Available online at www.sciencedirect.com
SCIENCE DIRECT®
Biochemical and Biophysical Research Communications 334 (2005) 61–68
BBRC
www.elsevier.com/locate/ybbrc

Expression of the C-terminal domain of novel human SR-A1 protein: Interaction with the CTD domain of RNA polymerase II

Maria E. Katsarou [a], Athanasios Papakyriakou [a], Nikos Katsaros [a], Andreas Scorilas [b,*]

[a] *Institute of Physical Chemistry, NCSR "Demokritos," 153 10 Ag. Paraskevi Attikis, Greece*
[b] *Department of Biochemistry and Molecular Biology, Faculty of Biology, University of Athens Panepistimioupolis, 15701 Athens, Greece*

Received 20 May 2005
Available online 22 June 2005

Abstract

We have recently cloned a new member of the human Ser/Arg-rich superfamily (SR) of pre-mRNA splicing factors, SR-A1. Members of the SR family of proteins have been shown to interact with the C-terminal domain (CTD) of the large subunit of RNA polymerase II, and participate in pre-mRNA splicing. The largest subunit of RNA polymerase II contains at the carboxy-terminus a peculiar repetitive sequence that consists of 52 tandem repeats of the consensus motif Tyr-Ser-Pro-Thr-Ser-Pro-Ser, referred to as the CTD. There is evidence that SR protein splicing factors are involved in cancer pathobiology through their involvement in alternative processing events. The CTD of human SR-A1 protein (aa 1187–1312), containing a conserved CTD-interaction domain and bearing a decahistidine (His10) tag was produced by DNA recombinant overexpression techniques in *Escherichia coli* from the vector pET16b and it was localized in the periplasmic space. The protein was further purified using a HiTrap chelating column and its circular dichroism spectra indicate that it assumes a defined structure in solution. Performing a pull-down assay we proved that the novel SR-A1 [1187–1312 His10] protein interacts with the CTD domain of RNA polymerase II.

Keywords: SR-A1; CTD binding domain; Splicing; Protein expression; Comparative modeling; SR proteins

The serine-arginine (SR) proteins are a group of highly conserved proteins in metazoans, that are required for constitutive splicing and also influence alternative splicing regulation [1,2]. SR proteins can complement splicing-inactive S100 fractions and alter splice-site selection in alternatively spliced pre-mRNAs [3] and they have also been shown to cooperate in the recruitment of multiple copies of TAP/NXF1, the primary receptor of general mRNA export, and for efficient mRNA export [4]. What is more, there is evidence that SR protein splicing factors are involved in cancer pathobiology through their effect on alternative processing decisions [5,6]. A new member of SR proteins was recently identified, SR-A1 gene which appears to be the human homologue of the rat A1 gene [7]. SR-A1 gene is located on chromosome 19q13.3, close to RRAS oncogene, BCL2L12, TSKS, and Kallikrein gene family locus. The expression of this novel gene in a series of human tumors was examined and it was found that it is overexpressed in a subset of ovarian cancers that are clinically more aggressive [7]. The predicted protein is highly homologous to the rat A1 SR protein [8] and it contains both signature domains present in the rat and human high molecular weight SR proteins, that is, the C-terminal domain (CTD) interacting domain and SR/RS dipeptides (Fig. 1). In rat and yeast, the SR proteins have been characterized as part of a large complex which is directly associated with the CTD of the large subunit of RNA polymerase II (RNAP II) [8].

The largest subunits of RNAP's I, II, and III are highly homologous, yet an unusual domain at the carboxyl terminus of this subunit is found only in RNAP

* Corresponding author. Fax: +30 210 727 4158.
E-mail addresses: ascorilas@biol.uoa.gr, scorilas@netscape.net (A. Scorilas).

0006-291X/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2005.06.053

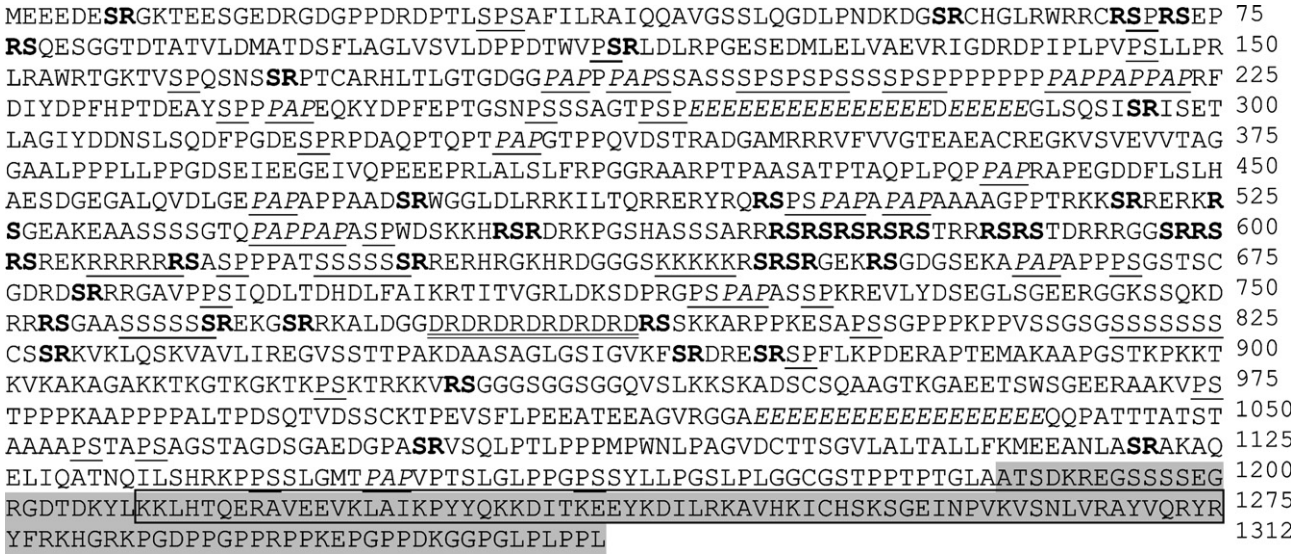

Fig. 1. Amino acid sequence of the SR-A1 protein. RS or SR dipeptides are shown in bold and the CTD-binding domain of the SR-A1 protein is boxed. The amino acid sequence that was expressed is high-lightened. The full genomic organization of the gene is reported in GenBank (Accession No. AF254411).

