Introducción a la Bioinformática BIO 267

Jessica Liliana Campo Giraldo

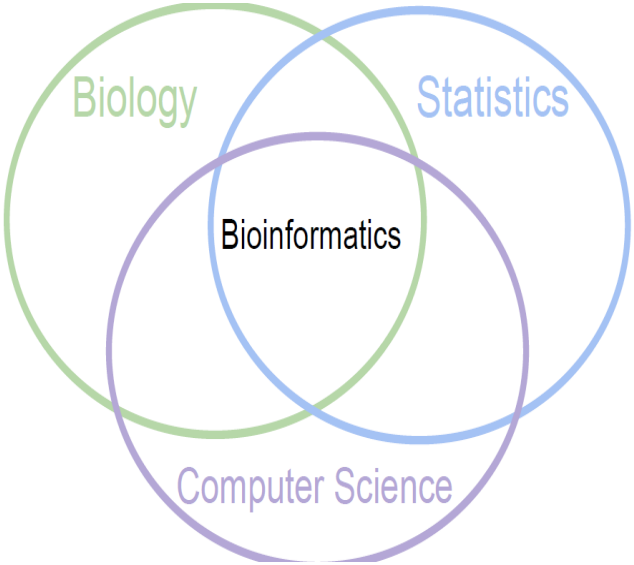
Bioinformática

Definición

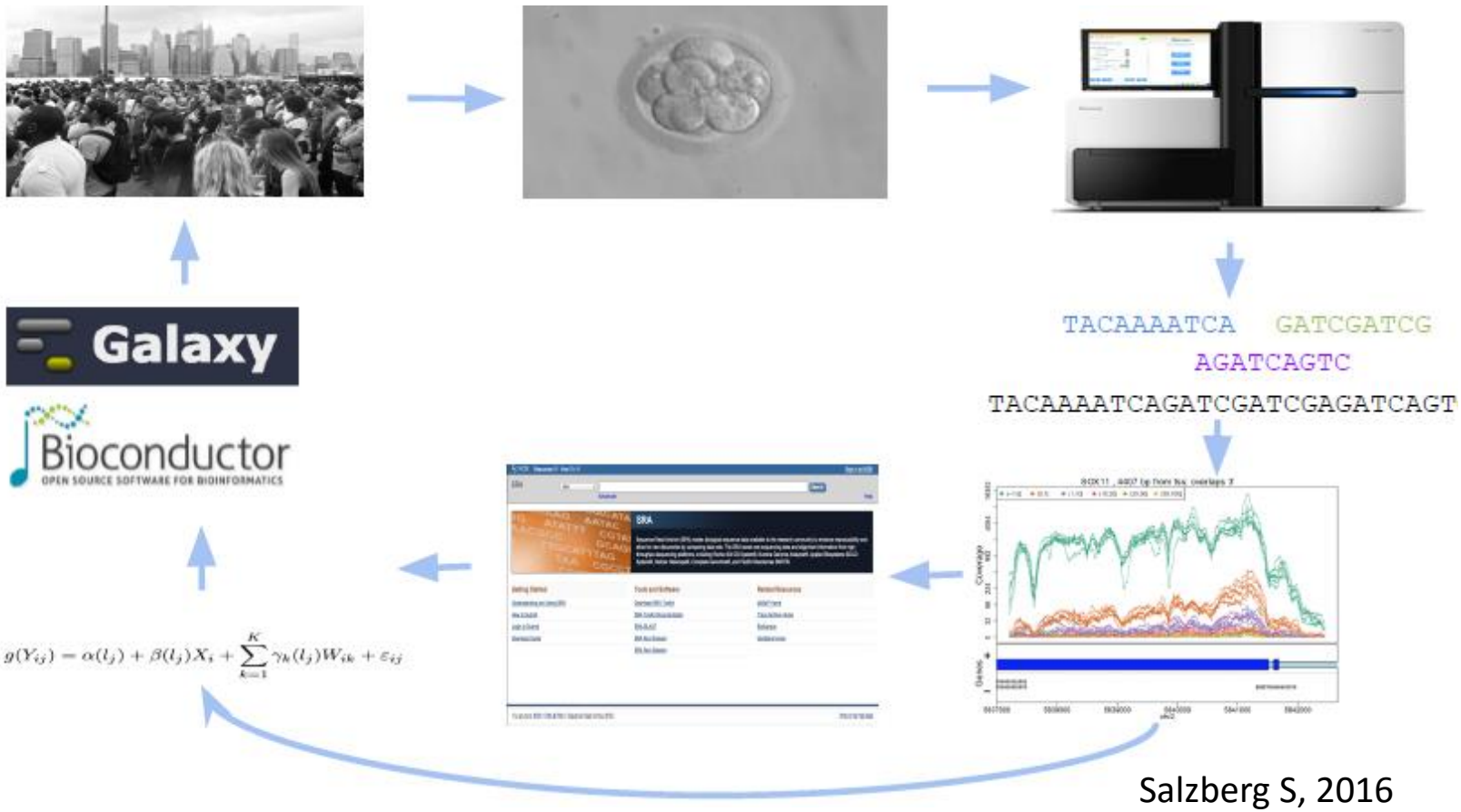
- Campo de la ciencia que se ocupa de la información y flujo de información en los sistemas biológicos, apoyado en el uso de métodos computacionales en genética y genómica(OED, 2015).

