GGG GGA GGU GUG <b>GA</b>
<b>SAA12</b>	GGG GGA GGU GUG <b>AA</b>
<b>SAA13</b>	GGG GGA GGU <b>GUA</b> <b>AG</b>
<b>SAA14</b>	GGG GGA <b>AAU</b> GUG GG
<b>SAA15</b>	GGG <b>AAA</b> GGU GUG GG
<b>SAA16</b>	G <b>GA</b> <b>AGA</b> GGU GUG GG
<b>SAA17</b>	<b>GAA</b> GGA GGU GUG GG
<b>SAA18</b>	<b>AAG</b> GGA GGU GUG GG
<b>SAA19</b>	G <b>GA</b> <b>GAA</b> GGU GUG GG
<b>SAA20</b>	G <b>GA</b> <b>AAA</b> GGU GUG GG

**Table 6** Apparent dissociation constants (nM) of hnRNP F and H' proteins to wtSVL and SAA20 probes.

	<i>hnRNP H'</i>	<i>hnRNP F</i>
<i>wtSVL</i>	37	121
<i>SAA20</i>	192	545

#### 4.4. DISCUSSION

The lack of a systematic mutation analysis had prevented a determination of the optimal hnRNP binding sequence for hnRNP F or H. By performing extensive mutational analysis and a multiple alignment with the experimentally proven hnRNP H and F sites we conclude that a run of five Gs and a single Adenine residue 8-9 nucleotide downstream of the Guanine core with another single upstream Guanine make up the consensus F/H binding site.

We have shown that hnRNP F and H bind to the same sequences albeit with different affinities. The extensive similarities of RRM of two proteins lead us to think that the observed differences in affinity may be due to the C-terminal-most domain present in hnRNP H but absent in hnRNP F. This highly charged part of hnRNP H may increase its affinity of binding without altering the sequence preference.

Many RNA binding proteins have multiple RRMs that are capable of binding RNA independently as separate units. In contrast, most of the DNA binding proteins have a single DNA binding domain responsible for sequence specific binding (Biggin and Tjian 2001). For DNA binding proteins there is typically a single defined optimal binding sequence; any type of mutation in this sequence decreases the affinity significantly. On the other hand, multi RRM containing RNA binding proteins, especially hnRNPs, could bind to RNA with each domain capable of RNA recognition. There is not a single optimal RNA binding sequence in the sense that a minor change is well tolerated and does not decrease the apparent dissociation constant dramatically. In addition, unlike many DNA binding proteins, hnRNPs are very abundant in the nucleus and there are no competing histones binding to RNA preventing the protein-RNA interactions. That is perhaps why the K<sub>d</sub>s of hnRNP F and H proteins are higher compared to a typical transcription factor K<sub>d</sub>, generally around single digit nM range.

The observed affinity of probe G (Figure 8b) to hnRNP F is lower compared to probe A (Figure 8a). The reason for this observation may be that the consensus downstream Adenine and upstream Guanine are not present in this short oligo both of which are present in the probe A.

The SAA20 mutant is diminished not only in the hnRNP F and H binding but also in the *in vitro* polyadenylation reaction. The effect of SAA20 mutation on *in vitro* polyadenylation is modest but reproducible. This may, in part, be due to the fact that polyadenylation *in vitro* is very processive once it starts unlike the situation in cells where the initial synthesis of about 10 Adenines is very slow (Wahle 1995). If an *in vitro* polyadenylation reaction is allowed to continue for extended periods, the length of the polyA tail goes well beyond the 200-300 nucleotides.

We have checked the potential secondary structures of the of the 86 nt wt (Probe B in Figure 7) and SAA20 mt SVL pre-mRNAs with the mfold software (Zuker 2003) and see no change of secondary structures by this criteria. However, the observed binding affinity of hnRNP F to the 86 nt probe B is higher than to the 224 nt probe A (Figure 8A) suggesting that secondary structure in probe A might inhibit the binding of hnRNP F. Potential secondary structures of probe A and B were determined by using mfold software (Zuker 2003). The Gibbs free energy difference corresponding to probe A's most stable secondary structure is significantly lower (-48.7 kcal/mole) than that of probe B (-14.1 kcal/mole) meaning that the former structure is noticeably more stable. This may contribute to the observed higher affinity of probe B than probe A to hnRNP F (Figure 8A). There is no secondary structure assigned to the 14 nt GRS transcripts; a result we expected since G-C basepairing is not possible. Multiple Guanines forms quarternary Guanine structures which was not reported by Zucker algorithm.

The effect of hnRNP F and H binding to the 3' UTR of gfp reporter on gene expression in vivo has been verified by multiple independent lines of evidence. First, we mutated RNA, i.e. we made SAA10 and SAA20 mutants which bind to hnRNP F and H with a higher and lower affinity compared to wt SVL pre-mRNA, respectively. These mutations were placed in the 3'UTR of gfp plasmid and transfection-based experiments showed that SAA10 and 20 have positive and negative effects on gfp expression, respectively. In other words, there is a correlation between the gfp expression and hnRNP F and H binding based on the sequence of the RNA. We further extended our results by keeping SVL RNA in the wild type sequence and overexpressing hnRNP F or H proteins. In this way, we showed increased gfp expression by modulating hnRNP F or H levels. Parallel to the previous SAA10 data, overexpression of either protein resulted in enhanced gfp expression. Protein: RNA binding therefore controls the expression and changing either modulates it. On the other hand, the overexpression of hnRNP F or H coupled with the SAA20 plasmid cotransfection could not modulate gfp expression (Figure 14B) since the overexpression levels are within the physiological protein levels (20-30% extra) instead of the 400-500% increase which would be needed to compensate for the SAA20 mutation. The 20-30% extra hnRNP F and H was not enough to increase the binding levels to SAA20 plasmid, hence we observed no change in its expression with only a moderate increase in hnRNP F or H.

The complementary evidence supporting our hypothesis is the RNAi mediated downregulation of hnRNP F and H. In the overexpression experiment, we wanted to increase the binding activity hence we have overexpressed the hnRNP F or H and cotransfected the cells with the pEGFP-SVLwt. The converse of this experiment would be the downregulation of hnRNP F or H coupled with pEGFP-SAA20 cotransfection. Not only do hnRNP F and H have about five

fold lower affinity to the SAA20 mutant than the wild type but there is also 4-5 fold less hnRNP F and H protein in this siRNA experiment (Figure 15A). We believe that we have knocked out the F/H binding to RNA in this situation. This resulted in a clear reduction of gfp expression (Figure 15B). Since the siRNA was very effective, there is also a slight reduction of gfp levels when SVLwt plasmid were cotransfected (Figure 15B).

Our *in vivo* experiments do not allow us to determine which is the more influential protein, hnRNP F or H. When we overexpress the hnRNP proteins (Figure 14), both increase GFP expression by greater than 2-fold, but hnRNP F appears to increase GFP expression slightly more than hnRNP H does. Nevertheless in 293T cells the hnRNP H levels are already higher than those of F, so hnRNP H could be close to maximal. Meanwhile, in the siRNA knockouts decreasing hnRNP H levels appears to decrease GFP expression slightly more than in the hnRNP F knock-downs. Since the levels of hnRNP F and H proteins were not changed to the same extent between the various manipulations definitive conclusions are difficult to draw. However, based on our *in vitro* binding assays with pure proteins, we believe that hnRNP H should have the more dominant effects. Based on the previously observed formation of complexes containing both proteins (Veraldi et al. 2001) (Markovtsov et al. 2000) the over-expression of hnRNP F might be acting positively on the SVL downstream region by driving complex formation with hnRNP H.

We have shown that hnRNP F and H proteins are modulators of gene expression through their sequence-specific effect on polyadenylation. We and others have shown that hnRNP F and H proteins are differentially expressed in normal and cancer tissues (Veraldi et al. 2001; Honore et al. 2004). The implications of our current results are that the hnRNP F/H amounts in the cell and

the binding affinity of the RNA sequence of a target gene may dramatically alter the use of a polyA site in any multi-polyA site gene and, in genes with multiple tandem polyA sites, like the immunoglobulin heavy chains, the choice of polyA sites can be altered.

#### 4.5. MATERIALS AND METHODS

##### PCR Mediated Constructs

DNA templates that were used to create RNA probe D and F (Figure 7) were synthesized by PCR as explained in (Wilusz et al. 1988). The pSP65-SVL plasmid has a wild type SVL polyA region. Using a 5' PCR primer with T7 promoter overhang and a 3' primer with mutant GRS overhang, we were able to synthesize PCR products having T7 promoter and mutant GRS. All the PCR primers we used to create GRS mutants in the context of 224 nt SVL are listed below. The 248-5'CC primer is the common forward primer and has a T7 promoter (in italic). The 248wt-3' primer was utilized as a reverse primer to create the wtSVL. On the other hand, 248A-3' and 248B-3' were used to amplify the SVL having SAA20 and SAA18 mutations, respectively. GRS mutations are underlined and written bold.

**248-5'CC:** *TAATACGACTCACTATAGAATACACGGAATTCGAGCT*,

**248A-3':**AAAAAACC TCCCACACCT**TTT**CCTGAACCTGAAACATAAAAATGAAT,

**248B-3':**AAAAAACCTCCCA CACCTCC**CTT**TGAACCTGAAACATAAAAATGAAT,

**248wt-3':**AAAAAACCTCCCACACC TCCCCCTGAACCTGAAACATAAAAATGAAT. The

PCR protocol used for each reaction consists of 30 cycles of 94<sup>0</sup>C for 1 min, 58<sup>0</sup>C for 30 sec, 72<sup>0</sup>C for 30 sec. Accupol DNA Polymerase (Genechoice) was used for each PCR reaction.

RNAs *in vitro* transcribed from these templates were used for polyadenylation assays.

##### Plasmid Constructs

Plasmids that were used to create RNA probes A, B, C, E (Figure 7) for gelshifts were a generous gift from Dr. Jeff Wilusz and described in (Qian and Wilusz 1991). The GFP open reading frame on pEGFP-C1 (Clontech) has an SV40 early polyA site. By using EcoRI and MluI sites, this default site was removed and then the inserts produced by PCR were cloned into the same sites. The 5' primer for each case is 248-5'CC whose sequence was given above. Wt and mt SVL having 5' EcoRI and 3' MluI flanking sites were synthesized by PCR as explained above using the following 3' primers.

**SAA20-3'CLONING:**

TACCGACGCGTAAAAAACCTCCCACACCTTTTCCTGAACCTGAAACATAAAATGAAT,

**SAA18-3'CLONING:**

TACCGACGCGTAAAAAACCTCCCACACCTCCTTTGAACCTGAA ACATAAAATGAAT,

**WT-3'CLONING:**

TACCGACGCGTAAAAAACCTCCCACACCT CCCCTGAACCTGAAACATAAAATGAAT,

**SAA10-3'CLONING:**

TACCGACGCGTAAAAAACCTCTCACACCTCCCCCTGAACCTGAAACATAAAATGAAT.

Each PCR product was double digested with EcoRI and MluI for 4 hr at 37<sup>0</sup>C. All the cloning experiments were performed as described in (Spear 2000).

