Structure of the Yeast Ribosomal 5 S RNA-binding Protein YL3*

Received for publication, November 5, 1990

Baozhang Tang and Ross N. Nazar
From the Department of Molecular Biology and Genetics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

A gene coding for the 5 S rRNA-binding protein (YL3) in yeast (Saccharomyces cerevisiae) was isolated using a polymerase chain reaction-amplified gene probe. The DNA sequence contains no introns and codes for a 297 amino acid (M₀ = 33,741) protein. Although the protein is just 1 residue longer than in rat, unlike the high sequence homology in the 5 S rRNAs from the same organisms, only about 45% of the amino acid residues are conserved with surprisingly little homology in the carboxy-terminal end. Nevertheless, comparative studies indicate that a number of structural features are conserved including small repeats in the primary structure and a number of helical estimates in the higher order structure. One of the sequence repeats also appears to be present in the carboxy-terminal end of the eukaryotic transcription factor TFIIIA suggesting an evolutionary relationship in these 5 S RNA-binding proteins.

The ribosome, a cell's "machinery" for the synthesis of proteins is a ribonucleoprotein complex consisting of approximately equal amounts of ribosomal RNA and protein. First reported by Blobel (1) almost two decades ago, when treated with EDTA, eukaryotic cytoplasmic ribosomes dissociate into their constituent subunits and a small ribonucleoprotein complex consisting of the 5 S rRNA and a single ribosomal protein of approximately equal molecular weight, YL3 in yeast (2, 3). Although the role of either the 5 S rRNA or its cognate binding protein is unclear, both the RNA and 5 S rRNA protein complex have been attractive models for the study of RNA structure and its interactions with ribosomal protein (see Refs. 4-6). In fact, because the protein/RNA mass is similar to the whole ribosome and the complex is both ubiquitous and easy to prepare, this complex may be a very good model for the ribosome structure, in general.

Although hundreds of 5 S RNA sequences have been determined, studies on 5 S RNA-protein complexes have been largely restricted to Escherichia coli (see Ref. 4), yeast (2, 3), and rat (1, 7), and complete protein sequences have been determined only in E. coli (8-10) and rat (7) with the gene sequence also reported in the toad (11). Any comparisons are further complicated by the fact that three RNA-binding proteins form an apparently equivalent complex in E. coli (see Ref. 3). In addition, only protein-binding sites in the RNA have been identified (4, 12) and little is known about the interacting sites in the protein.

To gain a better understanding of this protein and its interaction with the 5 S RNA, the YL3 ribosomal protein gene from Saccharomyces cerevisiae was isolated in the present study and used to determine the complete primary structure of this 5 S RNA-binding protein. Common structural features are considered as potential nucleic acid-binding sites.

EXPERIMENTAL PROCEDURES

S. cerevisiae (strain S288C) DNA was prepared essentially as described by Cryer et al. (13), and a genomic library was constructed using the λ-Charon 4A vector (14). A specific probe for the yeast 5 S RNA-binding protein (YL3) was prepared using a polymerase chain reaction (15) together with two synthetic deoxyribobioleoligonucleotides which were synthesized using a model 8800 automated DNA synthesizer (Milligen/Biosearch, Novato, CA) to be complementary to DNA-coding sequence which was predicted from previously determined partial protein sequence (2, 3). Mixed bases were used where the genetic code was ambiguous. Thirty amplification cycles were applied consisting of a 1-min denaturation step at 94 °C, a 1-min annealing step at 55 °C, and a 3-min elongation step at 72 °C. The resulting 804-base pair fragment was purified on a 0.8% agarose gel (16), eluted by electrophoresis (17), and its 5' end was labeled with [γ-32P]ATP and T₄ polynucleotide kinase (18). The entire gene was isolated using plaque hybridization (19) and the 7-kilobase pair fragment genomic insert was subcloned in the pRS316 plasmid vector (20) for DNA sequencing.

The entire coding sequence was determined by the dideoxy methods of Sanger and co-workers (21). Both single and double strand sequencing methods were applied using the two polymerase chain reaction primers, and three other oligonucleotides which were synthesized to be complementary to sequences determined with the initial two primers (see Fig. 1).