II and is highly conserved in widely diverged eukaryotes [9,10]. The CTD is unusual in that it is comprised of a tandemly repeated heptapeptide with the consensus Tyr-Ser-Pro-Thr-Ser-Pro-Ser [11]. The CTD appears to be structurally flexible in solution and it can assume different conformations depending on its binding partner [12]. The length of the CTD is variable among different eukaryotes, yet its length roughly correlates with the complexity of the organism. RNAP II of mammalian cells contains 52 copies of the consensus repeat, and yeast contains 26–27 copies, whereas other eukaryotes contain an intermediate number of repeats [13]. Each CTD heptad repeat possesses numerous potential phosphoreceptors and, in fact, two forms of the polymerase in vivo RNAP IIA (hypophosphorylated) and RNAP IIO (hyperphosphorylated), are present in cells [14]. The CTD is phosphorylated by multiple protein kinases including Kin28p/Cdk7 in the general transcription factor, TFIIH, Srb10p/Cdk8 in the Srb complex, and Ctk1p/Cdk9 [15]. These kinases are generally specific for either serine 2 or serine 5 of the individual repeats. Although all the details are not yet understood, this phosphorylation clearly plays an important role in the regulation of CTD function [13,16].

Reports have suggested the involvement of polymerase II (pol II) in the splicing and polyadenylation steps of mRNA production. Greenleaf [17] suggested that the negatively charged hyperphosphorylated CTD act as an ionic scaffold for the recruitment of arginine-rich (and thus positively charged) SR proteins involved in splicing. The CTD binding domain in complex with the SR proteins may be involved in the early post-initiation events that take place during mRNA synthesis, including splicing. What is more, the CTD binds capping enzymes, snRNPs and serine-arginine (SR)-like proteins [16,18–32].

Plasmid containing SR-A1 cDNA was used in PCR experiments to amplify a sequence corresponding to residues 1187–1312, representing the C-terminus of the novel protein which includes residues 1209–1275 that is believed to interact with the CTD-binding domain of RNAP II. The choice of the domain to be expressed was made using molecular modeling studies. The plasmid was subcloned into an expression vector containing an in-frame decahistidine (His10) tag. Substantial amounts of protein were obtained after expression of the target DNA, under the control of the T7 promoter and *araBAD* present on the vector was induced in *Escherichia coli* with L-arabinose. The overexpressed domain was purified using immobilized metal affinity column and purified fractions were used in examining the structural conformation of the novel protein in solution using circular dichroism (CD). The data showed that the expressed SR-A1 [1187–1312 His10] adopts a functional conformation in solution. Interaction studies with the CTD domain of RNA polymerase II showed that the expressed domain of SR-A1 interacts with the CTD of RNA Pol II.

Materials and methods

Materials. A HiTrap chelating metal affinity column was purchased from Amersham-Pharmacia Biotech. *Pfu* DNA polymerase and DH5α competent cells were purchased from Statagene. *E. coli* BL21-AI competent cells and Superscript^{II} for first strand RT-PCR analysis were purchased from Invitrogen Life Technologies. Trizol reagent, agarose, and Plasmid Miniprep kit was purchased from Gibco-BRL. T4 DNA ligase and restriction enzymes came from New England

Biolabs (NEB). The reagents used for the pull-down assay were purchased from PIERCE. All other reagents were of research grade and were obtained from commercial sources.

Bacteria, plasmids, and media. The bacteria used in this work were BL21-AI *E. coli* B/r strain (Invitrogen), genotype F-*ompT* $hsdS_B(r_{B^-}m_{B^-})$*gal dcm araB*::*T7RNAP-tetA*. The BL21-AI *E. coli* strain is derived from the BL21 strain and contains a chromosomal insertion of the gene encoding T7 RNA polymerase (T7 RNAP) into the *ara*B locus of the *araBAD* operon, placing regulation of the T7 RNAP gene under the control of the *araBAD* promoter. The *ara*B gene is deleted in this strain. In order to induce expression from the *araBAD* promoter, L-arabinose 20% is used. The plasmid made for this work was built from the vector pET-16b (Novagen), which allows expression of protein from a T7 promoter. Base media was Luria–Bertani (LB) medium supplemented with 0.1 g/ml ampicillin to maintain expression of plasmids.

Construction of pET-SRA1 plasmid. Total RNA was extracted from human histiocytic lymphoma (U937) [33] using Trizol reagent following the manufacturer's instructions. RNA concentration was determined spectrophotometrically. Two microgram of total RNA was reverse transcribed into first stand cDNA using the SuperscriptII preamplification system. The final volume was 20 μl. The gene was amplified by PCR using oligomers introducing *Nde*I and *Bam*HI one in each end of the amplified gene, encoding the carboxy-terminal portion (aa 1187–1312) of SR-A1 protein (Fig. 1). The sequence of the forward primer was 5′-GGGAATTC CATATG GCCACGTCTGACAAGA GAGA-3′ having an *Nde*I site containing the ATG starting codon (CAT**ATG**) and a sequence that enhance the activity of the restriction endonuclease (GGGAATTC). The sequence of the reverse primer was 5′- CGC GGATCC TCA GAGAGGGGGCAGGGGCA-3′, containing a *Bam*HI site (GGATCC), a sequence that enhance the activity of the restriction endonuclease (CGC) and a stop anticodon (TCA). PCR amplification was performed using *Pfu* DNA polymerase and samples were subjected to 30 cycles of 35 s of denaturation at 95 °C, 45 s of annealing at 60 °C, and 3 min of elongation at 72 °C in a Mastercycler Personal Eppendorf Model. The amplification products were analyzed by electrophoresis on a 1.5% agarose gel and stained with ethidium bromide. The resulting PCR product was then purified, double digested with *Nde*I and *Bam*HI and ligated into pET16b vector previously cut with these enzymes, resulting in the pET16b::SRA1 plasmid. The constructed pET16b::SRA1 plasmid was transformed into DH5α competent cells according to a transformation protocol. The positive colonies with the gene insert were grown in LB medium containing ampicillin (0.1 g/ml) and the plasmid was isolated from harvested bacterial cells using a plasmid extraction kit (MiniPrep DNA extraction Kit, Gibco-BRL). PCR amplification was performed using the pET16b::SRA1 plasmid as a template and the previously described primers. The amplification products were analyzed by electrophoresis on a 1.5% agarose gel and stained with ethidium bromide. Also, DNA sequencing analysis was performed on the plasmid and the inserted gene sequence was identified without any mutation. The isolated pET16b::SRA1 was then used to transform BL21-AI *E. coli* strain competent cells for expression purposes.

