Prediction of tertiary structure of NSSRs' RNA recognition motif and the RNA binding activity.

Kazuo Fushimi[1, 2], Shizuka Uchida[1], Ryousuke Matsushita[1] and

Toshifumi Tsukahara[1, 2] *

[1] Center for Nano Materials and Technology, Japan Advanced Institute of Science and Technology, 1-1, Asahidai, Tatsunokuchi, Ishikawa, 923-1292, Japan

[2] Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012, Japan.

Running Title: Function and Expression of NSSR in neurons

Key Words: Splicing; SR protein; Brain; Pre-mRNA; Tertiary structure; RNA recognition motif (RRM); cDNA microarry

*Correspondence author

Telephone: +81-761-51-1440

FAX: +81-761-51-1149

Email: tukahara@jaist.ac.jp

ABSTRACT

RNAs possess potentials to become excellent bio-material because of their biochemical and biological activities. For instance, most RNA splicings are catalyzed by machinery including their own RNAs or other RNAs. The eukaryote machineries for splicing of pre-mRNA, which are called spliceosomes, are flexible and accurate for separating substrate and catalytic RNAs. Although RNAs themselves catalyze the splicing, spliceosomes are supported by many proteins. Furthermore, a great accuracy is required for the alternative splicing because there are choices available, which must be regulated in tissue-specifically and developmentally manners. Neural-salient serine/arginine-rich (NSSR) proteins 1 and 2 are candidates for supporting the accuracy of the splicing. The features of their amino acid sequences suggest that NSSRs are SR proteins, which bind to pre-mRNA and determine the splicing site. Since SR proteins have a RNA recognition motif or motives (RRM or RRMs), which binds to RNA, we predicted the secondary and tertiary structures of NSSRs' RRM by comparing them to RRMs of other proteins. The predicted structure suggested that the RNA binding activity of NSSRs' RRM is similar to the poly A binding protein (PABP). Moreover, to detect the targets for NSSR, mRNAs were obtained by screening them from murine brains with bacterial recombinant NSSRs' RRM and microarray experiments were conducted using these mRNAs. The results suggested that NSSRs bind specifically to particular pre-mRNAs and regulate the alternative splicing of the binding pre-mRNA.

INTRODUCTION

Some RNAs have catalytic and biological activities; therefore, it is valuable to study them and to explore their applications for various fields, including material developments. Among these RNAs, self splicing is one of their catalytic activities. For example, many mRNAs in eukaryotic cells are produced when introns are spliced out from the pre-mRNAs, which are immediate copies of genomic DNA by transcription. This splicing is not an autocatalytic reaction, but catalytically-active RNA components splice the pre-mRNA [1]. In this mechanism, many proteins cooperate to activate the catalytic activity. Among these proteins, SR proteins are included. The characteristics of the amino acid sequences of SR proteins are the presence of the RNA recognition motif/motifs (RRM/RRMs) at the N-terminal and the serin-arginine dipeptide sequence at the C-terminal. They recognize exons and bring the splicing machinery to the correct position to perform splicing at exact sites [2]. In addition, more than one mRNAs are produced from a single gene due to different choices of exons on the gene in metazoan, which is known as alternative splicing. This alternative splicing is regulated in a tissue and developmental specific manner in living organisms. SR proteins also contribute to regulation of alternative splicing.

We previously reported that two genes for SR protein like polypeptides, neural salient serine/arginine –rich 1 and 2 (NSSR 1 and 2), were cloned from a subtraction library of neuroectodermal differentiated P19 cells by degenerative PCR [3]. To characterize these genes, the conserved domain search was performed, which suggested the presence of an RRM in the N-terminal region of the products.

Furthermore, their C-terminals consist of SR repetitive sequences. However, there exists a difference between them; that is, NSSR2 is shorter than NSSR1 due to alternative splicing. In the expression levels, mRNAs for NSSR1 and 2 are present at higher levels in the brain and testis than in other tissues. Unlike NSSR2, which is expressed before the differentiation of the cells, NSSR1 is expressed in the neural stage during the neuroectodermal differentiation of embryonal carcinoma cells, P19, suggesting that it is engaged in the production of neuronal splicing products. In addition, the homology search for NSSRs revealed that there exist high similarities with many RNA binding proteins, including SR proteins. Currently, it is known that the overexpression of NSSRs regulate alternative splicing of GluR-B gene [3]; however, it is not yet clear how the RRM of NSSRs contributes to the regulation of its splicing activities.

In recent years, the amount of protein structures analyzed by NMR and X-ray crystallography has increased tremendously. Among them, RRM is one of the well characterized polypeptide structures, which consists of four β-sheets strands and α -helix in between the strands (β / α / β / β / α / β) [4, 5, 6, 7, 8, 9, 10]. A typical RRM is 90-100 amino acid long with two conserved sequence motifs, RNP-1 [(K/R)-G-(F/Y)-(G/A)-F-V-x-(F/Y)] and RNP-2 [(L/I)-(F/Y)-(V/L)-(G/K)-(G/N)-(L/M)] [6]. The Aromatic amino acids on RNP-1 and 2 interact with bases of RNA by stacking on the surface of RRM. Its RNA binding activity is supported primarily on its β sheet surface.

Microarrays are a very useful tool to analyze relative amounts of mRNA in a high-throughput manner. They can be applied for analyses for mRNA expression to

understand biological phenomena, diseases or diagnoses. Their basic procedural scheme is that nucleic acids are labeled by fluorescent dyes, such as Cy3 and Cy5, and the labeled nucleic acids are hybridized with spotted probes on the microarray slide to give fluorescent intensities, which represent the amount of nucleic acids contained in the cell.

