AD_____

Award Number DAMD17-96-1-6172

TITLE: Mechanism of Splicing of Unusual Intron in Human
Proliferating Cell Nucleolor P120

PRINCIPAL INVESTIGATOR: Dr. Hongxiang Liu

CONTRACTING ORGANIZATION: Cold Spring Harbor Laboratory
                          Cold Spring Harbor, New York 11724

REPORT DATE: December 1998

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
               Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
                        Distribution Unlimited

The views, opinions and/or findings contained in this report are
those of the author(s) and should not be construed as an official
Department of the Army position, policy or decision unless so
designated by other documentation.

19990811 098

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE December 1998	3. REPORT TYPE AND DATES COVERED Annual Summary (1 Dec 97 - 30 Nov 98)

4. TITLE AND SUBTITLE
Mechanism of Splicing of Unusual Intron in Human Proliferating Cell Nucleolor P120

5. FUNDING NUMBERS
DAMD17-96-1-6172

6. AUTHOR(S)
Dr. Hongxiang Liu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Cold Spring Harbor Laboratory
Cold Spring Harbor, New York 11724

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Previously we identified three classes of exonic splicing enhancer (ESE) motifs recognized by the human SR proteins SF2/ASF, SRp40 and SRp55. This year I identified functional ESEs recognized by SC35 or SRp30c. The SC35 motifs, whose consensus is GGYYCSYR, are abundant in the IgM C4 exon, which follows an intron that splices in S100 extract complemented by SC35, but not by SF2/ASF. SF2/ASF motifs occur infrequently in this exon. In contrast, both SC35 and SF2/ASF motifs co-exist in the HIV Tat T3 exon, following an intron that splices in S100 extract plus SC35 or SF2/ASF. I investigated the relationship between SR protein consensus motifs and human mutations using computational and experimental approaches. A point mutation in exon 18 of BRCA1, which causes exon skipping and familial breast and ovarian cancer, disrupts an SF2/ASF motif. In vitro splicing showed that exon skipping results from disruption of the SF2/ASF-ESE interaction. Mutations in other human disease genes were also found to interrupt SR protein ESE motifs. Therefore, many human diseases caused by missense or nonsense exon mutations may result from splicing defects, which are often overlooked. Our approach provides a simple way to study the phenotypic basis of mutations in human disease genes.

14. SUBJECT TERMS Breast Cancer			15. NUMBER OF PAGES 10
			16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

_____ Where copyrighted material is quoted, permission has been obtained to use such material.

_____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

_____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

__X__ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

_____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

_____ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

_____ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

_____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

_____
PI - Signature   Date

Table of Contents

Cover

Report Documentation Page

Foreword

Table of Contents

Annual report for award number DAMD17-96-1-6172

The proposed goals of my research were to identify exonic splicing enhancer (ESE) motifs recognized by SC35 and SRp30c (aim 1), to study the distribution of SR-specific ESE motifs in natural genes and their implication to splicing and gene function (aim 2), to characterize a substrate-specific splicing factor (aim 3), and to identify and clone this splicing factor (aim 4). Aims 1, 2 and 3 have been completed with a slightly different emphasis. Preliminary experiments have been carried out towards aim 4.

Identification of ESE motifs recognized by SC35 and SRp30c under splicing conditions

A functional SELEX procedure similar to the one I reported earlier this year (Liu et al., 1998) was used to study the sequence specificity of the human SR proteins SC35 and SRp30c under splicing conditions. The winner sequences of SC35 have the consensus GGYYCSYR (Y represents pyrimidine; S represents G or C; R represents purine). The content of G, A, and T residues in the SC35 winner pool did not change significantly from the initial RNA pool. The C residue content increased from 19% to 23%. In agreement with our previous functional SELEX data, the SC35 consensus sequence is highly degenerate. A score matrix was generated according to the frequency of each nucleotide at each position of the consensus motif. These score matrices were used to search the high-score motifs in each winner sequence, including 18 nt of the 5' and 20 nt of the 3' flanking regions. Often more than one high score motif could be found in each winner sequence. The high scores of SC35 winner sequences ranged from 1.2 to 3.6, with a mean score of 2.6. The random pool gave a lower mean score of 1.8 when searched by the same score matrix. SC35 was previously shown to recognize two groups of sequences in SELEX procedures based on binding (Tacke and Manley, 1995). The reported sequences, AGSAGAGTA and GTTCGAGTA, are significantly different from, and simpler than, the consensus motif I found.

In the case of SRp30c, I carried out three rounds of functional SELEX. Thirty-six independent clones were sequenced. Many duplicates were found. One winner sequence, G1, was duplicated 17 times. The SRp30c-type ESEs have the consensus RRNCGUAACC (R represents purine; N represents any nucleotide).

