Expression cloning and characterization of a novel gene that encodes the RNA-binding protein FAU-1 from *Pyrococcus furiosus*

Akio KANAI[1], Hanako OIDA, Nana MATSUURA and Hirofumi DOI[2]

Doi Bioasymmetry Project, Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Corporation, 4-1-8 Kawaguchi, Saitama 332-0012, Japan

We systematically screened a genomic DNA library to identify proteins of the hyperthermophilic archaeon *Pyrococcus furiosus* using an expression cloning method. One gene product, which we named FAU-1 (*P. furiosus* AU-binding), demonstrated the strongest binding activity of all the genomic library-derived proteins tested against an AU-rich RNA sequence. The protein was purified to near homogeneity as a 54 kDa single polypeptide, and the gene locus corresponding to this FAU-1 activity was also sequenced. The *FAU-1* gene encoded a 472-amino-acid protein that was characterized by highly charged domains consisting of both acidic and basic amino acids. The N-terminal half of the gene had a degree of similarity (25 %) with RNase E from *Escherichia coli*. Five rounds of RNA-binding-site selection and footprinting analysis showed that the FAU-1 protein binds specifically to the AU-rich sequence in a loop region of a possible RNA ligand. Moreover, we demonstrated that the FAU-1 protein acts as an oligomer, and mainly as a trimer. These results showed that the FAU-1 protein is a novel heat-stable protein with an RNA loop-binding characteristic.

Key words: archaea, RNA-binding protein, stem–loop structure.

INTRODUCTION

It is currently proposed that RNA, not DNA, was present during the initial stages of life on earth [1]. Thus analysing RNA metabolism in a very ancient organism, such as the hyperthermophilic archaeon, could lead to new insights into the fundamental regulation of genes. On the other hand, recent progress in various genome projects has also revealed complete genomic DNA sequences for the archaea, especially *Pyrococcus* species [2–4], that grow in deep sea hydrothermal vents at approx. 100 °C. However, only half of the total proteins deduced from these genomic DNA sequences could be assigned a putative cellular role by searching their similarities against certain databases. Therefore a systematic analysis of proteins at the genomic level by function may be a valuable strategy for characterizing proteins for post-genome projects. In this respect, use of an expression cloning method in the test tube [5,6] or a biochemical genomics approach in yeast [7] may turn out to be very useful.

We have developed an efficient highly sensitive method for identifying new genes encoding for DNA/RNA-binding proteins in the hyperthermophilic archeon *Pyrococcus furiosus*. This system has several advantages: (a) *P. furiosus* has only approx. 2000 genes and its complete genomic nucleotide sequence is already available (http://www.genome.utah.edu/); (b) the encoded proteins are mostly heat-stable and easy to handle biochemically; and (c) most of the genes involved in nucleic acid metabolism in archaea are similar to those found in eukarya, and the regulation mechanisms are simpler in the former than in the latter. In the present study, we systematically screened an expression genomic library to isolate unidentified genes encoding for DNA/RNA-binding proteins from *P. furiosus*. Gel-shift assays detected approx. 50 DNA/RNA-binding activities. We isolated and characterized the FAU-1 protein, a gene product that can bind the $r(A\text{-}U)_{10}$ sequence from *P. furiosus*. These results indicate that FAU-1 is a novel RNA loop-binding protein that forms oligomeric structures in nature.

MATERIALS AND METHODS

Construction of an expression genomic DNA library from *P. furiosus* and systematic screening for genes that encode DNA/RNA-binding proteins

A library was constructed essentially as described by Kaiser and Murray [8], except that a plasmid vector was used instead of λ phage. Briefly, genomic DNA isolated from a culture of *P. furiosus* was partially digested with *Sau*3AI, and the resulting DNA fragments were fractionated by electrophoresis on a 0.6 % (w/v) agarose gel. Fragments with molecular size of approx. 7 kb were prepared from the gel and ligated into a *Bam*HI-cleaved pRSET A plasmid vector (Invitrogen). The ligated plasmid was transformed into the *Esherichia coli* strain DH5α, and the colonies thus obtained (approx. 9.6×10^4 colony forming units/μg of *P. furiosus* DNA) were used to screen for new genes.

To make the protein pools, we first collected 30 independent *E. coli* colonies and then incubated the pooled colonies in 15 ml of Luria–Bertani broth containing 50 μg/ml of ampicillin for 16 h at 37 °C. The culture was divided into three groups: 0.5 ml of the culture was used to produce a glycerol stock of pooled *E. coli*; 2.0 ml was used to prepare the plasmid pools; and 12.5 ml was used to make each of 80 protein pools. After the 12.5 ml of *E. coli* culture was collected by centrifugation (11 000 ***g*** for 10 min at 4 °C), the resulting pellet was dissolved in ice-cold

Abbreviations used: BS[3], bis(sulphosuccinimidyl) suberate; FAU-1, *P. furiosus* AU-binding; FAU-1H, His-tagged FAU-1; NCBI, National Centre for Biotechnology Information; ORF, open reading frame.

[1] Present address and address for correspondence: Institute for Advanced Biosciences, Keio University, Tsuruoka, Yamagata 997-0017, Japan (e-mail akio@sfc.keio.ac.jp).

[2] Present address: Celestar Lexico-Sciences, Inc., MTG D17, 1-3 Nakase, Chiba 261-0023, Japan.

The nucleotide sequence data for the FAU-1 protein from *Pyrococcus furiosus* has been deposited in the DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under the accession number AB055587.

© 2003 Biochemical Society

20 mM Tris/HCl (pH 8.0), and proteins were then extracted following a brief sonication (30 s). The extract was heat-treated at 85 °C for 15 min to denature endogenous *E. coli* proteins, and then centrifuged at 17 000 ***g*** for 10 min at 4 °C to separate the extract from the cellular debris. The supernatant was stored as a protein pool at − 80 °C until use. The 80 pools represented 2400 individual clones, which was expected to cover the 2 megabase genome for *P. furiosus* twice.

Expression and purification of recombinant FAU-1 protein

The plasmid clone pAL66-38Δ2, containing *P. furiosus* genomic DNA that codes for a product that has AU-rich RNA-binding activity (FAU-1 activity), was overexpressed in *E. coli* strain DH5α, and was then extracted by sonication (30 s) in a buffer containing 150 mM Tris/HCl (pH 7.5), 2 mM EDTA, 0.25 mM 4-amidinophenylmethane-sulphonyl fluoride ('APMSF') and 0.2 % Tween 20. This extract was heat-treated at 80 °C for 15 min to denature endogenous *E. coli* proteins and then dialysed against buffer A [50 mM Tris/HCl (pH 7.5), 1 mM EDTA, 0.2 % Tween 20, 7 mM 2-mercaptoethanol and 10 % (v/v) glycerol]. The dialysed lysate (approx. 10 ml per 500 ml of *E. coli* culture) was loaded on to a 6 ml RESOURCE-Q column (Amersham Biosciences) equilibrated with buffer A and eluted with a linear gradient of NaCl (from 0–1.0 M) in buffer A. Fractions containing peak RNA-binding activity (samples eluted between 0.2–0.5 M NaCl) were pooled. The pooled lysate was mixed with an equal volume of buffer A/2 M ammonium sulphate and was applied on to a 1 ml RESOURCE-PHE column (Amersham Biosciences) equilibrated with buffer A/1 M ammonium sulphate. Bound proteins were eluted with buffer A. The eluted fractions containing protein with RNA-binding activity were diluted with buffer A until their conductivity reached that of buffer A. The fractions were then loaded on to a 1 ml poly(U)-Sepharose column equilibrated with buffer A. The column was again eluted with a linear gradient of NaCl (from 0–1.0 M) in buffer A. Fractions containing proteins with peak RNA-binding activity (samples eluted between 0.5–1.0 M NaCl, fractions 15–25) were pooled. SDS/PAGE showed that the purity of the FAU-1 protein in the final preparation approached near homogeneity. The N-terminal sequence of the purified FAU-1 protein was determined by sequential Edman degradation (Takara).