In this paper, we predicted the structure of the RRM in NSSRs and showed its possibility of binding to a RNA. Furthermore, using combination of recombinant technology and microarray analysis, we successfully screened murine brain mRNAs to find NSSRs' targets.

EXPERIMENTAL PRECEDURE

Structure analyses of NSSRs' RRM

The comparison of amino acid sequences was performed using conserved domain database search at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml) [11]. For the prediction of the secondary structure of NSSRs' RRM, PSIPRED (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html) was utilized [12]. The tertiary structure of NSSRs' RRM was computed through SWISS MODEL (http://swissmodel.expasy.org//SWISS-MODEL.html) [13]. The structures were visualized in Swiss-PdbViewer (http://kr.expasy.org/)

Production and extraction of the NSSRs' RRM recombinant protein

To produce a NSSRs' RMM recombinant protein, the DNA fragments that code the RRM were amplified by high fidelity PCR (pyrobest, TAKARA) using primers, RRMF 5'-ctggaattcccatgtcccgatacctgcgcc-3' and RRMR 5'-ccgaattctatatcggtcataatcgtcata-3'. These fragments were subcloned into the *Eco* RI site of pYesTrp3. The obtained plasmids were digested with *Hind* III and *Xho* I to insert into pET-34b(+), which carries a cellulose binding domain tag (CBD tag, Novagen).

The NSSRs' RRM recombinant protein was produced in the 100ml culture of *E. coli*., Rosetta(DE3)pLysS, by induction with 1mM IPTG in LB medium for three hours at 37 ºC. The bacterial pellet was frozen once, resuspended with PBS containing the protease inhibitor cocktail (NACALAI TESQUE), and then lysed by sonication though the treatment on ice (VIBRA CELL, SONIC & MATERIAL). The lysate was concentrated by ultrafiltration using Amicon Ultra-4 (MWCO 10,000, MILLIPOR), and then applied onto cellulose resin to detect the RNA binding activity of NSSR's RRM-CBD fusion protein.

Messenger RNA binding assay

Total RNA was extracted from the bulk of murine brains obtained from 5-day-old to adult mice using TRIzol (Invitrogen) followed by purification using mTRAP total kit (ACTIVE MOTIF). The purified recombinant NSSRs' RRM-CBD fusion protein was mixed with mRNA fraction, then absorbed on 5 mg of cellulose resin (CBinD 100, Novagen). After four washes with 1xSSC containing 0.01% TRITON X-102 and

0.2 units/μl SUPERase•In (Ambion), the resin was resuspended in 1000 μl of the same buffer and then 20 μg of the murine brain mRNA were added. After incubation for ten minutes, the resin was washed four times with the same buffer. Next, the resin was resuspended with 100 μl of 0.5 % SDS and then, 500 μl of TRIzol were added into the resin to purify the mRNA on the resin.

SDS-PAGE and Immunoblot analysis

SDS-PAGE was performed as described by Laemmli [14]. The proteins fractionated by SDS-PAGE were stained with CBB (Rapid Stain CBB kit, NACALAI TESQUE) or transferred onto PVDF membranes. Anti-CBD-Tag rabbit polyclonal antibody (Novagen) was used as a primary antibody (x5000 dilution), and alkaline phosphatase linked anti-mouse IgG and anti-rabbit IgG (Chemicon) were used as secondary antibodies (x10000 dilution). The CBD-tagged proteins were detected using BICP-NBT solution kit (NACALAI TESQUE).

Microarry analysis

The procedure for the microarray experiment was followed as described in the products manual. Briefly, the purified RNA was reverse transcribed using the T7 tagged oligo-dT primer and then, labeled RNA with Cy3 or Cy5 (Low RNA Input Fluorescent Linear Amplification Kit, Agilent). These labeled cRNAs were hybridized onto Mouse Oligo Microarray Kit (Agilent). Then, the hybridized microarray was scanned using CRBIO IIe (Hitachi Software Engineering). The scanned image was

analyzed using ScanAlyze version 2.50 (http://rana.lbl.gov/EisenSoftware.htm).

RESULTS AND DISCUSSION

Secondary structure analyses

The secondary structure of NSSR 1 was predicted using PSIPRED. As a result, alpha helix structures were suggested in positions from 23 to 32 and from 61 to 71, while residues next to both helixes at the C- and N- terminals were likely to form β-strand structures (Figure 1). Because an RRM forms a β / α / β / β / α / β structure from its N-terminal and is about 100 amino acids long, the prediction by PSIPRED suggested that NSSRs have a similar structure to other RRMs (Figure 2). The sequence features of NSSRs were further explored by comparison with other proteins. The following proteins were selected for comparison: solution structures of U2AF65 (accession; 1U2FA), hnRNP D0 RBD 1 (accession; 1HD0) and 2 (accession; 1IQT), and mushashi1 RBD 1 (accession; 2MSS), all of which were analyzed by NMR [7, 8, 9, 10]. Their structures include four-stranded antiparallel β-sheets backed by two alpha helices, which are characteristics of RRMs. The comparison among the NSSRs' RRM and above mentioned proteins revealed that the amino acid sequences S11, F13 and R15 of the first β-strand from N-terminal (β1), V37, D38, Y40 and P42 of the second β-strand from N-terminal (β2), and F53, Y55 and Q57 of the third β-strand from N-terminal (β3) in the NSSRs' RRM are likely to be exposed to the solvent that surrounds them. β1 and 3 were similar to the corresponding β-sheets of the NSSRs' RRM predicted by PSIPRED, which are reflecting RNP-1 and 2 motifs in β3 and 1 of RRMs, respectively. The third

