Evidence for transcription and potential translation of the human 1.9 kb HindIII repetitive element

Bruce A. Citron, Prabhakara V. Chaudary, Desirazu N. Rao and Seymour Kaufman

Laboratory of Neurochemistry, National Institute of Mental Health and Molecular and Medical Genetics Section, Developmental and Metabolic Neurology Branch, National Institute of Neurological and Communicative Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

Received 15 December 1985; Revised 24 February 1986, Accepted 25 February 1986

ABSTRACT
Recombinant cDNA clones corresponding to the human 1.9kb HindIII repetitive element have been isolated from a cDNA library of liver cytoplasmic polyadenylated RNA. These cDNAs share 95% homology with the reported genomic DNA sequence and a similar amount of homology at the amino acid level with putative coding sequences (see preceding article by Mottez et al). They were isolated as two of four false positives from a human cDNA library in Agtll and were selected with an antibody to an unrelated enzyme. These results provide direct evidence that this repetitive element is transcribed to form poly(A)⁺ RNA which could be translatable. Also, these observations may add to our understanding of the sources of false positives which are frequently observed in screens of cDNA libraries with antibodies as probes.

INTRODUCTION
Repetitive DNA elements are sequences ranging in size from a few hundred bp to several kb that occur several thousand times in eukaryotic genomes (1). They are usually identified by the characteristic restriction fragments produced by their digestion and grouped into families named after the restriction enzymes. In spite of their ubiquitous occurrence, the function of these elements has yet to be elucidated.

Genomic clones representing a variety of families from different organisms have been studied (1). The assignment of a phenotype associated with these elements is made nearly impossible by their repetitive nature (where the loss of one copy would not be expected to have an effect) or their lack of a function at all (2). Since no protein product seemed important (at least in the species present today) researchers have not directed their attention, in detail, to cytoplasmic RNA molecules.

The 1.9 kb HindIII repeat, in humans, was discovered and characterized by Manuelidis (3) and found to be repeated about 3500 times (4,5). With the absence of any biochemical evidence
Figure 1: Orientation of cDNA clones with respect to humrsh3. Positions are noted in base pairs. "**"s indicate nonsense codons which are postulated to be absent in the fully functional copy of this repetitive element. In SHR4 the deletion of a T residue at position 258 would cause a switch to an open frame and avoid the opal codon at position 338. In humrsh3, the amber codon at position 878 would not be in frame if there were a T inserted at position 803. Also, there are three ochre codons between positions 158 and 208 in humrsh3. There are a variety of slight changes in the sequences that could keep these reading frames open in different ways. The drawing is not to scale.

for expression of this repeat, a structural role was first suggested although it was noted that there was a high degree of conservation of these elements during evolution (3).

Here, evidence concerning the expression of the HindIII repetitive element is reported. The two clones described here were obtained during routine screening of human cDNA libraries. The implications of detecting clones containing repeat sequences is discussed.

MATERIALS AND METHODS

A cDNA library, in λgt11, containing poly(A)+ mRNA from the human liver (Meloy Laboratories) was screened according to Young and Davis (6) with the use of sheep antibodies raised against purified phenylalanine hydroxylase. The procedures used in the manipulation of DNA were essentially as outlined by Maniatis et al. (7) and were performed in accordance with the NIH guidelines for recombinant DNA. The DNA from the inserts of the positive clones was sequenced with the use of the Maxam and Gilbert technique (8) and computer analyzed (9-11). The sequences have been submitted to GenBank (12).
RESULTS AND DISCUSSION

A cDNA library made in the expression vector λgt11 (6) with poly(A){sup+} RNA from human liver was screened with polyclonal antibodies to phenylalanine hydroxylase (6). Four positive clones were isolated and plaque purified. Selective parts of the DNA inserts were sequenced (figures 1 and 2). None of the sequences shared any homology with the published nucleotide/amino acid sequence for phenylalanine hydroxylase (13), the desired enzyme.

The sequences obtained were compared to the GenBank nucleic acid and Dayhoff protein databases (12). The only homologies that could be uncovered were between two of the clones that we isolated and the GenBank entry humrsh3, the 1.9 kb HindIII repetitive element (3). Figure 1 depicts the relations between our clones SHR3 and SHR4 and humrsh3. There was 95% homology at the DNA level between these two clones and humrsh3.