To express the His-tagged FAU-1 protein (FAU-1H protein), we subcloned the *FAU-1* gene into the pET-23b expression vector by PCR using region-specific oligonucleotides as primers. These oligonucleotides were designed to provide PCR products with *Nde*I and *Xho*I sites at the 5′- and 3′-termini respectively. The resulting construct (termed pET-FAU1H) contained the sequence coding for the FAU-1 protein (472 amino acids) and six consecutive histidine residues (His-tag) at the C-terminus to be used as an affinity tag for purification. The nucleotide sequence of the insert was confirmed to be the original sequence. Recombinant FAU-1H protein was overexpressed in *E. coli* strain BL21(DE3) under the control of the phage T7 promoter. Protein was then extracted by sonication (30 s) in a buffer containing 20 mM Tris/HCl (pH 8.0), 500 mM NaCl, 5 mM imidazole and 0.1 % Triton X-100. The extract was heat-treated at 80 °C for 15 min to denature endogenous *E. coli* proteins and then centrifuged at 12 000 ***g*** for 10 min at 4 °C to separate the cell debris from the extract. The FAU-1H protein was purified using a 5 ml Ni^{2+}-Sepharose column according to the manufacturer's instructions (Novagen). The protein with peak RNA-binding activity was purified further by RESOURCE-Q column chromatography equilibrated with buffer A at a flow rate of 1 ml/min. SDS/PAGE showed that the purity of the FAU-1H protein in the final preparation approached near homogeneity. The amount of protein in each fraction was determined by the Bradford assay using BSA as a standard.

Antibody production and immunoblot analysis

A polyclonal antibody was raised in rabbits against a synthetic peptide corresponding to the predicted amino acid sequence 430–446 of the FAU-1 protein (plus cysteine at the C-terminus). The immunogen peptide was conjugated to keyhole limpet haemocyanin and the antiserum was used for immunoblot analysis.

For immunoblot analysis, proteins were separated by SDS/PAGE and transferred on to nitrocellulose membranes (Schleicher und Schüll) using a semi-dry transfer cell (Marysol). Membranes were probed with anti-FAU-1 antibody and developed using an alkaline phosphatase-conjugated goat anti-rabbit immunoblot assay kit according to the manufacturer's instructions (Bio-Rad).

Gel-shift assay

Binding reactions containing ^{32}P-end-labelled RNA/DNA (30 000–50 000 c.p.m.; 1–2 $\times$ 10^6 c.p.m./pmol DNA or RNA) and either 8–620 ng of purified FAU-1 or approx. 1–3 μg of extract from *E. coli* expressing the *P. furiosus* genomic DNA were incubated at 75 °C or various temperatures for 15 min in 20 μl of RNA/DNA-binding buffer [10 mM Tris/HCl (pH 7.5), 50 mM NaCl, 0.5 mM EDTA, 2.5 mM $MgCl_2$, 5 % (v/v) glycerol, 1 mM dithiothreitol and 200 ng of poly(dI-dC)]. The RNA/DNA–protein complexes were analysed by electrophoresis on a non-denaturing 6 % (w/v) polyacrylamide gel. The gels were dried and the quantity of DNA/RNA–protein complexes was evaluated by scanning the radioactive image using a computerized image analyser (Fuji).

For competition analysis, an excess quantity of homopolymers (1 or 3 μg/reaction), poly(dT), poly(dA), poly(C), poly(U), poly(A) (Amersham Biosciences) or an excess amount of yeast tRNA was added to the reaction mixture. The sequences for the oligonucleotides used are as follows: $r(A\text{-}U)_{10}$, 5′-r(AUAU-AUAUAUAUAUAUAUAU)-3′; $r(G\text{-}C)_{10}$, 5′-r(GCGCGCGCG-CGCGCGCGCGC)-3′; $d(A\text{-}T)_{10}$, 5′-d(ATATATATATATATATAT-AT)-3′; $d(G\text{-}C)_{10}$, 5′-d(GCGCGCGCGCGCGCGCGCGC)-3′; and 5N-22D, 5′-r(ACGUCUAGGCUAUCUGCACUGGCCGU)-3′.

In vitro selection of RNA ligands

The linear N15 oligonucleotide library [5′-GGATACTTACC-TGGCAGGA$(N)_{15}$CACTGGCCGTCGTTTTACA-3′] was amplified by PCR using BS-T7 (5′-AATGTCGACTAATACGA-CTCACTATAGGGATACTTACCTGGCAGGA-3′) and BS-R (5′-TGTAAAACGACGGCCAGTG-3′) primers. The resulting PCR products were used to generate a pool of randomized ^{32}P-labelled RNAs by using an *in vitro* transcription kit (Ambion). RNA gel-shift assays were performed using purified FAU-1H protein to select binding sequences from the RNA pool. Selected RNAs, which corresponded to the positions of shifted RNA bands in the gel-shift assay, were eluted and extracted with phenol/chloroform (1:1, v/v), followed by ethanol precipitation. The pellets were resuspended in diethyl pyrocarbonate-treated water and reverse-transcribed by rTth DNA polymerase (Toyobo),

© 2003 Biochemical Society

generating a PCR template for the next round of selection. The selection process was carried out for five rounds.

RNase footprinting analysis

The 5′-terminus of the synthetic ribonucleotide 5N-22D was radiolabelled with [γ-^{32}P]ATP using the T4 polynucleotide kinase. Prior to RNase footprinting analysis, the labelled probes (200 000 c.p.m.; 1×10^6 c.p.m./pmol RNA) were incubated with various amounts of the FAU-1 protein (0–2 μg) or BSA in DNA/RNA-binding buffer at 22 °C for 15 min. RNases were diluted and added to the above reaction mixtures. After 15 min incubation at 22 °C, the digestion reactions were stopped by adding 40 μl of stop buffer [1 M Tris/HCl (pH 7.5) and 8 M urea]. For preparation of a sequence ladder, the 5N-22D RNA was incubated with 50 mM NaOH at 22 °C for 15 min. The cleavage products were separated by electrophoresis on denaturing 20 % (w/v) polyacrylamide gels containing 8 M urea and visualized using a computerized image analyser (Fuji).

Gel filtration of FAU-1 oligomers

Purified FAU-1H protein (150 μg), obtained from poly(U)-Sepharose column chromatography fractions, was preincubated in 200 μl of buffer B at 75 °C for 15 min. The reaction mixture was loaded on to a 24 ml Superdex 200 HR10/30 gel-filtration column (Amersham Biosciences) equilibrated in buffer B and run at 0.25 ml/min (0.25 ml/fraction). Protein peaks were identified by immunoblot analysis of the collected fractions using the anti-FAU-1 antibody.

Chemical cross-linking analysis

FAU-1H protein (0.4 μM) was incubated in buffer containing 50 mM Tris/HCl (pH 8.0), 10 mM NaCl, 2.5 mM $MgCl_2$ and various concentrations of bis(sulphosuccinimidyl) suberate (BS^3; 0–1.0 mM) at 22 °C for 20 min. Reactions were stopped by adding 1 M Tris/HCl (pH 8.0) and analysed by SDS/PAGE [2–15 % (w/v) gels]. Bands were visualized by immunoblot analysis using the anti-FAU-1 antibody. For estimating high-molecular-mass complexes, cross-linked phosphorylase b markers (Sigma) were used.