amino acids on β1 and β3 of RRMs are usually aromatic amino acids (F or Y) that are on the surface of the proteins and interact with bases of the target nucleic acids by stacking. The conservation of aromatic amino acids at the 13^{th} and 55^{th} positions from the N-terminal of the NSSRs' RRM (F13 and Y55) suggested that these amino acids are also located on the surface of the protein and stack with the bases of the target nucleic acids. However, β2 and β4 (the fourth β-strand from N-terminal) were not well-conserved among the compared RRMs, especially, the β4 of hnRNP D0 RRM2 has an unusual structure [8]. Although only one β-strand follows after α2 in a typical RRM, there are two possible regions to form β-strands in hnRNP D0 RRM2; in which, the N-terminal region (YHNV) forms an additional antiparallel strand (β4-) with β4 consisting of the C-terminal region (SKCEIKVA). According to the prediction by PSIPRED, a highly possible region to form β-strands next to C-terminal of α2 is divided into two by a random coil as well as hnRNP D0 RRM2 (Figure 1). Therefore, the NSSRs' RRM may fold its structure in a similar manner as hnRNP D0 RRM2 consisting of 5 stands, β1, 2, 3, 4 and 4- does.

Analysis of the tertiary structure of the NSSRs' RRM

The tertiary structure of the NSSRs' RRM was built using SWISS-MODEL. The structure of the domain 1 in human poly A binding protein chain A (accession; 1CVJ_A) was used as a template [5]. As a result, the identity of the amino acid sequences between NSSRs' RRM and 1CVJ_A was 34.6% (Figure 3), and the total final energy for NSSRs' RRM was –5046.952KJ / mol. From these results, NSSRs' RRM

may have a similar structure to a typical RRM (Figure 4 A). In Figure 4C, F13, F53 and Y55 in NSSRs' RRM, which correspond to the solvent exposed aromatic amino acids in the RNP-1 and 2 of PABP (Y14, Y54 and Y56 in Figure 4D), were faced on the outside of the β-sheet. The secondary structure suggest that the RRM of NSSRs has the additional β-strand at the C-terminal as well as the hnRNP D0 domain 1 structure. The upstream short strand of the β4 region (KWI) built β-sheet with the next β-strand (RQIEIQFA) which also built the β-sheet structure with β1. This structure was also observed in PABP domain 1. The tertiary structure analysis suggests that the NSSRs' RRM polypeptide is fold into a typical RRM structure in coincidence with the secondary structure prediction other than the additional β-strand next to β4 (Figure 4 C).

The electrostatic potential of NSSRs' RRM was calculated and shown in Figure 5. The highly negative region was found around the second α-helix of NSSR's RRM (Figure 5A, B), while PABP domain 1 did not have such a large negative area around its second helix region (Figure 5C). Furthermore, the highly negative region around the second α-helix was absent in other RRMs (Figures 2, 3). Most RRMs recognize RNA on the surface of their β–sheets. Therefore, this unique feature of NSSRs' RRM suggested that there exists a specific interaction with certain positively charged proteins or nucleic acids by interacting with their β–sheets. Since the surface of the β–sheet in NSSRs' RRM was neutral as well as that of PABP domain 1, the β–sheet in NSSRs' RRM may interact with bases of RNA by aliphatic and van der Waals interaction, base stacking, and hydrogen bond.

Next, the spatial positions of the amino acids which interact with RNA in

PABP domain 1 and 2 were compared with those of equivalent NSSRs' amino acids. In the case of PABP domain1 and 2, phosphates of RNA interact with K104, R89, Y14, Y54, and Y140 (Figure 3). Among the equivalent positions of NSSRs' RRM, only R15, which corresponds to K104 in the domain 2 of PABP, potentially interacted with phosphate of RNAs. Therefore, the manner in which the residue supports the phosphate of an RNA backbone seemed to be different between NSSR and PABP. In respect of amino acids recognizing adenines, counterparts of NSSRs to the amino acids supporting Ade2 and Ade6 in PABP showed considerable similarity suggesting NSSRs hold adenines at those positions (Figure 6 A, B, G, H, Table 1). Totally, the amino acids which stack with adenines in PABP were well conserved, but most of the counterparts for the amino acids specified adenines in NSSRs may not recognize adenines (Figure 6, Table 1). The equivalent amino acids of NSSRs to amino acids interact with Ade 5 were unclear, because the amino acids straddle both of RRM domain 1 and 2 in PABP.

RNA Binding activity of NSSRs' RRM

The structural observation prompted us to explore the RNA binding activity of NSSR and the targets for NSSR. Generally, SR proteins recognize exons of pre-mRNA to include the selected exons into the spliced mRNA. By using this mechanism, it is possible to screen for the target of the specific splicing. Therefore, we tried to identify target RNAs of NSSRs' RRM. The DNA fragment coding RRM of NSSR was fused with CBD to produce the recombinant protein. The NSSR's RRM-CBD fusion protein was obtained from E. coli lysate, and absorbed on 5 μg of

cellulose resin. We confirmed that 2.6 μg of the purified fusion protein were absorbed on the resin. As shown in Figure 7A, a band at about 46 KDa was detected as a major protein on the resin. In addition, it was reactive to the anti-CBD antibody (Figure 7B). These results suggested that the CBD fused NSSRs' RRM was successfully produced and trapped on the cellulose resin.

