

CIBSOFT: Un paquete de programas para el análisis de ácidos nucleicos y proteínas

RICARDO BRINGAS PÉREZ Y SERGIO PÉREZ TALAVERA

Centro de Investigaciones Biológicas, La Habana, Cuba

Centro de Ingeniería Genética y Biotecnología, La Habana, Cuba

20 de septiembre de 1986

RESUMEN

Presentamos un conjunto de programas en forma de menú para el análisis y la manipulación de secuencias de ADN, que contiene las opciones más útiles y de mayor uso de trabajos similares publicados como son: análisis de restricción, traducción a aminoácidos, uso de codones, búsqueda de marcos abiertos de lectura y un editor de secuencias de ADN. Los programas están al alcance de personas sin conocimientos de programación.

SUMMARY

We present a menu driven set of programs for analysis and handling of DNA sequences, containing the most useful and frequently used options of similar published works, as restriction analysis, translation to amino acids, codon usage, search for open reading frames and DNA sequences editor. Programs can be used for people without any knowledge of programming.

INTRODUCCION

En los últimos años se ha desarrollado gran cantidad de software para el almacenamiento y análisis de secuencias de ADN, lo que unido al desarrollo de los métodos de secuenciación y de bancos de datos como el GenBank y el EMBL Data Bank han dado un gran impulso a la aplicación de la computación a las investigaciones en ingeniería genética y la biología molecular en general, aspectos estos que cada día desempeñan un papel más importante en las investigaciones y que actualmente resulta impredecible el alcance que pueden tomar en el futuro.

En este trabajo presentamos un conjunto de programas que consideramos son los de uso más frecuente y de mayor utilidad para los investigadores, de acuerdo a trabajos similares reportados y nuestra experiencia acumulada en su uso; estos programas, representan un primer paso en la aplicación de la computación a las investigaciones en ingeniería genética en nuestro centro.

MATERIALES Y METODOS

CIBSOFT ha sido desarrollado sobre UCSD-Pascal en una microcomputadora NEC-9801F (actualmente cuenta con una versión desarrollada en Turbo Pascal para microcomputadoras compatibles con la IBM PC). Los programas son de muy fácil manejo por lo que están al alcance de personas con escasos o ningún conocimiento de computación. Su estructura es similar a la de otros paquetes publicados (Lagrimini *et al.*, 1984; Pustell *et al.*, 1984; Queen *et al.*, 1984; Schwindinger *et al.*, 1984). Las diferentes opciones que ofrece se presentan en forma de menú, dividido en dos partes o páginas, donde aparece una información muy breve de las funciones que realiza cada programa; también tiene opciones para cambiar ciertos parámetros de entrada-salida, de forma muy fácil. Uno de los programas es un editor de secuencias de ADN muy cómodo para la introducción y rectificación de secuencias.

Al ejecutar el programa aparecerá ante usted el menú de la página 1 de CIBSOFT, tal y como se lo presentamos en la figura 1.

Página 1		
	<i>Función</i>	<i>Descripción</i>
1.	Imprime	Imprime la secuencia en uso
2.	Restricción	Halla los sitios de restricción de la secuencia en uso
3.	Traducción 1	Traduce la secuencia en uso a un marco específico
4.	Traducción 3	Traduce la secuencia en uso a los tres marcos posibles
5.	Localiza	Localiza un segmento de la secuencia en uso
6.	Uso	Frecuencia de uso de codones en la secuencia en uso
7.	Hidrofilicidad	Dibuja un gráfico de hidrofilicidad en el ploter
8.	Marcos abiertos	Halla posibles marcos abiertos de lectura
9.	Página 2	Aparece la página 2 del menú
0.	Parámetros	Cambia los parámetros de e/s
F.	Fin	Termina el programa

(escoja las variantes tecleando su número)

En el extremo izquierdo aparece el número de cada uno de los programas u opciones, con el que usted seleccionará la opción a ejecutar; a continuación aparece el nombre de la función y a la derecha una breve descripción del programa. La página 2 de CIBSOFT presenta la misma estructura y se presenta como aparece en la figura 2.

Página 2		
	<i>Función</i>	<i>Descripción</i>
1.	Editor	Introducir o hacer cambios en una secuencia
2.	Compara	Compara dos secuencias
3.	Matriz de punto	Compara dos segmentos de secuencia con matriz de punto
4.	Transferencia	Transfiere un fichero a la consola o a la impresora
5.	GenBank	Acceso al GenBank
6.	Modifica	Modifica secuencias existentes
7.		(libre para instrumentar)
8.		(libre para instrumentar)
9.	Página 1	Aparece la página 1 del menú
0.	Parámetros	Cambia los parámetros de e/s
F.	Fin	Termina el programa

(escoja las variantes tecleando su número)

En la página 1 de CIBSOFT, en la opción *parámetros* se debe especificar cuál será la secuencia que pondremos en uso y esta será la secuencia que al elegir cada opción se someterá al análisis correspondiente, y que hemos llamado *secuencia en uso*.

Imprimir: Esta opción produce un listado de la secuencia que se encuentre en uso, la impresión se hará en segmentos de 10 nucleótidos y numerada para facilitar su lectura.

Restricción: En esta opción se pueden hallar los sitios de restricción de las diferentes enzimas de restricción que usted seleccione de las 76 que el programa le brinda. La salida de este programa ofrece la posición de cada sitio de restricción y la distancia entre sitios adyacentes o lo que es lo mismo los tamaños de los segmentos resultantes de la restricción.

