2990–2997 *Nucleic Acids Research, 1996, Vol. 24, No. 15*

Molecular cloning of a RNA binding protein, S1-1

Akira Inoue*, Kenichi Paulo Takahashi[2], Masatsugu Kimura[1], Takanori Watanabe and Seiji Morisawa[3]

Department of Biochemistry and [1]Laboratory of Biophysics, Osaka City University Medical School, 1-4-54 Asahimachi, Abeno-ku, Osaka 545, Japan, [2]Department of Anatomy and Physiology, Osaka Prefectural College of Health Sciences, 3-7-30 Habikino, Habikino, Osaka 583, Japan and [3]Seishin School of Nursing, 78-53 Kande-cho, Yoshinari, Nishi-ku, Kobe 651-23, Japan

Received March 28, 1996; Revised and Accepted June 17, 1996

DDBJ accession no. D83948

ABSTRACT

S1 proteins A–D constitute a nuclear protein family that are liberated rapidly in a set from chromatin by mild digestion with a DNA or RNA hydrolyzing enzyme. With an anti-S1-protein B antiserum that reacted with B2, C1 and D1, a cDNA clone, pS1-1, was obtained, which encoded a protein of 852 amino acids. The S1-1 protein, encoded within the cells by a mRNA of 3480 nt, was a novel protein and could be distinguished from the S1 proteins B, C and D by their amino acid sequences. The S1-1 protein synthesized by *in vitro* translation bound to RNA homopolymers, with a preference for G and U polyribonucleotides and little for poly(A). The protein contained two tandem RNP motifs and several intriguing sequences, such as a novel repeat of five octamers with a consensus sequence DP-S(Q/G)YYY and a potentially perfect amphipathic α-helix of five turns with basic and acidic amino acids positioned in an ordered way. The two RNP motif sequences were similar, although homologies were low, to the RNP motif sequences of yeast NSR1 protein, animal nucleolins, *Drosophila* hnRNP A1 and tobacco chloroplast RNP precursor protein, suggesting a functional uniqueness of the S1-1 protein in RNA metabolism and also the evolution of its RNP motif structure before plants and animals diverged. These results indicate that the S1-1 protein encoded by the cDNA is a new class of RNA binding protein.

INTRODUCTION

A group of proteins that are functionally related can often be extracted in a set under particular conditions and such a property helps further characterization of the proteins. Examples of such nuclear proteins are histones, HMG proteins and nuclear lamina pore complexes: they are isolated by extraction from nuclei with dilute mineral acids (for a review see 1), with 0.35 M NaCl and then 2% trichloroacetic acid (for a review see 2), and with Triton X-100 and high salt solutions as an insoluble residue from the nuclei digested with DNase I (3,4), respectively.

S1 proteins constitute another such group of nuclear proteins. They are extracted selectively at pH 4.9 from the supernatant of nuclei treated mildly with DNase or RNase (5–7). The S1 proteins are composed of proteins A, B, C and D, each separable into doublets by SDS–PAGE: A1 (80.0 kDa) and A2 (76.1); B1 (49.5) and B2 (48.2); C1 (45.2) and C2 (44.5); D1 (41.5) and D2 (39.5). They are liberated from nuclei with closely similar kinetics on DNase I digestion, suggesting that the S1 proteins are present in the same or very similar sites in the nucleus. They have been found in all rat tissues examined so far (5) and in mammals, a bird (chicken), a fish (carp), an amphibian (frog; unpublished data) and an echinoderm (starfish; 8). Polyclonal antibodies raised in two rabbits with protein B as immunogen both reacted with proteins B2, C1 and D1 (9,10). With these antisera, the S1 proteins were localized in the extranucleolar nucleoplasm, in the euchromatin bordering the heterochromatic areas (9), where most RNA is synthesized (for a review see 11). The S1 proteins constitute a family, as shown by their shared epitopes and primary structures: in addition to the fact that the polyclonal antibodies reacted with B2, C1 and D1, an anti-S1 protein monoclonal antibody produced by a cloned hybridoma reacted with C2 and D2, and all of these proteins have identical or very similar amino acid sequences, which have been partially analyzed (details to be published elsewhere). In addition, it was recently found that they occur in association with hnRNA in the cell nucleus; for example they sediment in association with RNase-sensitive complexes of heterogeneous sizes with *S* values up to 200 or more (details to be published elsewhere).

To understand their structures and functions, we undertook molecular cloning using the polyclonal antiserum and obtained a cDNA clone, pS1-1, from rat liver cDNA libraries. It was shown that while the S1-1 protein had similarities to the S1 proteins in immunoreactivity and in having RNA binding activity, it was an RNA binding protein not previously reported.

MATERIALS AND METHODS

Isolation of RNA

Total RNA was prepared from rat liver by the method of Chomczynski and Sacchi (12) and dissolved in DEPC-treated H_2O. Poly(A)$^+$ RNA was isolated from total RNA with

* To whom correspondence should be addressed

oligo(dT)-conjugated ceramic beads (Oligotex-dT30 Super; Roche, Japan).

Construction and screening of cDNA libraries

Double-stranded cDNA was synthesized from poly(A)$^+$ RNA, with random hexamers or oligo(dT) as a primer in the reverse transcription reaction, and cDNAs were inserted into the *Eco*RI site of λgt 11 or between the *Eco*RI and *Xho*I sites of the λZAP-II vector, with a kit from Amersham (λgt 11 cloning system RPN 1763) or a kit from Stratagene (ZAP-cDNA synthesis kit 200400, Gigapack II packaging extract no. 200214). Packaged recombinant phages were screened by the standard method (13) with an anti-S1 protein antiserum raised in a rabbit (10) or ^{32}P-labeled cDNA fragments. Color was developed in the immunoscreening with a horseradish peroxidase-conjugated goat anti-rabbit IgG secondary antibody and 3-amino-9-ethylcarbazole/H_2O_2. The primary and secondary antibodies for immunoscreening were pretreated with an extract from *Escherichia coli*. In DNA hybridization, blots on nylon membranes were hybridized at 65°C for 20 h in 5× SSPE (0.9 M NaCl, 50 mM sodium phosphate and 5 mM EDTA, pH 7.7) containing 5× Denhardt's solution, 1% SDS and 20 μg/ml sonicated and heat-denatured salmon sperm DNA and then washed. Final washes were in 0.1× SSPE containing 0.1% SDS at 65°C for 10 min and the membranes were analyzed by autoradiography.