The mRNAs from murine brains were added onto the immobilized NSSRs' RRM-CBD fusion protein/ cellulose column and washed four times with the binding buffer to eliminate unbound mRNAs. Next, the trapped mRNAs were extracted by denaturation. The recovered mRNAs were amplified and labeled with Cy5. The untreated murine brain mRNAs were labeled with Cy3 as a reference. Both labeled cRNAs were hybridized onto the mouse oligo microarray, and scanned with CRBIO IIe. The scanned image was analyzed by ScanAlyze.

Figure 8 shows the distribution of the expression ratio for the microarray data. Channel 1 was the Cy5 fluorescence intensities for the mRNA from NSSRs' RRM trapped on the resin, and Channel 2 was the Cy3 fluorescence intensities for the reference murine brain mRNA. A single peak was observed, but the distribution was skewed towards the higher ratio than towards the lower, which might suggest that some mRNAs were concentrated for NSSRs' RRM trapped on the resin. The most concentrated RNA was the clone IMAGE:3584181 (13-fold expression level increase compared to the reference), and 63 other RNAs were concentrated more than 4.9 times (Table 2). These RNAs coded many types of proteins, including structure proteins (e.g. neurofilament-L), kinases (e.g. Mapk12), secretory proteins (e.g. Chromogranin B) and RNA-binding

proteins (e.g. mHuC-L). Many hypothetical genes and pseudogenes were also found among them. Chromogranin B was listed many times therefore the microarray analysis was likely to be reliable. The Database searches for these genes through Ensembl Genome Browser (http://www.ensembl.org/) revealed that 27.5 % of the identified genes had alternative splicing isoforms. Although NSSRs were expected as an alternative splicing regulator, there were many genes which were not suggested as splicing isofoms in the list. Therefore, the splicing isoforms of these genes may be unidentified so far. Another possibility is that NSSRs participate in not only alternative splicing but also constitutive splicing in the brain.

CONCLUSIONS

In this paper, computational methods were applied to predict the secondary and tertiary structures of NSSRs' RRM from its amino acid sequence. Through the homology searches, it was suggested that the RNA binding activity of NSSRs' RRM is similar to that of a typical RRM, PABP. Furthermore, the NSSR's targets were screen from the murine brain by a combination of recombinant technology and microarray analysis. These results suggested that NSSRs contribute to choices of exons for the alternative splicing by directly binding to its target pre-mRNAs.

ACKNOWLEDGEMENTS

This works was supported by the Center for Nano Materials and Technology, JAIST.

REFFERENCES

[1] S. J. Melissa and J. M. Melissa, Pre-mRNA splicing: awash in a sea of proteins, Mol. Cell 12 (2003) 5-14.

[2] B. R. Graveley, Sorting out the complexity of SR protein functions RNA 6 (2000) 1197-1211.

[3] M. Komatsu, E. Kominami, K. Arahata, T. Tsukahara, Cloning and characterization of two neural-salient serine/arginine-rich (NSSR) proteins involved in the regulation of alternative splicing in neurones Genes Cells 4 (1999) 593-606.

[4] C. Oubridge, N. Ito, P. R. Evans, C. Teo, K. Ngai, Crystal structure at 1.92Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA heirpin, Nature 372 (1994) 432-438.

[5] R. C. Deo, J. B. Bonanno, N.Sonenberg, S. K. Burley, Recognition of polyadenylate RNA by the poly(A)- binding protein, Cell 98 (1999) 835-845.

[6] E. Birney, S. Kumar, A. R. Krainer, Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors, Nucleic Acid. Res. 21 (1993) 5803-5816.

[7] T. Nagata, R. Kanno, Y Kurihara, S. Uesugi, T. Imai, S. Sakakibara, H. Okano, M. Katahira, Structure, backbone dynamics and interactions with RNA of the C-terminal RNA binding domain of a mouse neural RNA-binding protein, Musashi 1, J. Mol. Biol. 287 (1999) 315-330.

[8] M. Katahira, Y. Miyanori, Y. Enokizono, G. Matsuda, T. Nagata, F. Ishikawa, S. Uesugi, Structure of the C-terminal RNA-binding domain of hnRNP D0 (AUF1), its

interactions with RNA and DNA, and change in backbone dynamics upon complex formation with DNA, J. Mol. Biol. 311 (2001) 973-988.

[9] T. Nagata Y. Kurihara, G. Matsuda, J. Saeki, T. Kohno, Y. Yanagida, F. Ishikawa, S. Uesugi, M. Katahira, Structure and interactions with RNA of the N-terminal UUAG-specific RNA-binding domain of hnRNP D0, J. Mol. Biol. 287 (1999) 221-237.

[10] C. L. Kielkopf, N. A. Rodionova, M. R. Green, S. K. Buriey, A novel peptide recognition mode revealed by the x-ray structure of a core U2AF35/U2AF65 heterodimer, Cell 106 (2001) 595-605.

[11] A. Marchler-Bauer, J. B. Anderson, C. DeWeese-Scott, N. D. Fedorova, L. Y. Geer, S. He, D. I. Hurwitz, J. D. Jackson, A. R. Jacobs, Lanczycki CJ, Liebert CA, Liu C, Madej T, Marchler GH, Mazumder A. N. Nikolskaya, A.R. Panchenko, B. S. Rao, B. A. Shoemaker, V. Simonyan, J. S. Song, P. A. Thiessen, S. Vasudevan, Y. Wang, R. A. Yamashita, J. J. Yin, S. H. Bryant, CDD: a curated Entrez database of conserved domain alignments, Nucleic Acid. Res. 31 (2003) 383-387.