Computer analysis indicated the presence of open reading frames in both of our clones as well as humrsh3. Where the sequences overlap, there were 25 base changes involving 22 codons detected in a total of 502 bp (see figure 2). Of these differences, 5 were silent, 10 were conservative, and 7 were not conserved.

The isolation of cDNA copies of this repetitive element provided evidence that this element was transcribed and polyadenylated. There are several potential roles for such a message. For example, the transcription could be part of a transposition mechanism (14), although the RNA in such a pathway would probably not be polyadenylated (15). The KpnI repetitive family in primates, the probable ancestor of the HindIII 1.9 kb repeats in humans (16), has been shown to be homologous to nuclear RNA molecules in monkey and human cells by northern analysis (17). Broad analysis of the distribution of these transcripts during development and in different tissues may help uncover their functions.

A more complex explanation or multiple roles seem to be indicated by the fact that the open reading frame in humrsh3 is in phase with the only open frame in clones SHR3 and SHR4, consistent with translation of the repetitive element. Furthermore, the translation frame in SHR3 is demonstrably in phase in the recombinant fusion protein. (The sequence data for SHR4 does not include the fusion protein junction.)
Figure 2: Sequence of SHR3 (A) and SHR4 (B). The one letter amino acid code and the DNA sequences of the cDNA inserts is shown in the top two lines. Only the nucleotides and amino acids that differ in humrsh3 are shown in the lower two lines. The "*"s denote insertions of T residues that were made in SHR4 (at 258) and in humrsh3 (at 803) that were made to maximize the homology between the coding sequences. It is suggested that a functional copy of these sequences would have one of these insertions.

The other question concerns the reason why clones of the repetitive sequences turned up in the antibody screening of the cDNA expression libraries at all. If the HindIII 1.9 kb repeat
was indeed translated in the liver then the phenylalanine hydroxylase preparation might have contained this protein as a contaminant and lead to antibodies that directly selected these clones.

At the other extreme, it is conceivable that any random transcript in a cDNA library could have six potential ways of producing a cross-reacting antigenic determinant. In other words, one of the five reading frames that are not used in the cell might occasionally match, for a short stretch, the three dimensional structure of a desired protein. However, there are two reasons that the use of these unnatural coding frames, as an explanation, may be limited. First, the alternate reading frames should not be translated well since there is a non-randomness to codon usage (correlated with tRNA availability). Therefore, synthesis of these proteins should be further retarded to a level that may be undetectable (18,19). Secondly, biologically unnatural peptides may be much more unstable because they are not derived from evolved folding structures (20).

Many false positives in these screenings may simply be due to coincidence. Antigenic determinants may be as small as five amino acids and if there are 10,000 genuine reading frames averaging 300 amino acids each in any one cell, and assuming that amino acid usage and RNA prevalence is uniform, then the probability of finding a random match would be 61% (using $P=1-(1-p)^n$).

A simpler possibility that might apply, at least, to this instance would invoke the very nature of repetitive elements—which are known to have portions scattered throughout the genome (1). Consequently, we could suppose that the phenylalanine hydroxylase gene contains a remnant of a HindIII 1.9 kb repetitive element and therefore shares an antigenic determinant. If this is the case, then the homology between the repetitive element translation product and the phenylalanine hydroxylase protein (or any known potential protein contaminant) necessary for the antibody response appears to be below the statistical background that can be examined by a variety of computer techniques.

For whatever reason, the two independent cDNA clones were discovered by serendipity. Their existence provides strong support for the statistical arguments used in the current hypotheses concerning the function of the 1.9 kb HindIII repetitive element (see Mottez, et al. [16]).
ACKNOWLEDGEMENTS

We thank Jeffrey Greene for help in purifying the antibody used in the study, Mike Jay, William Drohan, T. Venkat Gopal, and Arthur Nienhuis for stimulating discussions and special facilities, and Carol Kaufman for expert technical help. We are also grateful to Peter Rogan and Laura Manuelidis for useful discussion and providing their latest results prior to publication.

REFERENCES