DNA labeling

cDNA inserts were released from vector arms by digestion with a restriction enzyme(s) and separated on low melting temperature agarose gels. DNAs in the melted gel slices were labeled with random hexanucleotide primers, [α-^{32}P]dCTP (3000–6000 Ci/mmol) and Klenow I fragment (Multiprime DNA labeling system RPN 1601Y; Amersham). Labeled DNAs were purified by reverse phase chromatography on Nensorb 20 cartridges (NLP-022; NEN).

Oligonucleotide synthesis

Oligonucleotides were synthesized on a DNA synthesizer (model 381A; Applied Biosystems), purified by extraction with phenol/chloroform or chromatography on OPC columns (Applied Biosystems) and used as primers in PCR or in DNA sequencing.

DNA sequencing

Sequencing of cDNAs subcloned into the pUC118 or pBluescript vectors was performed after progressive unidirectional deletion of the cDNA insert with a kit (Kilo-sequence deletion kit; Takara Shuzo Co., Kyoto) based on the methods of Henikoff (14) and Yanisch-Perron *et al*. (15) or by sequence extension with synthetic oligonucleotide primers. Sequencing reactions of double-stranded DNA were by the dideoxynucleotide chain termination method (16) with a Taq Dideoxy terminator cycle sequencing kit (Applied Biosystems) and the reaction products were analyzed on an Applied Biosystems model 370A DNA sequencer.

RT–PCR of S1-1 mRNA

RT–PCR was carried out with a kit (Gene Amp RNA PCR Kit N808-0017; Perkin Elmer) according to the manufacturer's protocol with the following modifications. The rat liver total RNA (0.5–0.8 μg) was heat denatured at 70°C for 3 min in 4 μl H_2O in the presence of the 3′ primer and kept at 55°C for 1 min. To this total RNA was then immediately added 16 μl of the reaction mixture, which contained all other ingredients and had been kept at the same temperature for 5 min. Reverse transcription was done, with the temperature being lowered gradually to 42°C over 15 min. The 5′ primer in the following PCR was designed to break a putative hairpin secondary structure of 39 nt (see Fig. 2 legend).

Sequence similarity search and structural analysis

Sequence similarity was searched for in the protein database from SWISS-PROT and the alignment was optimized as described by Dayhoff (17). Structural analysis was with the Genetyx system, version 7.3 (Software Development Co. Ltd, Tokyo): protein secondary structures were predicted based on the methods of Chou and Fasman (18,19) and of Robson (20), the wheel models on that of Schiffer and Edmundson (21) and hydropathy profiles on Hopp and Woods (22), averaging each hydropathy value for three successive amino acid residues.

In vitro transcription and translation

The pBluescript plasmid containing S1-1 cDNA (pS1-1) was linearized with *Kpn*I, and S1-1 protein was synthesized by *in vitro* transcription with T3 RNA polymerase followed by translation in reticulocyte lysate with kits from Stratagene (mCAP mRNA Capping Kit and *In Vitro* Express Translation Kit). For good translation, the RNA transcript (8 μg) was incubated in H_2O (15 μl) at 68°C for 45 s and immediately mixed with 100 μl nuclease-treated rabbit reticulocyte lysate (Wako Pure Chemicals, Osaka) and l0 μl [^{35}S]methionine (100 μCi, 1000 Ci/mmol). The mixture was incubated at 37°C for 1 h.

RNA binding assays

The binding procedure was essentially the same as described by Siomi *et al*. (23). The [^{35}S]S1-1 protein produced by *in vitro* translation was incubated at 4°C for 15 min by gentle rocking with RNA homopolymers bound to agarose matrix [poly(A), poly(U) and poly(C); Sigma] or polyacrylhydrazido–agarose matrix [poly(U) and poly(G); Sigma] in a total of 0.5 ml binding buffer (10 mM Tris–HCl, pH 7.4, 2.5 mM $MgCl_2$ and 0.5% Triton X-100). The beads were pelleted by a brief spin at 4°C and washed five times with cold binding buffer before resuspension in 20 μl SDS–PAGE loading buffer at 100°C for 10 min. SDS–PAGE (7.0%) of the centrifuged supernatant was as described by Laemmli (24) and proteins were visualized by fluorography.

RESULTS

Cloning of S1-1 cDNA

A λgt11 library of rat liver cDNAs, synthesized with random hexanucleotides as primers, was screened with a polyclonal anti-S1 protein B antibody. A clone with an insert of 317 bp (clone 12) was isolated (Fig. 1). By rescreening with the ^{32}P-labeled cDNA, clone 32 was obtained. The clone 32 cDNA was 2012 bp, with an open reading frame (ORF) of 1668 nt and a start Met codon at nt 345. It did not, however, have a stop codon and the cDNAs of most other clones were shorter than sequence 12. A library was reconstructed in the ZAPII-λ phage vector, this time using oligo(dT) as primer. Its five clones, isolated with sequence 12 as probe, contained the cDNA inserts with a polyadenylation

