

Easy Bioinformatics Analysis (EBiAn): a package for manipulating and analysis of short biological sequences

Luiz Carlos Bertucci Barbosa, Saulo Santesso Garrido, Anderson Garcia, Davi Barbosa Delfino, Rodrigo Duarte Gonçalves, Reinaldo Marchetto*

Institute of Chemistry, UNESP – Univ Estadual Paulista, Department of Biochemistry and Technological Chemistry, Araraquara, São Paulo, Brazil. Reinaldo marchetto - E-mail: marchetto@iq.unesp.br; Phone: +55-16-3301.6670; Fax: 55-16-3301.6692; *Corresponding author

Received May 14, 2010; accepted June 08, 2010; published July 6, 2010

Abstract:

The work of biochemists and molecular biologists often is dependent or extremely favored by a preliminary computer analysis. Thus, the development of an efficient and friendly computational tool is very important. In this work, we developed a package of programs in Javascript language which can be used online or locally. The programs depend exclusively of Web browsers and are compatible with Internet Explorer, Opera, Mozilla Firefox and Google Chrome. With the EBiAn package it is can perform the main analysis and manipulation of DNA, RNA, proteins and peptides sequences. The programs can be freely accessed and adapted or modified to generate new programs.

Availability: <http://www.iq.unesp.br/EXTENSAO/EBiAn/html/ebian.html>

Keywords: bioinformatics, sequence manipulating

Background:

Computational tools to analyze nucleic acid sequences and primary structures of proteins are essential in the help of the work of biochemists and molecular biologists. Many tools for this purpose are available so far in sites of various research groups or in servers of biological data, such as GenBank [1], Protein Data Bank [2] and ExpASY [3]. Javascript is an object-oriented language that is interpreted by several current Web browsers. Together with the hypertext markup language (HTML) and cascading style sheets (CSS) languages, Javascript allows the development of several bioinformatics simple tools, with Web interface, to manipulate short sequences of nucleic acids, peptides or proteins, generating primers and extracting specific information contents from a biological data file. In this work, we developed a package of programs in Javascript, called Easy Bioinformatics Analysis (EBiAn), which enable the more routinely analysis used in preliminary computational approaches to experimental laboratory practices in the biochemistry and molecular biology area.

Description:

The package includes the following tools:

GenBank to Fasta:

Receives a GenBank file as input and returns the entire biological sequence in FASTA file. Use this tool to remove non-sequence information from a GenBank file.

EMBL to Fasta:

Receives an EMBL file as input and returns the entire biological sequence in FASTA file. Use this tool to remove non-sequence information from an EMBL file.

To three Letter Code:

Converts a protein sequence in one-letter code to a sequence in three-letter code.

ToOneLetterCode:

Converts a protein sequence in three-letter code to a sequence in one-letter code.

DNA/RNA-Param:

Receives a RNA or DNA sequence and returns its contents of nitrogen bases.

Complement:

Receives a DNA sequence and returns its complementary sequence. In this program there is the possibility of obtain the reverse complementary sequence.

Translate:

Receives a DNA sequence and performs the translation into amino acids, using the standard codon table. It is possible perform the translation for the three open reading frames (ORFs) in the direct strand (reading +1, +2 and +3) and in the reverse strand (reading -1, -2 and -3).

Transcription:

Receives a DNA sequence and returns its transcript sequence.

DNA Cutter:

Receives a DNA sequence and returns the possible cleavage sites of most used restriction endonucleases, giving the name of the enzyme, the number and locations of the cuts along of the sequence.

ORF Finder:

Receives a DNA sequence and returns all the possible open reading frames (ORFs). In this tool, its possible obtain the three open reading frames in the direct strand (reading +1, +2 and +3) and in the reverse strand (reading -1, -2 and -3). The user can discriminate the minimum size of coding sequences to be screened.

Primer Generator:

Receives a DNA sequence and returns its forward and reverse terminal primers. It is possible also generate the internal primer from the input sequence. The user can determine the length of the primers, as well as adding endonuclease restriction sites.

The screenshot shows the EBiAn web interface. The header is green with the EBiAn logo and the text 'Easy Bioinformatics Analysis'. Below the header, there is a navigation menu on the left with categories: 'Format conversion', 'DNA tools', 'Peptides and proteins tools', and 'Download'. The main content area is white and contains several tool descriptions under their respective category headers. On the right side, there are logos for USEAPP, Instituto de Química Araraquara, and UNESP.