To create pEGFP-nopolyA, the default polyA site of GFP open reading frame was removed from pEGFP-C1 plasmid by using EcoRI and MluI sites and two oligos were inserted to the same position. The sequence of both oligos are as following: Eco1:AATTCGATCGGATATCAGTACTA, Mlu1: CGCGTAGTACTGATATCCGATCG. Every cloned plasmid was verified by DNA sequencing. All the pEGFP and pDsredExpress vectors were linearized by SspI digestion for transfection. This only disrupts the fl origin in all the plasmids.

## **DNA Oligonucleotides, *In vitro* Transcription and Polyadenylation Assay**

All oligonucleotides were synthesized at the DNA Synthesis Facility, University of Pittsburgh. *In vitro* transcription reactions were performed as explained in (Milligan JF 1987) with the exception that no cap analog was used. The 14 nt oligoribonucleotides given in Table 1 were produced by *in vitro* transcription reactions with T7 RNA Polymerase according to (Milligan JF 1987). Polyadenylation assays were performed with A20 nuclear extracts as described in (Natalizio et al. 2002).

## **Protein Purification**

Plasmid pET-15b having the human hnRNP F ORF and the plasmid pGEX-2t, having human hnRNP H' ORF, were kind gifts of Dr. Douglas Black and Dr. Jeff Wilusz, respectively. Histidine tagged hnRNP F for gelshift experiments was purified as described (Chou et al. 1999). To obtain highly pure and nondenatured his-hnRNP F and GST-hnRNP H' for filter binding assay, a two step purification method was developed for each one. Escherichia coli (Rosetta, Novagen) cells were used to express both proteins. These cells have a plasmid coding for six tRNAs for mammalian codons that are rarely used in bacteria. Cells were grown at 37<sup>0</sup>C until the OD<sub>600</sub> reached 0.5. Then they were induced with 1 mM IPTG and allowed to grow at 28<sup>0</sup>C for 3 hr. The final OD<sub>600</sub> is about 2.0. In the case of the hnRNP F purification, cells were lysed with 50 mM Tris pH 8.0 and 1% Triton X-100 and 5 mM ATP were added to the lysed cells. Three burst of sonication for 20 sec at 10 Watts (RMS) were performed on ice. Cells were spun at 15,000 g for 15 min and the pellet was discarded. The supernatant was loaded on to a gravity operated 10 ml DEAE-Sepharose plastic column (15 mm diameter, Pierce) that was equilibrated with 50 mM Tris pH 8.0. Flowthrough was collected and purified over 2 ml metal affinity Talon (Clontech) plastic column (7 mm diameter, Pierce) according to manufacturer's protocol. Two step elutions

were performed with 10 mM and 300 mM Imidazole; the latter has the pure his-hnRNP F (Figure 16). In the case of hnRNP H', the induction was done with 0.1 mM IPTG to enhance the solubility of the protein. The cells were lysed with 1xPBS pH 7.3 with 1% Triton-X-100 and 5 mM ATP added. The sonication and spin were performed as described above for hnRNP F. The hnRNP H' recombinant protein was first enriched by using 2 ml Glutathione-Sepharose (Amersham Biosciences) column according to the manufacturer's protocol. Two-step elution with 1 and 100 mM Glutathione in 50 mM Tris pH 8.0 was performed and the latter contained the most enriched fraction. Glutathione was eliminated by buffer exchanging on a 10 ml medium G-25 Sephadex column. The final fractions in 1xPBS pH 7.3 were loaded to 10 ml DEAE-Sepharose column and two step elution was performed with 250 and 750 mM NaCl; the latter has the pure GST-hnRNP H' (Figure 17). Finally, both protein samples were buffer exchanged with Buffer D by using 10 ml medium G-25 Sephadex column and are at least 90% pure as assayed by Coomasiie staining.

### **Electromobility Shift Assay and Filter Binding Assay**

Each gelshift reaction has 12.5 ug purified yeast tRNA, 3 mM DTT, 1.2 mM ATP, 16.7 mM Creatine Phosphate, 0.75 mM MgCl<sub>2</sub>, 2.7% PVA, RNA and protein. 1 fmol radiolabeled RNA (5 fmol if RNA probe is <50nt), and the indicated amount of protein were added to each tube. Every tube had 20 µl in total; volumes were equalized with Buffer D (20 mM Hepes pH7.9, 100 mM KCl, 1 mM DTT, 20% Glycerol, 0.2 mM EDTA, 1x Protease cocktail (Calbiochem) . All the reactions were incubated for 10-20 min at 30<sup>0</sup>C.

Acrylamide gels of 4-8% with a 20:0.25 Acrylamide:bisacrylamide ratio were run at 30 mAmp, 300V and 15W for about 1-2 hours. Gel running buffer contains 25 mM Tris Base, 25

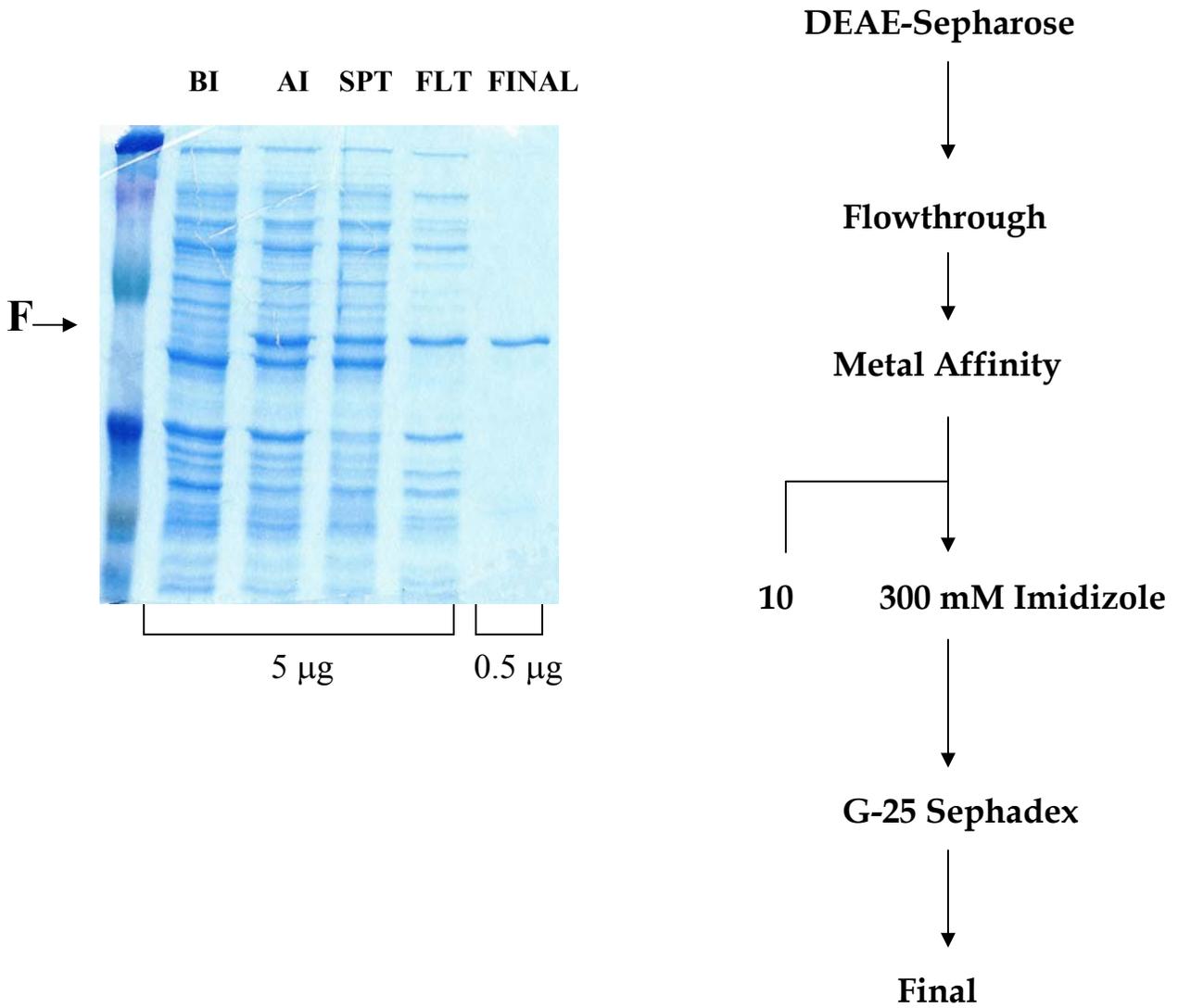
mM Boric acid, 1 mM EDTA. Gels were also prepared in this buffer. Gel runs were always started with both a cold gel and cold buffer and were performed at room temperature.

Gelshifts with competitors were done with a 2 min pre-incubation of protein with RNA competitor followed by 10 min incubation of the <sup>32</sup>P labeled probe B. The specific competitor was a tritium-labeled 86 nt probe B whereas the nonspecific competitor is polyA having the same length.

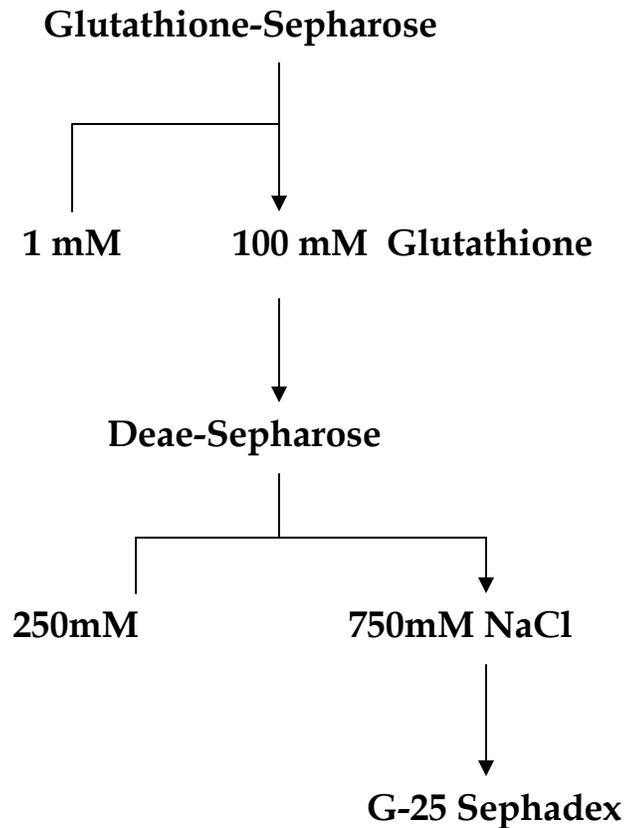
Filter binding assays were done essentially as described (Liu et al. 2002).

### **Transfection**

Cells were transfected with linearized plasmids by using Fugene 6 (Roche). pDsredExpress plasmid (Clontech) coding for DsredExpress protein; a variant of original Dsred protein (Baird et al. 2000) was used as a transfection marker. From 1.5 to 3.0x 10<sup>5</sup> cells were plated on 6 well plates 12-18 hours before transfection. About 2 µg equimolar plasmid, serum free media and Fugene 6 transfection agent with a fugene6:plasmid ratio of 3:2 (v:w) were combined in 100 ul volume. This mix were incubated 15 min at RT and added to the wells having 2 ml serum positive media.