RESULTS AND DISCUSSION

In the course of this study, three different probes were attempted to select the YL3 gene, complementary mixed base oligonucleotides (GC[G,T,A]TT[C,T]CA[G,A]AA[G,A]-GA[C,T]GC and GC[G,A]TA[C,T]TG[C,T]TC[C,T]TTGC, G,T,A[GT][G,A]AA[C,T,TT] which were synthesized using the known partial protein sequences (2, 9), a clone (7) containing the equivalent gene sequence in rat (kindly provided by Professor I. G. Wool, University of Chicago), and a specific gene fragment which was prepared by a polymerase chain reaction (Fig. 1). Despite numerous changes in the hybridization conditions lack of specificity in the first two methods made these entirely inadequate. Sufficient specificity in the polymerase chain reaction was obtained by adjusting the annealing temperature to 55 °C when the 804-base pair fragment became clearly the major component.

The nucleotide sequence encoding the YL3 protein was primarily determined using the polymerase chain reaction-amplified gene fragment and the initial two complementary oligonucleotides as sequencing primers. As indicated in Fig. 1, one additional primer was synthesized to complete this gene fragment sequence. While the protein sequence proximal and distal to the polymerase chain reaction-amplified fragment had been determined by protein sequencing (2, 3), this was confirmed and the gene sequence was also unambiguously determined using two additional synthetic primers which were
synthesized to be complementary to the polymerase chain reaction fragment but primed sequencing reactions into the external regions.

As shown in Fig. 2, the gene sequence was found to be entirely consistent with the short stretches of amino-acid carboxyl-terminal protein sequence which had previously been determined from protein fragments (2, 3). The entire protein sequence is 297 amino acids long, also closely consistent with nucleotides sequences for YL3 are presented together from rat and yeast.

The phylogenetic comparison which is included in Fig. 2 raises a number of interesting questions about this ribosomal protein and its interactions with the 5 S rRNA. While very few amino acid changes were found between the rat and toad sequence (11) with over 90% homology, the present comparison indicates that only about 45% of the amino acids are homologous in the yeast sequence. The lack of sequence homology is especially pronounced in the carboxyl-terminal where only 20% of the last 75 residues are identical when the sequences are aligned in a continuous fashion. Furthermore, even at the DNA level, the sequence remains surprisingly different. Some additional homology is evident when two insertions are allowed in the rat protein sequence (Fig. 3), but many differences clearly remain.

Assuming that any common structure is likely to be functionally important, the substantial differences in sequence were exploited as a means to identify potentially important regions.

Fig. 1. Isolation and determination of the nucleotide sequence for the YL3 protein. The expanded area indicates the region of the gene which was amplified by a polymerase chain reaction and used to probe a bank of EcoRI genomic fragment recombinants in X-Charon 4A. The lines below indicate the extent of DNA sequencing using the dideoxy methods of Sanger and co-workers (20); arrowheads indicate the sequencing primers.

Fig. 2. A comparison of ribosomal 5 S RNA-binding protein sequences from rat and yeast. The complete coding (upper) and protein (lower) sequences for YL3 are presented together with nucleotides (above) or amino acids (below) which are different in RL5. Identical amino acid residues are indicated with shading. The rat sequences are taken from Ref. 7.

Yeast Ribosomal 5 S RNA-binding Protein
Fig. 3. Maximized sequence homology in the carboxyl-terminal region of ribosomal 5 S RNA-binding proteins from rat and yeast. Differences in rat coding or protein sequences are identified above or below the YL3 sequences.

Fig. 4. Common structural features in estimates of the tertiary structure for ribosomal 5 S RNA-binding proteins from rat and yeast. Sequences were analyzed with DNAas (Hitachi Corp., San Francisco, CA) software using the predictive schemes of Chou and Fasman (12), Chou et al. (13), or Robson (14). The prediction of Chou et al. for the YL3 is presented. The structure which is predicted by the alternate schemes and which is also found in the rat protein is shaded. Solid circles indicate identical glycine residues in both proteins; solid lines indicate possible repeating elements as described in Fig. 5.