Expression of human SR-A1 protein. Overnight cultures were grown at 37 °C in 2 L of LB medium containing ampicillin (100 mg/L) to an OD_{600} of 0.6. At this point the culture was induced with 1 ml L-arabinose 20% and incubated at 37 °C for 6 h with shaking at 180 rpm. The cells were harvested by centrifugation at 9500 rpm for 20 min at 4 °C and the cell pellet was stored at −20 °C.

Purification of human SR-A1 protein. The cells were re-suspended in 25 ml of binding buffer (50 mM NaH_2PO_4, pH 8.0, 300 mM NaCl, and 10 mM imidazole) and the cell suspension was lysed by sonication 10 times for 30 s each time. The insoluble debris was removed by ultracentrifuge using a Beckman Optima LE-80K Ultracentrifuge. The HiTrap column was first washed with 25 ml of MilliQ water and then it was equilibrated with 0.1 M $ZnSO_4$ to charge the column with zinc ions followed by 25 ml of MilliQ water to remove unbound zinc ions from the column, and then by 25 ml of binding buffer to equilibrate the column. The clarified sample was applied to the HiTrap column after filtering with a 0.22 μm filter. The column was washed with 75 ml of wash buffer 1 (20 mM Na_2HPO_4, pH 8.0, 1 M NaCl, and 20 mM imidazole) followed by 50 ml of wash buffer 2 (20 mM Na_2HPO_4, pH 8.0, 1 M NH_4Cl, and 20 mM imidazole). The human SR-A1 protein was eluted with 25 ml of an elution buffer (20 mM NaH_2PO_4, pH 8.0, 0.9 M NaCl, and 100 mM EDTA). The purified protein was analyzed by SDS–polyacrylamide gel electrophoresis (SDS–PAGE) with 17% Tris gels and using a Tris/glycerol/SDS running buffer. The gels were stained with Coomassie brilliant blue. Protein concentration was determined at A_{280} using a spectrophotometer.

Circular dichroism spectroscopy. Circular dichroism (CD) spectroscopy was used to evaluate the conformational stability of SR-A1 protein. The spectra were recorded on a JASCO V-715 dichograph equipped with a temperature controller. The protein solutions at concentration of 65 μM in 20 mM sodium phosphate buffer, pH 8.0, and 0.1 M NaCl were transferred into a CD cuvette of 1.0 cm. All spectra were acquired at 20 °C from 180 to 280 nm at 0.2 nm resolution, with a bandwidth of 1.0 nm at 50 nm/min speed and response of 1.0 s.

Pull-down assay. RNAP II–CTD (RNA polymerase II, carboxy-terminal domain) human, recombinant, *E. coli* was purchased from Jena Bioscience GmbH, Germany and supplied in 20 mM Tris–HCl, pH 7.9, 100 mM KCl, 0.2 mM EDTA, 1 mM DTT, and 20% glycerol (Cat-No PR-714). The interaction of SR-A1 [1187–1312 His10] (100 μg) with CTD of RNAP II (2.5 μg) was examined using ProFound Pull-Down PolyHis Protein:Protein Interaction Kit (Pierce), according to the manufacturer's instructions. The concentration of SR-A1 protein used in the reaction was measured using BCA Protein Assay Kit (Pierce). The two proteins were dialyzed in BupH Tris buffered saline. Briefly, His-tagged SR-A1 [1187–1312] was immobilized in a cobalt chelate after 2 h rotation at 4 °C. The protein was washed, using ProFound Lysis Buffer/BupH Tris buffered saline (1:1), 40 mM imidazole 4 M (wash buffer). RNA pol II–CTD was added (2.5 μg) and the cobalt chelate was left rotated for 3 h at 4 °C. The column was washed three times with wash buffer. The bound proteins were eluted by imidazole elution buffer (ProFound Lysis Buffer/BupH Tris buffered saline (1:1), 290 mM imidazole 4 M). Experiment was performed using SR-A1 [1187–1312] as control. Proteins were detected by SDS–PAGE.

Comparative structure modeling and docking studies. Since there are no related proteins of known structure, we have identified four protein templates for the modeling of the CTD binding domain of SR-A1. Using a BLAST search for short nearly exact matches [34] we found proteins that are not related with the SR family, but alignment of specific fragments from them with four overlapping parts of the 1205–1277 aa domain of SR-A1 exhibits sequence identity over 40% (Table 1). The program MODELLER v8.0 was used for modeling by satisfaction of spatial restraints and optimization employing molecular dynamics with simulated annealing [35]. The structure with the lowest target function was subsequently refined with a 2000-step conjugate gradient minimization using the AMBER force field [36] in an implicit solvent representation with the generalized Born model [37]. Validation of the final model was carried out at the Structure Analysis and Verification Server of UCLA (http://www.doe-mbi.ucla.edu/Services/SV/) employing the programs PROCHECK, WHATIF, VERIFY3D, ERRAT, and PROVE. Overall, the energy-minimized model exhibits a medium quality with no errors, apart from a low backbone *Z*-score. Its helical content is 40% with the rest of the residues equally shared between coils and turns (Table 1).