**Figure 16** hnRNP F was purified by ion-exchange and metal affinity columns from *E. coli* extracts. BI, AI, SPT and FLT stand for Before Induction, After Induction, Supernatant and Flowthrough, respectively.



**Figure 17**  
**coli extracts.**

**hnRNP H was purified by glutathione-spearose and ion-exchange columns from E.**

**RNAi transfection:**

Approximately  $1.5 \times 10^5$  cells were seeded on each well of 6-well plates a day before the transfection. A total of 200 ul transfection mix including 0.1 nmol duplex siRNA, 8 ul Amine transfection reagent (Ambion) and serum free DMEM was added to each well having 800 ul media with serum, essentially according to the manufacturer's instructions. After 20-24 hour incubation the media was exchanged with 2 ml fresh serum positive media with serum. After a total of two days of incubation, cells were transfected with pEGFP based plasmids as explained above. They were then harvested after 24 hours of incubation for flow cytometry. The identical

region in hnRNP F and hnRNP H is targeted with the RNAi experiments. The siRNA sequences are given below. The region targeted at the hnRNP H1 message is identical in hnRNP H2 i.e. the same siRNA duplex is effective both for hnRNP H1 and H2. Furthermore three mismatches between hnRNP F and H transcripts in this region ensured that there is no cross reactivity. The differences between hnRNP F and H siRNAs are highlighted.

siRNA for hnRNP F (Ambion siRNA lot # 044786siB) :  
GGUGUCCAUUUCAUCUACAtt  
UGUAGAUGAAAUGGACACctg

siRNA for hnRNP H1 (Ambion siRNA lot # 048945si) :  
GGUAUUUGUUUCAUCUACAtt  
UGUAGAUGAAACGAAUACctt

### **Overexpression:**

Plasmids having the hnRNP F and hnRNP H2 ORF have pCMV-flag (Sigma) backbone and were described in (Veraldi et al. 2001). hnRNP F, H2 or empty vector plasmid were transfected to 293T cells with the Fugene 6 reagent and allowed to overexpress for 24 hours then the very same cells were retransfected with the pEGFP based plasmids for two days and they are harvested for flow cytometry or western blots.

### **Flow Cytometry and Western Blot**

After incubating 18-41 hours in media, cells were washed with 1xPBS and trypsinized with 0.5 ml 1xTrypsin for 1 min. Cells were spun down in 5 ml Falcon tubes. Cell pellets were sequentially washed with 1xPBS and then with 1xFACS buffer (5% FCS, 1xPBS, 0.01% NaAzide). Finally, cells were fixed in 1xParaformaldehyde in 1xPBS.

All the flow cytometry experiments were performed with a single 488 nm laser. Both GFP and DsredExpress proteins were excited with 488nm laser and emissions were recorded in FL1 (green) and FL2 (red) or FL3 (red) channel. A total of  $5-20 \times 10^3$  events were recorded for each sample. The analysis was done with the WinMdi software available at <http://facs.scripps.edu/software.html> (November 2004).

All the flow cytometry experiments were performed with the Coulter Epics XL flow cytometer operating with a single 488 nm laser. Both GFP and DsredExpress proteins were excited with 488nm laser and emissions were recorded in FL1 (green) and FL3 (red) channel. A modest compensation was done by subtracting 3.7% of the FL1 channel from FL3. A total of  $5-20 \times 10^3$  events were recorded for each sample.

Western blots were performed as described in (Martincic et al. 1998). Anti-GFP and anti-DsredExpress antibodies were purchased from Clontech whereas anti-Neomycin antibody was obtained from Upstate Biochemical Cell Signaling Solutions.

### **Fluorescent Microscopy**

Fluorescent images of monolayer live cells were captured from six well plates with a Nikon TS-100 microscope. GFP is excited with 425 nm and emission was recorded at 500 nm whereas DsredExpress was excited with 545 nm and the emission is recorded at 620 nm. The gain and exposure settings are identical for all images. Each original image has 1600x1200 pixels and 24 bit color.

5. MANUSCRIPT TWO: GLOBAL GENE EXPRESSION ANALYSIS REVEALS  
NOVEL B-CELL STAGE SPECIFIC GENES THAT ARE REGULATED BY HNRNP F.

**5.1. ABSTRACT**

In an attempt to identify the factors that play a role in the membrane to secretory specific Immunoglobulin (Ig) switch in Plasma cells, we performed gene expression analysis on a macro scale with Affymetrix oligonucleotide based murine chip U74Av2 and two pairs of samples. The first group consists of AxJ plasmacytoma and A20 B-cell lymphoma cell line. An extensive list of B-cell stage specific genes were obtained. While many genes are expected to be important or required for B-cell development, we are particularly interested in the candidate transcripts that may have a role in membrane to secretory Immunoglobulin (Ig) switch. In order to get the candidate transcripts that both have a role in Ig secretion and are regulated by hnRNP F protein, we also performed array experiments with hnRNP F overexpressing AxJ cells and compared its gene expression profile to empty vector transfected AxJ. Each experiment has been performed at least twice with a biological replicate. Extensive gene ontology analysis revealed three major functional protein classes that are important in secretory Ig switch: defense response, signal transduction and nucleic acid metabolism. Finally, ELL-related RNA polymerase II, elongation factor (ELL 2) is induced in Plasmacytoma cells whereas it is repressed upon forced expression of hnRNP F. A model regarding Ig secretion, involving ELL 2 is proposed.

## 5.2. INTRODUCTION

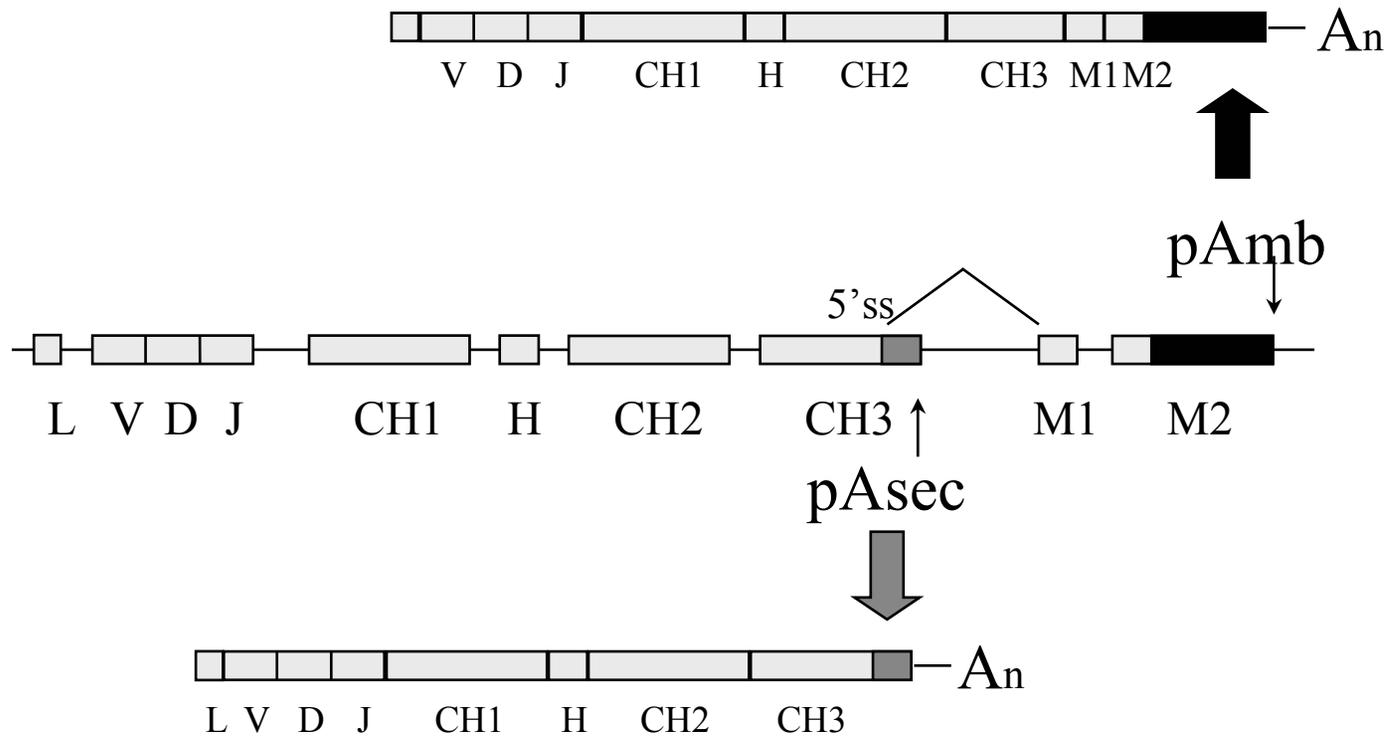
The immunoglobulin heavy chain genes represent one of the best studied examples of developmentally regulated polyadenylation. Ig  $\gamma$  heavy chain is diagrammed in Figure 18. The maturation of B cells to Ig-secreting plasma cells is associated with an increase in the ratio of secreted (sec) to membrane (mb) form Ig mRNAs which differ in their 3' termini. Since selection of the promoter proximal Ig  $\gamma$ -sec polyA site excludes the splicing downstream to the Ig  $\gamma$ -mb exons, early studies stressed the importance of the recognition of the suboptimal 5' splice site for maintaining Ig  $\gamma$  sec:mb ratio. However, there is now a strong support for the idea that an increase in Ig  $\gamma$ -sec polyA site selection is due to increase in polyadenylation efficiency.

The Ig  $\gamma$ -sec polyA site has been shown to be weaker than the Ig  $\gamma$  -mb polyA site. The observation that an increase in the cleavage/polyadenylation efficiency, but not splicing, correlated with the maturation of B-cells to plasma cells suggests that polyadenylation/cleavage at the proximal site is the regulated process (Peterson et al. 1991). While overexpression of CstF-64 in chicken B lymphoma cells shifts the polyA site selection balance to secretory site (Takagaki and Manley 1998), an increase in the binding of activity of CstF-64 but not the amount was observed in mouse plasmocytoma cell line AxJ compared to lymphoma cell line A20 (Edwalds-Gilbert and Milcarek 1995). Previously, an increase in the level CstF-64 was observed during the transition from G<sub>0</sub> to S phase (Martincic et al. 1998). However the fact that this change is not accompanied by increased binding activity of CstF-64 and increased polyadenylation efficiency in primary B-cells strongly suggests that the binding activity but not

the absolute level of CstF-64 is one of the determining factors on polyA selection on Ig  $\mu$  locus. The Ig  $\mu$ -sec polyA site was replaced by other weak polyA sites without disturbing the sec:mb ratio in mouse B-cell lymphoma A20 and mouse plasmacytoma AxJ cells (Matis et al. 1996). This suggests that the regulation is accomplished without regulatory cis-elements (Matis et al. 1996). An increase in CstF-64 binding activity was also observed in Adenovirus (Mann et al. 1993) and Herpesvirus (McGregor et al. 1996) infections without a change in the level protein. These findings strongly argue for the existence of accessory, tissue-specific factors that influence the binding activity of CstF-64 and/or activities of other basal factors, which influence hence polyadenylation efficiency.

