Volume 313, number 2, 95–97 FEBS 11752 November 1992

Hypothesis

Segments of *Escherichia coli* genome similar to the exons of human prothymosin α gene

Andrey Vartapetian, Nina Chichkova, Ilya Lyakhov, Tatyana Makarova, Alexandra Evstafieva and Alexey Bogdanov

Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119899, Russia

Received 10 September 1992

Identification of the putative prothymosin α homolog in *Escherichia coli* cells prompted the search for a prothymosin α-coding gene in the *E. coli* genome. A set of interspersed DNA segments was identified, which match various parts of the human prothymosin α molecule. Their location in the *E. coli* genome and high degree of similarity with the appropriate regions of the human prothymosin α gene suggest that some kind of *trans*-splicing should exist in *E. coli*, which could be responsible for bringing these putative bacterial prothymosin α-coding exons together.

Prothymosin α; Nucleotide sequence similarity; *Trans*-splicing; *Escherichia coli*

1. INTRODUCTION

Among immunoactive proteins derived from thymus is prothymosin α, an extremely acidic protein of 13 kDa initially classified as a precursor of the putative thymic hormone, thymosin α1 [1]. However, its high evolutionary conservation, from yeast to man [2], as well as its wide tissue distribution [3], argues for some general and fundamental role of prothymosin α, rather than an exclusively immunological one. Evidence is accumulating that prothymosin α participates in the mammalian cell division process [4], and expression of its gene is regulated by the MYC oncoprotein [5].

Here we report data concerning the existence and peculiar organization of potential prothymosin α-coding sequences in the genome of *Escherichia coli*. Our data suggest that a process resembling eukaryotic *trans*-splicing is required to produce mature prothymosin α mRNA in eubacteria.

2. RESULTS AND DISCUSSION

In the course of a search for a prothymosin α-related protein in bacteria, we have identified, in *E. coli* cells, a protein with molecular weight, physico-chemical properties and peptide map very similar to that of human prothymosin α (T.M. and A.V., unpublished observations). Its negligible amount in the bacterial cell, however, hindered structural investigation of the protein. On the other hand, blot hybridization analysis of *E. coli* DNA revealed a number of DNA segments which hybridized, although only moderately, with the human prothymosin α cDNA probe, suggesting the existence of a number of short prothymosin α-like sequences in the *E. coli* genome, rather than one continuous sequence.

Three of these prothymosin α-like sequences were analyzed in detail. One of them, termed region of similarity #1 (see Fig. 1a), is 31 bp long and has 87% nucleotide identity with the central portion of the human prothymosin α-coding sequence. The sequences flanking this region of similarity had nothing in common with either the prothymosin α sequence, or any other sequence in the EMBL database. Most notably, the left boundary of the identified region of similarity corresponds precisely to the point of the intron–exon junction in the human prothymosin α gene [6], making it tempting to speculate that some kind of splicing may also operate in *E. coli* to produce mature prothymosin α mRNA, in case other 'exons' also exist.

275 bp upstream from region #1 yet another short prothymosin α-related sequence was identified. This region of similarity #2 (see Fig. 1b) has 82% nucleotide identity with the N-terminal prothymosin α sequence, including the initiator ATG codon. It turned out, however, that the direction of transcription of region #2 should be opposite to that of region #1 (see Fig. 1c). In accordance with this, region #2 is preceded by putative

Correspondence address: A. Vartapetian, Belozersky Institute of Physico-Chemical Biology, Moscow State University, Moscow 119899, Russia. Fax: (7) (095) 939 3181.