In order to probe the interaction of CTD with SR-A1 we used AUTODOCK v3.0 [38] with the Lamarckian genetic algorithm. One hundred docking sites of the YSPTS^{phos}PS heptapeptide were initially evaluated with respect to their binding affinity. Two out of the five highest binding energy conformations were found to be close in space, so that they could be linked together. Based on the coordinates of these

Table 1
Sequence alignment of the CTD binding domain of SR-A1 [1205–1277] with four protein templates that were used for comparative modeling: 1O4U [158–196], 1TLJ [7–23], 1OFW [244–263], and 1N81 [71–87]

1O4U	DNHLKMYGSAERAVQEVRKII-PFTTKIEVEVENLEDALR----------
1TLJ	----------EE----LREKALNKIYHDK--EI----------
1OFW	----------CHTK--EIDKANPGR-PNLMAAY----------
1N81	----------VS--IRRYVRKNQNRYN-YF
SRA1	**DKYLKKLHTQERAVEEVKLAIKPYYQKKDITKEEYKDILR-KAVHKICHSKSGEI---NPVKVSNLVRAYV---Q-RY-RYF**
Strc	*HHHHHHHTTHHHHHHHHHHHCCTTTTCCCCCCCCCHHHHHC-HHHHHHHHTTTTTC---CCCCCCTTTTTTT---T-TT-TCC*

Similar residues are marked in bold and positive hits are underlined. The secondary structure (Strc) of the energy-minimized model is noted as *H* for α-helix, *T* for turn, and *C* for coil.

two phosphopeptides we prepared the $Y_{1a}S_{2a}P_{3a}T_{4a}S_{5a}{}^{phos}P_{6a}S_{7a}$ $Y_{1b}S_{2b}P_{3b}T_{4b}S_{5b}{}^{phos}P_{6b}S_{7b}$ CTD segment using the XLEaP module of AMBER v8. After a 500-step steepest descent minimization of the peptide, a short 100 ps molecular dynamics equilibration was performed in implicit solvent increasing the temperature of the system to 300 K. A snapshot of the SR-A1 [1205–1277]/CTD complex from the last 10 ps is shown in Fig. 8A and a surface representation at the last step is shown in Fig. 8B.

Results

Construction of cDNA of SR-A1 protein

The pET16b vector was selected for overexpression of the SR-A1 protein because it has several advantages. The recombinant protein has the N-terminal 10× His-tag, allowing affinity purification using the HiTrap chelating column and pull-down protein–protein interaction assays. After PCR, the length of the fragment that codes for the SR-A1 protein is 381 bp and it contains the cleavage sites for the restriction endonucleases *Nde*I and *Bam*HI. Fig. 2 shows the 1.5% agarose gel of the cDNA fragment encoding the CTD of the only clone from the five that were tested, having the correct orientation. The vector with the insert was transformed into *E. coli* BL21-AI cells for large-scale expression.

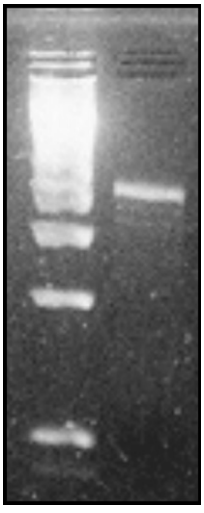

Fig. 2. The cDNA of C-terminal domain (aa 1187–1312) of SR-A1 protein after ligation in 1.5% agarose gel. This clone was transformed in *E. coli* BL21-AI cells for large-scale expression.

Expression of SR-A1 protein in LB^{amp} medium

For the expression of SR-A1 protein the fermentation of *E. coli* cells was performed at 37 °C. Fig. 3 shows the dependence of SR-A1 production on induction time. It can be seen that the maximum amount of protein is produced after 6 h of induction with L-arabinose 20%. The cells were harvested by centrifugation and resuspended in 25 ml binding buffer. The insoluble pellet was removed by ultracentrifuge and the soluble phase was added to the HiTrap chelating column for the isolation of SR-A1 protein. Fig. 4 shows the expression of the protein in the soluble lysate.

Purification of SR-A1 protein using HiTrap chelating column

In this study, the 10× His-tag was introduced in the N-terminal of SR-A1 protein in order to be able to purify the protein using a zinc metal-affinity resin column. The protein was collected in an elution buffer (20 mM

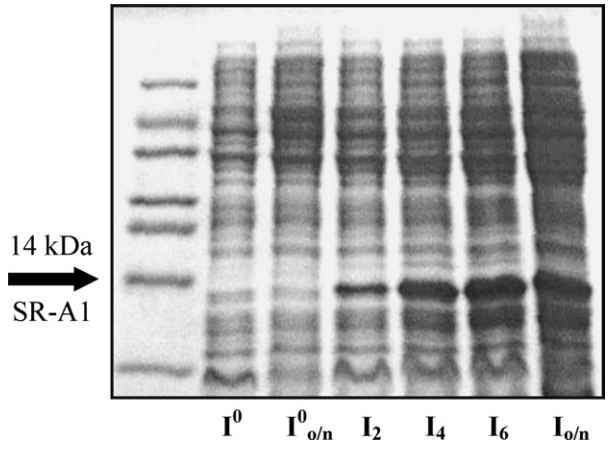

Fig. 3. Expression of C-terminal domain of SR-A1 protein in *E. coli* cells. SDS–PAGE 17% of whole-cell lysates at different induction times. I^0 and $I^0_{o/n}$ represent lysates derived from no induction and no induction left overnight at 37 °C, respectively, I_2, I_4, I_6, $I_{o/n}$ represent 2, 4, 6, and overnight induction times. The first lane represents molecular mass standards.