[12] L. J. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (2000) 404-405.

[13] T. Schwede, J. Kopp, N. Guex, M. C. Peitsch, SWISS-MODEL: an automated protein homology-modeling server, Nucleic Acid. Res. 31 (2003) 3381-3385.

[14] J. Sambrook, D. W. Russell, Molecular Cloning, third ed., Cold Spring Harbor, New York, 2001.

FIGURE LEGENDS

Figure 1 Predicted secondary structure of NSSR.

The predicted secondary structure of NSSR 1 was produced from its full primary sequence by the PSIPRED program using its default settings. E and yellow arrows indicate β-strands, and α–helices are shown by H and green columns. C indicates possibility of random coils. From S11 to Q84, β/α/β/β/α/β alignment was observed for NSSR1 as well as typical RRMs. In the C-terminal of NSSRs' RRM, 2 β-strands were predicted (β β).

Figure 2 Comparison of primary structure of NSSRs' RRM with other RRMs characterized by NMR.

The RRMs, whose structures were determined by NMR, were aligned with NSSRs' RRM. Gray shaded regions and underlined regions form α-helices and β-strands, respectively. Yellow shaded parts indicate amino acids whose exposed solvent surrounded the RRMs' in the β-sheets. RNP-1 and 2 motifs are conserved on third and first β-strands, respectively, and aromatic amino acids indicated by asterisks are exposed to the outside of the proteins.

Figure 3 Comparison of primary structure of NSSRs' RRM with PABP domain 1 and 2.

To construct the tertiary structure of NSSRs' RRM, PABP domain 1 was chosen as a model. The primary sequences of NSSRs' RRM, PABP domain 1, and 2 were

aligned to compare their amino acids that are contributing to the RNA binding. Asterisks indicate aromatic amino acids conserved in RNP-1 and 2. Residues of amino acids shaded by gray support RNAs. The yellow shaded amino acid backbone participates in holding RNA. Amino acids contributing to highly negative potential on the surface of NSSRs' RRM are shaded by green.

Figure 4 The tertiary structure of NSSRs' RRM.

The tertiary structure of NSSRs' RRM was constructed through SWISS MODEL. The cartoons of the structures for NSSRs' RRM and PABP domain 1 are shown in A and B, respectively. The aromatic amino acids conserved in RNP-1 (F53 and Y55) and RNP-2 (F13) of NSSRs were exposed to the solvent surrounding them (C) as well as those of PABP domain 1 (D)

Figure 5 Electrostatic potential of NSSRs' RRM.

Electrostatic potentials of NSSR's RRM and PABP domain 1 were computed and shown in A, B (NSSR) and C (PABP). The second α-helix region of NSSRs' RRM was highly negatively charged (A and B), but only two negative charged amino acids on the surface of same region of PABP were observed (C).

Figure 6 The details of the tertiary structure of NSSRs' RRM.

Amino acids contributing to holding RNA in PABP (A, C, E, G, I and K) and their positions were well conserved in NSSRs' RRM (B, D, F, H, J and L). N105 and

Y140, K104, N100 and S127, backbone of W86 and Q88, D45, and S12 and N58 in PABP are specified adenine, Ade2 (A), Ade3 (C), Ade4 (E), Ade6 (G), Ade7 (I) and Ade8 (K), respectively.

Figure 7　Production of CBD tagged NSSRs' RRM and absorption onto cellulose.

The CBD-tagged NSSRs' RRM was produced by *E. coli*, purified by using a cellulose column and analyzed with SDS-PADE (A) and immunobloting (B). The *E. coli* lysate contained the CBD anti-body reactive protein at 46kDa (B, the left lane). It was confirmed that the recombinant proteins were successfully pulled-down with 5 μg of CBinD resin (A, right lane and B, middle lane).

Figure 8　The histogram for the M-values of the microarray data.

The M-value for each spot was calculated using the following formula: $M = \log_2 (CH2/CH1)$, where both CH1 and CH2 are the background-subtracted intensity values for Channel 1 and Channel 2, respectively. The histogram was drawn based on the M-values to observe the overall tendency of the microarray data. The x-axis indicates the M-values, and the y-axis shows the frequency distribution. As it can be observed from the figure, there is a tendency of the distribution of the M-values towards higher M-values (compare the frequencies around M=-2 and M=1 and higher). This tendency corresponds to the fact that some gene transcripts are concentrated due to the attachment of the recombinant NSSRs' RRM.