1

The frequency of duplication did not correlate with the magnitude of the high score motif in each winner sequence. The mean score of the SRp30c winner pool was 2.62. When searched by the same score matrix, the initial RNA pool gave a mean score 0.4. The duplication of SRp30c winners indicates that the selection procedure was overperformed. A loss of information is therefore, expected for the ESE motifs selected by this protein.

The selected ESE motifs of SC35 and SRp30c are functional

I next tested whether the selected sequences could function as true splicing enhancers. A number of sequences with varied scores were chosen from the winner pool of SC35, and subjected to splicing in nuclear extract or in S100 extract complemented by SC35. Four out of five SC35 winner sequences activated IgM pre-mRNA splicing very efficiently in nuclear extract. They promoted IgM splicing in S100 extract complemented by SC35 but not in S100 extract alone. One winner sequence of the SC35 winner pool, D5, did not appear to be an efficient splicing enhancer. In general, the splicing efficiency correlated well with the highest score in the winner sequence. However, the correlation between splicing efficiency and scores is not linear. This could be explained by the fact that some winner sequences have more than one high-score motif. Several sequences from the random RNA pool were also examined for in vitro splicing. All of them spliced poorly in nuclear extract. In many cases the RNA precursors were degraded, suggesting that spliceosomal complexes did not form with these RNAs.

The enhancer activity of a few SRp30c winner sequences was also examined. These sequences also activated IgM pre-mRNA splicing in nuclear extract very efficiently. They also promoted IgM splicing in S100 extract supplement with recombinant SRp30c but not in S100 alone, although the efficiency was low. In S100 extract, the SRp30c winner sequences appeared to be weaker enhancers than the SC35 winner sequences. However, they were as potent as the SC35 winners when tested in HeLa nuclear extract.

The selected SC35 ESE motifs are specific and biologically significant

Next I studied the SR protein specificity of the SC35-selected ESEs. This experiment was carried out in S100 extract complemented by SC35, SF2/ASF, SRp40 or SRp55. All the tested SC35 winners could splice in S100 extract when

complemented by SC35, SRp40 or SRp55. When complemented by SF2/ASF, the splicing efficiency was lower.

To ask the question if the selected ESE motifs are relevant to splicing of natural pre-mRNA substrates, I conducted a search of SC35 high-score motifs in natural genes. Only the scores higher than the lowest score of the SC35 winner pool were considered. For comparison, I also searched the high-score motifs of SF2/ASF in these genes. The first gene I examined was the M2 exon of the IgM gene. The search result indicated that there are many SC35 ESE motifs in the characterized natural ESE. The distribution of SC35 high-score motifs is different from that of SF2/ASF. SF2/ASF-specific motifs exist at a much higher density in the natural ESE than the flanking region. In contrast, SC35 high-score motifs have a relatively even distribution across the M2 exon.

To address the issue of whether the identified ESE motifs are specific to SC35, I searched two additional pre-mRNA substrates that have different SR specificity, encoded by the IgM gene C4 exon and the Tat gene T3 exon. Splicing of transcripts from the IgM C3-C4 mini-gene (note the difference from the IgM M1-M2 mini-gene which I used for the SELEX study) is activated in S100 extract when complemented by SC35 but not by SF2/ASF, as shown in recent work from our lab (Mayeda *et al.*, in press). In contrast, the Tat T2-T3 mini-gene is only activated by SF2/ASF but not by SC35 in S100 extract (Chandler *et al.*, 1997); Mayeda *et al.*, in press). After the deletion of an SC35-specific silencer in the 3' region of the T3 exon, both SF2/ASF and SC35 can activate the T2-T3 splicing in S100 extract. Detailed analysis of the splicing of these two pre-mRNAs showed that the C4 exon and T3 exon determine the SR protein specificity (Mayeda *et al.*, in press). My motif search results match the experimental data. Many high-score motifs matching the consensus of SC35 are present in the C4 exon. Only two SF2/ASF motifs are present in this exon. High-score motifs for both SF2/ASF and SC35 are located in the T3 exon of the Tat gene.

Finally, I studied the distribution of high-score motifs of the SC35-type ESE in human exons versus introns. A total of 570 genes, representing 2626 exons (426 kb) and 2079 introns (1,295 kb), were extracted from the ALLSEQ database (Burset and Guigo, 1996) and analyzed. Scores equal to or higher than the mean score of the winner pool were taken into account. High-score motifs appeared more frequently in exons than in introns. An average of 9 high-score motifs were found per kilobase of exon but only 5.9 high-score motifs were found per

3

kilobase of intron. This comparison was statistically significant since a large database was used and the p-value is less than 10^{-10}.

