Prime Scholars Library

Advances in Life Science and Biotechnology

Editorial

Vol. 9 (2), pp.04 – 04 June2021

Available online at
https://primescholarslibrary.org/

Prediction of the coding sequences

Sato Watanabe*

Department of Biotechnology, Tohoku University, Katahira, Aoba Ward, Sendai, Miyagi, Japan.

DESCRIPTION

Protocol has been established for the prediction of the coding sequences of unidentified human genes supported the double selection and sequence analysis of cDNA clones with inserts carrying unreported 5'-terminal sequences and with insert sizes akin to nearly full-length transcripts. By applying the protocol, cDNA clones with inserts longer than 2 kb were isolated from a cDNA library of human immature myeloid cell line KG-1, and therefore the coding sequences of 40 new genes were predicted. A computer search of the sequences indicated that 20 genes contained sequences like known genes within the GenBank/EMBL databases. The sequences of the remaining 20 genes were entirely new, and characteristic protein motifs or domains were identified in 32 genes. Other sequence features noted were that the coding sequences of 23 genes were followed by relatively long stretches of 3'-untranslated sequences which 5 genes contained repetitive sequences in their 3'-untranslated regions. The chromosomal location of those genes has been determined. By increasing the size of the above analysis, the coding sequences of the many unidentified genes is predicted.

An expression profile of genes active within the human colonic mucosa was obtained by collecting 959 partial sequences from a 3’-directed cDNA library. Seven genes were found to provide mRNA each of which comprized quite 1% of total mRNA. Four of those genes are novel, and are likely to be uniquely expressed within the colonic mucosa, and also the other three are identified as genes for carboxylic acid binding protein, immunoglobulin lambda chain, and carcinoma-associated antigen GA733-2. Within the remaining 952 clones, 310 were composed of 118 species occurred recurrently but but 1%, and 533 clones appeared one time. Because the 3’-directed cDNA library faithfully represents the mRNA population within the source tissue, these numbers represent the relative activities of the organic phenomenon.

Altogether 156 gene species were identified in GenBank, and a big portion of those genes encode proteins found in Golgi apparatus and lysosomes, chromosome-encoded mitochondrial proteins, cell surface proteins, and components within the protein synthesis machinery. the categories and proportions of genes identified is in keeping with the known major activities of the colonic mucosa like mucous protein production, energy-dependent water absorption, and rapid cell proliferation and turnover.

Another research was conducted. There the Sequence features of the expected coding regions for 40 unidentified human genes. Gene numbers are given on the left side. The horizontal scale represents organic compound residues from the N-terminus. Occurrence of hydrophobic organic compound residues is indicated by blue vertical bars (long: aromatic, short: aliphatic) which of hydrophilic by either green (Arg, Lys and His: long, medium and short upward bars; Glu and Asp: long and short downward bars) or red vertical bars (long: amides, short: alcohols). Pro is shown by black bars and therefore the positions of Ala, Gly and Cys were left unmarked. Above the lines of aminoalkanoic acid residues, the regions that showed similarities to known genes are indicated by horizontal lines with arrowheads at both ends, the situation of protein motifs that matched those within the PROSITE motif database by black blocks, and significant hydrophobic regions that were identified by the methods of Engelman et al. and of Kyte and Doolittle2 by red blocks. Abbreviations used are; B, bovine; Ch, Chinese hamster; D, drosophila, H, human; M, mouse; N, Neurospora crassa; P, pig; R, rat; Sc, Saccharomyces cerevisiae.