RESULTS AND DISCUSSION

Systematic identification of DNA/RNA-binding proteins by expression cloning

We made an expression genomic DNA library for *P. furiosus*. After preparing the *P. furiosus* genomic DNA, partially digested DNA fragments (average size approx. 7 kb) were ligated into a pRSET A plasmid vector. Although the vector contained the T7 RNA polymerase promoter sequence, gene expression from *P. furiosus* genomic DNA was observed in *E. coli* without induction of T7 RNA polymerase. Because *P. furiosus* promoter sequences may also work in *E. coli*, we chose to use the *E. coli* strain DH5α, which does not endogenously contain the T7 RNA polymerase gene. Next, we prepared protein pools consisting of 30 independent colonies of *E. coli* in the library. These protein pools were heat-treated to denature endogenous *E. coli* proteins. This heat-treatment step is very effective for detecting the DNA/RNA-binding activities from proteins derived from *P. furiosus* with a low background. Because the genome of *P. furiosus* has 1 908 256 bases, screening of 1200 clones (40 pools of 30 clones) should cover the whole genome [9]. Thus 12 h is enough to screen a genome of this size.

Since we are interested in nucleic acid metabolism in archaea, we performed a gel-shift assay in order to systematically detect DNA/RNA-binding activities in the library using a series of probes, such as $r(A\text{-}U)_{10}$, $r(G\text{-}C)_{10}$, $d(A\text{-}T)_{10}$ and $d(G\text{-}C)_{10}$. As a result, we identified approx. 50 nucleic-acid-binding activities and found that many of these activities were $r(A\text{-}U)_{10}$-binding activities. In the present study, we focused on one of the $r(A\text{-}U)_{10}$-binding proteins and named it FAU-1 (*P. furiosus* AU-binding). We chose this name because the protein possesses the strongest $r(A\text{-}U)_{10}$-binding activity that we could find in the genome, and because $r(A\text{-}U)_{10}$-rich sequences correspond well with several biological activities, including RNA stabilization [10,11] and degradation [12–14].

Identification of the *FAU-1* gene

The extracts of 40 pools (1200 independent clones) were screened with $r(A\text{-}U)_{10}$ probes, and 12 positive pools containing $r(A\text{-}U)_{10}$-binding activity were obtained. Among the positive pools, three (plasmid pools 66, 67 and 80) had exactly the same band-shift pattern, as analysed by gel-shift assay, and possessed the strongest binding activity to the $r(A\text{-}U)_{10}$ probe at 75 °C (FAU-1 activity). We isolated one plasmid clone, termed 66-38, in the second screening from the 66 plasmid pool. The protein product of clone 66-38 possessed two bands (B1 and B2) in a gel-shift assay (Figure 1B), identical with the result from the original plasmid pool 66. Plasmid clone 66-38 was found to contain a 7.8 kb insert (Figure 1A), and we mapped the DNA fragment using restriction enzymes. In order to determine the location of the *FAU-1* gene, we prepared three deletion mutants derived from the original 66-38 insert, using the appropriate restriction enzymes. We tested for FAU-1 activity by overexpressing each deletion mutant and analysing each extract. Figure 1(B) demonstrates that the 3.2 kb insert in plasmid 66-38Δ2 was sufficient to detect FAU-1 activity. However, the other mutant (plasmid 66-38Δ1) completely abolished the FAU-1 activity, suggesting that the *FAU-1* gene or the promoter sequence responsible for its expression is located in the 1.3 kb region between *Hin*dIII and *Pst*I. We then determined the nucleotide sequence for the insert of plasmid 66-38Δ2 (GenBank® accession number AB055587) and found a region that contains a possible open reading frame (ORF) starting with GTG as the first codon. This putative gene encodes a hypothetical protein consisting of 472 amino acids with a calculated molecular mass of 54 kDa.

To determine whether or not the ORF identified by the deletion analysis truly represents FAU-1 activity, we purified the protein produced from plasmid 66-38Δ2 from the *E. coli* extract to near homogeneity (Table 1 and Figure 2). Purification yielded approx. 0.1 mg of the FAU-1 protein from 1 litre of *E. coli* culture. The product from a final poly(U)-Sepharose column yielded a purified protein that had an apparent molecular mass of 54 kDa as determined by SDS/PAGE (Figure 2B). This finding is in agreement with the value determined by the deduced amino acid sequences for the gene. Protein from the column also possessed RNA-binding activity (Figure 2B and 2C), indicating that the purified 54 kDa protein is indeed FAU-1. Furthermore, we made a new construct encoding the FAU-1 protein fused to a His_6-tag (FAU-1H), expressed the plasmid in *E. coli* and purified the resulting product to near homogeneity. The FAU-1H protein also possessed $r(A\text{-}U)_{10}$-binding activity at high temperature. However, the FAU-1H protein produced only band B1 and not band B2 when analysed by gel-shift assay (see Figure 4A). Our preliminary results suggested that the partial degradation of the

© 2003 Biochemical Society

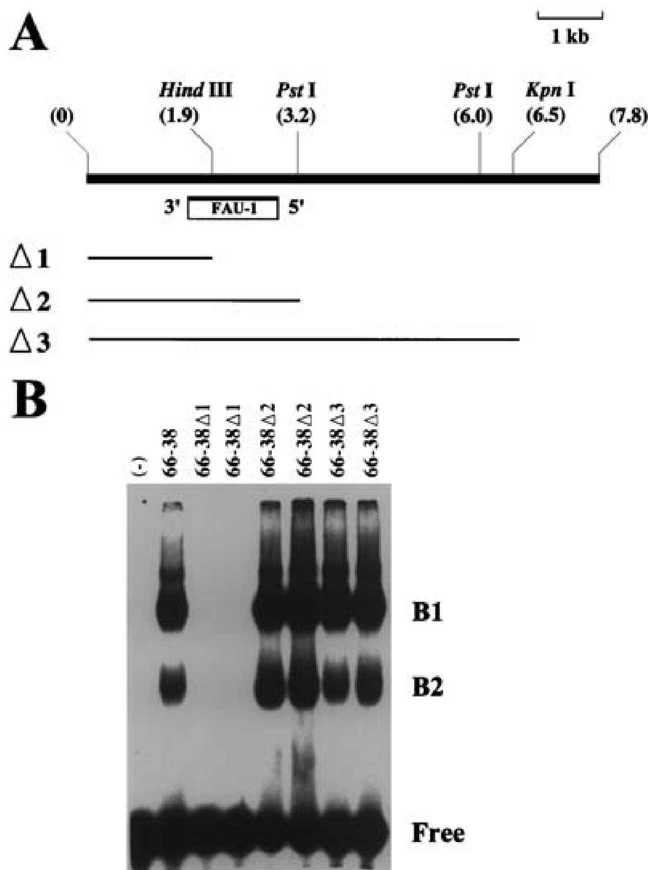

Figure 1 Determination of *FAU-1* gene locus

(**A**) Restriction enzyme map of the *FAU-1* genomic clone and its deletion derivatives (Δ1–3). The 7.8 kb insert from plasmid pAL66-38 is shown. The box indicates the position and direction of the *FAU-1* gene. The numbers show positions of the restriction enzyme sites in kb. (**B**) RNA gel-shift assay using extracts from overexpression of the product of the deletion mutants in *E. coli*. Heat-treated fractions of *E. coli* extracts expressing pAL66-38 or its deletion mutants (Δ1–3) were used for the gel-shift assay. The ^{32}P-end-labelled oligonucleotide r(A-U)$_{10}$ was incubated with approx. 3 μg of extract in RNA/DNA-binding buffer. RNA–protein complexes were analysed by electrophoresis on a non-denaturing 6 % (w/v) polyacrylamide gel. Samples were analysed in duplicate. B1 and B2 denote the positions of the RNA–protein complexes.

Table 1 Purification steps for the FAU-1 protein

A unit is defined as the amount of the fraction required to shift approx. 1 pg of a 20-mer r(A-U)$_{10}$ oligonucleotide probe.