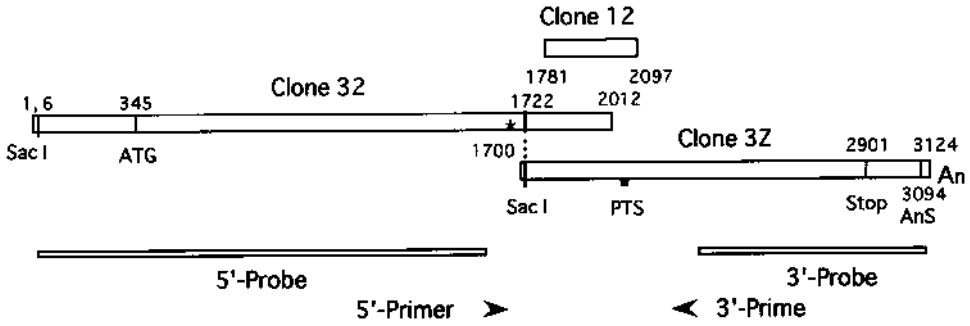

Figure 1. Cloning of S1-1 cDNA. From a λgt11 library of rat liver cDNAs, clone 12 was isolated with an anti-S1 protein antiserum; with ^{32}P-labelled sequence 12 as probe, the clones 32 and 3Z were isolated. The 32 and 3Z sequences overlapped over 313 bp. ATG, start Met codon of the ORF; Stop, TGA stop codon at nt 2901; Ans, polyadenylation signal (AAUAAA) located 25 nt upstream from a poly(A) addition site; *Sac*I, *Sac*I sites; PTS, sites of premature termination in reverse transcription; asterisk, a putative 39 nt hairpin region that inhibited RT–PCR. Probes and primers indicated were those used in the Figure 2 experiments.

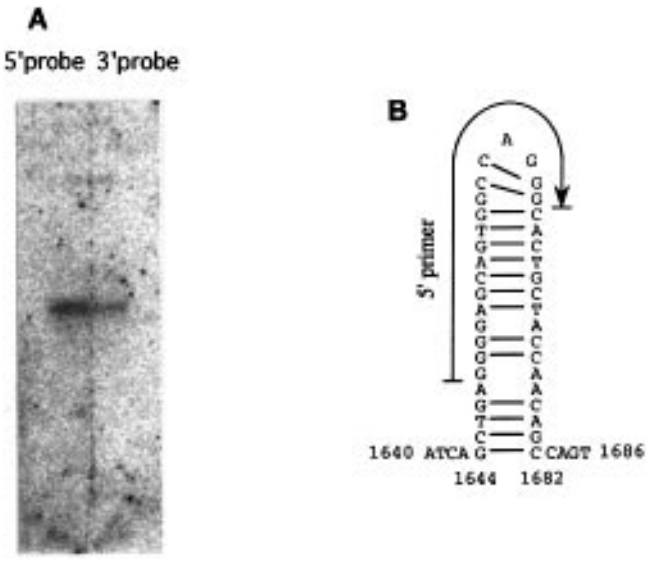

Figure 2. Occurrence of the S1-1 sequence in liver. (**a**) Northern analysis. Total RNA from rat liver was electrophoresed and blotted to a nylon membrane. The blot was cut longitudinally in the middle of the sample lane and each strip was hybridized with 5′ (nt 6–1600 of the S1-1 sequence) and 3′ (nt 2284–3124) probes, each representing non-overlapping regions of the 32 and 3Z sequences, respectively (Fig. 1). The washed membranes were placed side by side and analyzed on an image analyzer. (**b**) 5′ Primer in RT–PCR. The 5′ (nt 1649–1666) and 3′ (nt 2151–2129) primers interposing the overlapping region of the 32 and 3Z sequences (Fig. 1) were used in RT–PCR. For a successful RT–PCR, the 5′ primer was designed to break a putative 39 nt hairpin structure at 1644–1682 (indicated by an asterisk in Fig. 1). This hairpin was estimated to have a minimum free energy of –24.4 kcal/mol by the Genetyx algorithm.

signal, AAUAAA, a poly(A) tail and identical upstream sequences. The 5′-ends of these clones started, however, at 11–33 nt downstream of sequence 32, except for clone 3Z, indicating premature termination in reverse transcription. Sequence 3Z (1423 bp) overlapped the 32 sequence, perfectly over 313 nt. The composite complete coding sequence was named S1-1 (Fig. 1).

Construction of complete S1-1 cDNA

The overlapping region had a single *Sac*I site (Fig. 1). The subclones pAy32 and pAt3, containing sequences 32 and 3Z in pUC118 and pBluescript respectively, were digested with *Sac*I restriction enzyme. A 1717 bp fragment from the former clone, corresponding to the 5′ two thirds of the S1-1 sequence, and the remaining one third linked at its 3′-end to the pBluescript plasmid vector from the latter were ligated and a clone, pS1-1, containing a complete S1-1 coding sequence was isolated. The cDNA construct was confirmed by the insert size, by digestion with various restriction enzymes, which yielded fragments at expected molar ratios, and by sequencing, which confirmed the correct orientation of the ligated fragments.

Occurrence of S1-1 mRNA in rat liver cells

Northern analysis of total rat liver RNA demonstrated the presence of S1-1 mRNA. It was 3480 nt long (Fig. 2a).

Northern analysis also verified the genuineness of the composite S1-1 sequence: the 5′ and 3′ probes, which interposed the overlapping region of the 32 and 3Z sequences (Fig. 1), gave images of the same molecular size, and the band intensity obtained with the 5′ probe was ~2-fold, as expected (this probe was 2-fold larger in size than and had almost the same specific radioactivity as the 3′ probe).

RT–PCR of total rat liver RNA and sequencing of its products also confirmed the S1-1 sequence (not shown). RT–PCR of the interposing overlapping region amplified an expected 503 bp fragment (Fig. 1). To obtain successful amplification, however, the 5′ primer was specifically designed to break a putative hairpin structure of 39 nt located immediately upstream of the overlapping region (Fig. 2b).

Thus, it was concluded that the S1-1 mRNA represented a true sequence present in the mRNA population in the liver cells. S1-1 cDNA encodes a novel protein of 852 amino acids, with a molecular mass of 94.3 kDa. It is composed of 14.4 and 13.3 mol% basic and acidic amino acids respectively.

