Volume 16 Number 5 1988 Nucleic Acids Research

Electronic fingerprinting of RNA

Peter Gegenheimer

Departments of Biochemistry and Botany, and Molecular Genetics Program, University of Kansas, Lawrence, KS 66045, USA

Submitted August 17, 1987

ABSTRACT

Software has been developed to assist RNA fingerprinting analysis. One program generates, from a DNA sequence data file, the oligonucleotides resulting from digestion of an RNA transcript labeled with any specified nucleotide(s). Oligonucleotides are sorted according to their position on the fingerprint. Expected molar yields and products of secondary redigestion are also indicated. A second program facilitates calculation of experimental molar yields of oligonucleotides.

RNA "fingerprinting" is a powerful technique which can be used to determine RNA termini, splice junction locations, or structures of RNA splicing intermediates (1-4). Although current technology allows ready analysis of RNAs under one hundred to over 1000 nucleotides in length, a major hindrance to the design and interpretation of experiments requiring fingerprinting is the tedium of manually cataloguing all the primary and secondary products expected from RNA labeled with particular nucleotides and digested with a given enzyme.

To facilitate fingerprint analysis, a simple "electronic fingerprinting" program, **FINGERS**, was developed for personal computers running under the MS-DOS operating system. Sequences are read from a standard DNA sequence text file. The user specifies the regions of RNA to splice together, the origin of splice junction phosphates, and whether 5' or 3' termini are phosphorylated. Further choices are the enzyme used for digestion and the α-^{32}P-NTP(s) with which the RNA is to be labeled. The program first generates internally an array of oligonucleotides produced by the enzyme chosen (RNAase T1 or A) and which would be ^{32}P-labeled with the nucleotides used. The base code 'N' may be used in the sequence file to represent a residue resistant to RNAase cleavage. This array is sorted, if desired, using a recursive shell sort, into the order of appearance on a two-dimensional fingerprint: first by size, then by U + G content, and finally by A content. 5'- and 3'- terminal

oligonucleotides, and oligonucleotides with identical sequence but different labeled nearest neighbors, are considered separately. The final listing indicates the position code (size, U+G content, and A content), the number of labeled phosphates, the sequence of each oligonucleotide including the labeled products of opposite-enzyme redigestion, and the nearest neighbor nucleotide. Output may be directed to the screen, a printer, or a disk file (with optional embedded word processing commands). The same sequence may be analyzed repeatedly with different combinations of labeling, splicing, and digestion.

A companion program, **OLIGOS**, serves as a specialized database for storage, retrieval, and manipulation of experimental data resulting from quantitation of radioactivity in each oligonucleotide. The user specifies oligonucleotides to be used for normalization; the program then calculates molar yields given chain lengths, or chain length assuming a single mole of an oligonucleotide. Formatted results are displayed on the screen, printed, or stored as a text file on disk. Raw data is saved separately. Records may be analyzed and edited at any time.

FINGERS and **OLIGOS** have been used together to manage, for example, a fingerprinting project involving multiple RNAs labeled *in vitro* with several different nucleotides (5).

Programs are available from the author upon receipt of an MS-DOS formatted 3.5 inch or 5.25 inch disk (double or high density). Some customization for printers, word processors, and modified nucleotide codes is possible if specifications are provided.

REFERENCES

1. Barrell, B. G. (1972) Prog. Nucl. Acids Res. **2**, 751-779.
2. Volckaert, G., Min Jou, W., and Fiers, W. (1976) Anal. Biochem. **72**, 422-466.
3. Gegenheimer, P. and Apirion, D. (1980) J. Mol. Biol. **143**, 227-258.
4. Gegenheimer, P., Gabius, H.-J., Peebles, C. L., and Abelson, J. (1983) J. Biol. Chem. **258**, 8365-8373.
5. Engelke, D. R., Gegenheimer, P., and Abelson, J. (1985) J. Biol. Chem. **260**, 1271-1279.