Figure 1

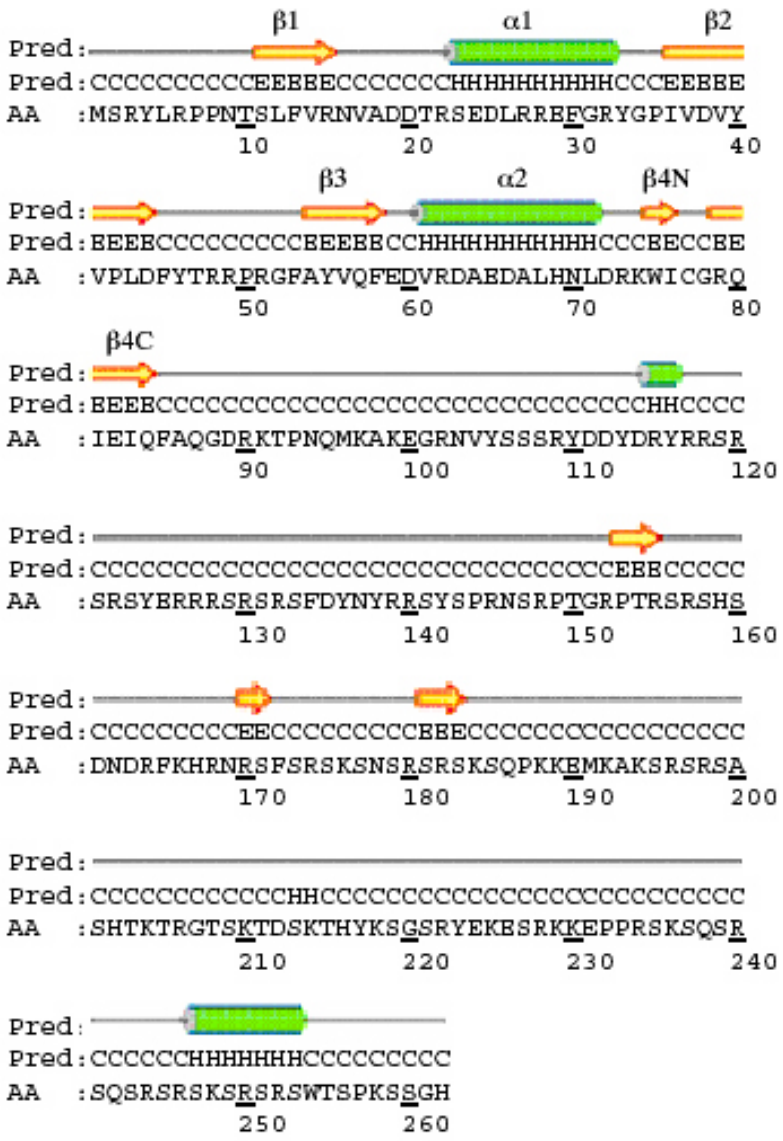
β1
α1
β2
Pred:
Pred:CCCCCCCCCCCEEEEECCCCCCCHHHHHHHHHHCCCEEEEE
AA :MSRYLRPPNTSLFVRNVADDTRSEDLRREFGRYGPIVDVY
10 20 30 40
β3
α2
β4N
Pred:
Pred:EEEECCCCCCCCCEEEEECCHHHHHHHHHHCCCEECCEE
AA :VPLDFYTRRPRGFAYVQFEDVRDAEDALHNLDRKWICGRQ
50 60 70 80
β4C
Pred:
Pred:EEEECCCCCCCCCCCCCCCCCCCCCCCCCCCCCHHCCCC
AA :IEIQFAQGDRKTPNQMKAKEGRNVYSSSRYDDYDRYRRSR
90 100 110 120
Pred:
Pred:CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEECCCCC
AA :SRSYERRRSRSRSFDYNYRRSYSPRNSRPTGRPTRSRSHS
130 140 150 160
Pred:
Pred:CCCCCCCCCEECCCCCCCCCEEECCCCCCCCCCCCCCCCC
AA :DNDRFKHRNRSFSRSKSNSRSRSKSQPKKEMKAKSRSRSA
170 180 190 200
Pred:
Pred:CCCCCCCCCCCCHHCCCCCCCCCCCCCCCCCCCCCCCCC
AA :SHTKTRGTSKTDSKTHYKSGSRYEKESRKKEPPRSKSQSR
210 220 230 240
Pred:
Pred:CCCCCCHHHHHHHCCCCCCCCC
AA :SQSRSRSKSRSRSWTSPKSSGH
250 260

Figure 2

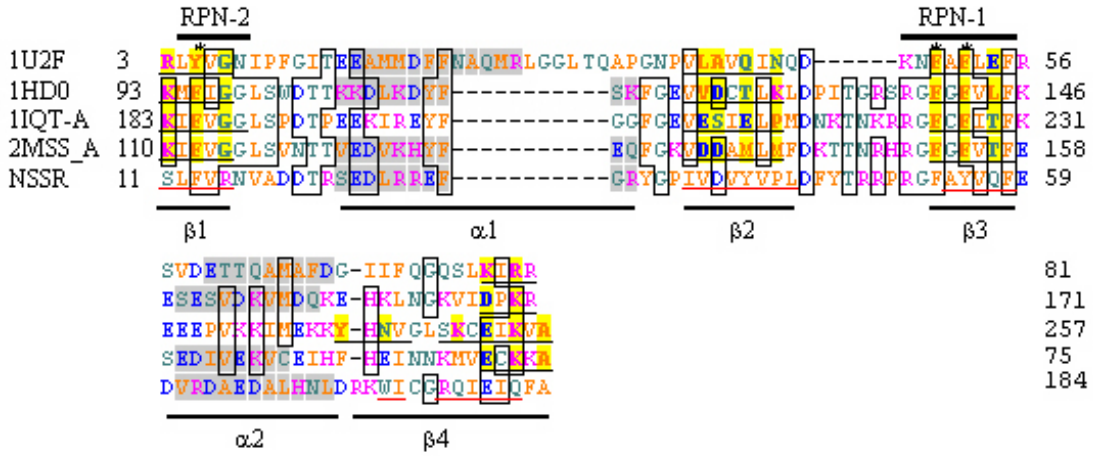
RPN-2
RPN-1
1U2F 3
1HD0 93
1IQT-A 183
2MSS_A 110
NSSR 11
56
146
231
158
59
β1
α1
β2
β3
81
171
257
75
184
α2
β4