Relationship between S1-1 protein and S1 proteins B, C and D

The amino acid sequences were compared. The S1 proteins B, C and D (all N-termini blocked) were digested with *Staphylococcus aureus* V8 protease and microsequencing of the peptides was performed by Edman degradation. The S1 proteins B2, C1 and D1 and proteins C2 and D2 constitute two groups respectively, as judged by very similar peptide maps on SDS gels and their similar amino acid sequences. In addition, the two groups are related: for example both groups had similar pentadecamer sequences, which

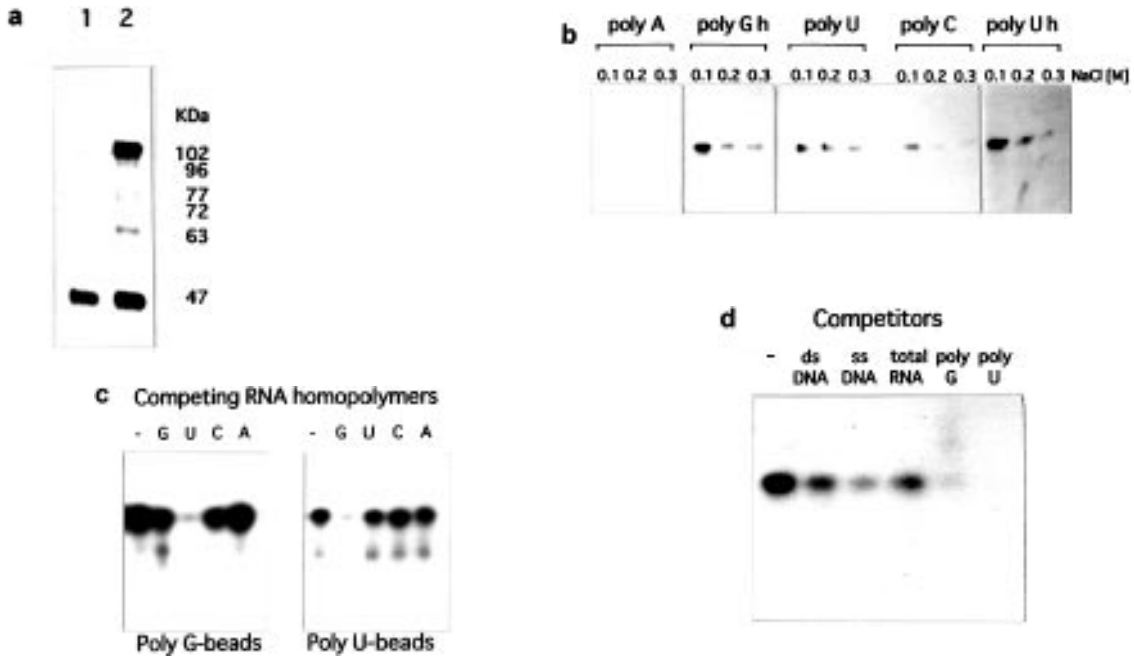

Figure 3. RNA binding properties of S1-1 protein. (**a**) *In vitro* synthesis of S1-1 protein. pS1-1, linearized with *Kpn*I, was transcribed with T3 RNA polymerase. The S1-1 mRNA (lane 2) or the control (lane 1; minus S1-1 mRNA) was translated in a rabbit reticulocyte lysate in the presence of [^{35}S]methionine (1000 Ci/mmol) and the reaction mixtures were analyzed by 7% SDS–PAGE. The numbers are molecular masses in kDa estimated with marker proteins. (**b**) RNA binding. The [^{35}S]S1-1 protein (4 µl of the *in vitro* translation mixture, 5×10^4 c.p.m.) was incubated with RNA homopolymer (25 µg) attached to a solid matrix at various NaCl concentrations and the bound proteins were analyzed by SDS–PAGE. RNA homopolymers were attached to agarose beads [poly(A), poly(U) and poly(C)] or polyacrylhydrazido–agarose beads [poly(Gh) and poly(Uh)]. (**c**) Competition with free RNA homopolymers. [^{35}S]S1-1 protein was incubated with poly(G) attached to polyacrylhydrazido–agarose beads (left panel) or poly(U)–agarose beads (right panel) in the presence or absence of a 2.5-fold excess (by weight) of competitor RNA homopolymers [poly(G), poly(U), poly(C) and poly(A)] at 0.1 M NaCl. Proteins bound to the beads were analyzed as above. (**d**) Competition with double-stranded and heat-denatured DNAs and with total RNA. Binding of [^{35}S]S1-1 protein to poly(G) attached to polyacrylhydrazido–agarose beads was competed at 0.1 M NaCl with a 6-fold excess of competitors. ds, rat liver double-stranded DNA; ss, heat-denatured (100°C for 7 min) single-stranded DNA; total RNA, rat liver total RNA.

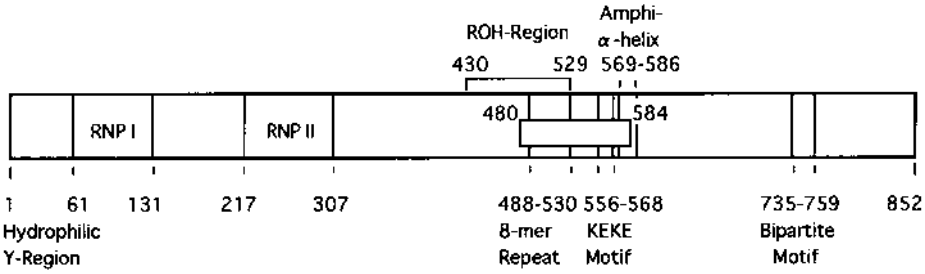

Figure 4. Structural features of S1-1 protein. From the 2556 nt ORF sequence, the amino acid sequence of the S1-1 protein was deduced. N-terminal hydrophilic Y region contains 75% hydrophilic amino acids as well as seven tyrosines. After this region, there are two RNP motif regions I and II. ROH region contains 42% hydroxyl amino acids (S, T and Y) and a novel tandem repeat of five octamers on its C-terminal side. Region T569–K586 is predicted to form a perfect amphipathicα-helix. The KEKE and the bipartite motifs may be the signals for protein association and nuclear translocation respectively. The inside rectangle (P480–L584) corresponds to an epitope region encoded by the cDNA that was initially isolated with the anti-S1 antibody. The C-terminal fifth of the molecule (K680–Q852) has a net positive charge of +14.

differed by only three amino acids in the middle and one at the C-terminus. The S1 proteins were shown to associate with hnRNAs in the cell nucleus (details to be published elsewhere).