In an effort to characterize B-cell stage specific polyadenylation factors, an *in vitro* purification method of polyadenylation complexes was developed (Veraldi et al. 2001). Veraldi et al adapted a HeLa spliceosome purification scheme (Reed 1990) to purify polyadenylation complexes formed on various polyadenylation sites. While polyadenylation complexes were successfully isolated with this method, a larger complex devoid of polyadenylation factors was reproducibly observed. hnRNP F and hnRNP H/H' proteins were found in this complex when it is formed by A20 B-cell lymphoma but not AxJ Plasmacytoma cells.

Overexpression of hnRNP F in AxJ cells resulted in the less frequent usage of the secretory specific polyA site on Immunoglobulin pre-mRNA. In other words, hnRNP F was able “revert” the B-cell clock. We did global expression analysis of the hnRNP F overexpressing cells and compared the expression profile with empty vector transfected AxJ cells. Moreover, we are particularly interested in the hnRNP F target genes that are also differentially expressed in a B-cell stage specific manner. To obtain these candidate transcripts, we also have performed microarray experiments with AxJ Plasmacytoma and A20 B-cell lymphoma cells.



**Figure 18** Alternative processing of Ig  $\gamma$  chain pre-mRNA result either in secretory or membrane specific IG. If promoter proximal and distal polyA site is used, secretory and membrane specific mature transcript will be synthesized, respectively.

### 5.3. RESULTS

There are many B-cell stage specific genes that were identified before (Underhill et al. 2003). By using microarray technology, we wished to get an extensive list of transcripts whose expression varies in our model system, the A20 lymphoma/ memory cell line and the AxJ plasma cell we derived from it. The affymetrix murine U74Av2 chip has probes that represent more than 12000 transcripts. We have performed separate array experiments with A20 and AxJ samples and compared the gene expression profile. The signal values of each chip was normalized before comparison. We found 207 transcripts that are differentially expressed in AxJ plasmacytoma cells; there are about equal number of induced and repressed transcripts. There are about an equal number of induced and repressed genes. Gene ontology analysis was performed with this group of genes ([www.affymetix.com](http://www.affymetix.com), December 2004). By making this analysis, we were able to get the potential functional class, molecular activity and cellular localization of each transcript. Of our particular interest are the transcripts that are important for DNA and RNA metabolism. The transcripts that are induced in AxJ cells that have a role in DNA and RNA metabolism are given in Table 7. A similar table was also established with the transcripts repressed in AxJ (Table 8). There are more transcripts in the latter table. Moreover, transcription factors make up the vast majority in both cases. Among the polyadenylation factors CPSF-2 (CPSF-100) is repressed in AxJ cells.

The same Affymetrix U74Av2 arrays were used with hnRNP F transfected AxJ samples. The hnRNP F was stably transfected into AxJ plasmacytoma cells; the gene expression profile of the transfected cells was compared with the empty vector transfected AxJs. There are 31 differentially expressed genes upon forced expression of hnRNP F (Table 9); all of them are

repressed. This is in parallel to hnRNPF's experimentally shown negative effect on polyadenylation (Veraldi et al. 2001) When the sets of differentially expressed transcripts in Plasmacytoma cells and hnRNP F overexpressing AxJ cells were intersected, three genes were obtained (Figure 19). Since hnRNP F somehow reverts the B-cell clock with respect to Ig secretion, we are particularly interested in genes with the opposite expression patterns i.e. transcripts that are induced in one case and repressed in the other. Elongation factor RNA Polymerase II 2 (ELL 2) is induced in Plasmacytoma cells whereas it is repressed about 3 fold by hnRNP F overexpression according to microarray results (Table 10). Purified ELL2 protein was shown to enhance the elongation rate of RNAP II (Shilatifard et al. 1997). An RT-PCR was performed to verify the array data corresponding to ELL2: This gene is highly induced in AxJ and repressed slightly in hnRNP F overexpressing AxJs (Figure 20). Ribosomal protein S16 was used the control for the assay. Not only does expression of S16 not change in any of the samples used in this study but it is also expressed at very high levels.

**Table 7** Transcripts that are induced in AxJ Plasmacytoma cells.

<b>Probe Set ID</b>	<b>Gene Title</b>	<b>PC vs B</b>
102764_at	tumor rejection antigen P1A	400.011
93573_at	metallothionein 1	309.973
93728_at	transforming growth factor beta 1 induced transcript 4	287.109
96597_at	plasmacytoma expressed transcript 2	231.070
99420_at	immunoglobulin heavy chain 4 (serum IgG1)	215.381
101883_s_at	X-linked lymphocyte-regulated 3b	185.845
93833_s_at	histone 1, H2bc	162.241
100583_at	immunoglobulin heavy chain, (J558 family)	143.662
99436_at	sialophorin	78.357
99065_at	casein kappa	60.573
101561_at	metallothionein 2	55.896
93287_at	Bcl2-interacting killer-like	54.912
94288_at	histone 1, H1c	52.532
104063_at	Src activating and signaling molecule	47.622
96146_at	B-cell translocation gene 3	47.372
101082_at	malic enzyme, supernatant	47.018
99667_at	cytochrome c oxidase, subunit VI a, polypeptide 2	40.164
102313_at	GTP cyclohydrolase 1	38.509
160623_at	cyclin-dependent kinase-like 2 (CDC2-related kinase)	37.984
92904_at	PR domain containing 1, with ZNF domain	33.809
93039_at	plasma glutamate carboxypeptidase	32.185
98435_at	adenylosuccinate synthetase 1, muscle	29.504
101056_at	radixin	28.329
93638_s_at	immunoglobulin lambda chain, variable 1	26.281
96831_at	protein disulfide isomerase-related	23.752
104299_at	NEW1 domain containing protein	23.087
96134_at	deleted in polyposis 1-like 1	21.248
97819_at	glutathione S-transferase omega 1	19.109
99935_at	tight junction protein 1	17.659
103029_at	programmed cell death 4	13.772
92871_at	Sel1 (suppressor of lin-12) 1 homolog (C. elegans)	13.258
98007_at	ribosomal protein S6 kinase, 90kD, polypeptide 2	12.843
102060_at	golgi autoantigen, golgin subfamily a, 4	12.640
95419_at	H1 histone family, member 0	12.576
95673_s_at	brain abundant, membrane attached signal protein 1	12.376
104618_at	retinoblastoma binding protein 9	12.232
100286_at	macrophage stimulating 1 (hepatocyte growth factor-like)	12.079
101359_at	laminin, beta 2	11.896
102221_at	synaptogyrin 1	11.835
94247_at	E26 avian leukemia oncogene 2, 3' domain	11.253

161666_f_at	growth arrest and DNA-damage-inducible 45 beta	10.888
104627_at	CDP-diacylglycerol synthase 2	10.717
102237_at	CD28 antigen	10.422
104256_at	pleckstrin homology, Sec7 and coiled/coil domains, binding protein	10.404
93593_f_at	epithelial membrane protein 3	9.740
101040_at	calpain 2	9.482
161486_f_at	immunoglobulin heavy chain, (J558 family)	9.131
161754_f_at	galactosidase, beta 1	8.994
101900_at	cyclin-dependent kinase inhibitor 2B (p15, inhibits CDK4)	8.869
95675_at	germinal center kinase-like kinase-like	8.631
160499_at	tumor rejection antigen gp96	8.106
92870_at	Sel1 (suppressor of lin-12) 1 homolog (C. elegans)	7.536
104257_g_at	pleckstrin homology, Sec7 and coiled/coil domains, binding protein	7.325
160520_at	yes-associated protein	7.315
104712_at	myelocytomatosis oncogene	7.235
93320_at	carnitine palmitoyltransferase 1, liver	6.992
92891_f_at	homeo box C9	6.738
101123_at	integral membrane protein 2B	6.678
99582_at	tumor-associated calcium signal transducer 1	6.671
104476_at	retinoblastoma-like 1 (p107)	6.648
103723_at	interleukin 13 receptor, alpha 1	6.406
102955_at	nuclear factor, interleukin 3, regulated	5.718
160119_at	matrix gamma-carboxyglutamate (gla) protein	5.696
103895_at	expressed sequence AW549877	5.570
97876_at	vacuolar protein sorting 29 (S. pombe)	5.360
92611_at	GPI-anchored membrane protein 1	5.288
103892_r_at	ELL-related RNA polymerase II, elongation factor	5.246
99532_at	transducer of ErbB-2.1	5.236
100606_at	prion protein	5.047
96650_at	AU RNA binding protein/enoyl-coenzyme A hydratase	4.993
95423_at	calcium binding protein, intestinal	4.992
93083_at	annexin A5	4.869
99475_at	suppressor of cytokine signaling 2	4.798
103647_at	galactosidase, beta 1	4.748
93495_at	peroxiredoxin 4	4.630
100136_at	lysosomal membrane glycoprotein 2	4.527
102655_at	integrin alpha 4	4.427
102316_at	calpain 5	4.373
104538_at	prostaglandin I2 (prostacyclin) synthase	4.353
100622_at	anti-oxidant protein 2	4.341
96633_s_at	MORF-related gene X	4.225
93870_at	Braf transforming gene	4.210

104006_at	epidermal growth factor receptor pathway substrate 15	4.165
92737_at	interferon regulatory factor 4	4.091
96752_at	intercellular adhesion molecule	3.566
104065_at	ER degradation enhancing alpha mannosidase-like	3.537
98059_s_at	lamin A	3.353
100572_at	transmembrane 9 superfamily member 2	3.329

**Table 8** Transcripts that are repressed in AxJ Plasmacytoma cells.

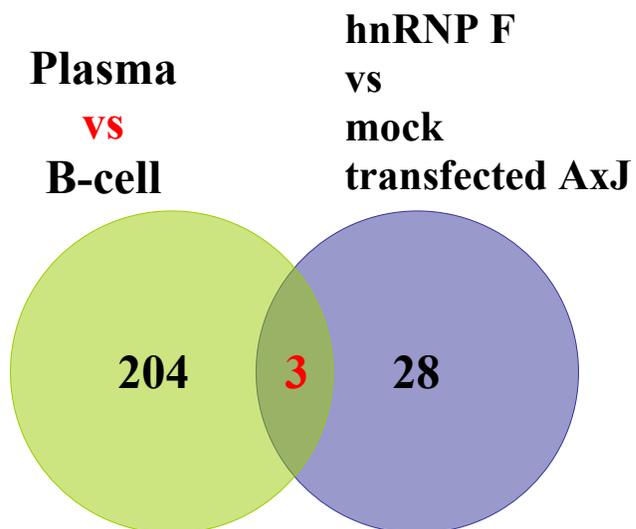
<b>Probe Set ID</b>	<b>Gene Title</b>	<b>PC vs B cells</b>
92564_at	leucine rich repeat (in FLII) interacting protein 1	0.349
93797_g_at	ATPase, Na <sup>+</sup> /K <sup>+</sup> transporting, alpha 1 polypeptide	0.242
101869_s_at	hemoglobin, beta adult major chain	0.237
94970_at	hypothetical protein LOC232314	0.232
96753_at	B-cell CLL/lymphoma 7C	0.224
102337_s_at	Fc receptor, IgG, low affinity IIb	0.224
160611_at	expressed sequence AW111961	0.214
97335_at	T-cell immunoglobulin and mucin domain containing 2	0.213
99461_at	hematopoietic cell specific Lyn substrate 1	0.213
99478_at	protein tyrosine phosphatase, receptor type, C polypeptide-associated protein	0.210
92678_at	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 25	0.205
102796_at	nucleoplasmin 3	0.195
100564_at	D-dopachrome tautomerase	0.193
93626_at	ATP-binding cassette, sub-family G (WHITE), member 2	0.193
94161_at	protein kinase C, epsilon	0.184
99945_at	CD19 antigen	0.181
100065_r_at	gap junction membrane channel protein alpha 1	0.169
104603_at	glutathione S-transferase, theta 2	0.165
102208_at	sialyltransferase 10 (alpha-2,3-sialyltransferase VI)	0.165
101147_at	germinal center expressed transcript	0.164
160545_at	cyclin D3	0.159
101845_s_at	component of Sp100-rs	0.152
93633_at	guanine nucleotide exchange factor (RCC1 related)	0.150
92644_s_at	myeloblastosis oncogene	0.149
99071_at	macrophage expressed gene 1	0.147
98962_at	lipase, hepatic	0.143
103959_at	hypothetical protein LOC230936	0.142
103462_at	dedicator of cyto-kinesis 2	0.140
160246_at	expressed sequence AA987150	0.140
160064_at	syntaxin 7	0.133