Traducción 1: Nos ofrece la impresión de la secuencia, con la traducción a aminoácidos del exón o exones que se le especifica.

Traducción 3: Traduce la secuencia en sus tres posibles marcos de lectura.

Localiza: Halla un segmento de secuencia en la secuencia en uso, con cierto grado de homología que le proporciona el usuario.

Uso: Nos ofrece una tabla con cada uno de los codones y aminoácidos y la frecuencia con que aparecen en la secuencia en uso, así como el peso molecular de la proteína, para lo que debemos especificar el inicio y el final de cada uno de los exones presentes en la secuencia.

Hidrofilicidad: Esta opción nos ofrece un gráfico de hidrofiliidad, según el método de Hopp y Woods (1981), este es producido en un plotter Graphtec MP-1000.

Marcos abiertos: Nos presenta los posibles marcos abiertos de lectura de la secuencia en uso, analizando los tres marcos y la secuencia complementaria.

En la página 2 de CIBSOFT en la opción *parámetros* no se incluye la secuencia en uso, ya que debido a las características de los programas que la componen, en cada uno de ellos se seleccionan las secuencias a analizar.

Editor: Este programa facilita la introducción de secuencias a la computadora así como su arreglo o modificación. Este editor está estructurado en páginas de 300 nucleótidos cada una en 5 filas de 60 nucleótidos y en columnas de 10 nucleótidos enumerados de 10 en 10, en la parte superior de la pantalla aparecen las diferentes opciones que nos ofrece, similar al editor de UCSD-Pascal. Posee opciones para insertar y eliminar nucleótidos, las opciones de movimiento con las teclas respectivas y una opción para dirigirse a la página anterior.

Compara: Esta opción compara dos secuencias, hallando zonas de homología, el porcentaje de homología así como los tamaños de los segmentos los proporciona el usuario. La salida de este programa nos ofrece las dos secuencias comparadas una encima de la otra separadas por asteriscos en donde coincidan nucleótidos de ambas secuencias. Este programa funciona adecuadamente con secuencias que tengan un nivel de homología aceptable.

Matriz de punto. Compara dos secuencias por homología creando una matriz de puntos; esta opción que es frecuente en otros programas reportados (Lagrimini *et al.*, 1984; Larson *et al.*, 1983; Pustell *et al.*, 1984; Schwindinger *et al.*, 1984), facilita grandemente la determinación de zonas de homología. El usuario debe proporcionar al programa el ancho de la ventana de comparación a utilizar, así como el mínimo de homología a considerar, las zonas de homología de las dos secuencias comparadas aparecerán en la matriz de puntos como diagonales.

Transferencia: Esta opción permite transferir ficheros alojados en los discos a la consola o a la impresora.

GenBank: Esta opción permite el acceso a ficheros y textos que contengan la información que ofrece el banco de secuencias GenBank, permitiendo crear un nuevo fichero con la

información de la secuencia que se desee y a partir de este crear un nuevo fichero que contenga solo la secuencia nucleotídica, quedando listo para su uso por CIBSOFT.

Modifica: Permite armar nuevas secuencias a partir de otras ya existentes. Es posible con esta opción insertar o eliminar segmentos en una secuencia. Como datos de entrada, se le proporciona al programa la posición inicial y final del segmento a eliminar, o el nombre de la secuencia que contiene el segmento a insertar y la posición inicial y final que delimita este segmento y por último el sitio donde se insertará.

CONCLUSIONES

Se ha elaborado un paquete de programas que contiene un conjunto de opciones básicas para el análisis de secuencias nucleotídicas, que facilita y agiliza el trabajo de investigación en un laboratorio de Ingeniería Genética. Es de fácil manejo y rápido aprendizaje. Su efectividad ha sido demostrada por el uso durante dos años por el Departamento de Ingeniería Genética del Centro de Investigaciones Biológicas y el Centro de Ingeniería Genética y Biotecnología.

La comparación de las características de este paquete con otros analizados por Korn L. J. y Queen C. (1984), nos muestra que para los parámetros principales como: tamaño máximo de la secuencia a analizar, velocidad de búsqueda de sitios de restricción y diversidad de opciones, este conjunto de programas es similar a otros que se encuentran comercializados.

REFERENCIAS

- HOOP, T. P y K. R. WOODS (1981). *Prediction of Protein Antigenic Determinants from Amino Acid Sequence*. Proc. Natl. Acad. Sci. USA., 78 (6): 3824-3828.
- KORN, L. J. y C. QUEEN (1984). *Analysis of Biological Sequences in Small Computers*. DNA, 3: 421-436.
- LAGRIMINI, L. M.; S. T. BRENTANO y J. E. DONELSON (1984). *A DNA Sequence Analysis Package for the IBM Personal Computer*. Nucl. Acid. Res. 12: 605-614.
- LARSON, R. y J. MESSING (1983). *Apple II Computer Software for DNA and Protein Sequence Data*. DNA, 2 (1): 31-36.
- PUSTELL, J. y F. KAFATOS (1984). *A Convenient and Adaptable Package of Computer Programs for DNA and Protein Sequence Management, Analysis and Homology Determination*. Nucl. Acid. Res., 10: 4765-4782.
- QUEEN, C y L. J. KORN (1984). *A Comprehensive Sequence Analysis Program on IBM Personal Computer*. Nucl. Acid. Res. 12: 581-599.
- SCHWINDINGER, W. F. y J. R. WARNER (1984). *DNA Sequence Analysis on the IBM-PC*. Nucl. Acid. Res. 12: 601-604.