Purification step	Activity (units)	Protein (mg)	Specific activity (units/mg)	Purification (-fold)
Heat-extract	260 000	15.3	17 000	1
RESOURCE-Q	230 000	4.1	56 000	3
RESOURCE-PHE	450 000	0.49	920 000	54
Poly(U)-Sepharose				
Fraction 20	55 000	0.012	4 600 000	271
Fraction 21	55 000	0.009	6 100 000	359

FAU-1 protein results in the generation the B2 band (results not shown).

We determined the N-terminal amino acid sequence of the purified protein to be Ser-Thr-Glu-Ser-Glu-Ile-Ala (Figure 3). This sequence is located just four amino acids downstream from the putative initiation amino acid (GTG codon for Met) and 22 amino acids upstream from the first methionine (ATG codon for Met). Three of the first four amino acids in the possible ORF are encoded by GTG codon in the frame. All four of these amino acids are missing from the purified FAU-1 protein. The reason for this deletion of four amino acids is unknown.

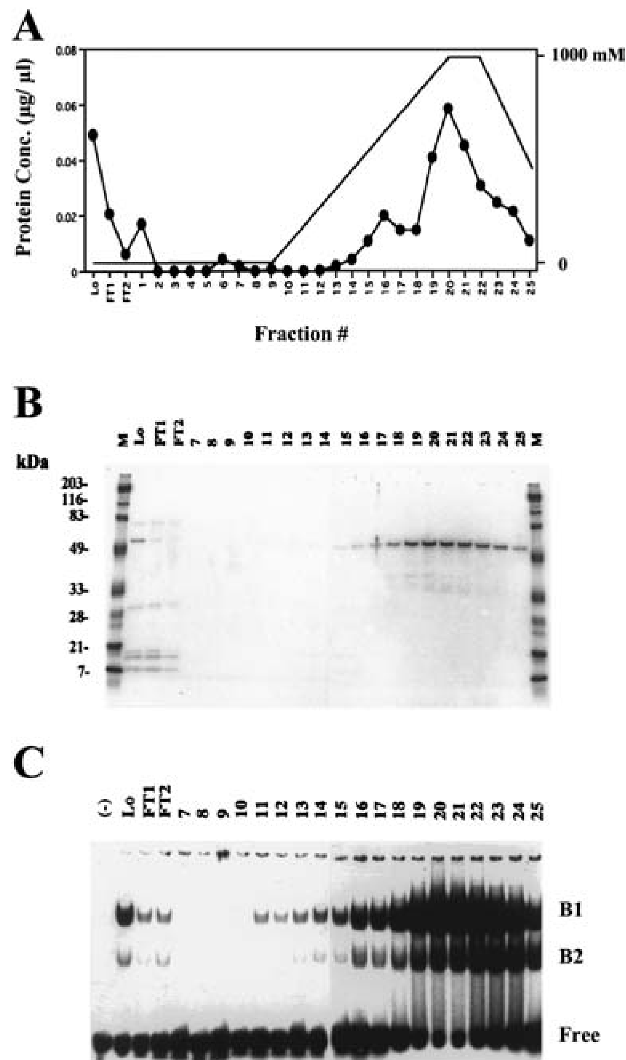

Figure 2 Purification of the recombinant FAU-1 protein

(**A**) An elution profile for poly(U)-Sepharose column chromatography. The fraction showing peak RNA-binding activity from RESOURCE-PHE column chromatography was loaded on to a poly(U)-Sepharose column (Amersham Biosciences) equilibrated with buffer A; the flow rate was set at 0.2 ml/min and 0.2 ml fractions were collected. Lo and FT denote the fraction load and flow-through respectively. (**B**) SDS/PAGE of proteins from the poly(U)-Sepharose column chromatography. Aliquots of the fractions from the poly(U)-Sepharose column were subjected to SDS/PAGE [10–20 % (w/v) gels), and the gels were stained with Coomassie Brilliant Blue (Wako). Lane M, molecular-mass markers in kDa. (**C**) RNA gel-shift assay using aliquots of the fractions from the poly(U)-Sepharose column. The ^{32}P-end-labelled oligonucleotide r(A-U)$_{10}$ was incubated with the purified FAU-1 protein on a poly(U)-Sepharose column in RNA/DNA-binding buffer. RNA–protein complexes were analysed by electrophoresis on a non-denaturing 6 % (w/v) polyacrylamide gel. B1 and B2 denote the positions of the RNA–protein complexes.

Primary structure of the FAU-1 protein

From the nucleotide sequence analysis of the genomic clone (plasmid 66-38Δ2), we deduced that the complete amino acid sequence for the FAU-1 protein consists of 472 amino acids (Figure 3). After searching the PSI-BLAST search engine

© 2003 Biochemical Society

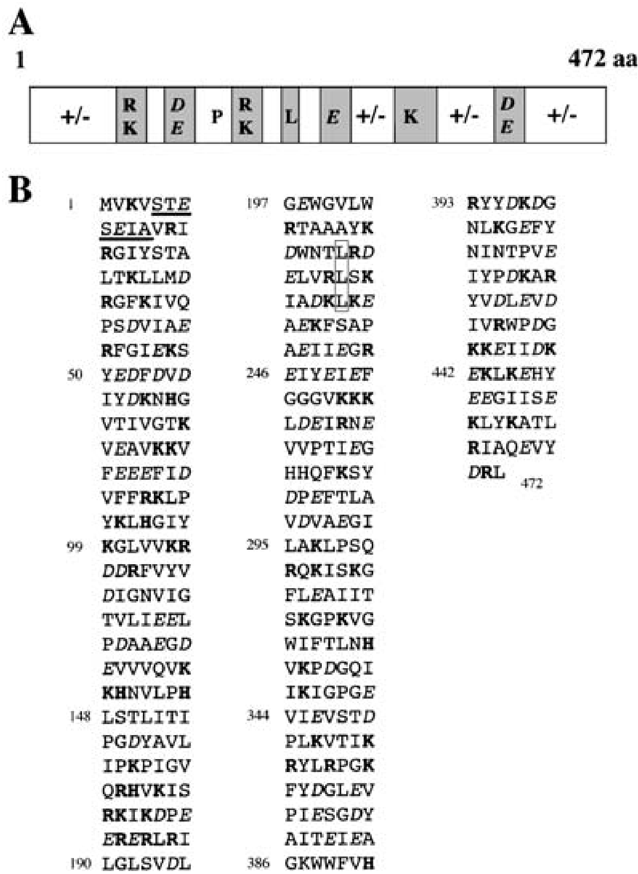

Figure 3 Amino acid sequence of the FAU-1 protein

(**A**) Schematic representation of the structure of the FAU-1 protein. The protein can be approximately divided by its amino acid content into acidic [aspartate (D) or glutamate (E)], basic [arginine (R) or lysine (K)], mixed-charge regions (+/−), and proline (P)-rich region. A putative leucine (L) zipper is found in the middle of the protein. (**B**) Amino acid sequence of the FAU-1 protein. The deduced amino acid sequence for the FAU-1 polypeptide was aligned every seven amino acids (where single-letter amino-acid notation has been used). Numbers indicate the amino acid position. Basic amino acids (K, R and H) are shown in bold, acidic amino acids (D, E) are in italics, and leucines (L) in the putative leucine zipper are boxed. The N-terminal amino acid sequence of the purified protein is underlined.