93104_at	B-cell translocation gene 1, anti-proliferative	0.132
92895_at	single-minded 1	0.127
94332_at	hypothetical protein LOC270138	0.127
100467_at	lymphoblastomic leukemia	0.127
102644_at	kidney cell line derived transcript 1	0.126
101019_at	cathepsin C	0.119
92359_at	B lymphoid kinase	0.118
93285_at	dual specificity phosphatase 6	0.116
95129_at	nuclear receptor co-repressor 2	0.115
102037_at	microtubule-associated protein, RP/EB family, member 2	0.114
98002_at	interferon concensus sequence binding protein	0.114
99486_at	centromere autoantigen B	0.112
98992_at	endothelial precursor protein B9	0.105
94285_at	histocompatibility 2, class II antigen E beta	0.104
102778_at	CD79A antigen (immunoglobulin-associated alpha)	0.103
160101_at	heme oxygenase (decycling) 1	0.103
103571_at	leukocyte specific transcript 1	0.099
101834_at	mitogen activated protein kinase 3	0.098
101918_at	transforming growth factor, beta 1	0.098
102292_at	growth arrest and DNA-damage-inducible 45 alpha	0.096
98588_at	fumarylacetoacetate hydrolase	0.093
161899_f_at	Williams-Beuren syndrome chromosome region 5 homolog (human)	0.091
100998_at	histocompatibility 2, class II antigen A, beta 1	0.086
95511_at	integrin alpha 6	0.083
103795_f_at	T-cell immunoglobulin and mucin domain containing 2	0.083
96648_at	coronin, actin binding protein 1A	0.081
103288_at	nuclear receptor interacting protein 1	0.080
93372_at	acidic (leucine-rich) nuclear phosphoprotein 32 family, member A	0.074
98398_s_at	apolipoprotein B editing complex 1	0.073
104652_at	potassium channel, subfamily K, member 2	0.068
101851_at	antigen identified by monoclonal antibody MRC OX-2	0.067
93567_at	profilin 2	0.067
101426_at	ceramide kinase	0.065
92962_at	tumor necrosis factor receptor superfamily, member 5	0.064
95893_at	B lymphoid kinase	0.060
98508_s_at	phosphatidic acid phosphatase 2a	0.060
102851_s_at	hemopoietic cell phosphatase	0.056
97963_at	signal-induced proliferation associated gene 1	0.055
92407_at	myomesin 1	0.053
92535_at	early B-cell factor 1	0.053
104173_at	membrane-spanning 4-domains, subfamily A, member 1	0.052
104174_at	ectonucleotide pyrophosphatase/phosphodiesterase 1	0.049
98790_s_at	myeloid ecotropic viral integration site 1	0.048
103690_at	Williams-Beuren syndrome chromosome region 5	0.045

	homolog (human)	
99841_at	Burkitt lymphoma receptor 1	0.044
94834_at	cathepsin H	0.038
94056_at	stearoyl-Coenzyme A desaturase 1	0.038
92867_at	polyhomeotic-like 2 (Drosophila)	0.036
102798_at	adrenomedullin	0.035
101441_i_at	inositol 1,4,5-triphosphate receptor 5	0.034
94432_at	sialyltransferase 1 (beta-galactoside alpha-2,6-sialyltransferase)	0.033
101972_at	kidney-derived aspartic protease-like protein	0.033
92415_at	tumor necrosis factor (ligand) superfamily, member 9	0.031
102904_at	histocompatibility 2, class II antigen E alpha	0.031
92496_at	vesicle-associated membrane protein 5	0.029
92760_at	Wiskott-Aldrich syndrome homolog (human)	0.029
102629_at	tumor necrosis factor	0.028
99511_at	protein kinase C, beta	0.028
94354_at	ATP-binding cassette, sub-family A (ABC1), member 1	0.027
102209_at	nuclear factor of activated T-cells, cytoplasmic 1	0.026
96591_at	reelin	0.024
102939_s_at	CD22 antigen	0.023
98129_at	thymosin, beta 10	0.022
99446_at	membrane-spanning 4-domains, subfamily A, member 1	0.021
162287_r_at	chloride channel calcium activated 3	0.019
102940_at	lymphotoxin B	0.018
96736_at	expressed sequence AA959454	0.018
92494_at	annexin A10	0.017
102630_s_at	lymphotoxin A	0.017
99510_at	protein kinase C, beta	0.017
100064_f_at	gap junction membrane channel protein alpha 1	0.016
101878_at	CD72 antigen	0.016
98525_f_at	erythroid differentiation regulator	0.016
93657_at	Mus musculus Ets transcription factor Spi-B, partial cds.	0.015
93013_at	inhibitor of DNA binding 2	0.012
92642_at	carbonic anhydrase 2	0.012
94057_g_at	stearoyl-Coenzyme A desaturase 1	0.010
92866_at	histocompatibility 2, class II antigen A, alpha	0.009
93842_at	hypothetical protein MGC6998	0.008
93086_at	immunoglobulin kappa chain variable 8 (V8)	0.008
103015_at	B-cell leukemia/lymphoma 6	0.004
101368_at	placentae and embryos oncofetal gene	0.002
96426_at	thymosin, beta 4, X chromosome	0.002
103794_i_at	T-cell immunoglobulin and mucin domain containing 2	0.001

**Table 9 Differentially expressed transcripts upon forced expression of hnRNP F in AxJ Plasmocytoma cells.**

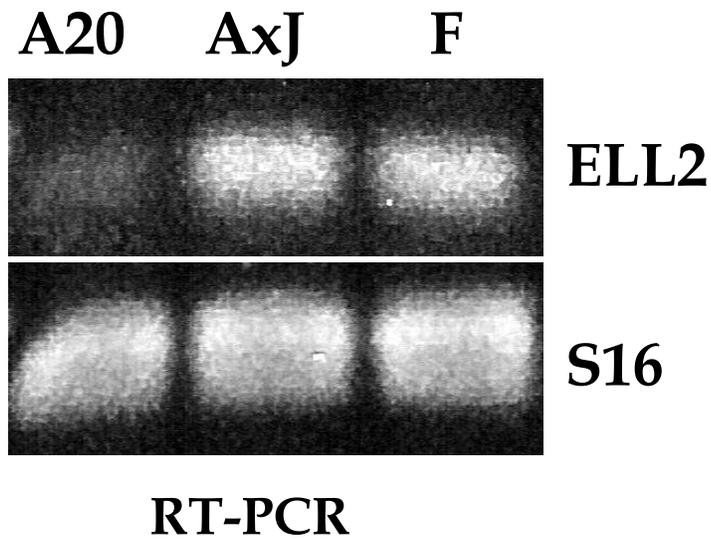
Probe set ID	Gene Title	F vs AxJ
96651_at	SWI/SNF related , actin dependent regulator of chromatin, subfamily e, member 1	0.271
102655_at	integrin alpha 4	0.288
103892_r_at	ELL-related RNA polymerase II, elongation factor	0.343
92579_at	Sjogren syndrome antigen B	0.364
102981_at	GA repeat binding protein, alpha	0.370
161076_at	hypothetical protein MGC36997	0.389
103654_at	nucleosome binding protein 1	0.398
160626_at	myelin basic protein expression factor 2, repressor	0.402
93646_at	PTK9 protein tyrosine kinase 9	0.412
98048_at	neural-salient serine/arginine-rich	0.413
93228_at	helicase, lymphoid specific	0.416
93878_at	myeloid/lymphoid or mixed lineage-leukemia translocation to 10 homolog (Drosophila)	0.423
103045_at	mature T-cell proliferation 1	0.425
100903_at	ADP-ribosyltransferase (NAD <sup>+</sup> ; poly(ADP-ribose) polymerase)-like 2	0.431
93806_at	SH3-binding domain glutamic acid-rich protein like	0.446
99343_at	retinoblastoma binding protein 7	0.446
160539_at	splicing factor, arginine/serine-rich 1 (ASF/SF2)	0.446
104476_at	retinoblastoma-like 1 (p107)	0.448
95606_at	NS1-associated protein 1-like	0.448
104070_at	p300/CBP-associated factor	0.449
93309_at	fibroblast growth factor inducible 14	0.450
161980_f_at	Bcl2-associated athanogene 3	0.452
161648_at	complement receptor related protein	0.452
97182_at	cyclin E2	0.472
161126_at	ESTs, Moderately similar to enhancer of invasion 10 [Homo sapiens] [H.sapiens]	0.473
93773_f_at	zinc finger protein 265	0.479
97102_at	YME1-like 1 (S. cerevisiae)	0.479
97541_f_at	histocompatibility 2, D region locus 1	0.480
101370_at	karyopherin (importin) alpha 1	0.483
100560_at	platelet-activating factor acetylhydrolase, isoform 1b, beta1 subunit	0.483



**Figure 19** Venn diagram showing that three transcripts are differentially expressed both in Plasma cells and hnRNP F overexpressing AxJ cells.

**Table 10** B-cell specific genes that are regulated by forced hnRNP F expression.

<b>Probe Set ID</b>	<b>Plasma-Bcell ratio</b>	<b>F-AxJ ratio</b>	<b>Gene Title</b>
102655_at	4.43	0.29	integrin alpha 4
103892_r_at	5.25	0.33	elongation factor RNA polymerase II 2
104476_at	6.65	0.45	retinoblastoma-like 1 (p107)



**Figure 20** ELL2 is induced in AxJ cells compared to A20 cells as assayed by RT-PCR

#### 5.4. CONCLUSIONS

Development from B-cell to a plasma cell includes radical morphological and functional changes. Therefore, it is not surprising that the expression of more than 400 genes were modified in AxJ Plasmacytoma cells versus its lymphoma predecessor. The candidate genes that have a role in membrane to secretory Immunoglobulin switch are broadly classified into an RNA metabolism class which include splicing and polyadenylation factors. Among the transcripts repressed in AxJ cells is the Cleavage and Polyadenylation Specific Factor 2 (CPSF-2 aka CPSF-100). However, the CPSF2 protein levels do not seem to change between different B-cell stages (Edwards-Gilbert and Milcarek 1995) a fact that might indicate a negative posttranscriptional control for this transcript.