The S1-1 protein did not, however, contain any sequences found in these S1 proteins. It is thus concluded that, while the S1-1 protein had similarities to the S1 proteins in immunoreactivity and in having RNA binding activity (as described below), it was a RNA binding protein distinct from the S1 proteins.

In vitro transcription and translation

The S1-1 protein had RNA binding activity. This was shown using [^{35}S]Met-labeled S1-1 protein synthesized from the pS1-1 cDNA by *in vitro* transcription and translation. The estimated molecular mass of the translation product was 102 kDa (Fig. 3a), 8 kDa larger than the theoretical value. Since S1-1 protein contains a number of potential phosphorylation sites (18 sites for

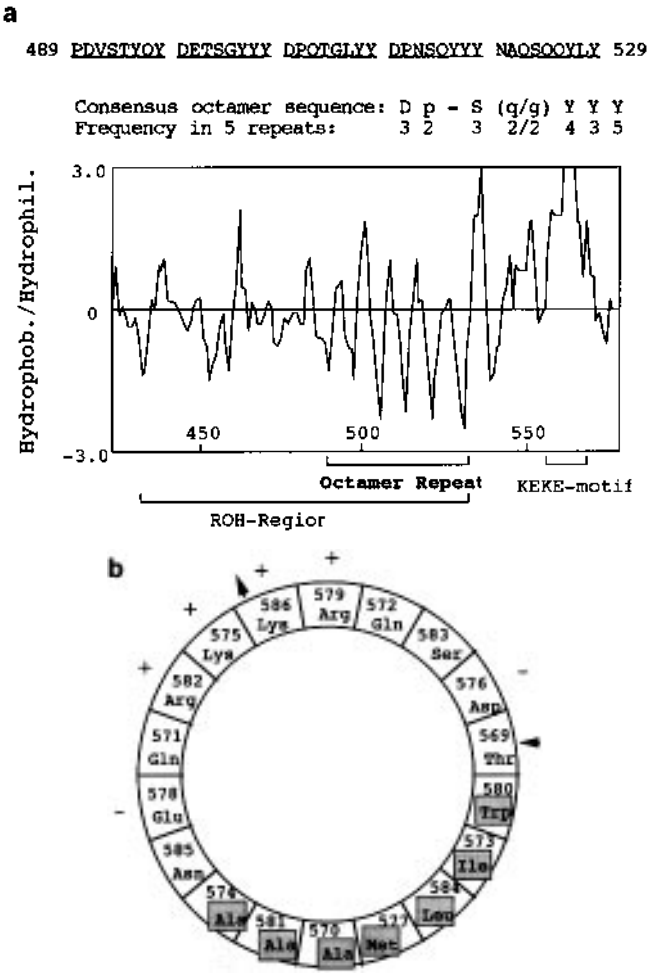

Figure 5. (**a**) Octamer repeat. (i) Consensus sequence. Below the repeat sequence of five octamers (P489–Y529), a consensus sequence is shown, where amino acids appearing more than three times are indicated in capital letters and twice in small letters. (ii) Hydropathy profile. The hydropathy profile shows that the octamer repeat has a path of five up-and-downs, corresponding to the five repeating units. (**b**) Amphipathic α-helix. The Edmundson wheel model showed a typical amphipathic structure in the region T569–K586. Its α-helical structure of five turns was predicted by the Chou–Fasman method. Except for the last two residues (N585 and K586), the Robson method also predicted an α-helical structure. Plus and minus indicate basic and acidic amino acids respectively and the arrows the start and end residues in the helical structure.

casein kinase II, five for C-kinase, two for tyrosine kinase and one for A-kinase), the larger molecular mass probably resulted from phosphorylation, which causes slower migration on SDS–PAGE. Truncated products at 77, 72 and 63 kDa are thought from the molecular weights to have been synthesized from the internal methionines at 127, 236 and 312 respectively. The 47 kDa protein was a product of an endogenous RNA in a rabbit reticulocyte lysate (Fig. 3a).

RNA binding assays

The [^{35}S]S1-1 protein bound various RNA homopolymers attached to solid matrix beads, strongly to poly(G) and poly(U), less to poly(C) and little to poly(A) (Fig. 3b). Binding was affected by increasing ionic strength and abolished at NaCl concentrations >0.4 M, suggesting that it involved an ionic interaction.

In the presence of competing free RNA homopolymers, binding of S1-1 protein to poly(U) beads was competed more strongly by free poly(G), while that to poly(G) beads by free poly(U) (Fig. 3c). The results not only suggest that S1-1 protein interacts with both poly(G) and poly(U), but also that binding of the S1-1 protein to RNA occurs by some unique mechanism.

DNA and total RNA also had competing activities, but their activities were low even at concentrations at which free poly(U) and poly(G) caused almost complete competition. Single-stranded DNA had a stronger activity than double-stranded DNA or total RNA (Fig. 3d).

Structural features

The S1-1 protein contained two RNP motif sequences, I and II (Fig. 4), at amino acids 61–131 and 217–307. Besides this, it had quite a few characteristic amino acid sequences. Some of them are as follows.

The central ROH region of 100 amino acids (Y430–Y529) is rich in hydroxyl amino acids, 42% of the residues being Ser, Thr and Tyr (Fig. 4). Besides this, it has a novel tandem repeat of five octamers in its C-terminal half, with a consensus sequence of DP-S(Q/G)YYY [Fig. 5a(i)]. Corresponding to the repeated structure, its hydrophilicity/hydrophobicity profile shows a characteristic path of five up-and-downs, as shown in Figure 5a(ii).

