2166–2167 Nucleic Acids Research, 1994, Vol. 22, No. 11

A novel RNA-binding motif in omnipotent suppressors of translation termination, ribosomal proteins and a ribosome modification enzyme?

Eugene V.Koonin*, Peer Bork[1,2] and Chris Sander[1]
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA, [1]Molecular Biology Laboratory, Meyerhofstrasse 1, D-69012 Heidelberg and [2]Max-Delbruck-Center for Molecular Medicine, D-13189 Berlin, Germany

Received March 15, 1994; Accepted April 25, 1994

ABSTRACT

Using computer methods for database search, multiple alignment, protein sequence motif analysis and secondary structure prediction, a putative new RNA-binding motif was identified. The novel motif is conserved in yeast omnipotent translation termination suppressor SUP1, the related DOM34 protein and its pseudogene homologue; three groups of eukaryotic and archaeal ribosomal proteins, namely L30e, L7Ae/S6e and S12e; an uncharacterized *Bacillus subtilis* protien related to the L7A/S6e group; and *Escherichia coli* ribosomal protein modification enzyme RimK. We hypothesize that a new type of RNA-binding domain may be utilized to deliver additional activities to the ribosome.

Mutations in at least five essential yeast genes result in the 'omnipotent suppressor phenotype', i.e. suppression of translation termination at all three stop codons (1). *SUP44* encodes ribosomal protein S4 (2); *SUP46* encodes ribosomal protein S13 (ref. 3; synonyms Ys11 and Yp28); and *SUP35* encodes a protein containing a domain related to elongation factor EF1α (4). *SUP1* (synonyms *SAL4, SUP45, YBR11.20*) gene product is associated with the 40S ribosome subunit and is thought to be involved in translation termination (5,6). Genes encoding highly conserved homologues of SUP1 have been identified in man, *Xenopus laevis* and Arabidopsis (7–9). SUP1 proteins are distantly related to the product of yeast gene *DOM34* (chromosome XIV) and its pseudogene homologue on chromosome III (10,11). Alignment of the DOM34 and SUP1 sequences (11) scored 8.4 SD on a 237 amino acid residue overlap, which is indicative of a genuine relationship (12).

A non-redundant amino acid sequence database (National Center for Biotechnology Information, NIH) was searched for similarity to the SUP1-related proteins using the BLAST program (13). The output of the BLAST search was analyzed for mutually consistent alignments; multiple alignment blocks constructed from conserved segments of such alignments were converted to position-dependent matrices and used for further database search (R.L.Tatusov, S.F.Altschul and E.V.K., unpublished). A conserved counterpart to one of these motifs was identified in three groups of eukaryotic and archaeal ribosomal proteins, namely L30e, L7Ae/S6e and S12e; and in *Escherichia coli* protein RimK. We also found that an uncharacterized *Bacillus*

```
sec. structure          1111     hhhhhhhhhhh1111bbbbb111
DOM34 yeast        216  NKDDDKAWYGEKEVVKAAEYGAISYLLLTDKV   63 P33309
DOM34R(III) yeast    ?  SKDDNKAWYGAEETERAAKLDAIETLLITDSV   ? X59720G
SUP1  yeast        291  SQDTGKFCYGIDDTLKALDLGAVEKLIVFENL 114 P12385
SUP1  Arabidopsis  294  SQDTGKYVFGVEDTLKALEMGAVETLIVWENL 109 S31328P
SUP1  Xenopus      294  SQDTGKYCFGVEDTLKALEMGAVEILIVYENL 111 S31633P

RPL30 human         20  VMKSGKYVLGYKQTLKMIRQGKAKLVILANNC  63 P04645
RPL32 yeast         16  VIKSGKYTLGYKSTVKSLRQGKSKLIIIAANT  57 P14120
RPL30 T.cruzi       10  AQDTGKIVMGARKSIQYAKMGGAKLIIVARNA  59 P29160
RPL30 M.vannielii   15  AVDTGNVVLGTKQAIKNIKHGEGKLVIIAGNC  58 P14025
RPL30 S.acidocald.  11  LLRSGKVILGTRKTLKLLKTGKVKGVVVSSTL  61 P11522

RPS6  H.marismortui 23  ARDTGAVKKGTNETTKSIERGSAELVFVAEDV  62 P12743
NHP2  yeast         66  ASKAKNVKRGVKEVVKALRKGEKGLVVIAGDI  75 P32495
RPL4A yeast        125  SPKPYAVKYGLNHVVALIENKKAKLVLIANDV 113 P17076
YIF4  B.subtilis    12  ANRARKVVSGEDLVIKEIRNARAKLVLLTEDA  56 P32729
RPL7A rice         121  AKKPIVVKYGLNHVTYLIEQSKAQLVVIAHDV 105 D12631G
RPL7A human        129  TKRPPVLRAGVNTVTTLVENKKAQLVVIAHDV 105 P11518

RPS12 T.brucei      39  ARETNGLICGLSEVTRALDRRTAHLCVLADDC  73 S24781P
RPS12 rat           24  ALIHDGLARGIREAAKALDKRQAHLCVLASNC  76 P09388

RimK  E.coli       117  TSDLIDMVGGAPLVVKLVEGTQGIGVVLAETR 143 P17116
consensus               ......O..G.......OJ......OUU..Z.
```