The ELL protein family contains ELL, ELL2 and ELL3. They all the enhance elongation rate of RNAP II (Shilatifard 1998). With promoter that varied in their RNAP II elongation rates and processivity properties weak alternative splice sites are more often skipped and the reverse statement is also true: the slower and less processive the RNAP II, the higher the chance that it will recognize suboptimal splice sites (Nogues et al. 2003) (Nogues et al. 2003). However, recognition of the optimal splice sites, i.e constitutive splice sites, is not effected by the RNAP II elongation rate (Nogues et al. 2003). A model for membrane to secretory Immunoglobulin switch that includes ELL2 is proposed (Figure 21). We propose that when ELL2 is induced in Ig-secreting cells, RNAP II becomes more processive. This will result in the skipping of a weak membrane specific 5' splice site. In early B-cells, the level of ELL2 is low and RNAP II would be expected to be less processive. In this way, the splicing machinery will have more time to recognize the suboptimal membrane specific splice site imbedded in CH3 and use it more frequently.

Interestingly, the ELL2 primary transcript has two polyadenylation signals that are 2 kb apart; it is possible that the mechanism of repression of ELL2 by hnRNP F includes alternative polyadenylation. Splicing factor SF2/ASF was also repressed by hnRNP F in the array data. Preliminary RT-PCR results showed similar levels of SF2/ASF in both hnRNP F and empty-vector transfected AxJs. However, SF2/ASF has two functional polyA signals, too, and the regulation of SF2/ASF expression by hnRNP F may also include alternative polyadenylation which does not change the overall levels of the transcript. These two sites are about 100 nt apart. Both ELL2 and SF2/ASF full-length transcripts have consecutive Guanines in their 3' UTR. Preliminary binding studies with recombinant hnRNP F and portions of 3' UTR of ELL2 and SF2/ASF confirmed that hnRNP F indeed binds to their 3' UTR (data not shown)

Evaluation of array data should be interpreted with caution for several reasons. First, there are many transcripts that have multiple representative probes on the chip. The expression values do not always correlate perfectly. Although this may indicate alternative processing of the messages, it may also result from imperfect design and/or hybridization of the probes. In addition, there are many probes which partially represent the corresponding transcript. For example, if the total length of the probe (oligonucleotides) is 250 nucleotides, not all of these sequence are specific to its designed transcript. However the expression data is obtained from the full probe.

The probability of getting three transcripts that are differentially expressed both in Plasma cells and hnRNP F overexpressing plasma cells is about  $1.52 \times 10^{-5}$ . Assuming that the selection of 207 transcripts in Plasma vs B-cell set and 31 genes in hnRNP F vs Empty vector set are random, the probability of getting three common genes is about  $1.52 \times 10^{-5}$ ; confirming that getting three transcripts is not an artificial result.

All the unidentified cDNAs, EST probes for DNA segments were removed for our analysis. First, there were already a very high number of transcripts to deal with. Second, most of these cDNAs do not match to any protein, motif or domain already known to have a role in RNA metabolism.

Extensive analysis of the array experiments performed in this study resulted in a novel gene (ELL2) and a unique mechanism for the membrane to secretory Ig expression.

membrane-specific  
(mature and memory B-cells)

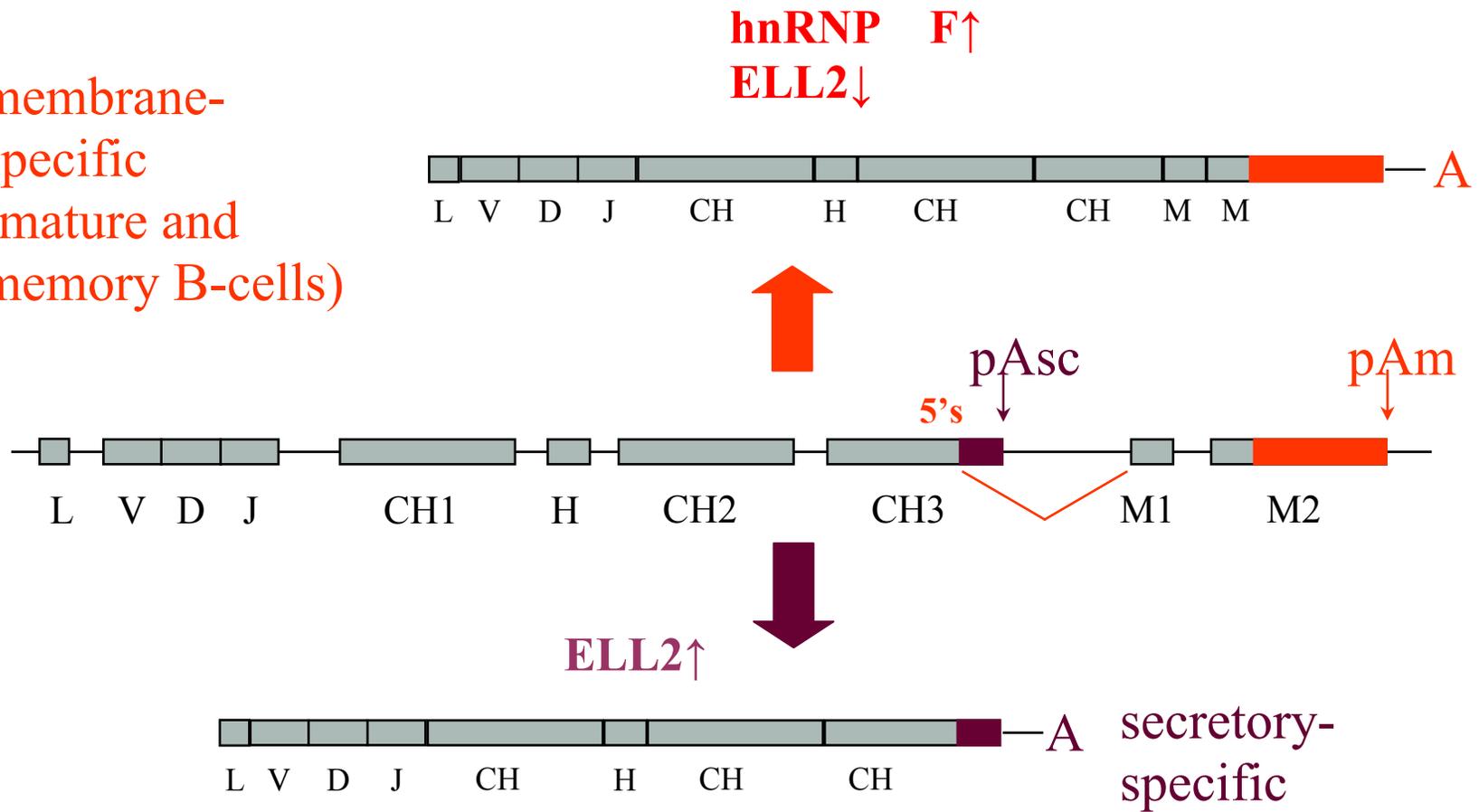


Figure 21  
5' Splice Site

Model: ELL2 Expression Results In More Processive RNA Polymerase II In Ig Secreting Cells So That It Skips the Membrane Specific

## 5.5. METHODS

Two sets of array experiments were performed in this study. For the first part, biotin labeled cRNAs were synthesized from polyA plus RNA samples in our lab according to Affymetrix instruction. These samples were sent to Center for Applied Genomics at University of Medicine and Dentistry of New Jersey to be hybridized with Affymetrix U74Av2 murine oligonucleotide chips. Each cRNA sample was hybridized to chips separately and each data comparison is a result of two separate hybridization. After normalization of signals among different chips, each probe is scored either “Present” or “Absent” by the Affymetrix software depending on the overall signal strength. There are three different analysis steps that were taken: First we performed Significance Analysis of Microarray (SAM) which is basically a unpaired t-test on a large scale (Tusher et al. 2001). Second, the induction and repression cutoff point was taken as two. Last but not least, we exclude the transcripts that are induced in one sample and absent at the same time.

Total RNA were isolated from cells corresponding to second set and submitted for cRNA synthesis and hybridization to the University of Pittsburgh Genomics and Proteomics Laboratory Array facility. The same U74Av2 chips were used to ensure one to one comparison with the first data set. The same analysis explained above were done also with this set. All the raw data analysis was performed with Microsoft Excel software. Gene ontology analysis was performed at [www.affymetrix.com](http://www.affymetrix.com).

### RT-PCR

Reverse transcription was performed according to manufacturer’s (InvitroGen) protocol. PCR program used for each amplification was: 30 cycles of 95<sup>0</sup> for 30 sec, 58<sup>0</sup> for 30 sec, 72<sup>0</sup> for 45

sec . PCR products were fractionated on a 1% Agarose with a 75 min run at 100V in 0.5x TBE Buffer.

## 6. DISCUSSION

Gene expression control could be achieved at multiple different stages. While most of the control is exerted at the level of transcription, splicing, polyadenylation, mRNA stability and RNA editing offers the cell various gene control mechanisms. Most hnRNP proteins associate with RNA polIII transcripts rapidly and increase the overall message stability in general. The hnRNP proteins bind to RNA via RRM or RGG domains. In most of the cases, RRM domains have been shown to have sequence specificity. The hnRNP H protein family displays a preference for Guanine rich sequences. The *in vitro* SELEX experiments with hnRNP H family have not been performed. In order to get the consensus hnRNP H binding sequence, a multiple alignment was performed with all the experimentally proven hnRNP H binding sequences. The minimal binding sequence element was used in each case. While several Guanines are present in every sequence, multiple alignment results showed a conserved Adenine (Figure 2) that is about 10 nt downstream of consecutive Guanines. The presence of this Adenine is not necessary since there are sequences that did not conserve this Adenine and still bind to hnRNP H (Figure 8). However, it may be required for optimal or high affinity binding. Interestingly, hnRNP F as well as hnRNP H2 bound the G-A mutants SAA10 and SAA11 with a higher affinity compared to wt GRS (Figure 9). Not only is the location of these mutations similar to consensus Adenine but also the full length SVL probe (Table 2) has this conserved Adenine residue. Other authors (Caputi and Zahler 2001) have implicated GGGA as the binding site for the hnRNP F/H? protein family. Although our data points out the importance of an Adenine residue, it is about 10 nt downstream of Guanine stretch, not immediately downstream. Furthermore, five, not three consecutive Guanines were required to reach the optimal affinity in our study.

Regardless of binding sequence, hnRNP H binds to RNA with a higher affinity compared to hnRNP F. Since this phenomenon was observed with the E. coli produced recombinant proteins, it may not result from posttranslational modifications. Because individual RRM between hnRNP H and F are at least 90% identical, this difference also may not be due to primary RRM sequence. It is possible that the very C-terminal part of hnRNP H that is not present in hnRNP F is responsible for this observation. It will be useful to design chimera protein having N-terminal part of hnRNP F and C-terminal part of hnRNP H and assay sequence specific binding. Another useful construct maybe the one expressing a C-terminally truncated hnRNP H.