Region T569–S584 was predicted to form a perfect α-helix. It is remarkable that both the Chou–Fasman and Robson rules matched this α-helix structure perfectly. Moreover, an Edmundson wheel analysis indicated that this region (amino acids 569–586) consisted of a typical amphipathic α-helix of about five turns (Fig. 5b). This five turn helical structure was noticeable in having an orderly positioning of charged amino acids: basic amino acids are placed every three or four residues, clustered at the center of the hydrophilic surface, and two acidic amino acids (D576 and E578) occur in the middle of the helix on each side of the amphipathic boundary. It is expected that this amphipathic region is important for the S1-1 molecule in inter- and/or intramolecular interactions.

The N-terminal region of the molecule (amino acids 1–69, the hydrophilic Y region; Fig. 4) has 75% hydrophilic amino acids, including 34 charged amino acids (49%). This region also has a high density of tyrosine residues (seven out of 69 amino acids). Also, the C-terminal 173 amino acid region (K680–Q852), representing a fifth of the molecule, is noticeable for having a potential net positive charge of 14 (Fig. 4).

DISCUSSION

Cloned sequence S1-1

Upon screening of an expression library using a polyclonal anti-S1 protein antiserum, a positive clone was isolated. With this cDNA as probe, libraries were re-screened and a composite clone (pS1-1) containing a complete coding sequence was obtained. The genuineness of the sequence was verified by Northern blotting, RT–PCR and sequencing of the amplified product. The S1-1 cDNA coded a new protein of 852 amino acid residues with RNA binding activity.

The S1-1 protein was ~2-fold larger than the S1 proteins B, C and D. The possibility that the S1-1 protein is a precursor of these proteins is unlikely, because it did not match the amino acid sequences of the S1 proteins. On the other hand, the S1-1 protein had similarities to the S1 proteins in that it had RNA binding

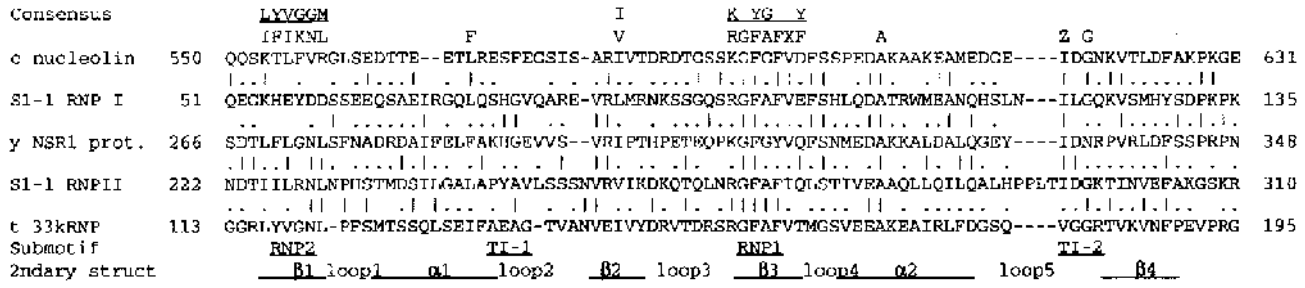

Figure 6. Sequence similarities of the RNP I and RNP II to known RNP motifs. The RNP motifs I and II of S1-1 protein were aligned with similar sequences found in chicken nucleolin, yeast NSR1 protein and tobacco chloroplast 33K RNP precursor protein respectively, according to the method of Dayhof, in the regions specified for 70 RNP proteins (25). The positions of start and end residues are given on the left and right of the alignments. Identical amino acids are indicated by vertical lines, similar ones with a Dayhoff score >0 by dots. Gaps are indicated by dashes. Core amino acid residues of the consensus RNP motif as well as the positions of the submotifs (RNP1 and RNP2) are listed above and below. The secondary structures in the RNP motifs (α-helices, β-sheets, loops and tight turns TI-1 and TI-2) have been characterized (40–42) and reviewed (25,26,43). These regions were cited for reference. Z = Ile, Ieu or Val.

activity and immunoreactivity to the anti-S1 protein antiserum. The latter indicates that the S1-1 protein has a common or closely similar epitope structure to those of the S1 proteins B2, C1 and D1. For these reasons we named the protein coded by the cDNA S1-1.

Premature termination sites in cDNA synthesis

A cDNA clone containing a complete coding sequence could not be isolated because of premature terminations occurring in the region of nt 1700–2100. When cDNA synthesis was primed with random primers, MuLV reverse transcriptase stopped around nt 1780 in most cases. Similarly, with an oligo(dT) primer, the enzyme stopped mostly at nt 2024–2046. Computer-assisted examination for intramolecular base pairings indicated that there were a number of possible long base pairings, particularly densely concentrated in the regions nt 1460–1825 and 1935–2078. In fact, *in vitro* translation of S1-1 mRNA and RT–PCR to amplify this region (nt 1700–2100) were difficult. To obtain efficient *in vitro* translation, the S1-1 mRNA had to be used immediately after heat treatment (if it was left longer than 5 min on ice it resulted in poor translation). Similarly, the RT–PCR conditions were the best ones that could be achieved among many attempts using various primers (Fig. 2). Whether or not there is a functional significance in these stable structures is intriguing.

RNP motif structures

In accord with the *in vitro* RNA binding activity, S1-1 protein had two RNP motifs (I and II, Fig. 4) at amino acids 61–131 and 217–306 respectively (Fig. 6). These RNP motif regions seemed to be responsible for the RNA binding activity of S1-1 protein, as truncated S1-1 proteins synthesized from internal methionines and lacking the RNP motifs did not usually show RNA binding activity (Fig. 3).

RNP sequences are generally composed of ~90 amino acids and contain two short submotif sequences, RNP1 and RNP2 (reviewed in 25,26). The S1-1 RNP I had a typical RNP1 submotif (RGFAFVEF), but lacked a RNP2 submotif on the N-terminal side. In fact, RNP2 is generally less strictly conserved (27) and many RNP proteins lack this submotif (25,28). On the other hand, the second RNP II was more similar to the canonical structure (26,29) and had the submotifs RNP1 (RGFAFIQL) and RNP2 (IILRNL).

