Vol. 198, No. 1, 1994
January 14, 1994

BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
Pages 288-291

EXPRESSED SEQUENCE TAGS IDENTIFY A HUMAN ISOLOG OF THE *SUI1* TRANSLATION INITIATION FACTOR

Chris Fields and Mark D. Adams

The Institute for Genomic Research
932 Clopper Road
Gaithersburg, MD 20878

Received November 22, 1993

The complete cDNA sequence of a human isolog of the yeast *sui1* translation initiation factor gene was obtained by assembling over 40 expressed sequence tags (ESTs) for this gene obtained from a variety of tissue-specific cDNA libraries. The human *sui1iso1* gene product is a 113 amino-acid polypeptide similar to proteins known from yeast, rice, mosquito, and *Methanococcus*. The identification of *sui1iso1* illustrates the utility of assemblies of independent ESTs for deriving full-length cDNA sequences for new human genes.

The *sui1* gene of *Saccharomyces cerevisiae* encodes a novel translation initation factor that appears to interact with eIF2 to register the initator tRNA-Met at an AUG codon (1). Elements of the initiation complex are expected to be well-conserved throughout eukaryotes; however, only two apparent isologs of *sui1* have been reported, in rice (2) and mosquito (3). Here we describe a human isolog of *sui1*, and show that *sui1* and its isologs are related to the product of an uncharacterized open-reading frame in the ribosomal protein operon of the archae *Methanococcus vannielii* (4).

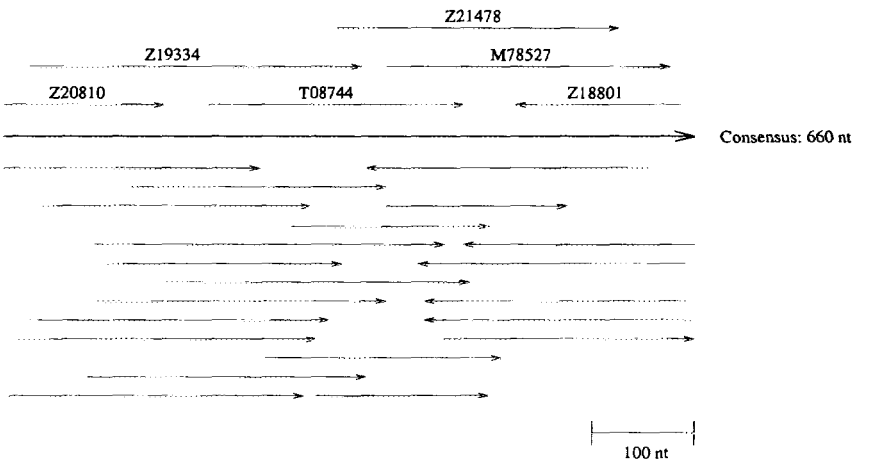
Methods

Expressed sequence tags (ESTs) were sequenced from cDNA clones obtained from 20 human tissue-specific cDNA libraries as described (5). Additional EST sequences obtained by other laboratories were identified by searching GenBank or dbEST (6) using BLAST (7) with ESTs sequenced in this laboratory. The sequences were assembled using an alignment editor, and a six-frame translation compared with the protein sequence database using BLAST. The inferred amino-acid sequence of the *sui1iso1* product was also compared with the protein database using BLAZE (8).

Results and Discussion

The human *sui1iso1* cDNA sequence, its predicted translation, and a schematic representation of the overlapping EST sequences from which it was derived are shown in Fig. 1. The

0006-291X/94 $5.00




```
  1 gccgccgcga ggattcagca gcctcccccc tgagcccccc cgcttcccga cgttccgttc

 61 ccccctgccc gccttctccc gccaccgccg ccgccgcctt ccgcagccgt ttccaccgag

                        M   S   A   I   Q   N   L   H   S   F   D   P   F   A
121 gaaaaggaat cgtatcgtat gtccgctatc cagaacctcc actctttcga ccccttgct

      D   A   S   K   G   D   D   L   L   P   A   G   T   E   D   Y   I   H   I   R
181 gatgcaagta agggtgatga cctgcttcct gctggcactg aggattatat ccatataaga

      I   Q   Q   R   N   G   R   K   T   L   T   T   V   Q   G   I   A   D   D   Y
241 attcaacaga gaaacggcag gaagaccctt actactgtcc aagggatcgc tgatgattac

      D   K   K   K   L   V   K   A   F   K   K   K   F   A   C   N   G   T   V   I
301 gataaaaaga aactagtgaa ggcgtttaag aaaaagtttg cctgcaatgg tactgtaatt

      E   H   P   E   Y   G   E   V   I   Q   L   Q   G   D   Q   R   K   N   I   C
361 gagcatccgg aatatggaga agtaattcag ctacagggtg accaacgcaa gaacatatgc

      Q   F   L   V   E   I   G   L   A   K   D   D   Q   L   K   V   H   G   F   *
421 cagttcctcg tagagattgg actggctaag gacgatcagc tgaaggttca tgggttttaa

481 gtgcttgtgg ctcactgaag cttaagtgag gatttccttg caatgagtag aatttccctt

541 ctctcccttg tcacaggttt aaaaacctca cagcttgtat aatgtaacca tttggggtcc

601 gcttttaact tggactagtg taactccttc atgcaataaa ctgaaaagag ccatgcaaaa
                                    ------                    ----
```

Fig. 1. Upper panel: Assembly of the human *suilisol* cDNA sequence from overlapping ESTs. ESTs drawn above the line representing the consensus sequence are in GenBank and are labelled by their accession numbers; those below the consensus were subsequently obtained in our laboratory. An additional 14 ESTs covering the 5' half of the sequence are not shown in the figure. Lower panel: Consensus sequence of human *suilisol* cDNA and predicted polypeptide translation. The AATAAA polyadenylation signal and the first 4 nucleotides of polyA are underlined. ESTs including over 20 nucleotides of polyA (e.g., Z18801) were included in the assembly.

average redundancy of the assembly is greater than 10-fold over most of the sequence. The assembled *suilisol* cDNA sequence has been assigned Genome Sequence DataBase accession number L26247.