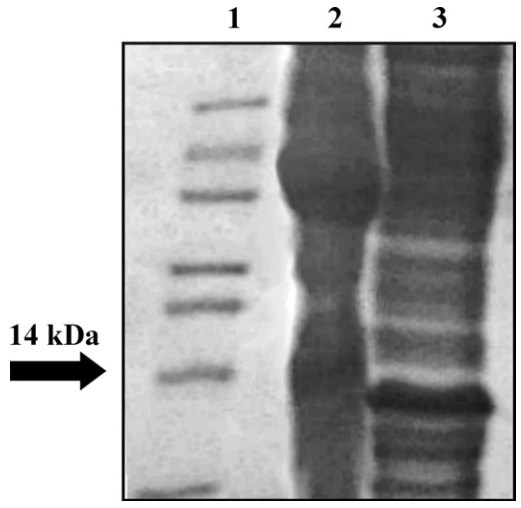

Fig. 4. Coomassie blue-stained SDS–PAGE of the human SR-A1 protein before purification as expressed in the soluble lysate. Lane 1, molecular mass standards; lane 2 bacterial pellet after sonication; and lane 3 crude lysate before purification.

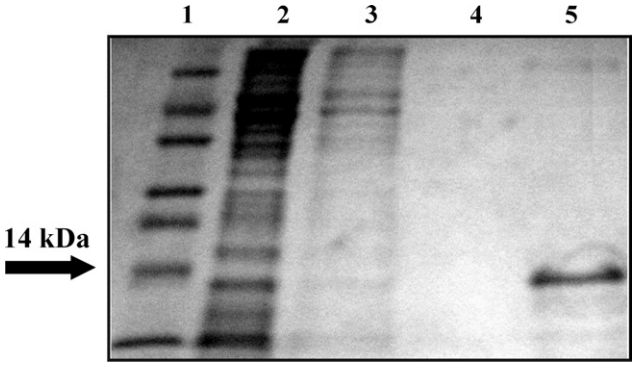

Fig. 5. SDS–PAGE 17% of purified SR-A1 protein using HiTrap chelating column. Lane 1, molecular mass standards; lane 2, protein lysate flowthrough from column; lane 3, wash buffer 1; lane 4, wash buffer 2; lane 5, SR-A1 in elution buffer.

NaH_2PO_4, pH 8.0, 0.9 M NaCl, and 100 mM EDTA). The purity of the protein was examined by SDS–PAGE and one single band corresponding to the 14 kDa protein was observed with >95% purity (Fig. 5). The amount of the protein produced was calculated using a spectrophotometer. It was 32 mg of purified human SR-A1 protein from 2 L of culture. The His-tagged protein can be stored at 4 °C in the elution buffer for more than one week without degradation as indicated by SDS–PAGE electrophoresis.

Conformational study of CTD of SR-A1 protein

Circular dichroism spectroscopy was used to examine the conformational properties of the expressed domain of SR-A1 protein. The CD spectrum of the protein in sodium phosphate buffer at pH 8.0 (Fig. 6) reveals that this domain contains a high amount of α-helix. The program CDPro [39,40] was used to analyze the CD data, in order to calculate the amounts of secondary structural elements comprising this domain. The percentage of α-helix is calculated to be 60%. On the other hand, the percentage of β-sheet was calculated from the CD data to be 17%. The structural content in turn is 5% and random coil 18%. These results give a primary indication of the structure of the expressed domain. It seems that this part of the protein is α-helical.

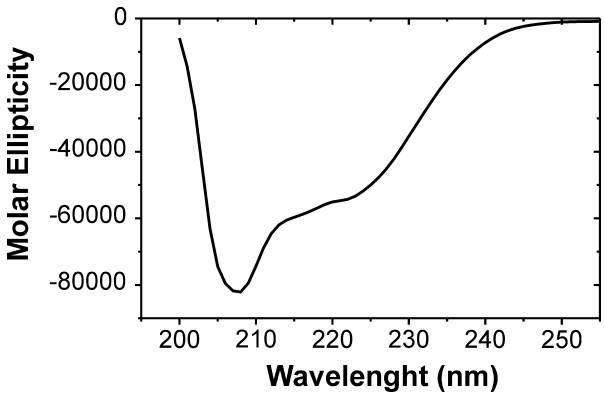

Fig. 6. CD spectrum of the C-terminal domain of SR-A1 in sodium phosphate buffer 20 mM, pH 8.0, and 0.1 M NaCl at 20 °C.

SR-A1–CTD direct interaction

The carboxy-terminal domain of human SR-A1 protein, containing the conserved domain that interacts with CTD, was expressed as a His-tagged protein and binding to pol II was assayed. There are two forms of RNA pol II in vivo, designated IIO, which is extensively phoshorylated at the CTD, and IIA, which is not phosphorylated. *E. coli* RNA pol II–CTD was purchased phosphorylated, according to the manufacturer. Fig. 7A demonstrates that the SR-A1 [1187–1312 His10] protein binds to *E. coli* pol II, since both proteins are retained in the cobalt-chelating column and are eluted together using a buffer with high concentration of imidazole. A control experiment was also performed using only SR-A1 protein (Fig. 7B).

Molecular modeling of SR-A1–CTD complex

Despite its relatively homogenous primary structure, the CTD is able to interact with many proteins that display no obvious structural relatedness. The most appealing hypothesis is that the CTD is structurally disordered in solution, but takes on defined structures that are templated by its particular binding partner. Different secondary structures were observed for the CTD bound to peptidyl-proline isomerase Pin1 [41] versus RNA guanylyl-transferase Cgt1 [42]. In the first case, the single consensus CTD repeat phosphorylated at Ser2 and Ser5 assumes an extended coil conformation, one that