The RNP motif I was similar to those of chicken nucleolin (30,31), yeast nuclear localization sequence binding protein NSR 1 (32), and *Drosophila* hnRNP A1 (33,34) and the RNP II to those of the yeast NSR 1 (32) and tobacco chloroplast RNP precursor proteins (35) (Table 1 and Fig. 6). Some of these proteins possess two tandem RNP motifs and others four (Table 1). Functionally similar RNP-containing proteins have similar domain organizations with respect to copy number of the motif and auxiliary domains (25,36). This suggests that the S1-1 protein represents a new class of RNA binding protein. The characteristic of the S1-1 protein RNP motifs is that they resemble both animal and plant RNPs. It is likely that this S1-1 RNP motif structure evolved in the early stage of evolution, before plants and animals appeared.

The range of sequence similarity between two unrelated RNP motifs has been estimated at 10–20% identity by Birney and others (25, figs 1 and 4 therein); in the same regions as those analyzed by them, the S1-1 RNP motif I (amino acids 51–135) showed only 29% identity to that of chicken nucleolin (amino acids 550–631), and RNP motif II (amino acids 222–310) 26% identity to that of yeast NSR 1 (amino acids 266–348) (Table 1). Interestingly, the identity between RNP I and II of the S1-1 protein was also low, at 25%. These rather low homologies suggest that the S1-1 protein should have a special function in RNA metabolism.

RNA binding properties

The S1-1 protein had RNA binding activity. It bound to poly(G) and poly(U) more strongly than to other RNA homopolymers. Accordingly, it is suggested that the S1-1 protein is a RNA binding protein that preferentially binds to regions rich in G and U ribonucleotides.

In the competition experiments, binding of S1-1 protein to poly(U) beads was competed strongly by free poly(G), and that to poly(G) beads by free poly(U). Possible duplex formation of poly(U) beads with free poly(G) or of poly(G) beads with free poly(U) could not explain these results, since free poly(A) or poly(C), which can form more stable duplexes with poly(U) or poly(G) beads, exerted only small inhibitory effects under the assay conditions used. These results imply that some unique mechanism operates in the binding of S1-1 protein to RNA. The binding could occur in a random manner, a concerted manner or a stepwise manner. The present results seem to favor a mechanism in which multiple RNA binding sites participate in an ordered manner.

Table 1. Similarity of known RNP sequences to the RNPs I and II of S1-1 protein

RNP of S1-1	Protein with similar RNP sequence	No. of RNPs	Similar RNP in the repeat	Optimum score	Identity (%)	Reference
RNP I	Chicken nucleolin	4	4th RNP	102	29	30
	Yeast NSR1 protein	2	2nd RNP	102	25	32
	Drosophila hnRNP A1	2	2nd RNP	95	26	33, 34
	Mammalian nucleolins	4	4th RNP	92	21	44–46
RNP II	Yeast NSR1 protein	2	2nd RNP	93	26	32
	Tabacco 33 kDa RNP	2	1st RNP	87	25	35
	Xenopus nucleolin	4	4th RNP	86	26	47
	Human HuD protein	3	3rd RNP	85	24	48

Sequences similar to the RNP I (amino acids 51–135) or II (amino acids 222–306) of S1-1 protein were searched in the SWISS-PROT database. Analyzed RNP regions correspond to those specified in the alignment of 70 RNP proteins (25). Optimum score was obtained according to Dayhoff (17). The top four proteins with the highest scores, together with their RNP copy numbers and the position of the similar RNP in the repeats, are cited. Identical amino acids (%) are also shown. Mammalian nucleolins are those of mouse (44), rat (45) and golden hamster (46). The (possible) functions of these proteins are: nucleolins, in pre-rRNA transcription and ribosome assembly; NSR1 protein, in pre-rRNA processing; hnRNP A1, as a component of heterogeneous nuclear RNPs; tobacco 33 kDa RNP precursor protein, in chloroplast RNA processing; HuD protein, in neuron-specific RNA processing. HuD protein has been characterized as a paraneoplastic encephalomyelitis antigen (48). The score for tobacco 33 kDa RNP protein was given for the amino acid 141–191 region (amino acids 252–306 of the S1-1 protein), where the score was highest.

The S1-1 protein had binding affinity for single-stranded DNA. RNA binding proteins often show a similar activity *in vitro*; the significance of the S1-1 affinity for single-stranded DNA is unknown.

Other structural features

Another characteristic feature of the S1-1 protein was a novel tandem repeat of five octamers. This repeat sequence may be functionally important by playing a role as a structural element in the molecular architecture.

In addition, a positively charged region, K735–K759, had a bipartite structure, which satisfies a motif sequence for a nuclear localization signal (37,38). Also, the Lys/Glu-rich sequence at K556–K568 satisfies the so-called KEKE motif, proposed by Realini and others (39) as a motif for promotion of association between proteins (Fig. 4).

We conclude that the cloned cDNA encoded a protein, S1-1, which was a novel RNA binding protein with characteristic and unique structural features.

ACKNOWLEDGEMENTS

AI thanks the fourth year students of Osaka City University Medical School for their participation and enthusiasm, who chose the S1-1 study in their laboratory course in November and December. They are (year): Shunnsuke Fujimoto (1990), Yuko Asechi, Kinuko Kono and Yamazaki Shioko (1991), Sinnichiro Ishino, Shuji Iwai and Rie Shibata (1992), Kazumitsu Ueda and Hidetada Fukushima (1993), Hideo Kuniba and Tomoaki Morioka (1994) and Takanori Watanabe (1995).