Figure 3


NSSR
domain1
domain2
1 -----MSRYLRPPNTSLFVRNVADDTRSEDLRREFGRYGPIVDVYVPLDFYTRRPRGFAY 55
12----------SLYVGDLNPDVTRAMLYEKFSPAGPILSIRVCRDMITRRSLGYAY 56
85MWSQRDPSLRKSGVGNIFIKNLDKSIDNKALYDTFSAFGNILSCKVVCDENGSK--GYGF 142
NSSR
domain1
domain2
56 VQFEDVRDAEDALHNLDRKWICGRQIEIQFAQGDRKTPNQMKAKEGRNVY 105
57 VNFDQPADAERALDTMNFDVIKGKPVRIMWSQRDPSLR---------- 83
143VHFETQEAAERAIEKMNGMLLMDRKVFVGRFKSRKER---------- 179

Figure 4

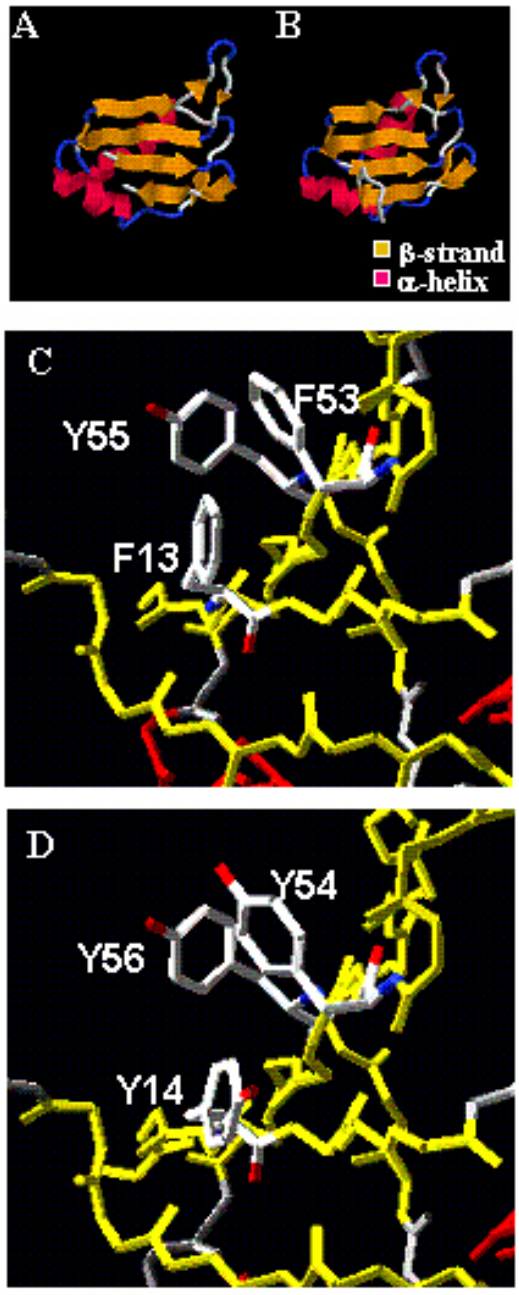

