Molecular Biology

Lab 6

Introduction to Bioinformatics

Analysis of DNA sequence

Designing PCR primers

Bioinformatics

From Wikipedia, the free encyclopedia

**Bioinformatics** is a branch of [biological](http://en.wikipedia.org/wiki/Biological) science which deals with the study of methods for storing, retrieving and analyzing biological data, such as [nucleic acid](http://en.wikipedia.org/wiki/Nucleic_acid) (DNA/RNA) and protein [sequence](http://en.wikipedia.org/wiki/Sequence_%28biology%29), [structure](http://en.wikipedia.org/wiki/Protein_structure), function, [pathways](http://en.wikipedia.org/wiki/Signal_transduction) and [genetic interactions](http://en.wikipedia.org/wiki/Epistasis).

------------------------------------

Bioinformatics allows one to understand genetic changes associated with disease, to analyze genome organization, to see at the molecular level, evolutionary relationships, and to model bio-molecular interactions. Bioinformatics combines the field of biology, with a computer power! Typically bioinformatics experts have more computer science background than knowledge of biological processes. The huge amount of data (particularly DNA sequence data) that has been generated in the past 20+ years demanded development of a whole knew set of technologies for handling and analysis. The field is one of the fastest growing and changing in the science area.

Today we will return to an old friend, pGLO. We will be analyzing the plasmid’s entire DNA sequence using publically available databases.

* Using computers we will access the pGLO sequence from the NIH/NCBI GenBank database.
* You will be given a set of tasks and allowed to explore with the sequence.
* In the end you will be asked to design PCR primers for our lab next week.

Procedure:

1. (Spend the first hour of lab doing this). Using a US government supported database---NCBI :
2. Access the Internet and load the ***NCBI*** website. What does NCBI stand for?
3. Explore the site---try not to feel overwhelmed, but discuss the many aspects of the database! Name 3 ways you might use this database.
4. On the left side choose DNA and RNA
5. Choose ***GenBank***—What is GenBank?
6. Choose ***BLAST***—what does BLAST stand for? What is it used for?
7. Choose nucleotide Blast.
8. Using the PDF linked to the website, cut and paste some of the pGLO sequence into the search box. Explain to your partner the detail of producing this sequence with the Sanger method of sequencing.
9. Set parameters as you think best. Click the big blue BLAST button.
10. Allow the site to refresh multiple times as the database is searched for your sequence. When the site seems stable again, choose summary.
11. Explore and discuss the results. Choose the best match for your sequence. What is the real name of pGLO in the NCBI data base? Click on hypertext to see different links to this plasmid sequence.

How big is the plasmid. Did you determine the same/similar size in Lab 4? Explain.

1. When you have accessed the summary, I will give you a printed copy of the data you should use.
2. Use the website to find the 3 genes encoded in pGLO that we discussed in Labs one and 2. Click on each gene description and find the sequence within the larger plasmid DNA. Find the protein sequences for each gene.
3. Designing primers for PCR.
4. Find the gene for GFP within the plasmid.
5. How long is the gene? How many Amino Acids should be in the protein it encodes? Using what your remember from genetics, check to see that the genetic code was used properly in making the protein.
6. Choose some sequence that spans 200-400 nucleotides. Rewrite ~ 25 nucleotides at the ends of this section of the gene, showing both strands.
7. Design 20-22 nucleotide primer pairs for PCR. These primers should prime DNA synthesis of the entire DNA region between them encoded on both strands. Assume a 5’-3’ direction for DNA synthesis. Play around with the sequence.
8. Determine the approximate melting/annealing temperature (Tm ) for your primers using the following formula.

# A or T bases X 2 degrees C + # C or G bases X 4 degrees C

For example---the very short sequence AATTGCCAT would have a Tm of 2 X 6 + 4 X 3 = 24 degrees C.

You would expect a primer such as this to begin annealing to a complementary sequence at about 24 degrees C.

1. If your primers have Tm that differs more than 10 degrees, try another nearby sequence that more closely matches the primer’s sequence.
2. If possible, try to let NCBI/Primer-BLAST make your primers for you. Cut and paste some of the sequence from the pGLO plasmid into the program. Set parameters. Give time for the program to generate possibilities.
3. Submit your primer sequences to Super. She will check them and order 2 or 3 of the 5 submitted.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Amino Acid** | **3-Letter**[106] | **1-Letter**[106] | **Side-chain polarity**[106] | **Side-chain charge (pH 7.4)**[106] | [**Hydropathy index**](http://en.wikipedia.org/wiki/Hydropathy_index)[107] | [**Absorbance**](http://en.wikipedia.org/wiki/Absorbance) **λmax(nm)**[108] |
| [Alanine](http://en.wikipedia.org/wiki/Alanine) | Ala | A | nonpolar | neutral | 1.8 |  |
| [Arginine](http://en.wikipedia.org/wiki/Arginine) | Arg | R | polar | positive | −4.5 |  |
| [Asparagine](http://en.wikipedia.org/wiki/Asparagine) | Asn | N | polar | neutral | −3.5 |  |
| [Aspartic acid](http://en.wikipedia.org/wiki/Aspartic_acid) | Asp | D | polar | negative | −3.5 |  |
| [Cysteine](http://en.wikipedia.org/wiki/Cysteine) | Cys | C | nonpolar | neutral | 2.5 | 250 |
| [Glutamic acid](http://en.wikipedia.org/wiki/Glutamic_acid) | Glu | E | polar | negative | −3.5 |  |
| [Glutamine](http://en.wikipedia.org/wiki/Glutamine) | Gln | Q | polar | neutral | −3.5 |  |
| [Glycine](http://en.wikipedia.org/wiki/Glycine) | Gly | G | nonpolar | neutral | −0.4 |  |
| [Histidine](http://en.wikipedia.org/wiki/Histidine) | His | H | polar | positive(10%)neutral(90%) | −3.2 | 211 |
| [Isoleucine](http://en.wikipedia.org/wiki/Isoleucine) | Ile | I | nonpolar | neutral | 4.5 |  |
| [Leucine](http://en.wikipedia.org/wiki/Leucine) | Leu | L | nonpolar | neutral | 3.8 |  |
| [Lysine](http://en.wikipedia.org/wiki/Lysine) | Lys | K | polar | positive | −3.9 |  |
| [Methionine](http://en.wikipedia.org/wiki/Methionine) | Met | M | nonpolar | neutral | 1.9 |  |
| [Phenylalanine](http://en.wikipedia.org/wiki/Phenylalanine) | Phe | F | nonpolar | neutral | 2.8 | 257, 206, 188 |
| [Proline](http://en.wikipedia.org/wiki/Proline) | Pro | P | nonpolar | neutral | −1.6 |  |
| [Serine](http://en.wikipedia.org/wiki/Serine) | Ser | S | polar | neutral | −0.8 |  |
| [Threonine](http://en.wikipedia.org/wiki/Threonine) | Thr | T | polar | neutral | −0.7 |  |
| [Tryptophan](http://en.wikipedia.org/wiki/Tryptophan) | Trp | W | nonpolar | neutral | −0.9 | 280, 219 |
| [Tyrosine](http://en.wikipedia.org/wiki/Tyrosine) | Tyr | Y | polar | neutral | −1.3 | 274, 222, 193 |
| [Valine](http://en.wikipedia.org/wiki/Valine) | Val | V | nonpolar | neutral | 4.2 |  |