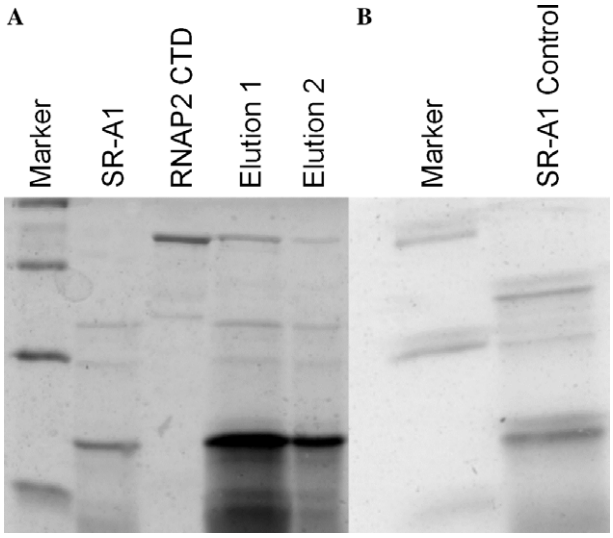

Fig. 7. Direct interaction between SR-A1 [1187–1312 His10] and RNA pol II. (A) Retention of RNA pol II on a SR-A1 CTD-binding domain column. The positions of the proteins before the assay are presented in the first two lanes SR-A1 and RNAP2 CTD. Elution 1 and elution 2 present the two proteins retained in the column and eluted using high concentration of imidazole. Bound proteins were eluted and electrophoresed on a SDS gel. (B) Control experiment using only SR-A1 protein: SR-A1 control: SR-A1 protein retained in the column.

resembles the conformation of a type II polyproline helix. In contrast, the 17 aa ordered segment within the Ser5 phosphorylated CTD 28-mer peptide bound to Cgt1 monomer B adopts a β-like extended conformation within two docking sites.

In the model of SR-A1 [1205–1277]/CTD complex, the phosphopeptide displays an "S-like" extended coil conformation with two turns at P_{3a} and Y_{1b} (Fig. 8A). The C-termini S_{7b} is anchored at the N-terminal domain of SR-A1 utilizing a hydrogen bond with Lys1206 and the N-termini tyrosine hydroxyl of Y_{1a} forms two hydrogen bonds with Ser1255 and Glu1257. The S_{5a}^{phos} phosphate is positioned within a positively charged pocket comprising His1248, His1252, and Lys1262, while the second S_{5b}^{phos} participates in hydrogen bonding interactions with the backbone nitrogen atoms of Asp1205, Tyr1207, Leu1208, and Thr1235. All three proline P_{3a}, P_{6a}, and P_{3b} residues utilize van der Waals contacts through their interaction with SR-A1, whereas P_{6b} is oriented away from the protein surface, likewise both threonine T_{4a} and T_{4b} residues. Y_{2a} is buried in a hydrophobic pocket and does not participate in electrostatic or hydrogen-bonding CTD interactions with SR-A1. Comparing our model structure with those of Pin1 [41] and Ctg1 [42] most CTD interactions diverge significantly and no conserved binding motif is exhibited. Apparently, the CTD is structurally plastic and can assume markedly different conformations depending on its binding partner.

Discussion

The purpose of the experiments described here was to demonstrate that the C-terminal domain of a newly identified protein, SR-A1, could be produced via recombinant DNA techniques, in a properly folded conformation and in quantities sufficient for further analysis by biophysical techniques. Using standard expression and purification conditions, overexpression of the hexahistidine-tagged construct in *E. coli* BL21-AI rapidly yielded ~15 mg of pure SR-A1 [1187–1312 His10] recombinant

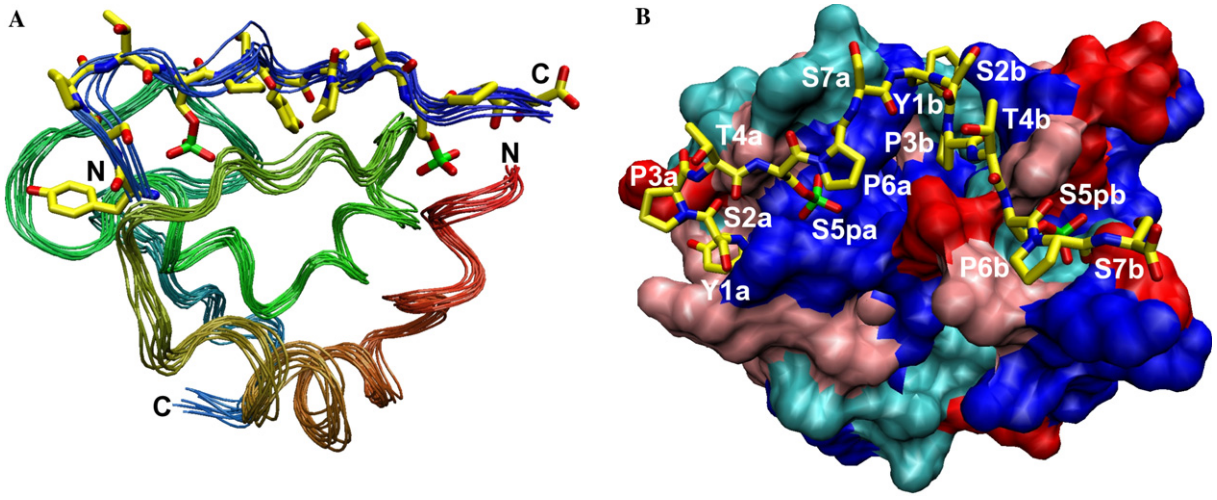

Fig. 8. Models of SR-A1 [1205–1277]/CTD complex. (A) Ribbon diagram of an ensemble of 10 structures from the last 10 ps of MD simulation. The CTD peptide (blue ribbons) is also shown by solid bond representation with carbons colored yellow and phosphorus green. The N- and C-termini of SR-A1 and CTD are indicated. (B) Surface view of the complex with CTD depicted as in (A). Basic residues of SR-A1 are colored blue, acidic are red, hydrophobic are cyan and polar are pink. Images generated using VMD v1.8 [46]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this paper.)

protein per liter of culture. The gene that encodes for the expressed protein, namely SR-A1 gene, was recently identified by our associates [7] and since little is known about SR-A1 protein, it was very important that one of its domain could be expressed, using standard DNA recombinant techniques. Previous bioinformatic techniques revealed that there were some regions in the full SR-A1 sequence that could exhibit well-known functions of the proteins that are members of the SR proteins [7]. The CTD-binding domain of ratA1, ratA9, and SRrp129 protein is the sequence corresponding to residues 1209–1275 in SR-A1 protein. Also, two areas with negatively polyglutamic acid (E) stretches that are present in rat A1 but not in other known SR proteins are also present in the expressed protein. In addition, an Arg/Asp-rich motif (RD) is present in the new protein. This motif is also present in a number of RNA-binding proteins such as the U1-70K [43] the RD RNA-binding protein [44] and the 68 kDa human pre-mRNA cleavage factor Im [45]. In addition to the above described motifs, this protein is also proline-rich and contains a large number of PAP trinucleotides (Fig. 1).