The *suilisol* gene is expressed in a wide variety of human tissues, as measured directly by EST sequencing from tissue-specific libraries. The previously-published ESTs derive from fetal brain (M78527; 9), testis (Z21478 and Z21478), fibroblast (Z18801), liver (Z20810), and

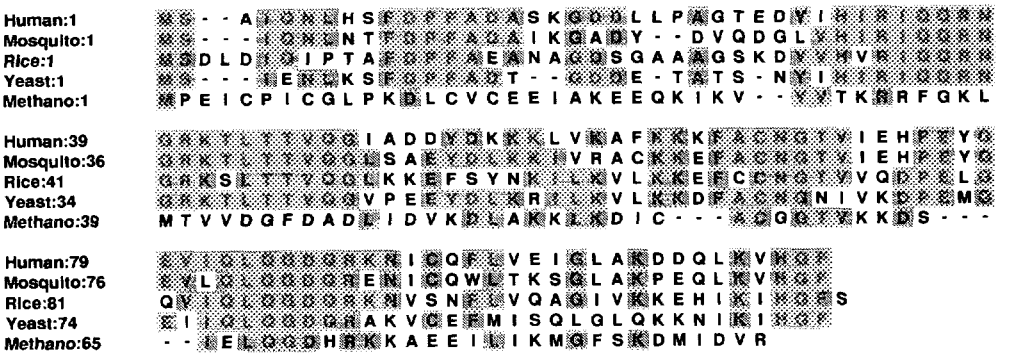


Fig. 2. Multiple alignment of sequences of the predicted protein products of human, mosquito (3), and rice (2) isologs of *sui1* with the sequences of the yeast *sui1* gene product (1) and the product of the *A* open-reading from from the ribosomal protein operon of *Methanococcus vannielii* (4). Amino acids conserved in at least 3 of the 5 sequences are shaded.

skeletal muscle (Z19334). The additional ESTs obtained here represent 20 different whole embryo or fetal, infant, or adult tissue-specific libraries from brain, skin, heart, fat, and endocrine tissues. These data suggest that *sui1iso1* may be a ubiquitously-expressed housekeeping gene.

The sequence of the predicted protein product of the *sui1iso1* gene is aligned with the sequences of the predicted products of *sui1* (1), the rice (2) and mosquito (3) isologs of *sui1*, and the predicted product of the uncharacterized *A* open-reading frame of the *Methanococcus vannielii* ribosomal protein operon (4) in Fig. 2. These are the only known protein sequences that are significantly similar to the *sui1iso1* product, as detected with either BLAST or BLAZE. The four eukaryotic sequences are essentially equally similar over their entire lengths. The amino acids DQ in the context LQGDQR (positions 83 and 84 in *sui1*) are altered in *sui1* mutant alleles (1). These positions are conserved in all the eukaryotic sequences; Q(84) is replaced by H in the *Methanococcus* sequence, but D(83) and much of the context is conserved. The high degree of similarity between the eukaryotic isologs of *sui1* suggests that they may all function as initiation factors; the conservation of what appears to be a functionally-critical context in the *Methanococcus* protein, and the location of its gene in the ribosomal protein operon, suggest that the *Methanococcus* isolog of *sui1* may encode an initiation or similar translation-facilitating factor as well.

References

1. Yoon, H., and Donahue, T. F. (1992) Mol. Cell. Biol. 12, 248-260.
2. de Pater, B. S., Schilperoort, R. A. (1991) GenBank Accession X51910 (unpublished).
3. Besansky, N. J., Mills Hamm, D., and Collins, F. H. (1993) GenBank Accession U02588 (unpublished).
4. Auer, J., Spicker, G., and Bock, A. (1989) J. Molec. Biol. 209, 21-36.

5. Adams, M., Kelley, J., Gocayne, J., Dubnick, M., Polymeropolous, M., Xiao, H., Merril, C., Wu, A., Olde, B., Moreno, R., Kerlavage, A., McCombie, W., and Venter, J. C. (1991) Science 252, 1651-1656.

6. Boguski, M., Lowe, T., and Tolstoshev, C. (1993) Nature Genet. 4, 332-333.

7. Altschul, S., Gish, W., Miller, W., Myers, E., and Lipman, D. (1990) J. Molec. Biol. 215, 403-410.

8. Brutlag, D. L., Dautricourt, J. P., Diaz, R., Fier, J., Moxon, B. and Stamm, R. (1993) Comp. Chem. 17, 203-207.

9. Adams, M., Dubnick, M., Kerlavage, A., Moreno, R., Kelley, J., Utterback, T., Nagle, J., Fields, C., and Venter, J. C. (1992) Nature 355, 632-634.