Figure 1. Conserved motif in omnipotent suppressor gene products, three groups of ribosomal proteins and RimK. The alignment of the 32 residue protein segments was generated by database search with a position-dependent weight matrix as described in the text. Several highly conserved sequences of ribosomal proteins are omitted. The sequence of the human homologue of SUP1 (7) is not shown as it appeared to contain multiple frameshift errors in the region coding for the conserved motif. Distinct groups of proteins, namely the SUP1 family; the L30e family; the L7Ae/S6e family; the S12e family; and RimK, are separated by blank lines. In the L30e family, the conserved region includes the previously derived PROSITE signature (PS00709; ref. 20). In the consensus, O designates a hydrophobic residue (I,L,V,M,F,Y,W,C,A), U designates a bulky hydrophobic residue (I,L,V,M,F), Z designates a polar residue (K,R,D,E,S,T,N,Q,H), J designates a charged residue (K,R,D,E). The secondary structure is based on the information in all sequences in the multiple alignment (14). h designates α-helix, b designates β strand, and 1 designates loop; no symbol is shown for positions where the prediction was uncertain. For each sequence, the distances from the protein ends are indicated by numbers. The sequences were from the current databases and are accompanied by their accession numbers in SwissProt, PIR (P) or GenBank (G). DOM34(III) is the pseudogene homologue of DOM34 encoded on chromosome III (10,11). NHP2 is a nuclear protein related to the L7Ae/S6e family of ribosomal proteins (21).

* To whom correspondence should be addressed

subtilis protein belongs to the L7Ae/S6e group (Fig. 1). Only one glycine residue is strictly conserved in all these proteins, but several positions are occupied by physico–chemically related residues. The region shown in Fig. 1 was independently identified as the most conserved block in multiple alignments constructed separately for the SUP1-related proteins and three groups of ribosomal proteins using the MACAW program (15). The similarity between SUP1 and ribosomal proteins is evident from both a pairwise comparison and a motif search: the probability of the SUP1 homologue from Arabidopsis and archaeal S6 protein matching by chance is about 3.5×10^{-3}; and, a position-dependent matrix constructed from the ribosomal protein alignment selectively retrieved all the SUP1-related sequences from the database (data not shown). These observations suggest that despite the scarcity of invariant amino acids, the alignment in Fig. 1 represents a conserved structural unit.

Secondary structure prediction (14) suggested conservation of an α-helix and a β-strand in SUP1 and the related proteins (Fig. 1). In those ribosomat proteins whose three-dimensional structure has been resolved, the RNA-binding domain has a mixed α/β structure, with characteristic, conserved hydrophobic β-strands (16–18). The ribosomal association of SUP1 may be through RNA binding and the alignment in Fig. 1 may therefore represent a new type of RNA-binding motif that is shared by SUP1 proteins and several groups of ribosomal proteins. The structure and function of the distinct N-terminal domain of SUP1 that is not related to any known proteins remain to be determined. Another omnipotent suppressor gene product, SUP35 also has a two-domain organization (4).

RimK is an enzyme that catalyzes addition of glutamic acid residues to the N-terminus of *E.coli* ribosomal protein S6 (19). Our finding of a motif that is conserved between RimK and several groups of ribosome-associated proteins raises the intriguing possibility that RimK has to be specifically positioned on the ribosome through binding to rRNA, in order to modify S6.

These findings indicate that all known yeast omnipotent suppressor genes encode ribosomal proteins or ribosome-associated factors and that a new type of RNA-binding domain may be used to deliver additional activities to the ribosome.

REFERENCES

1. Eustice,D.C. *et al.* (1986) *J. Mol. Biol.* **188**, 207–214.
2. All-Robyn,J.A. *et al.* (1990) *Mol. Cell. Biol.* **10**, 6544–6553.
3. Vincent,A. and Liebman,S.W. (1992) *Genetics* **132**, 375–386.
4. Didichenko,S.A., Ter-Avanesyan,M.D. and Smirnov,V.N. (1991) *Eur. J. Biochem.* **198**, 705–711.
5. Breining,P. and Piepersberg,P. (1986) *Nucleic Acids Res.* **14**, 5187–5197.
6. Stansfield,I. *et al.* (1992) *Mol. Microbiol.* **6**, 3469–3478.
7. Grenett,H.E., Bounelis,P. and Fuller,G.M. (1992) *Gene* **110**, 239–243.
8. Tassan,J.P. *et al.* (1993) *Mol. Cell. Biol.* **13**, 2815–2821.
9. Quigley,F.R. (1992) GenBank X69375.
10. Lalo,D. *et al.* (1993) *C. R. Acad. Sci. Paris, Sciences de la vie* **316**, 367–373.
11. Koonin,E.V., Bork,P. and Sander,C. (1994) *EMBO J.* **13**, 493–503.
12. Gorbalenya,A.E. *et al.* (1989) *J. Mol. Evol.* **28**, 256–268.
13. Altschul,S.F. *et al.* (1990) *J. Mol. Biol.* **215**, 403–410.
14. Rost,B. and Sander,C. (1993) *J. Mol. Biol.* **232**, 584–599.
15. Schuler,G.D., Altschul,S.F. and Lipman,D.J. (1991) *Proteins* **9**, 180–190.
16. Wilson,K.S. *et al.* (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7251–7255.
17. Leijonmark,M. and Liljas,A. (1987) *J. Mol. Biol.* **195**, 555–580.
18. Leijonmark,M. *et al.* (1988) *Proteins* **3**, 243–251.
19. Kang,K.W. *et al.* (1989) *Mol. Gen. Genet.* **217**, 281–288.
20. Bairoch,A. (1993) *Nucleic Acids Res.* **21**, 3097–3103.
21. Kolodrubetz,D. and Burgun,A. (1991) *Yeast* **7**, 79–90.