Figure 5

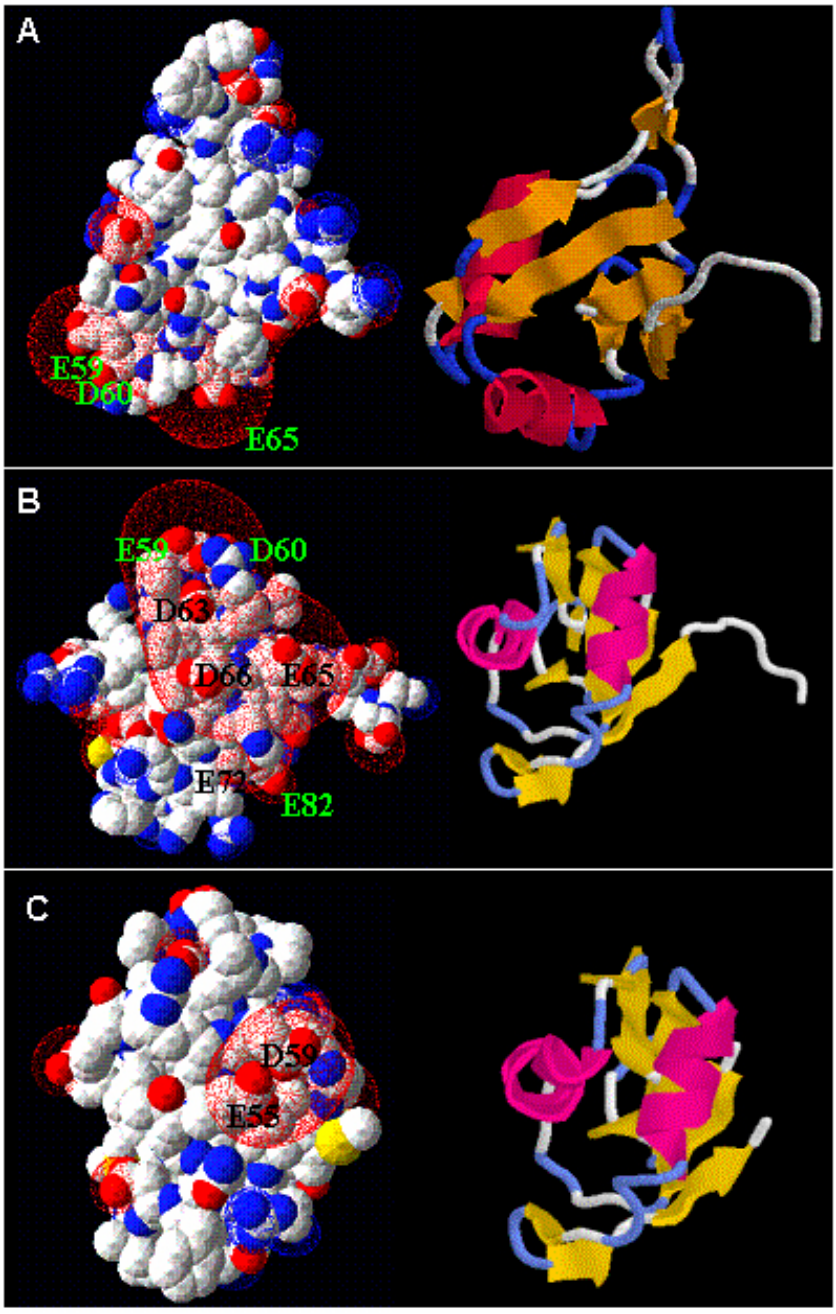

Figure 6

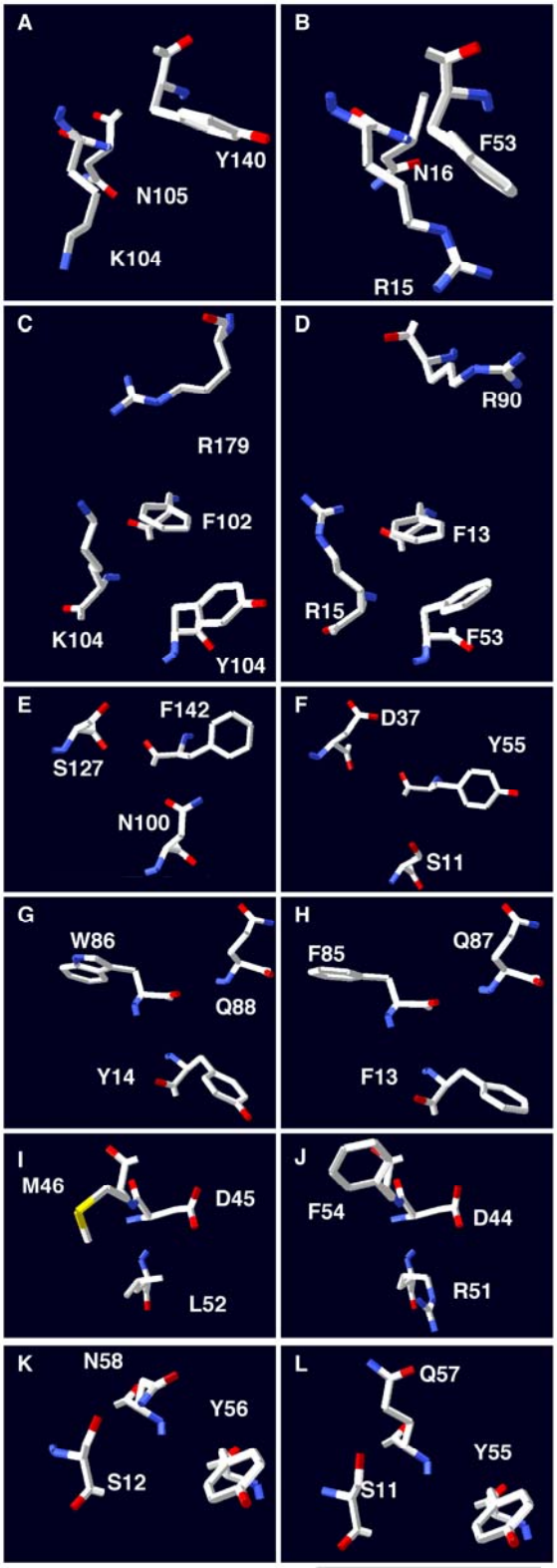
A
Y140
N105
K104
B
F53
N16
R15
C
R179
F102
K104
Y104
D
R90
F13
R15
F53
E
F142
S127
N100
F
D37
Y55
S11
G
W86
Q88
Y14
H
F85
Q87
F13
I
M46
D45
L52
J
F54
D44
R51
K
N58
Y56
S12
L
Q57
Y55
S11

Figure 7

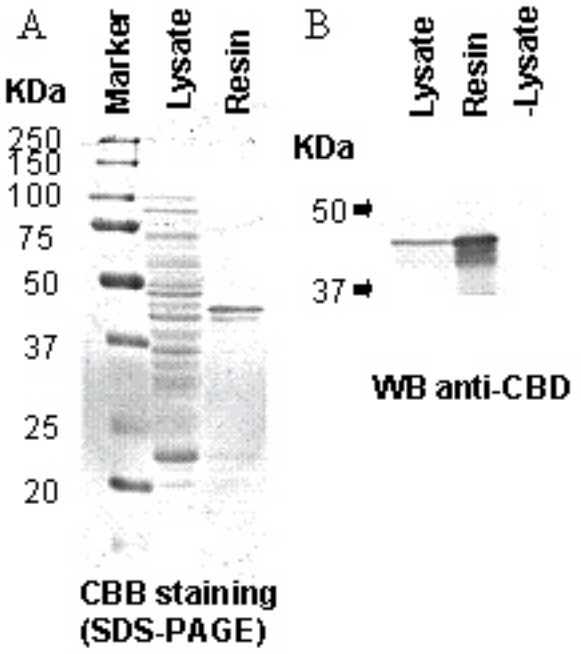
A
KDa
Marker
Lysate
Resin
250
150
100
75
50
37
25
20
CBB staining
(SDS-PAGE)
B
Lysate
Resin
-Lysate
KDa
50
37
WB anti-CBD

Figure 8