Word-Processor Macro for Restriction Endonuclease Analysis

BioTechniques 27:1228-1231 (December 1999)

Nelson Cabrera León

Centro de Ingeniería Genética y Biotecnología, La Habana, Cuba

ABSTRACT

This paper describes a Microsoft Word 97 macro designed for restriction endonuclease analysis. Selected DNA fragments in the active Word document can be analyzed through a dynamic dialog box that formats the enzyme restriction lists for further analysis. The results can be obtained in a new Word document with the name of the enzymes, number of cuts and positions. This macro has several advantages: the results can be printed in a format suitable for record keeping, no additional programs are required and it is simple to use.

INTRODUCTION

The restriction endonuclease analysis of a DNA sequence is one of the basic tools in molecular biology for the development of cloning strategies. Although numerous computer programs are available for this kind of analysis, some are expensive and require some familiarity with them. Those programs are able to perform diverse tasks, but, in their daily practice, researchers only use a few of them. An additional difficulty of other computer programs is that the user must change the output format to include it in text documents.

DNA sequences, in general, are stored in a format that can be read by many word processors, but in some of them, the user can develop applications (macros) to carry out tedious or time-consuming tasks. In this work, we report the use of a Word 97 macro to make endonuclease restriction analyses.

MATERIALS AND METHODS

Computer Programs Used

The Windows™ version of Microsoft® Word 97™ (Microsoft, Redmond, WA, USA) is the only program used in this work.

Files Used

Restriction Sites Analyzer.dot. The macro is contained in a file called Restriction Sites Analyzer.dot. It has three procedures: AutoExec, AutoExit and ResAnalyzer (the latter executes the restriction analysis). A copy of this file in the StartUp directory of Word 97 (the default path to this folder is C:\Program Files\Microsoft Office\Office\StartUp) is necessary to run this macro properly because it has a procedure (AutoExec) that inserts the Restriction Analysis option in the Tools menu when Word 97 starts. The AutoExit procedure deletes the Restriction Analysis option when the user exits Word.

If users do not want to load the macro the next time that Word starts, they simply exit Word and the Restriction Sites Analyzer.dot file from the StartUp folder.

To view or edit the macro, the user must open the Restriction Sites Analyzer.dot file in the StartUp directory and choose Macro in the Tools menu. Then, the user must select the Macros option, then the AutoExec, AutoExit or ResAnalyzer procedure in the dialog box and then click on the Edit button.

Default Enzymes List.txt. To run this macro, the user needs another file: Default Enzymes List.txt. The user must put a copy of this file in the My documents folder (the default path to this folder is C:\My documents). This file has an alphabetical list of almost all the restriction endonucleases available commercially and includes the restriction enzyme name for each enzyme, its cut positions and all its recognition sites separated by commas (Table 1). The function of the macro is to use the FIND function to search different recognition sites in a DNA sequence that have been typed in a Word document. The user can use the symbols ^\$ to represent any nucleotide because the code is understood by the function. Table 1 shows how the code was written for the *Bgl*II, *Hae*II, *Hga*I and *Hind*III enzymes.

Enzymes with two different cut positions must be typed separately (see *Hga*I in Table 1). *Hga*I is a non-palindromic endonuclease, so the sequence that it recognizes in a DNA chain is different from the sequence in the complementary DNA chain. In this case, the user must type the same enzyme twice, one following the other, with its respective cut positions and its recognition sites. The user can edit the Default Enzymes List.txt by adding or deleting restriction enzymes to create new lists for specific purposes. These new lists should

Table 1. Some Restriction Enzymes Written in a Format Suitable for the Macro

<i>Bgl</i> II, 7, GCC^\$^\$^\$^\$^\$GGC
<i>Hae</i> II, 5, GGCGCC, GGCGCT, AGCGCC, AGCGCT
<i>Hga</i> I, 10, GACGC^\$^\$^\$^\$^\$
<i>Hga</i> I, 0, ^\$^\$^\$^\$^\$^\$^\$^\$^\$^\$^\$GCGTC
<i>Hind</i> III, 1, AAGCTT

be saved using different names with the .txt format extension.

DNA sequences in the Word document should contain no numbers, spaces or paragraph mark symbols, which will result in the sequence written as shown in Figure 1.

To run the macro, the user only needs to select the DNA sequence in the active document, then select Restriction Analysis in the Tools menu and a message box will appear. If the user selects Yes, the macro will use the Default Enzymes List.txt file to execute the restriction analysis; if the user selects No, the Open dialog box appears, which allows the user to select another enzyme list. The user can exit the macro automatically by clicking Cancel.

As the macro runs, a new document is created (Figure 2). This document contains the results of the restriction analysis and shows a table with the name of the enzymes, number of cuts and positions, the DNA length and a list with non-cutter enzymes.

RESULTS AND DISCUSSION

The analysis of the macro was validated for several DNA sequences with BioSOS (1) and Gene Runner v3.02 (Hastings Software) programs with optimal results.

The method described here is a helpful tool for the molecular biologist. It is easy to use and avoids the use of expensive commercial software for the restriction analysis of DNA sequences. Furthermore, the report created is concise and easy to format for record keeping. The Visual Basic for Application code of this macro is a valuable guide for those who want to develop new applications in Word 97.

AVAILABILITY

The files referred to in this article can be downloaded from the **Software Library** on the *BioTechniques* Web site (www.biotechniques.com).

ACKNOWLEDGMENTS

I thank Tanilo Rivero, Alejandro Soto, Luis Ríos and Dr. Javier Menéndez for their critical reading of the manuscript.

REFERENCES

1. Bringas, R., R. Ricardo, J. Fernández de Cossio, M.E. Ochagavía, A. Suárez and R. Rodríguez. 1992. BioSOS: A program package for the

analysis of biological sequences in the microcomputer. *Biotechnología Aplicada* 9:180-185.

Received 7 June 1999; accepted 6 October 1999.

Address correspondence to:

Dr. Nelson Cabrera León
División de Bioindustria
Centro de Ingeniería Genética y Biotecnología
Ave. 31 e/ 158 y 190. Aptdo 6162, C.P. 10600
La Habana, Cuba
Internet: nelson.cabrera@cigb.edu.cu