Among the well-known motifs present in the majority of the proteins belonging to the SR protein family, the expressed domain of the newly identified protein contains the CTD-binding domain of RNA polymerase II and the proline-rich domain. At the beginning, we attempted to express only the CTD-binding domain, using the same recombinant technique. The domain was expressed in *E. coli* using a different vector, but the yield was poor (data not shown). Spectroscopic studies of the CTD of the new SR-A1 [1187–1312 His10] revealed that the domain is predominately α-helical (60%).

The newly expressed protein was used in a protein–protein interaction assay, in order to determine whether it binds to the CTD RNAP II, like other known SR proteins [8]. Taking into consideration that the carboxy-terminal domain of human SR-A1 protein was expressed as a polyhistidine-tagged protein, we employed a pull-down protein–protein interaction kit and assayed the expressed protein with *E. coli* CTD RNAP II, which was purchased for that matter. The assay indicated that indeed the expressed carboxy-terminal domain of human SR-A1 protein binds to RNAP II.

The CTD of SR-A1 protein appears to interact with the CTD via a CTD binding region that is very similar to that described for rA1 and rA9, two SR proteins isolated by virtue of their CTD interaction [8]. CTD-binding proteins may act to directly link the nascent transcript to the CTD or they may interact with other SR proteins that bind the transcript independent of the CTD. Functional studies of CTD-binding proteins will be important in determining the role of the CTD in linking the processes of transcription and splicing.

This study represents the expression of a domain of the human SR-A1 protein for the first time. The protein is stable in solution and it binds to human CTD RNA polymerase II. Further biological and structural experiments will reveal more about this new protein, which is also clinically important, since it is reported to alternate mRNA processing in aggressive ovarian cancers.

Acknowledgments

M.K. thank Prof. Ivano Bertini for his hospitality at CERM (University of Florence). The knowledge of protein expression gained at CERM was critical for the fulfillment of this work. We also thank the Institute of Radioisotopes and Radiodiagnostic Products, NCSR "Demokritos" and especially Persefoni Klimentzou. This work was supported by the National Secretariat of Research and Technology (NSRT) in Greece, Grant PENED 2001 01ED625.

References

[1] X.D. Fu, The superfamily of arginine/serine-rich splicing factors, RNA 1 (1995) 663–680.
[2] B.R. Graveley, Sorting out the complexity of SR protein functions, RNA 6 (2000) 1197–1211.
[3] A.M. Zahler, K.M. Neugebauer, J.A. Stolk, M.B. Roth, Human SR proteins and isolation of a cDNA encoding SRp75, Mol. Cell. Biol. 13 (1993) 4023–4028.
[4] Y. Huang, R. Gattoni, J. Stevenin, J.A. Steitz, SR splicing factors serve as adapter proteins for TAP-dependent mRNA export, Mol. Cell 11 (2003) 837–843.
[5] E. Stickeler, F. Kittrell, D. Medina, S.M. Berget, Stage-specific changes in SR splicing factors and alternative splicing in mammary tumorigenesis, Oncogene 18 (1999) 3574–3582.
[6] C. Gobert, L. Bracco, F. Rossi, M. Olivier, J. Tazi, F. Lavelle, A.K. Larsen, J.F. Riou, Modulation of DNA topoisomerase I activity by p53, Biochemistry 35 (1996) 5778–5786.
[7] A. Scorilas, L. Kyriakopoulou, D. Katsaros, E.P. Diamandis, Cloning of a gene (SR-A1), encoding for a new member of the human Ser/Arg-rich family of pre-mRNA splicing factors: over-expression in aggressive ovarian cancer, Br. J. Cancer 85 (2001) 190–198.
[8] A. Yuryev, M. Patturajan, Y. Litingtung, R.V. Joshi, C. Gentile, M. Gebara, J.L. Corden, The C-terminal domain of the largest subunit of RNA polymerase II interacts with a novel set of serine/arginine-rich proteins, Proc. Natl. Acad. Sci. USA 93 (1996) 6975–6980.
[9] L.A. Allison, J.K. Wong, V.D. Fitzpatrick, M. Moyle, C.J. Ingles, The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function, Mol. Cell. Biol. 8 (1988) 321–329.
[10] E. Barron-Casella, J.L. Corden, Conservation of the mammalian RNA polymerase II largest-subunit C-terminal domain, J. Mol. Evol. 35 (1992) 405–410.
[11] L.A. Allison, M. Moyle, M. Shales, C.J. Ingles, Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases, Cell 42 (1985) 599–610.
[12] G.D. Meredith, W.H. Chang, Y. Li, D.A. Bushnell, S.A. Darst, R.D. Kornberg, The C-terminal domain revealed in the structure of RNA polymerase II, J. Mol. Biol. 258 (1996) 413–419.