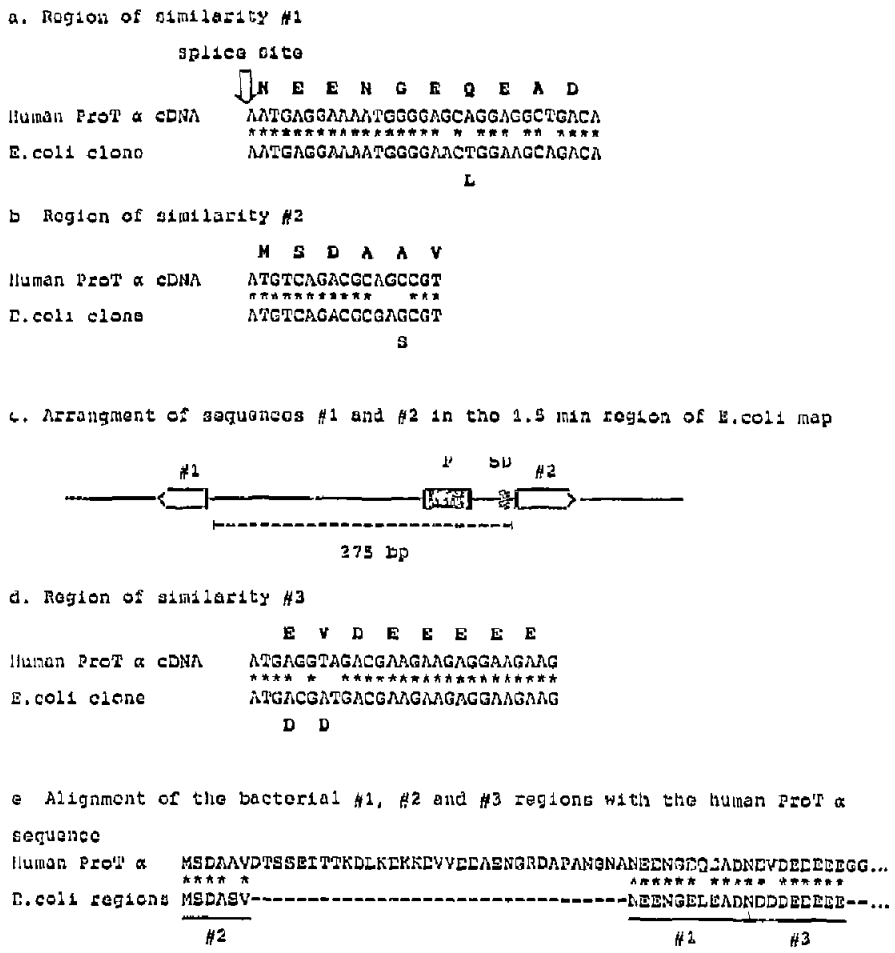

Fig. 1. Comparison of human prothymosin α cDNA and protein sequences [3] with nucleotide and deduced amino acid sequences of *E. coli* (MRE600 strain) segments #1, #2, #3. Identical nucleotide and amino acid residues are indicated by asterisks. P and SD denote putative promoter and Shine–Dalgarno sequences, respectively. The amino acids deduced from bacterial sequences that differ from those in human prothymosin α are shown below the nucleotide sequences in **a**, **b** and **d**. Hyphens in **e** indicate the regions of human prothymosin α sequence for which cognate bacterial counterparts have not yet been identified; conservative change is indicated by a dot. Protein sequences are shown in the single-letter amino acid code.

promoter-like and Shine–Dalgarno (ribosome binding) sequences, which may contribute to transcription of segment #2 and translation of the resultant transcript. Again, the sequence downstream from region #2 has nothing in common with the prothymosin α sequence. Regions of similarity #1 and #2 were located at 1.5 min on the *E. coli* map with the help of Gene Mapping Membrane (Takara Shuzo Co.).

The third potential prothymosin α-coding segment, termed region of similarity #3, has 89% nucleotide identity with the human prothymosin α-coding sequence immediately downstream from that coincident with region #1 (see Fig. 1d and e). Moreover, linkage of regions #1 and #3 reconstitutes the codon for asparagine at the site of the junction (Asp^{50}, Fig. 1e), which is present in human prothymosin α. As if the situation were not sufficiently astonishing, region #3 was located at 86 min on the *E. coli* map, making it highly unlikely that regions #1 and #3 could be contained within a single primary transcript.

Regions of similarity #1, #2 and #3 together comprise about 23% of the total human prothymosin α length with only four amino acid mismatches (see Fig. 1e).

Two possibilities are feasible to account for the above observations. One interpretation is that the identified prothymosin α-like sequences may not belong to the genuine prothymosin α homolog, but instead to some kind of pseudogene or anonymous bacterial protein which have a domain in common with prothymosin α. This, however, does not explain the origin of the prothymosin α-like protein identified in *E. coli* cells. Alternatively, the regions of similarity may belong to an authentic bacterial prothymosin α homolog, the gene of which is arranged in some peculiar way so that the prothymosin α-coding sequences are interspersed in the

genome. Since no chromosomal protein-coding genes of eubacteria arranged in this eukaryotic-like fashion are known, this explanation seems rather queer. Nevertheless, location of the identified segments in the *E. coli* genome and high degree of similarity with the corresponding regions of the human prothymosin α gene suggest that some kind of *trans*-splicing should exist in *E. coli*, which could be responsible for bringing putative bacterial prothymosin α-coding exons together.

Acknowledgements: We thank A. Noda (Takara Shuzo Co., Japan) for providing a sample of the Gene Mapping Membrane This work was supported by a grant from the Russian 'Human Genome' Programme.

REFERENCES

[1] Haritos, A.A., Blacher, R., Stein, S., Caldarella, J. and Horecker, B.L. (1985) Proc. Natl. Acad. Sci. USA 82, 343–346.
[2] Makarova, T., Grebenshikov, N., Egorov, T., Vartapetian, A. and Bogdanov, A. (1989) FEBS Lett. 257, 247–250.
[3] Eschenfeldt, W. and Berger, S.L. (1986) Proc. Natl. Acad. Sci. USA 83, 9403–9407.
[4] Sburlati, A.R., Manrow, R.E. and Berger, S.L. (1991) Proc. Natl. Acad. Sci. USA 88, 253–257.
[5] Eilers, M., Schirm, S. and Bishop, J.M (1991) EMBO J. 10, 133–141.
[6] Eschenfeldt, W.H , Manrow, R.E., Krug, M.S. and Berger, S.L. (1989) J. Biol. Chem. 264, 7546–7555.