Format conversion

- GenBank to Fasta
- EMBL to Fasta
- ToThreeLetterCode
- ToOneLetterCode

DNA tools

- DNA/RNA-Param
- Complement
- Translation
- Transcription
- DNA Cutter
- ORF finder
- PrimerGenerator

Peptides and proteins tools

- Prot-Param
- ProteinDigest
- MS-Digest
- ProtColourer
- PepBuilder

Download

- Download EBiAn

Format conversions

GenBank to Fasta: receives a GenBank file as input and returns the entire biological sequence in FASTA file.

EMBL to Fasta: receives a EMBL file as input and returns the entire biological sequence in FASTA file.

ToThreeLetterCode: converts a protein sequence in one-letter code to a sequence in three-letter code.

ToOneLetterCode: converts a protein sequence in three-letter code for a sequence in one-letter code.

DNA tools

DNA/RNA-Param: receives a RNA or DNA sequence and returns its contents of nitrogen bases.

Complement: receives a DNA sequence and returns its complement sequence.

Translate: receives a DNA sequence and performs the translation into amino acids, using the standard codon table.

Transcription: receives a DNA sequence and returns its transcript sequence.

DNA Cutter: receives a DNA sequence and returns the possible cleavage sites of restriction endonucleases.

ORF Finder: receives a DNA sequence and returns all the possible open reading frames (ORFs).

PrimerGenerator: receives a DNA sequence and returns its forward and reverse terminal primers. It is possible also generate internal primer from the input sequence.

Peptides and proteins tools

Prot-Param: receives a protein sequence and returns the amount of each type of amino acid.

ProteinDigest: receives a protein sequence and returns the position of its potential sites of proteolytic cleavage by an enzyme selected by user.

MS-Digest: receives a protein sequence and returns all the teoric fragments generated by proteolytic cleavage by an enzyme selected by user. The program also generates the monoisotopic and average mass, simulating monoprotonated fragments.

ProtColourer: receives a protein or peptide sequence and highlights a group of amino acids selected by user.

PepBuilder: allows the user to create textual representations of peptide sequences. The program automatically calculates the average and monoisotopic mass, and returns its molecular formula.

USEAPP
Unidade de Síntese, Estrutura e Aplicações de Peptídeos e Proteínas

Instituto de Química
Araraquara

UNESP

Copyright© EBiAn 2010
IQ-UNESP Araraquara

Figure 1: Web interface of the EBiAn package.

Prot-Param:

Receives a protein sequence and returns the amount of each amino acid type. Moreover, returns the amount of hydrophobic, hydrophilic, positive, negative and aromatic amino acids residues and the amount of amino acids with hydroxyl group.

ProteinDigest:

Receives an amino acid protein sequence and returns the position of its potential sites of proteolytic cleavage [4, 5] of a user selected enzyme.

MS-Digest:

Receives an amino acid protein sequence and returns all theoretic peptide fragments generated by proteolytic cleavage [4, 5] of a user selected enzyme. The program also generates the monoisotopic and average mass, simulating monoprotonated fragments.

ProtColourer:

Receives a protein or peptide sequence and highlights a group of amino acids selected by user. The program contains pre-defined groups of amino acids to be marked: hydrophobic, hydrophilic, positive, negative and aromatic residues and amino acids with hydroxyl group. In addition, the user can choose the color to be used in the highlights and define its own group of amino acids.

PepBuilder:

Allows the user to create textual representations of peptide sequences. The program automatically calculates the average and monoisotopic mass, and returns its molecular formula.

The Easy Bioinformatics Analysis package has an intuitive and easy navigation. The programs can be accessed online by the main page accessible from <http://www.iq.unesp.br/EXTENSAO/EBiAn/html/ebian.html> (Figure 1). Moreover, it can also be used offline after downloaded it from the main page. The user with programming skills can access the source code and change it to create new programs or adapting them. The programs depend only of Web browsers and are compatible with Internet Explorer, Opera, Mozilla Firefox and Google Chrome, having better performance for the last two. In future versions of the package, the included programs will be improved and new programs may be included. Additional questions or comments regarding the Easy Bioinformatics Analysis package can be sent by e-mail to l Luiz_cbb@hotmail.com.

Acknowledgments

We gratefully acknowledge CNPq and CAPES for financial support. R. Marchetto is recipient of research fellowship from CNPq.

References:

- [1] <http://www.ncbi.nlm.nih.gov>
- [2] <http://www.pdb.org>
- [3] <http://br.expasy.org>
- [4] B Keil, *Protein Seq Data Anal.* **1**: 13 (1987) [PMID: 3447153]
- [5] I Schechter & A Berger, *Biochem Biophys Res Commun.* **27**: 157 (1967) [PMID: 6035483]

Edited by P. Kanguane

Citation: Barbosa *et al.* Bioinformatics 5(2): 46-48 (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.