Blocking the hnRNP H/F binding site on SVL polyA region has a negative effect on gene expression as assayed by Gfp-based reporter assays. This effect was verified by western blot, flow cytometry and fluorescent microscopy. The fact that it was observed with single polyA constructs implies that the relative effect of hnRNP F/H binding for alternative polyA site choice would be much greater.

Development from B-cell to a plasma cell involves major changes of the levels of transcription factors including but not limited to Blimp-1, c-myc, Spi-B (Shapiro-Shelef and Calame 2004). This change brings major morphological and functional alterations. Whereas the memory B-cells have a 1:1 sec:mb Ig mRNA ratio, typical plasma cells have an elevated sec:mb ratio. The choice between sec and mb specific message involves alternative splicing and alternative polyadenylation. In order to understand this B-cell stage specific switch, we compared the gene expression profile of A20 lymphoma with AxJ plasmacytoma cells. More than 400 transcripts are differentially expressed in AxJ cells. There are many genes related to DNA and RNA metabolism. hnRNP A1 and hnRNP C proteins are induced more than two fold in AxJ cells

(Table 7). It would be worthwhile to try to show an interaction between these two proteins and Ig polyA region.

There are no induced splicing factors in AxJ cells; Splicing factor 3a, SRp20, splicing factor proline/glutamine rich, and CUG triplet repeat RNA binding protein 2 are all repressed (Table 8). Whether the repression of these four splicing factor is decisive for mb:sec switch needs to be experimentally shown.

The level of the CPSF-2 transcript is repressed in AxJ cells (Table 8) however its protein level has been shown not to change between A20 and AxJ cells (Edwalds-Gilbert and Milcarek 1995). The level of another polyA factor, CstF-64, does not change cite papers?(data not shown).

There are many transcription factors repressed or induced in AxJ cells reflecting the major differences between these two cell lines. Interestingly, ELL2 mRNA is induced two-three fold in plasmacytoma cells (Table 7). There are two different probes that represent ELL2 in Table 7 and they are both induced. Purified ELL2 protein was shown to enhance the elongation rate of RNAP II. Recent studies implicate a relationship between the processivity of RNAP II and the usage of suboptimal splice sites. Hence the change in ELL2 levels may be very significant for Ig expression.

A complex of HnRNP F and H was shown to be formed on Ig polyA regions by early B-cell extracts but not plasma cell extracts. Overexpression of hnRNP F in plasmacytoma cells resulted in decreased usage of the secretory pA site compared to mock transfected cells. In order to understand the underlying mechanism of this effect, we performed microarray analysis of hnRNP F transfected plasmacytoma cells. There are about 30 differentially expressed transcripts; most of them are repressed genes. This is parallel to previously shown negative effect of hnRNP F on polyadenylation (Veraldi et al. 2001). SF2/ASF is one of the repressed transcripts upon

hnRNP F overexpression. However preliminary RT-PCR results showed no significant changes between hnRNP F and mock transfected cells. Since hnRNP F was shown to revert the B-cell clock to some extent, we are particularly interested in genes that have opposite expression patterns between plasmacytoma vs lymphoma and hnRNP F overexpressing AxJ vs AxJ sets. There are three transcripts that fit into this definition (Table 10) you only show 3. messenger RNA for ELL2 is induced in Plasma cells whereas it is repressed in hnRNP F overexpressing AxJ cells. The RT-PCR results verified the induction of ELL2 RNA in AxJ cells compared to A20 cells (Figure 20), however the repression is marginal in hnRNP F transfected cells. A novel model for mb to sec Ig switch has been proposed (Figure 21). If ELL2 induction in plasma cells results in a more processive RNAP II, the splicing machinery can skip the weak mb specific splice site. In memory and mature B-cells, ELL2 levels decrease and RNAP II becomes less processive and the mb specific splice site may be used more frequently. The immediate obvious experiment can be the overexpression of ELL2 in early B-cells or downregulation of ELL2 in Plasma cells and then assaying the sec:mb transcript ratio.

## 7. GENERAL DISCUSSION AND CONCLUSION

This study identified the consensus binding sequence of hnRNP F and H proteins. First, an extensive mutational analysis has been performed with the 14 nt GRS region on SVL RNA. The *in vitro* binding data channeled us to perform a multiple sequence alignment with all the experimental H/F binding sites. The consensus sequence includes a Guanine stretch, an Adenine that is 8-9 nucleotide downstream and an upstream Guanine.

In order to show that hnRNP F and H binding to the 3'UTR has an effect on polyadenylation, polyadenylation assays were performed with the binding mutants. There is a subtle but reproducible decrease in polyadenylation efficiency. If the same mutation (SAA20) is cloned into the 3'UTR of *gfp* reporter gene on a plasmid, the GFP protein levels were diminished in parallel to the *in vitro* results. In addition, the SAA10 mutant SVL oligo bound to both hnRNP F and H better than the wt sequence. The similar cloning and transfection experiments yielded the expected increase in GFP levels.

The hnRNP F and H levels were modulated by plasmid mediated overexpression and siRNA mediated RNAi and GFP levels were up and down, respectively. In summary, we have shown that hnRNP F and H has positive effect on gene expression with the SVL 3' UTR.

We performed two sets of microarray experiments; one set was designed to get the gene expression profile of A20 B-lymphoma and AxJ Plasmacytoma cell lines. In the other, we stably overexpressed hnRNP F in AxJ cell line and compared the gene expression profile with the empty vector transfected cells. There are over 200 differentially expressed transcripts in Plasma cells. The hnRNP F effected the expression of 31 transcripts and there are only three differentially expressed genes in both sets. Interestingly, all the 31 hnRNP F target genes are

down regulated. Overexpression of hnRNP F in 293T cells had a positive effect when GFP expression plasmid has SVL sequences in its 3'UTR. This can be explained in multiple ways: First, the stable expression of hnRNP F resulted in the activation of feedback loop that overcompensated for the the effect of hnRNP F. explain more Second, stable transfection does not always result in direct effects i.e. the modulation of a protein levels may change the expression of another gene that in turn may be responsible for the observed phenotype. For example downregulation of ELL2 or ASF/SF2 may be responsible for the 31 downregulated ?? genes. Third, since both hnRNP F and H can bind to single stranded G-rich DNA, they are capable of binding also G-rich sequences in the promoter regions. Fourth, the hnRNP H and F are usually associated with splicing silencers so that the observed negative effect in the microarray study may be mediated by splicing.

The primarily downregulation of genes of the hnRNP F-array experiments might be explained in several ways. While the hnRNP F array data remains to be studied further, the unquestionable result from the first part of this study is that with the SVL constructs hnRNP F and H had a positive effect on protein expression through binding to the 3' UTR.

This study showed that polyadenylation is an important step for gene expression and hnRNP F and H modulates polyadenylation and hence can exert major influences on gene expression.

## 8. FUTURE DIRECTIONS

When (Veraldi et al. 2001) isolated polyadenylation complexes from nuclear extracts formed on polyA site containing RNA, a larger complex containing hnRNP H/F proteins was observed. However this “complex” was eluted in the void volume of the column. The void volume is the elution volume of molecules which are too large to enter even the largest pores of the filtration gel. This indicates that there are many unspecific proteins in this fraction and questions the nature and specificity of the H/F complexes. In order to improve the purification, size-exclusion chromatography should be repeated with a resin of higher resolution capacity such as S-1000 Sephacryl (Amersham-Pharmacia) media with 400 nm spherical particles instead of S-500. The S-500 media has a fractionation range (Mr) of only  $4 \times 10^4 - 2 \times 10^7$  while the S-1000 media has a much larger (Mr of  $5 \times 10^5 - > 10^8$ ) which would help to resolve the much larger protein complex that includes the hnRNP H/F. Apart from the resolution of the chromatography, the differential association hnRNP H/F on the polyA containing RNA should be verified with multiple different B-cell lines representing early or memory B-cell stage and Plasma cells. The confirmation of the results with sorted primary B-cell and Plasma extracts would complete the picture. The specificity of the H/F complexes would be ensured by mutating their binding regions on RNA and let the H/F complex be formed on the mutant RNA.

Once the identity and nature of the H/F and polyadenylation complexes are verified in multiple cell lines, then the analytical analysis of the proteins by mass spectrometry present in these complexes would be helpful to try to isolate other auxiliary modifiers of polyadenylation. The functionally more important proteins that have the potential to modify polyadenylation should be in the polyadenylation complex not in the H/F complex. However given the experimental ground in this study showing the effect of hnRNP F and H on polyadenylation, it is worthwhile to verify the specific nature of larger complex having hnRNP F/H.

Since this study successfully applied siRNA mediated RNAi for hnRNP F and H, one can downregulate these two proteins in A20 B-lymphoma cell line and assay Ig sec:mb message ratio. Furthermore, to complement the hnRNP F array data present in this study, one can downregulate hnRNP F in AxJ cells and assay the levels of 31 hnRNP F target genes by RT-PCR or Northern blot without the need of doing the array experiment. If the 31 genes are direct hnRNP F targets, the levels of their expression will go in the opposite direction by hnRNP F RNAi as compared to overexpression.

Having isolated the optimal hnRNP F and H binding site, one can search for these sequences in the 3'UTRs of mammalian transcripts on global scale. Before this study, practically this was not possible, since not only the consensus sequence was not defined but also the proposed H/F binding site (just a G-stretch) yielded many common matches. A profile of potential hnRNP F and H target genes would be formed after this computational analysis.

A model that helps to explain the sec:mb ratio in B and Plasma cells was suggested in this study (Figure 21). In this model, ELL2 increases the elongation rate of RNAP II and this in turn results in skipping of the membrane specific weak splice site in Plasma cells. The situation is completely opposite in early/memory B-cells. If the model is true, then modulation of the ELL2 levels would result in the expected switch in sec:mb ratio. One can perform RNAi for ELL2 in Plasma cells or overexpress it in B-cells and if the sec:mb phenotype is reversed, our model would gain some validity. Independent of ELL2, the elongation rate of RNAP II can be decreased by a drug called DRB. In addition, there is alpha-amanitin resistant "slow" RNAP II construct (de la Mata et al. 2003), one can make use of it to diminish the elongation rate if RNAP

II in Plasma cells. But before pursuing this path, convincing data from ELL2 protein expression and deletion experiments have to be obtained.