(http://www.ncbi.nlm.nih.gov/BLAST/), we found that the top six similar proteins were all from archaea. The data for all these proteins came from archaeal genome projects, and all of the proteins were classified as hypothetical proteins. In the *Pyrococcus* species, the FAU-1 protein shares 86 % identity with the 472-amino-acid hypothetical proteins from both *P. horikoshii* [National Centre for Biotechnology Information (NCBI) protein database accession number BAA29186] and *P. abyssi* (NCBI protein database accession number CAB49039) [2,3]. Although the FAU-1 protein showed no significant similarity to any protein whose functions are well known at either the nucleotide or the amino acid level, weak similarity was observed between the FAU-1 protein and the RNase E/RNase G protein family at the N-terminal site [25–30 % identity with RNase E from *E. coli* (NCBI protein database accession number S27311), RNase E from *Porphyra purpurea* (NCBI protein database accession number AAC08097), RNase G from *Chlamydia pneumoniae* (NCBI protein database accession number D72015) and the 454-amino-acid hypothetical protein TM1606 from *Thermotoga maritima* (NCBI protein database accession number B72235)]. The C-terminal half of *E. coli* RNase E can be approximately divided by amino acid content into basic and acidic amino acid regions. The same feature was observed for the FAU-1 protein. The secondary structure prediction for the FAU-1 protein was based on the computer programs developed by Chou and Fasman [15] and Garnier et al. [16]. This prediction suggested that approximately half the amino acids of the whole FAU-1 protein contribute to an α-helix structure (results not shown). We therefore arranged the amino acid sequence for the FAU-1 protein into a series of seven amino acid residues and found that the charged amino acids are regionally assembled (Figure 3). This suggested that the acidic and basic amino acid regions are located as clusters in the tertiary structure of the FAU-1 protein. A putative leucine zipper is found in the middle of the FAU-1 protein, providing the possibility that the FAU-1 protein can oligomerize. The oligomeric structure of the FAU-1 protein is discussed later (see below).

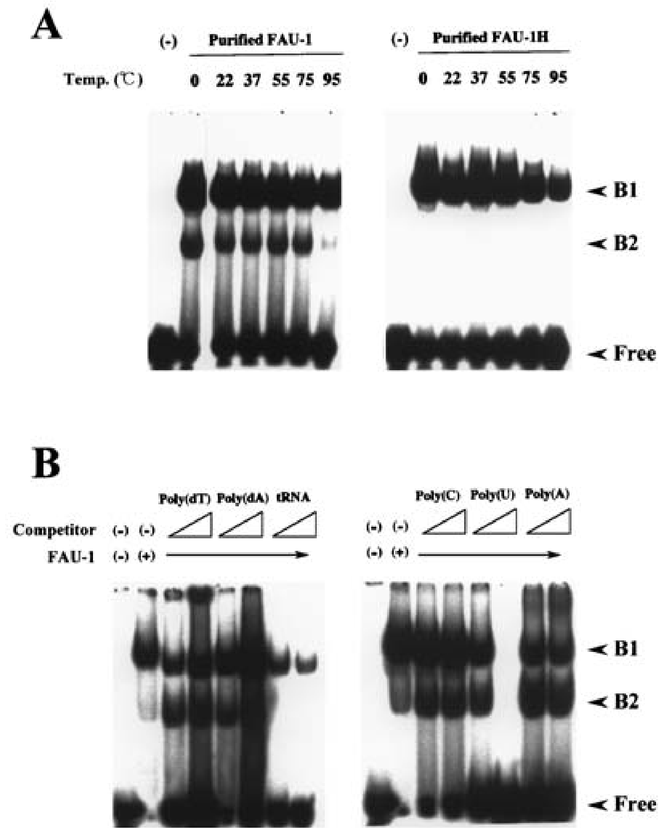

Figure 4 Characterization of RNA-binding activities

(**A**) Temperature-dependent RNA-binding activities of recombinant FAU-1 and FAU-1H proteins were examined by gel-shift analysis. The ^{32}P-end-labelled oligonucleotide r(A-U)$_{10}$ (50 000 c.p.m.; 1–2 × 10^{6} c.p.m./pmol RNA) was incubated with 60 ng of the purified FAU-1 [a fraction from the poly(U)-Sepharose column chromatography] or the His-tagged FAU-1, FAU-1H (a fraction from the RESOURCE-Q column chromatography) at various temperatures. (**B**) Competitive RNA gel-shift assay. The ^{32}P-end-labelled oligonucleotide r(A-U)$_{10}$ (50 000 c.p.m.; 1–2 × 10^{6} c.p.m./pmol RNA) was incubated with 50 ng of purified FAU-1 in the presence of 1 or 3 μg of competing unlabelled homopolymers or tRNA.

Characterization of RNA-binding activity of the FAU-1 protein

We carried out several experiments to determine the biochemical characteristics of the RNA-binding ability of the purified FAU-1 protein. Initially, RNA-binding activities at various temperatures were examined using both purified FAU-1 and FAU-1H proteins (Figure 4A). RNA–FAU-1 protein complexes were observed at

© 2003 Biochemical Society

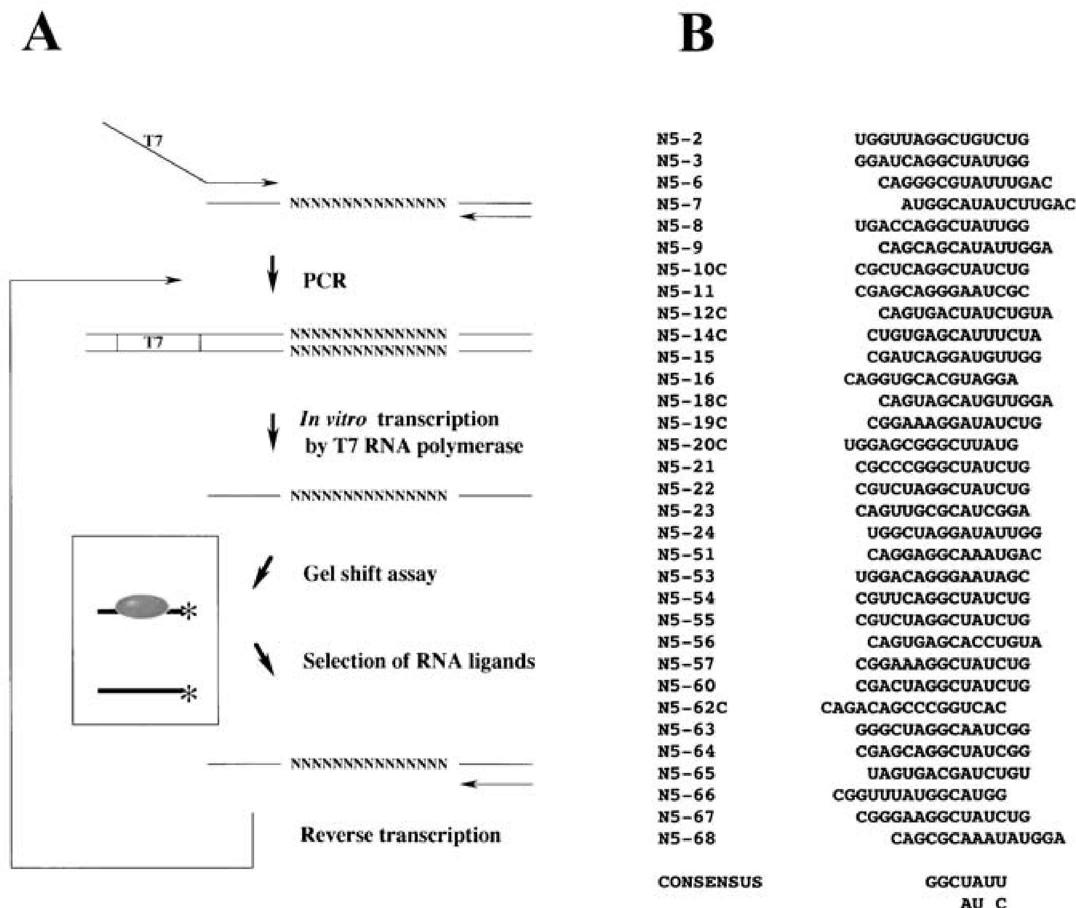

Figure 5 ***In vitro* selection of the RNA ligand for the FAU-1 protein**

(**A**) Schematic representation of the *in vitro* selection protocol used to select for the RNA ligand recognized by the FAU-1 protein. (**B**) Sequences remaining (33 in total) after five rounds of selection are shown. A 7 nt consensus region was found in these sequences, consisting of GGC(U/A)(A/U)U(U/C).

every temperature tested (0–95 °C). Poly(U) and tRNA competitively blocked binding between FAU-1 protein and r(A-U)$_{10}$ (Figure 4B), suggesting that the FAU-1 protein mainly recognized uridine and that perhaps a specific secondary structure is necessary for efficient binding.