Survey of natural ESEs in human disease genes

A total of 23 natural ESEs have been compiled and analyzed using the score matrices for SF2/ASF, SRp40, and SRp55. Our analysis was consistent with the experimental data of others and of our own lab. Many examples were described in last year's report and published (Liu *et al.*, 1998). Here I will focus on the survey results of a few human disease genes.

A nonsense mutation (G to T) in the breast cancer susceptibility gene, BRCA1, exon 18 causes skipping of the entire exon (Mazoyer et al., 1998). This mutation was found in a family with four cases of breast cancer and four cases of ovarian cancer. The skipping of exon 18 retains the same reading frame and removes 26 amino acids, disrupting the first BRCT domain of BRCA1. The skipping mechanism was interpreted to be related to nonsense codon-mediated disruption of splicing. I searched exon 18 with the SF2/ASF score matrix and found a high-score motif that was eliminated by the point mutation. I then constructed a mini-gene containing the sequence from exon 17 to exon 19 of the BRCA1 gene. Mini-gene transcripts were spliced in HeLa nuclear extract. Remarkably, exon 18 was fully included in the wild-type context but fully skipped in the nonsense mutant. Preliminary gel-shift analysis indicates that the mutant RNA binds SF2/ASF with reduced efficiency. I conclude that exon 18 skipping is due to the failure of SF2/ASF to recognize a specific ESE during the early stages of splicing. I further showed that a missense mutation at the same position that lowers the SF2/ASF score has the same effect, indicating that this phenomenon is not linked to the recognition of in-frame nonsense codons.

Another example is the skipping of exon 51 in the fibrillin-1 (FBN1) gene exon 51, which is caused by a silent mutation. Using the SRp55 score matrix, I found two high-score motifs that were eliminated by this single mutation. I am currently testing splicing of the mini-fibrillin gene in vitro. I have also searched other human disease genes. The results correlate well with experimental data. These SR protein-score matrices are very useful for characterizing human gene mutations. To understand the mechanism of a certain mutation in a disease gene, a major effort is required when using traditional methods. The score matrices can point to the relevant mechanism much faster when the mutations involve SR protein-exon sequence interactions. My work also suggests that ESEs are

extremely prevalent in metazoan exons, and that the phenotypes of many missense and nonsense mutations may reflect aberrant splicing, rather than a change in a single amino acid, premature termination of translation, or nonsense-mediated mRNA decay.

Characterization of substrate-specific splicing factor(s)

Most pre-mRNAs tested can splice in HeLa S100 extract complemented with SR proteins. However, a few pre-mRNAs require one or more additional nuclear factors. To ask if the missing activity is an snRNA, I treated nuclear extract with micrococal nuclease. The nuclease-treated extract still complemented the S100 extract plus SR proteins for splicing of a bovine growth hormone (bGH) mini-gene transcript, indicating that the complementing activity is proteinaceous. I then asked what kind of cis-elements is associated with this activity. When I optimized the 5' splice site or swapped the upstream exon with another exon from a different gene, the nuclear activity was still required. This result suggested that this activity might be functionally associated with a weak 3'splice site or with the exonic splicing enhancer in the bGH last exon.

I am trying to identify this activity. Preliminary results indicate that it fractionates in a 20-50% ammonium sulfate cut. It separated from nucleic acids on a $CsCl_2$ gradient and bound to a Poros-HQ column. Further purification is in progress.

Publications resulted from this funding
1. Liu, H.-X., Zhang, M., and Krainer A. R. (1999) Distinct and degenerate substrate specificity of human SC35 and SF2/ASF. Manuscript in preparation.
2. Chew S. L., Liu H.-X., Mayeda A. and Krainer A. R. (1999) An exonic enhancer functions in the second step of pre-mRNA splicing. Manuscript submitted.
3. Liu, H.-X., Zhang, M., and Krainer A. R. (1998) Identification of Functional Exonic Splicing Enhancer Motifs Recognized by Individual SR Proteins. Genes & Development, 12: 1998-2012

References
M. Burset and R. Guigo (1996) *Genomics*, **34**, 353-67.
S. D. Chandler, A. Mayeda, J. M. Yeakley, A. R. Krainer and X. D. Fu (1997) *Proc Natl Acad Sci U S A*, **94**, 3596-601.

A. Mayeda, G. Screaton, S. D. Chandler, X.-D. Fu and A. R. Krainer (1999) Substrate specificities of SR proteins in constitutive splicing are determined by their RRMs and composite pre-mRNA exonic elements. *Mol. Cell. Biol.*, **in press**.

S. Mazoyer, A. Puget, L. Perrin-Vidoz, H. T. Lynch, O. M. Serova-Sinilnikova and G. M. Lenoir (1998) *Am. J. Hum. Genet.* **62**, 713-15

R. Tacke and J. L. Manley (1995) *Embo J*, **14**, 3540-51.