REFERENCES

1 Johns,E.W. (1971) In Phillips,D.M.P. (ed.), *Histones and Nucleohistones.* Plenum, New York, NY, pp. 2–45.
2 Bustin,M., Lehn,D.A. and Londsman,D. (1990) *Biochim. Biophys. Acta,* **1049**, 231–243.
3 Aaronson,R.P. and Blobel,G. (1975) *Proc. Natl. Acad. Sci. USA,* **72**, 1007–1011.
4 Dwyer,N. and Blobel,G. (1976) *J. Cell Biol.*, **70**, 581–591.
5 Inoue,A., Higashi,Y., Hasuma,T., Morisawa,S. and Yukioka,M. (1983) *Eur. J. Biochem.*, **135**, 61–68.
6 Inoue,A., Higashi,Y., Hasuma,T., Yukioka,M. and Morisawa,S. (1986) *J. Biochem.*, **100**, 747–751.
7 Higashi,Y., Uda,M., Yukioka,M., Morisawa,S. and Inoue,A. (1984) *Biochem. Int.*, **9**, 697–704.
8 Emura,T., Fujimoto,S., Takahashi,K.P., Higashi,Y., Morisawa,S. and Inoue,A. (1992) *Biochem. Int.*, **26**, 531–535.
9 Takahashi,K.P., Higashi,Y., Uchimoto,T., Hasuma,T., Yamamoto,N., Morisawa,S. and Inoue,A. (1990) *Histochemistry,* **93**, 381–384.
10 Higashi,Y., Takahashi,K.P., Uchimoto,T., Morisawa,S. and Inoue,A. (1990) *Cell Struct. Function,* **15**, 113–116.
11 Fakan,S. and Puvion,E. (1980) *Int. Rev. Cytol.*, **65**, 255–299.
12 Chomczynski,P. and Sacchi, N. (1987) *Anal. Biochem.*, **162**, 156–159.
13 Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual,* 2nd Edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
14 Henikoff,S. (1984) *Gene,* **28**, 351–359.
15 Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene,* **33**, 103–119.
16 Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA,* **74**, 5463–5467.
17 Dayhoff,M. (1978) In *Atlas of Protein Sequence and Structure,* Vol. 5. National Biomedical. Research Foundation, Silver Spring, MD.
18 Chou,P.Y. and Fasman,G.D. (1978) *Adv. Enzymol. Related Areas Mol. Biol.*, **47**, 45–138.
19 Chou,P.Y. and Fasman,G.D. (1978) *Annu. Rev. Biochem.*, **47**, 251–276.
20 Garnier,J., Osguthorpe,D.J. and Robson,B. (1978) *J. Mol. Biol.*, **120**, 97–120.
21 Schiffer,M. and Edmundson,A.B. (1967) *Biophys. J.*, **7**, 121–135.
22 Hopp,T.P. and Woods,K.R. (1981) *Proc. Natl. Acad. Sci. USA,* **78**, 3824–3828.
23 Siomi,H., Choi,M., Siomi,M.C., Nussbaum,R.L. and Dreyfuss,G. (1994) *Cell,* **77**, 33–39.
24 Laemmli,U.K. (1970) *Nature,* **227**, 680–685.
25 Birney,E., Kumar,S. and Krainer,A.R. (1993) *Nucleic Acids Res.*, **21**, 5803–5816.
26 Burd,C.G. and Dreyfuss,G. (1994) *Science,* **265**, 615–621.
27 Dreyfuss,G., Swanson,M.S. and Piñol-Roma,S. (1988) *Trends Biochem. Sci.*, **13**, 86–91.
28 Piñol-Roma,S., Swanson,M.S., Gall,J.G. and Dreyfuss,G. (1989) *J. Cell Biol.*, **109**, 2575–2587.
29 Kenan,D.J., Query,C.C. and Keene,J.D. (1991) *Trends Biochem. Sci.*, **16**, 214–220.
30 Maridor,G. and Nigg,E.A. (1990) *Nucleic Acids Res.*, **18**, 1286.
31 Maridor,G., Krek,W. and Nigg,E.A. (1990) *Biochim. Biophys. Acta,* **1049**, 126–133.
32 Lee,W.-C., Xue,Z. and Mélèse,T. (1991) *J. Cell Biol.*, **113**, 1–12.
33 Haynes,S.R., Rebbert,M.L., Mozer,B.A., Forquignon,F. and Dawid,I.B. (1987) *Proc. Natl. Acad. Sci. USA,* **84**, 1819–1823.
34 Haynes,S.R., Raychaudhuri,G. and Beyer,A.L. (1990) *Mol. Cell. Biol.*, **10**, 316–323.

35 Li,Y. and Sugiura,M. (1990) *EMBO J.*, **9**, 3059–3066.
36 Fukami-Kobayashi,K., Tomoda,S. and Go,M. (1993) *FEBS Lett.*, **335**, 289–293.
37 Dingwall,C. and Laskey,R.A. (1991) *Trends Biochem. Sci.*, **16**, 478–481.
38 LaCasse,E.C. and Lefebvre,Y.A. (1995) *Nucleic Acids Res.*, **23**, 1647–1656.
39 Realini,C., Rogers,S.W. and Rechsteiner,M. (1994) *FEBS Lett.*, **348**, 109–113.
40 Nagai,K., Oubridge,C., Jessen,T.H., Li,J. and Evans,P.R. (1990) *Nature*, **348**, 515–520.
41 Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 2495–2499.
42 Wittekind,M., Görlach,M., Friedrichs,M., Dreyfuss,G. and Mueller,L. (1992) *Biochemistry*, **31**, 6254–6265.
43 Nagai,K., Oubridge,C., Ito,N., Avis,J. and Evans,P. (1995) *Trends Biochem. Sci.*, **20**. 235–240.
44 Bourbon,H., Lapeyre,B. and Amalric,F. (1988) *J. Mol. Biol.*, **200**, 627–638.
45 Bourbon,H. and Amalric,F. (1990) *Gene*, **88**, 187–196.
46 Lapeyre,B., Bourbon,H. and Amalric,F. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1472–1476.
47 Rankin, M.L., Heine,M.A., Xiao,S., LeBlanc,M.D., Nelson,J.W. and DiMario,P.J. (1993) *Nucleic Acids Res.*, **21**, 169.
48 Szabo,A., Dalmau,J., Manley,G., Rosenfeld,M., Wong,E., Henson,J., Ponser,J.B. and Furneaux,H.M. (1991) *Cell*, **67**, 325–333.