In order to know the most suitable RNA sequence for recognition by the FAU-1 protein, we performed *in vitro* selection experiments for RNA ligands (Figure 5). After five rounds of selection and amplification, all 33 of the analysed sequences contained an AU-rich sequence. Interestingly, the AU-rich sequence is followed by a GC-rich sequence. Therefore the consensus sequence obtained by the selection was found to be GGC(U/A)(A/U)U(U/C). According to a computer prediction of possible secondary structures for the selected RNA ligands, the GC-rich sequences may contribute to the formation of possible stem structures by hybridizing with the conserved primer-annealed sequences (discussed later). We determined the apparent equilibrium dissociation constant (K_d) of the FAU-1 protein for both the r(A-U)$_{10}$ probe and 5N-22D, a selected probe that is a synthetic ribo-oligonucleotide containing the 5N-22 sequence and part of a conserved primer-annealed sequence. Figure 6 shows the normalized binding curves. The affinity of the 5N-22D RNA ligand ($K_d = 7 \times 10^{-8}$ M) is three times that of the r(A-U)$_{10}$ ligand ($K_d = 2 \times 10^{-7}$ M). These results suggested that the FAU-1 protein recognizes a specific sequence, probably an AU-rich sequence, in the loop region of a tertiary RNA ligand.

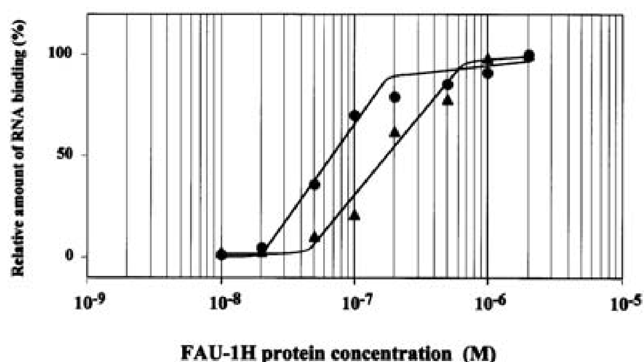

Figure 6 Binding affinity of FAU-1H for r(A-U)$_{10}$ (▲) and 5N-22D (●) RNA

The data for each curve were normalized to the saturation point for the FAU-1H protein. The K_d value is equal to the protein concentration at which 50 % of the highest RNA-binding activity is shown. The 5N-22D RNA ligand has three times the affinity ($K_d = 7 \times 10^{-8}$ M) of the r(A-U)$_{10}$ ligand ($K_d = 2 \times 10^{-7}$ M). Similar results were obtained in two independent experiments.

© 2003 Biochemical Society

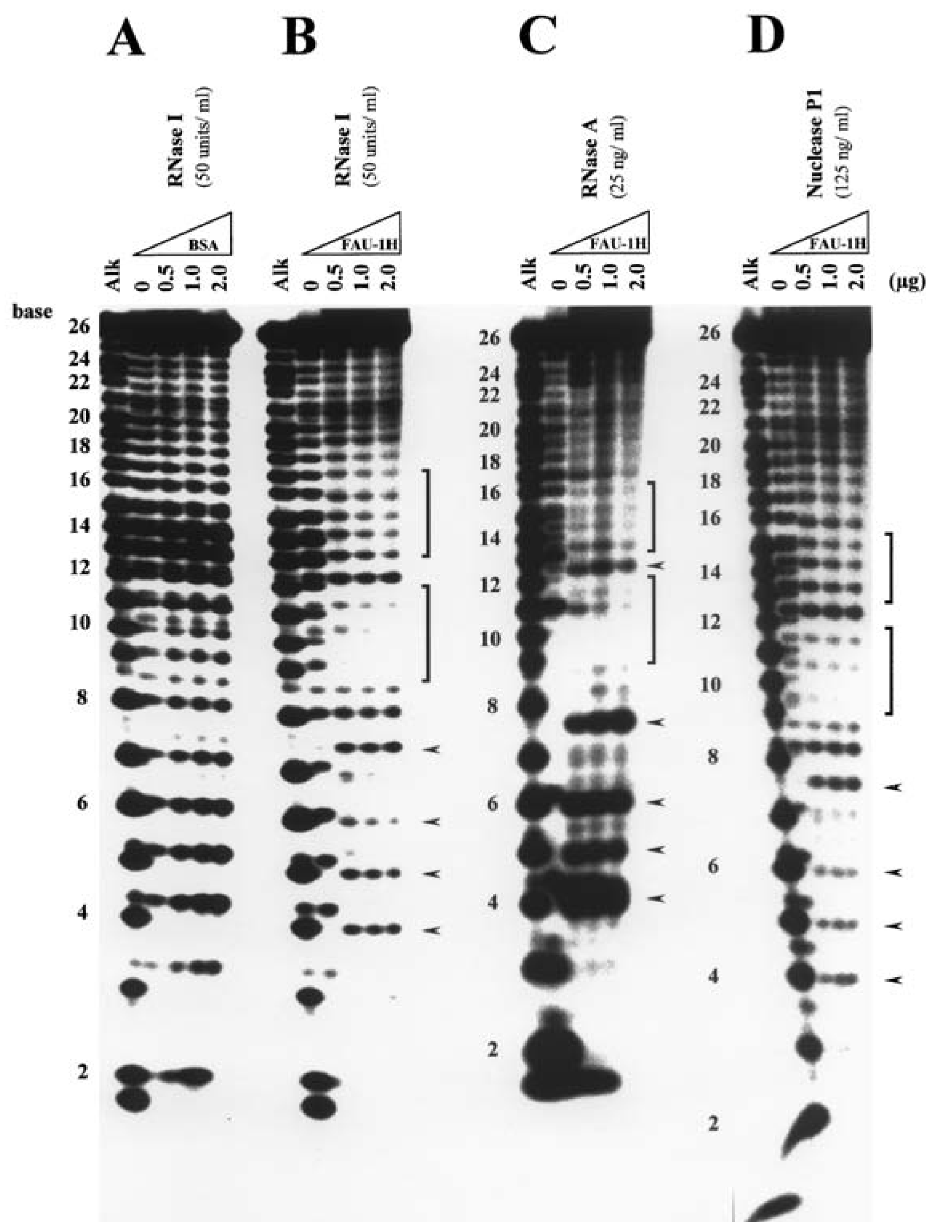

Figure 7 RNase footprinting analysis of the 5N-22D RNA in the absence or presence of the FAU-1H protein

The ^{32}P-end labelled 5N-22D RNA was digested with RNase I (**A** and **B**), RNase A (**C**) or nuclease P1 (**D**) in the absence or presence of the FAU-1H protein. In (**A**), BSA was used instead of the FAU-1H protein as a control. Alk denotes that RNA was partially alkaline hydrolysed. Brackets delineate sequences protected against RNase digestion. Arrowheads represent major hypersensitive sites by RNases.

In order to clarify the FAU-1-binding sequence on the 5N-22D RNA, we performed RNA footprinting analysis (Figure 7). The addition of the FAU-1 protein to the binding reaction mixtures at three different concentrations showed the protection of specific bases against nuclease attack (Figures 7B–7D). However, BSA did not protect those same sequences. The strongest binding was observed between the positions C_9 and A_{12}; these positions are located in the loop region in the 5N-22D probes. A second, weakly protected site was found between C_{14} and G_{16}, positions that are also located in the loop region of a possible stem–loop structure. Several hypersensitive sites were detected against the nucleases used. These results showed that the FAU-1 protein is a loop-binding protein and can specifically recognize AU-rich sequences in this region (Figure 8).