Fig. 5. A comparison of repeated sequence elements in YL3 and other 5 S RNA-binding proteins. The sequences for the rat, E. coli, and toad protein molecules were taken from Refs. 7, 8, and 26, respectively.

used estimates of tertiary structure (22-24) were applied to both the rat and yeast sequences using the DNAas software versions (Hitachi Corp., San Francisco, CA), a number of common features particularly potential helices were predicted with all three approaches. Some of these (e.g. the helical structure at the carboxyl-terminal end) involve poorly conserved sequence regions presumably allowing for the substantial sequence differences in these two functionally equivalent proteins. The most intriguing of the conserved features in the primary structure itself is the distribution of certain amino acid residues, particularly glycine. As indicated in Fig. 5, a high glycine density in the yeast protein at positions is conserved in rat and two more are intermediate neighbors in the respective primary sequences. Most are concentrated in the center of the molecule closely associated with much of the conserved tertiary structure.

In addition to similarities in the tertiary structure and glycine residues, several small but repeating features in the primary structure also appear to be conserved. As shown in Fig. 5, an MIV sequence in the rat 5 S RNA-binding protein (residues 51-53) is repeated 20 residues later while in the yeast protein an LVV sequence (residues 51-53) repeats as VVL, 20 residues later. Furthermore, in one of the E. coli 5 S RNA-binding proteins, EL18, an LVV sequence also repeats as EVL, 20 residues later. Similarly, a neighboring CAAY sequence in the rat 5 S RNA-binding protein (residues 76-79) is repeated as an AAAY sequence, 26 residues later, while AAAY is repeated as AAAY in the yeast protein. Again, in EL18, a VAAS sequence in the equivalent position is repeated as AAAS.

Because the eukaryotic transcription factor TFIIIA has an affinity for both the 5 S RNA gene as well as the 5 S RNA transcript (25,26) this sequence was also compared with YL3. Much of the TFIIIA sequence has been shown to consist of repeating units or "zinc fingers" which interact with the 5 S rDNA (26) and bear no sequence homology to the 5 S rRNA (26). In one of the 5 S RNA-binding proteins, EL18, an LVV sequence also repeats as EVL, 20 residues later. Similarly, a neighboring CAAY sequence in the rat 5 S RNA-binding protein (residues 76-79) is repeated as an AAAY sequence, 26 residues later, while AAAY is repeated as AAAY in the yeast protein. Again, in EL18, a VAAS sequence in the equivalent position is repeated as AAAS.

The similarities to EL18 also raises further questions regarding the relationship of YL3 and EL18 to the multiple 5 S RNA-binding proteins which have been reported in bacteria. Based on partial protein sequences and the similarities in the chemical and physical properties of the 5 S RNA containing ribonucleoprotein complexes, we previously suggested (3) that the three smaller proteins in E. coli (EL5, EL18, and EL25) may be equivalent to the two larger proteins in an archaebac-
Fig. 6. Putative sequence homology between the YL3 ribosomal 5 S RNA-binding proteins from yeast and EL18 from E. coli. The complete sequence for EL18 (8) is presented (lower) together with putatively homologous regions in the YL3 protein (above). Identical residues are shaded.

Yeast Ribosomal 5 S RNA-binding Protein


REFERENCES


The fact that substantial additional sequence homology was not identified is not surprising given the significant differences between the rat and yeast protein sequences and further sequences from divergent origins clearly will be required to firmly establish any evolutionary relationships.

While the present study did not determine RNA-binding sites or functional domains in the eukaryotic ribosomal 5 S RNA-binding protein, the comparisons underline potentially important features. Because yeast can be genetically transformed much more readily, the YL3 protein provides for an additional experimental approach toward the structure and function of this protein. Recently, in vivo site-specific mutagenesis has been effectively applied to studies on the structure and function of the yeast 5 S rRNA (27) and can now be applied, in parallel, to the yeast 5 S RNA-binding protein.