[13] J.L. Corden, Tails of RNA polymerase II, Trends Biochem. Sci. 15 (1990) 383–387.
[14] M.E. Dahmus, Reversible phosphorylation of the C-terminal domain of RNA polymerase II, J. Biol. Chem. 271 (1996) 19009–19012.
[15] M. Carlson, Genetics of transcriptional regulation in yeast: connections to the RNA polymerase II CTD, Annu. Rev. Cell. Dev. Biol. 13 (1997) 1–23.
[16] D. Bentley, Coupling RNA polymerase II transcription with pre-mRNA processing, Curr. Opin. Cell. Biol. 11 (1999) 347–351.
[17] A.L. Greenleaf, Positive patches and negative noodles: linking RNA processing to transcription? Trends Biochem. Sci. 18 (1993) 117–119.
[18] J.L. Corden, M. Patturajan, A CTD function linking transcription to splicing, Trends Biochem. Sci. 22 (1997) 413–416.
[19] A.L. Greenleaf, A positive addition to a negative tail's tale, Proc. Natl. Acad. Sci. USA 90 (1993) 10896–10897.
[20] K.M. Neugebauer, M.B. Roth, Transcription units as RNA processing units, Genes Dev. 11 (1997) 3279–3285.
[21] S. Shuman, Origins of mRNA identity: capping enzymes bind to the phosphorylated C-terminal domain of RNA polymerase II, Proc. Natl. Acad. Sci. USA 94 (1997) 12758–12760.
[22] E.J. Steinmetz, Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog? Cell 89 (1997) 491–494.
[23] B. Chabot, S. Bisotto, M. Vincent, The nuclear matrix phosphoprotein p255 associates with splicing complexes as part of the [U4/U6.U5] tri-snRNP particle, Nucleic Acids Res. 23 (1995) 3206–3213.
[24] E.J. Cho, T. Takagi, C.R. Moore, S. Buratowski, mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain, Genes Dev. 11 (1997) 3319–3326.
[25] C.K. Ho, V. Sriskanda, S. McCracken, D. Bentley, B. Schwer, S. Shuman, The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II, J. Biol. Chem. 273 (1998) 9577–9585.
[26] E. Kim, L. Du, D.B. Bregman, S.L. Warren, Splicing factors associate with hyperphosphorylated RNA polymerase II in the absence of pre-mRNA, J. Cell Biol. 136 (1997) 19–28.
[27] S. McCracken, N. Fong, E. Rosonina, K. Yankulov, G. Brothers, D. Siderovski, A. Hessel, S. Foster, S. Shuman, D.L. Bentley, 5′-Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated carboxy-terminal domain of RNA polymerase II, Genes Dev. 11 (1997) 3306–3318.
[28] M.J. Mortillaro, B.J. Blencowe, X. Wei, H. Nakayasu, L. Du, S.L. Warren, P.A. Sharp, R. Berezney, A hyperphosphorylated form of the large subunit of RNA polymerase II is associated with splicing complexes and the nuclear matrix, Proc. Natl. Acad. Sci. USA 93 (1996) 8253–8257.
[29] M. Patturajan, X. Wei, R. Berezney, J.L. Corden, A nuclear matrix protein interacts with the phosphorylated C-terminal domain of RNA polymerase II, Mol. Cell. Biol. 18 (1998) 2406–2415.
[30] Z. Yue, E. Maldonado, R. Pillutla, H. Cho, D. Reinberg, A.J. Shatkin, Mammalian capping enzyme complements mutant *Saccharomyces cerevisiae* lacking mRNA guanylyltransferase and selectively binds the elongating form of RNA polymerase II, Proc. Natl. Acad. Sci. USA 94 (1997) 12898–12903.
[31] J.P. Bourquin, I. Stagljar, P. Meier, P. Moosmann, J. Silke, T. Baechi, O. Georgiev, W. Schaffner, A serine/arginine-rich nuclear matrix cyclophilin interacts with the C-terminal domain of RNA polymerase II, Nucleic Acids Res. 25 (1997) 2055–2061.
[32] D.P. Morris, A.L. Greenleaf, The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II, J. Biol. Chem. 275 (2000) 39935–39943.
[33] K. Nilsson, C. Sundstrom, Establishment and characteristics of two unique cell lines from patients with lymphosarcoma, Int. J. Cancer 13 (1974) 808–823.
[34] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410.
[35] M.A. Marti-Renom, A.C. Stuart, A. Fiser, R. Sanchez, F. Melo, A. Sali, Comparative protein structure modeling of genes and genomes, Annu. Rev. Biophys. Biomol. Struct. 29 (2000) 291–325.
[36] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules, J. Am. Chem. Soc. 117 (1995) 5179–5197.
[37] A. Onufriev, D. Bashford, D.A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized born model, Proteins 55 (2004) 383–394.
[38] G.M. Morris, D.S. Goodsell, R.S. Halliday, R. Huey, W.E. Hart, R.K. Belew, A.J. Olson, Automated docking using a Lamarckian genetic algorithm and and empirical binding free energy function, J. Comput. Chem. 19 (1998) 1639–1662.
[39] N. Sreerama, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set, Anal. Biochem. 287 (2000) 252–260.
[40] W.C. Johnson, Analyzing protein circular dichroism spectra for accurate secondary structures, Proteins 35 (1999) 307–312.
[41] M.A. Verdecia, M.E. Bowman, K.P. Lu, T. Hunter, J.P. Noel, Structural basis for phosphoserine-proline recognition by group IV WW domains, Nat. Struct. Biol. 7 (2000) 639–643.
[42] C. Fabrega, V. Shen, S. Shuman, C.D. Lima, Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II, Mol. Cell 11 (2003) 1549–1561.
[43] R.A. Spritz, K. Strunk, C.S. Surowy, S.O. Hoch, D.E. Barton, U. Francke, The human U1-70K snRNP protein: cDNA cloning, chromosomal localization, expression, alternative splicing and RNA-binding, Nucleic Acids Res. 15 (1987) 10373–10391.
[44] C.S. Surowy, G. Hoganson, J. Gosink, K. Strunk, R.A. Spritz, The human RD protein is closely related to nuclear RNA-binding proteins and has been highly conserved, Gene 90 (1990) 299–302.
[45] U. Ruegsegger, D. Blank, W. Keller, Human pre-mRNA cleavage factor Im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits, Mol. Cell 1 (1998) 243–253.
[46] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38, pp. 27–28.