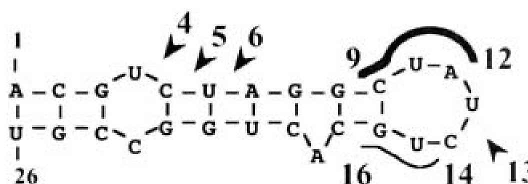

Figure 8 Possible secondary structure of the selected RNA ligand 5N-22D and the RNase footprints of the FAU-1H-binding regions

Protected sequences are indicated by lines (thick line, strong binding; thin line, weak binding). The arrowheads represent major RNase hypersensitive sites.

© 2003 Biochemical Society

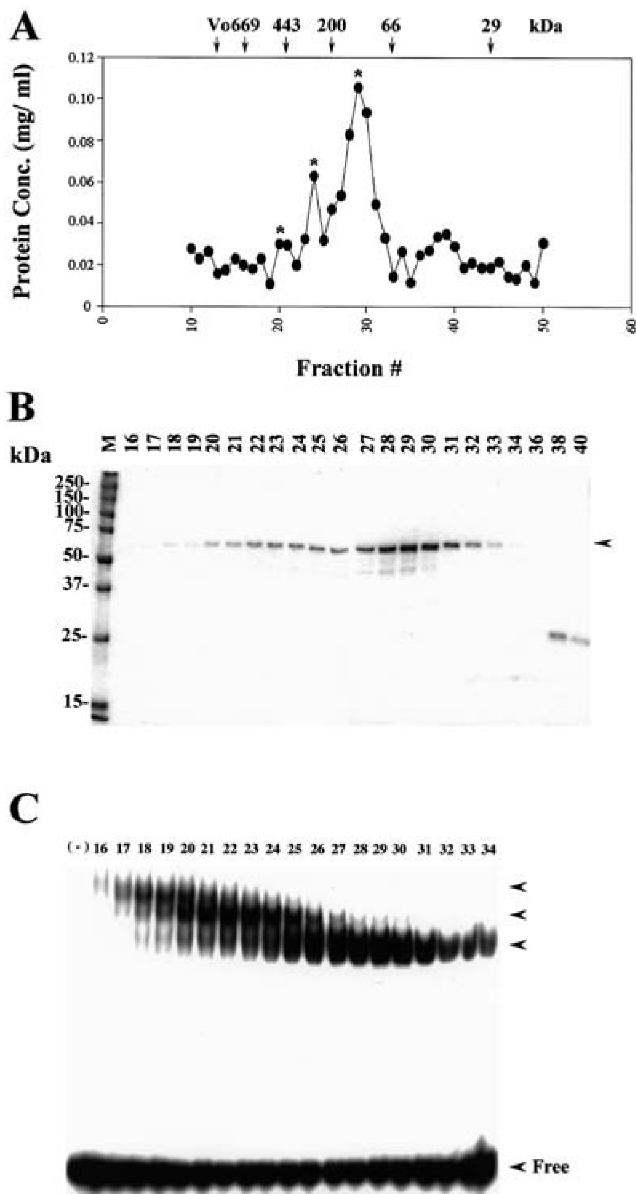

Figure 9 FAU-1H protein exists in nature with oligomeric structures

(**A**) Superdex 200 HR10/30 gel-filtration column chromatography elution profile. Fractions showing peak RNA-binding activity from poly(U)-Sepharose column chromatography were loaded on to a Superdex 200 HR10/30 gel-filtration column (Amersham Biosciences); the flow rate was set at 0.25 ml/min and 0.25 ml fractions were collected. (**B**) SDS/PAGE of proteins collected from a Superdex 200 HR10/30 gel-filtration column. Samples of the fractions from the column were subjected to SDS/PAGE [10–20 % (w/v) gels], and the gels were stained with Coomassie Brilliant Blue (Wako). Lane M, molecular-mass markers (in kDa). The arrowhead indicates the position of the FAU-1 protein. (**C**) RNA gel-shift assay using samples of fractions obtained from a HR10/30 gel-filtration column. The 5′-^{32}P-end-labelled oligonucleotide r(A-U)$_{10}$ was used as a probe. RNA–protein complexes were analysed by electrophoresis on a non-denaturing 6 % (w/v) polyacrylamide gel. Arrowheads indicate the positions of the RNA–FAU-1H complexes.

Oligomeric structure of the FAU-1 protein

To determine the native size of the FAU-1 protein, we initially carried out size-exclusion chromatography. The FAU-1 protein was eluted in three peaks, fractions 20, 24 and 29 (Figure 9A). Comparison with molecular-mass standards showed that these peaks correspond to molecular masses of approx. 150, 300 and 450 kDa respectively. SDS/PAGE showed that each peak contained mainly the 54 kDa purified FAU-1 polypeptide (Figure 9B). These results showed that the FAU-1 protein is made up of homotrimer, homohexamer and homononamer structures. The majority of the oligomeric structures were shown to be trimers. Each fraction was also assayed for its RNA-binding activity in gel-shift assays. Figure 9(C) shows that the three protein peaks also resulted in three band-shifting positions, suggesting that each oligomeric structure has RNA-binding capabilities.

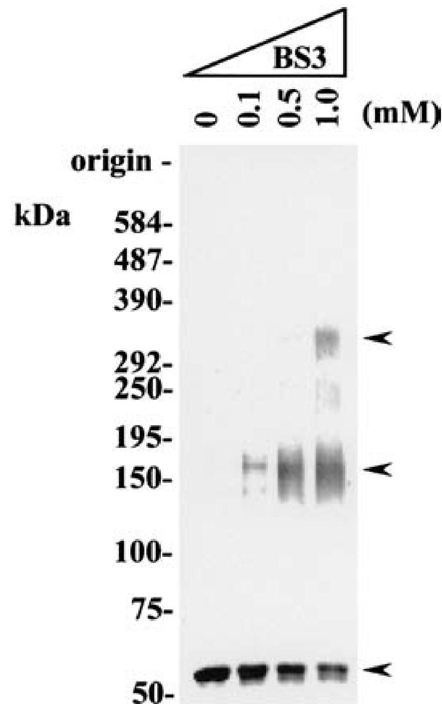

Figure 10 Chemical cross-linking analysis of the FAU-1H protein

FAU-1 protein (0.4 μM) was incubated with various concentrations of BS^3 (0–1.0 mM) at 22 °C for 20 min. Reactions were stopped and then analysed by SDS/PAGE [2–15 % (w/v) gels]. Bands were visualized by immunoblot analysis. Arrowheads indicate the oligomers formed. Molecular-mass markers (in kDa) are shown on the left.

To confirm the oligomeric structures of the FAU-1 protein, we conducted chemical cross-linking analysis. As shown in Figure 10, three distinct polypeptide species were found by cross-linking the purified FAU-1 protein with the BS^3 cross-linking reagent, and these were shown to be a monomer, trimer and hexamer. A band corresponding to the FAU-1 nonamer was very faint. In this cross-linking experiment, the trimer was shown to be the most abundant species.