## BIBLIOGRAPHY

- Arhin, G. K., M. Boots, et al. (2002). "Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals." Nucleic Acids Res **30**(8): 1842-50.
- Bagga, P., G. Arhin, et al. (1998). "DSEF-1 is a member of the hnRNP H family of RNA-binding proteins and stimulates pre-mRNA cleavage and polyadenylation in vitro." Nucleic Acids Research **26**(23): 5343-50.
- Bagga, P. S., L. P. Ford, et al. (1995). "The G-rich auxiliary downstream element has distinct sequence and position requirements and mediates efficient 3' end pre-mRNA processing through a trans-acting factor." Nucleic Acids Res **23**(9): 1625-31.
- Baird, G. S., D. A. Zacharias, et al. (2000). "Biochemistry, mutagenesis, and oligomerization of DsRed, a red fluorescent protein from coral." Proc Natl Acad Sci U S A **97**(22): 11984-9.
- Biggin, M. D. and R. Tjian (2001). "Transcriptional regulation in Drosophila: the post-genome challenge." Funct Integr Genomics **1**(4): 223-34.
- Birney, E., S. Kumar, et al. (1993). "Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors." Nucleic Acids Res **21**(25): 5803-16.
- Buratti, E., M. Baralle, et al. (2004). "hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes." Nucleic Acids Res **32**(14): 4224-36.
- Caputi, M. and A. M. Zahler (2001). "Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family." J Biol Chem **276**(47): 43850-9.
- Caputi, M. and A. M. Zahler (2002). "SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D." Embo Journal **21**(4): 845-855.
- Chen, C. D., R. Kobayashi, et al. (1999). "Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing in the ratbeta-tropomyosin gene." Genes & Dev **13**: 593-606.
- Chou, M. Y., N. Rooke, et al. (1999). "hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells." Mol Cell Biol **19**(1): 69-77.
- Conley, E. C. and J. R. Saunders (1984). "Recombination-dependent recircularization of linearized pBR322 plasmid DNA following transformation of Escherichia coli." Mol Gen Genet **194**(1-2): 211-8.
- Cooke, C. and J. C. Alwine (2002). "Characterization of specific protein-RNA complexes associated with the coupling of polyadenylation and last-intron removal." Mol Cell Biol **22**(13): 4579-86.
- Cooke, C., H. Hans, et al. (1999). "Utilization of splicing elements and polyadenylation signal elements in the coupling of polyadenylation and last-intron removal." Mol Cell Biol **19**(7): 4971-9.
- de la Mata, M., C. R. Alonso, et al. (2003). "A slow RNA polymerase II affects alternative splicing in vivo." Mol Cell **12**(2): 525-32.
- Edwalds-Gilbert, G. and C. Milcarek (1995). "The binding of a subunit of the general polyadenylation factor cleavage polyadenylation specificity factor (CPSF) to polyadenylation sites changes during B cell development." Nucleic Acids Symposium Series **33**: 229-233.

- Edwalds-Gilbert, G. and C. Milcarek (1995). "Regulation of poly(A) site use during mouse B-cell development involves a change in the binding of a general polyadenylation factor in a B-cell stage-specific manner." Molecular and Cellular Biology **15**: 6420-6429.
- Fogel, B. L. and M. T. McNally (2000). "A cellular protein, hnRNP H, binds to the negative regulator of splicing element from Rous sarcoma virus." Journal of Biological Chemistry **275**(41): 32371-32378.
- Gorlach, M., M. Wittekind, et al. (1992). "Interaction of the RNA-binding domain of the hnRNP C proteins with RNA." Embo J **11**(9): 3289-95.
- Grabowski, P. J. (2004). "A molecular code for splicing silencing: configurations of guanosine-rich motifs." Biochem Soc Trans **32**(Pt 6): 924-7.
- Hastings, M. L., C. M. Wilson, et al. (2001). "A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA." Rna-a Publication of the Rna Society **7**(6): 859-874.
- Hirose, Y., R. Tacke, et al. (1999). "Phosphorylated RNA polymerase II stimulates pre-mRNA splicing." Genes Dev **13**: 1234-1239.
- Honore, B. (2000). "The hnRNP 2H9 gene, which is involved in the splicing reaction, is a multiply spliced gene." Biochim Biophys Acta **1492**(1): 108-19.
- Honore, B., U. Baandrup, et al. (2004). "Heterogeneous nuclear ribonucleoproteins F and H/H' show differential expression in normal and selected cancer tissues." Exp Cell Res **294**(1): 199-209.
- Honore, B., H. Rasmussen, et al. (1995). "Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes." Journal of Biological Chemistry **270**(48): 28780-28789.
- Hyun, Y. L., D. B. Lew, et al. (2000). "Enzymic methylation of arginyl residues in -gly-arg-gly-peptides." Biochem J **348 Pt 3**: 573-8.
- Jacquet, S., A. Mereau, et al. (2001). "A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H." Journal of Biological Chemistry **276**(44): 40464-40475.
- Johnson, E. S. (2004). "Protein modification by SUMO." Annu Rev Biochem **73**: 355-82.
- Kim, E., L. Du, et al. (1997). "Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA." J Cell Biol **136**: 19-28.
- Krecic, A. M. and M. S. Swanson (1999). "hnRNP complexes: composition, structure, and function." Curr Opin Cell Biol **11**(3): 363-71.
- Lacroix, L., H. Lienard, et al. (2000). "Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro." Nucleic Acids Res **28**(7): 1564-75.
- Liu, H., W. Zhang, et al. (2002). "Mutations in RRM4 uncouple the splicing repression and RNA-binding activities of polypyrimidine tract binding protein." Rna **8**(2): 137-49.
- Mann, K. P., E. A. Weiss, et al. (1993). "Alternative Poly(A) Site Utilization during Adenovirus Infection Coincides with a Decrease in the Activity of a Poly(A) Site Processing Factor." Molecular and Cellular Biology **13**(4): 2411-2419.
- Markovtsov, V., J. M. Nikolic, et al. (2000). "Cooperative assembly of an hnRNP complex induced by a tissue-specific homolog of polypyrimidine tract binding protein." Mol Cell Biol **20**: 7463-7479.

- Martincic, K., R. Campbell, et al. (1998). "Increase in the 64-kDa subunit of the polyadenylation/cleavage stimulatory factor during the G<sub>0</sub> to S phase transition." Proc Natl Acad Sci USA **95**(19): 11095-11100.
- Matis, S. A., K. Martincic, et al. (1996). "B-lineage regulated polyadenylation occurs on weak poly(A) sites regardless of sequence composition at the cleavage and downstream regions." Nucleic Acids Research **24**: 4684-4692.
- McBride, A. E. and P. A. Silver (2001). "State of the arg: protein methylation at arginine comes of age." Cell **106**(1): 5-8.
- McGregor, F., A. Phelan, et al. (1996). "Regulation of herpes simplex virus poly(A) site usage and the action of immediate-early protein IE63 in the early-late switch." Journal of Virology **70**: 1931-1940.
- Milligan JF, G. D., Witherell GW, Uhlenbeck OC (1987). "Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates." Nucleic Acids Research **15**(21): 8783-98.
- Mnaimneh, S., A. P. Davierwala, et al. (2004). "Exploration of essential gene functions via titratable promoter alleles." Cell **118**(1): 31-44.
- Natalizio, B. J., L. C. Muniz, et al. (2002). "Upstream elements present in the 3'-untranslated region of collagen genes influence the processing efficiency of overlapping polyadenylation signals." J Biol Chem **277**(45): 42733-40.
- Nogues, G., M. J. Munoz, et al. (2003). "Influence of polymerase II processivity on alternative splicing depends on splice site strength." J Biol Chem **278**(52): 52166-71.
- Pagani, F., E. Buratti, et al. (2003). "Missense, nonsense, and neutral mutations define juxtaposed regulatory elements of splicing in cystic fibrosis transmembrane regulator exon 9." J Biol Chem **278**(29): 26580-8.
- Peterson, M. L., E. R. Gimmi, et al. (1991). "The developmentally regulated shift from membrane to secreted mu mRNA production is accompanied by an increase in cleavage-polyadenylation efficiency but no measurable change in splicing efficiency." Molecular and Cellular Biology **11**: 2324-2327.
- Qian, Z. and J. Wilusz (1991). "An RNA binding protein specifically interacts with a functionally important domain of the downstream element of the SV40 late polyadenylation signal." Molecular and Cellular Biology **11**: 5312-5320.
- Rawal, N., R. Rajpurohit, et al. (1995). "Structural specificity of substrate for S-adenosylmethionine:protein arginine N-methyltransferases." Biochim Biophys Acta **1248**(1): 11-8.
- Reed, R. (1990). "Protein composition of mammalian spliceosomes assembled in vitro." Proceedings of the National Academy of the Sciences USA **87**: 8031-8035.
- Romano, M., R. Marcucci, et al. (2002). "Regulation of 3' splice site selection in the 844ins68 polymorphism of the cystathionine Beta -synthase gene." J Biol Chem **277**(46): 43821-9.
- Shapiro-Shelef, M. and K. Calame (2004). "Plasma cell differentiation and multiple myeloma." Curr Opin Immunol **16**(2): 226-34.
- Shilatifard, A. (1998). "Factors regulating the transcriptional elongation activity of RNA polymerase II." FASEB J, **12**: 1437-1446.
- Shilatifard, A., D. R. Duan, et al. (1997). "ELL2, a new member of an ELL family of RNA polymerase II elongation factors." Proc. Natl. Acad. Sci. USA **94**: 3639-3643.
- Sironi, M., G. Menozzi, et al. (2004). "Silencer elements as possible inhibitors of pseudoexon splicing." Nucleic Acids Res **32**(5): 1783-91.

- Smith, J. J., K. P. Rucknagel, et al. (1999). "Unusual sites of arginine methylation in Poly(A)-binding protein II and in vitro methylation by protein arginine methyltransferases PRMT1 and PRMT3." J Biol Chem **274**(19): 13229-34.
- Spear, M. A. (2000). "Efficient DNA subcloning through selective restriction endonuclease digestion." Biotechniques **28**(4): 660-+.
- Takagaki, Y. and J. Manley (1998). "Levels of polyadenylation factor CstF-64 control IgM heavy chain mRNA accumulation and other events associated with B cell differentiation." Molecular Cell **2**: 761-771.
- Tusher, V. G., R. Tibshirani, et al. (2001). "Significance analysis of microarrays applied to the ionizing radiation response." Proc Natl Acad Sci U S A **98**(9): 5116-21.
- Underhill, G. H., D. George, et al. (2003). "Gene expression profiling reveals a highly specialized genetic program of plasma cells." Blood.
- Vassileva, M. T. and M. J. Matunis (2004). "SUMO modification of heterogeneous nuclear ribonucleoproteins." Mol Cell Biol **24**(9): 3623-32.
- Veraldi, K. L., G. K. Arhin, et al. (2001). "hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells." Mol Cell Biol **21**(4): 1228-38.
- Wahle, E. (1995). "Poly(A) tail length control is caused by termination of processive synthesis." Journal of Biological Chemistry **270**(6): 2800-2808.
- Wilusz, J., D. I. Felig, et al. (1988). "The C proteins of heterogeneous nuclear ribonucleoprotein complexes interact with RNA sequences downstream of polyadenylation cleavage sites." Molecular and Cellular Biology **8**: 4477-4483.
- Yoshida, T., K. Kokura, et al. (1999). "Heterogeneous nuclear RNA-ribonucleoprotein F binds to DNA via an oligo(dG)-motif and is associated with RNA polymerase II." Genes Cells **4**: 707-719.
- Yoshida, T., Y. Makino, et al. (1999). "Association of the rat heterogeneous nuclear RNA-ribonucleoprotein F with TATA-binding protein." FEBS Letters **457**: 251-254.
- Zuker, M. (2003). "Mfold web server for nucleic acid folding and hybridization prediction." Nucleic Acids Res **31**(13): 3406-15.