Possible function for the FAU-1 protein

In the present study, we identified and characterized the RNA loop-binding protein FAU-1. The FAU-1 protein is characterized by highly charged domains consisting of both acidic and basic amino acids. Its N-terminal half shares a degree of similarity to RNase E from *E. coli* (25 %). In 1994, Wang and Cohen [17] reported that the human gene *ard-1* complements a temperature-sensitive mutant of the *rne* gene in *E. coli*. The *ard-1* gene product possesses site-specific single-stranded endoribonuclease activity that functionally resembles RNase E [18]. On the other hand, NIPP-1 was identified from bovine nuclear extracts as an

© 2003 Biochemical Society

inhibitor of protein phosphatase-1. cDNA cloning of NIPP-1 and analysis of its gene structure revealed that the ARD-1 and NIPP-1 peptides are isoforms encoded by a single gene and that they share the same, albeit alternatively spliced, precursor RNA [19,20]. We found that all four proteins, FAU-1, RNase E, ARD-1 and NIPP-1, contained highly charged amino acid clusters. However, these proteins had no significant amino acid similarities with one another, except between ARD-1 and NIPP-1. Moreover, it has been reported that NIPP-1 preferentially binds to AU-rich sequences [21] and is associated with pre-mRNA splicing factors [22,23]. A recent study [24] suggested that RNomics in archaea reveals a further link between splicing of archaeal introns and rRNA processing. In this respect, the FAU-1 protein may be involved in RNA metabolic processes, such as the degradation, processing or structure-dependent recognition of RNA. Because a genetic engineering system for *P. furiosus* is not yet available, analysis of FAU-1 protein function *in vivo* remains very difficult. However, our present findings may provide a unique example of a heat-stable RNA-binding protein.

We thank Dr Gota Kawai and Dr Taichi Sakamoto for critical reading of the manuscript. We also thank Professor Masaru Tomita for scientific encouragement.

REFERENCES

1 Gesteland, R. F., Cech, T. R. and Atkins, J. F. (1999). The RNA World, 2nd ed., Cold Spring Harbor Laboratory Press, NY

2 Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A. et al. (1998). Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. DNA Res. **5**, 55–76

3 Chinen, A., Uchiyama, I. and Kobayashi, I. (2000). Comparison between *Pyrococcus horikoshii* and *Pyrococcus abyssi* genome sequences reveals linkage of restriction-modification genes with large genome polymorphisms. Gene **259**, 109–121

4 Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B. and Dunn, D. M. (2001). Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology. Methods Enzymol. **330**, 134–157

5 Stukenberg, P. T., Lustig, K. D., McGarry, T. J., King, R. W., Kuang, J. and Kirschner, M. W. (1997). Systematic identification of mitotic phosphoproteins. Curr. Biol. **7**, 338–348

6 King, R. W., Lustig, K. D., Stukenberg, P. T., McGarry, T. J. and Kirschner, M. W. (1997). Expression cloning in the test tube. Science (Washington, D.C.) **277**, 973–974

7 Martzen, M. R., McCraith, S. M., Spinelli, S. L., Torres, F. M., Fields, S., Grayhack, E. J. and Phizicky, E. M. (1999). A biochemical genomics approach for identifying genes by the activity of their products. Science (Washington, D.C.) **286**, 1153–1155

8 Kaiser, K. and Murray, N. E. (1985). The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries. In DNA Cloning, A Practical Approach (D. M. Glover, ed.), pp. 1–47, IRL Press, Oxford

9 Clarke, L. and Carbon, J. (1976). A colony bank containing synthetic Col El hybrid plasmids representative of the entire *E. coli* genome. Cell (Cambridge, Mass.) **9**, 91–99

10 D'Orso, I. and Frasch, A. C. (2001). Functionally different AU- and G-rich *cis*-elements confer developmentally regulated mRNA stability in *Trypanosoma cruzi* by interaction with specific RNA-binding proteins. J. Biol. Chem. **276**, 15783–15793

11 Balmer, L. A., Beveridge, D. J., Jazayeri, J. A., Thomson, A. M., Walker, C. E. and Leedman, P. J. (2001). Identification of a novel AU-rich element in the 3′ untranslated region of epidermal growth factor receptor mRNA that is the target for regulated RNA-binding proteins. Mol. Cell. Biol. **21**, 2070–2084

12 Tebo, J. M., Datta, S., Kishore, R., Kolosov, M., Major, J. A., Ohmori, Y. and Hamilton, T. A. (2000). Interleukin-1-mediated stabilization of mouse KC mRNA depends on sequences in both 5′- and 3′-untranslated regions. J. Biol. Chem. **275**, 12987–12993

13 Baker, D. M., Wang, S. L., Bell, D. J., Drevon, C. A. and Davis, R. A. (2000). One or more labile proteins regulate the stability of chimeric mRNAs containing the 3′-untranslated region of cholesterol-7α-hydroxylase mRNA. J. Biol. Chem. **275**, 19985–19991

14 Xu, N., Chen, C. Y. and Shyu, A. B. (2001). Versatile role for hnRNP D isoforms in the differential regulation of cytoplasmic mRNA turnover. Mol. Cell. Biol. **21**, 6960–6971

15 Chou, P. Y. and Fasman, G. D. (1975). The conformation of glucagon: predictions and consequences. Biochemistry **14**, 2536–2541

16 Garnier, J., Osguthorpe, D. J. and Robson, B. (1978). Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J. Mol. Biol. **120**, 97–120

17 Wang, M. and Cohen, S. N. (1994)). *ard-1*: a human gene that reverses the effects of temperature-sensitive and deletion mutations in the *Escherichia coli rne* gene and encodes an activity producing RNase E-like cleavages. Proc. Natl. Acad. Sci. U.S.A. **91**, 10591–10595

18 Claverie-Martin, F., Wang, M. and Cohen, S. N. (1997). ARD-1 cDNA from human cells encodes a site-specific single-strand endoribonuclease that functionally resembles *Escherichia coli* RNase E. J. Biol. Chem. **272**, 13823–13828

19 Chang, A. C., Sohlberg, B., Trinkle-Mulcahy, L., Claverie-Martin, F., Cohen, P. and Cohen, S. N. (1999). Alternative splicing regulates the production of ARD-1 endoribonuclease and NIPP-1, an inhibitor of protein phosphatase-1, as isoforms encoded by the same gene. Gene **240**, 45–55

20 Van Eynde, A., Wera, S., Beullens, M., Torrekens, S., Van Leuven, F., Stalmans, W. and Bollen, M. (1995). Molecular cloning of NIPP-1, a nuclear inhibitor of protein phosphatase-1, reveals homology with polypeptides involved in RNA processing. J. Biol. Chem. **270**, 28068–28074

21 Jagiello, I., Beullens, M., Vulsteke, V., Wera, S., Sohlberg, B., Stalmans, W., von Gabain, A. and Bollen, M. (1997). NIPP-1, a nuclear inhibitory subunit of protein phosphatase-1, has RNA-binding properties. J. Biol. Chem. **272**, 22067–22071

22 Trinkle-Mulcahy, L., Ajuh, P., Prescott, A., Claverie-Martin, F., Cohen, S., Lamond, A. I. and Cohen, P. (1999). Nuclear organisation of NIPP1, a regulatory subunit of protein phosphatase 1 that associates with pre-mRNA splicing factors. J. Cell Sci. **112**, 157–168

23 Boudrez, A., Beullens, M., Groenen, P., Van Eynde, A., Vulsteke, V., Jagiello, I., Murray, M., Krainer, A. R., Stalmans, W. and Bollen, M. (2000). NIPP1-mediated interaction of protein phosphatase-1 with CDC5L, a regulator of pre-mRNA splicing and mitotic entry. J. Biol. Chem. **275**, 25411–25417

24 Tang, T. H., Rozhdestvensky, T. S., d'Orval, B. C., Bortolin, M. L., Huber, H., Charpentier, B., Branlant, C., Bachellerie, J. P., Brosius, J. and Huttenhofer, A. (2002). RNomics in Archaea reveals a further link between splicing of archaeal introns and rRNA processing. Nucleic Acids Res. **30**, 921–930

Received 19 December, 2002/3 March 2003; accepted 4 March 2003
Published as BJ Immediate Publication 4 March 2003, DOI 10.1042/BJ20021968

© 2003 Biochemical Society