Objetivo

- Facilitar el descubrimiento de nuevas ideas biológicas así como crear perspectivas globales a partir de las cuales se puedan discernir principios unificadores en biología (NCBI, 2001).



¿Cuál es el potencial de esta área?



Salzberg S, 2016

¿Qué hemos aprendido a la fecha?

TACAAAATCA GATCGATCG
AGATCAGTC
TACAAAATCAGATCGATCGAGATCAG

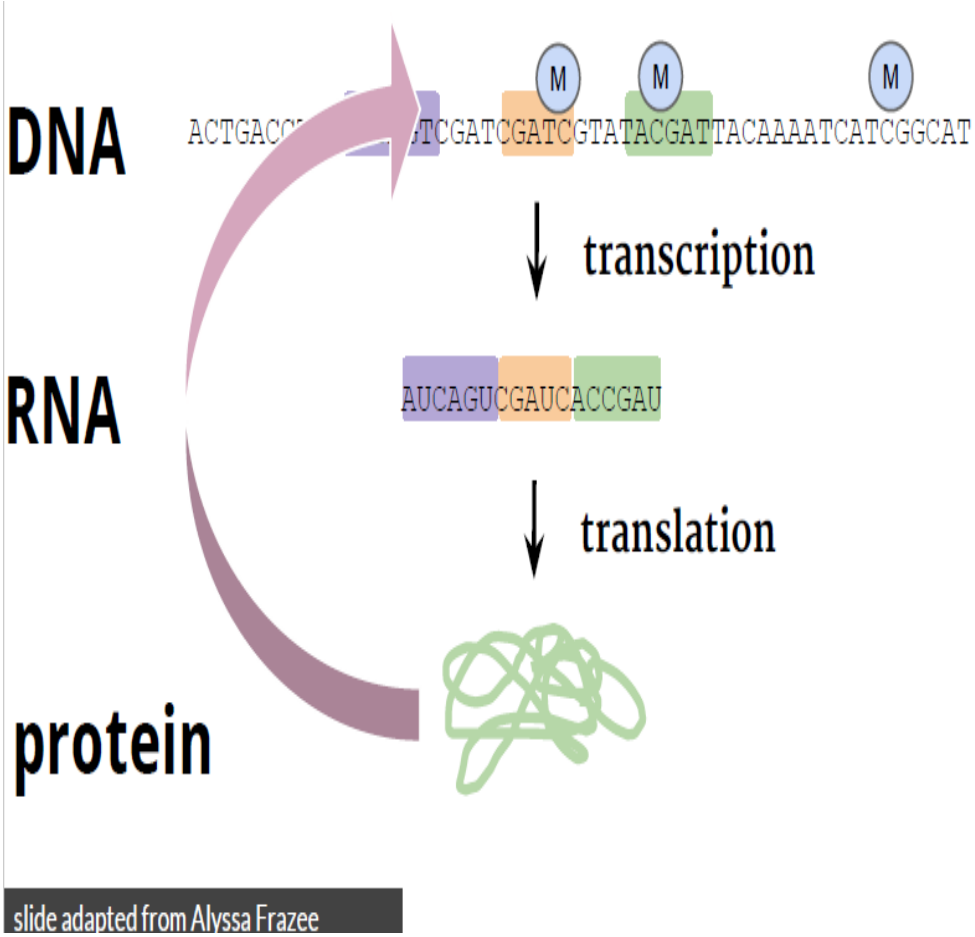
Desarrollo de hipótesis

Diseño experimental

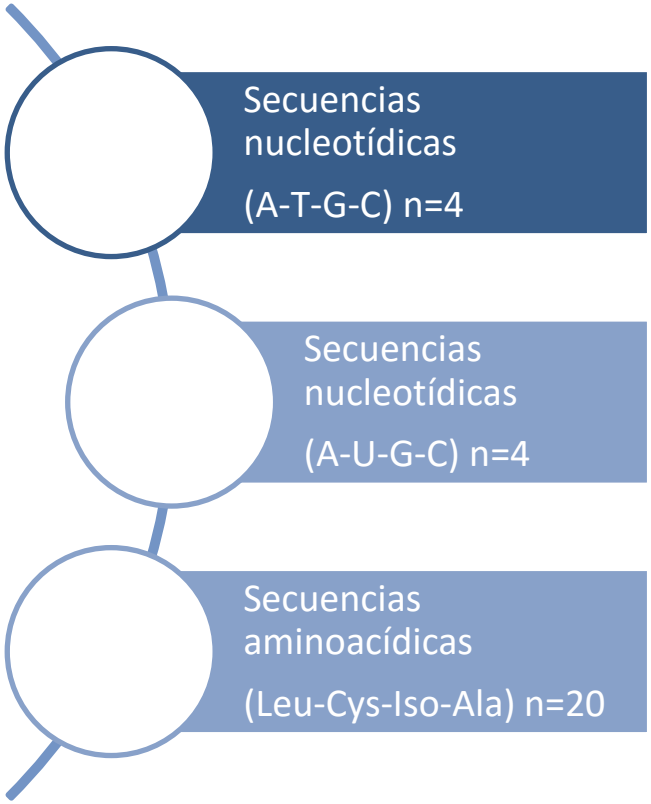
Análisis de resultados

Salzberg S, 2016

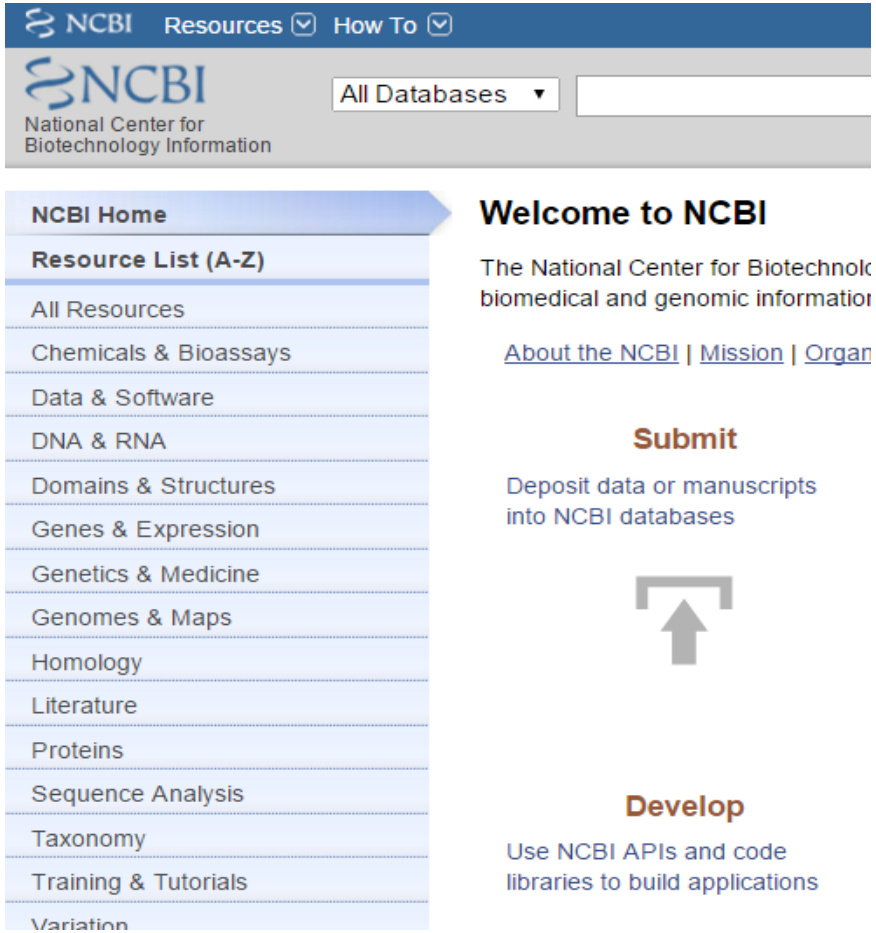
Dogma central de la biología



Flujo de la información



¿Qué nos falta aprender?



Uso de bases de datos

Herramientas de comparación de secuencias

Búsqueda de propiedades en una secuencia

Bases de datos

Bases de datos

ScienceDirect

Springer Link

EBSCO Information Services

WEB OF SCIENCE™

PubMed

PubMed comprises more than 26 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Ingresar a:

http://www.ncbi.nlm.nih.gov/

NCBI Resources How To Sign in to NCBI

NCBI
National Center for Biotechnology Information

All Databases Search

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News | Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

Popular Resources
PubMed
Bookshelf
PubMed Central
PubMed Health
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI Announcements
HTTPS at NCBI: Guidance for NCBI web API users
27 Jul 2016
As originally announced on June 10, NCBI will be moving all web services to

Obteniendo secuencia nucleotídica

http://www.ncbi.nlm.nih.gov/nuccore

NCBI Resources How To

Nucleotide | Nucleotide | glutamate dehydrogenase

Create alert Advanced

NCBI is phasing out sequence GI numbers in September 2016. Please use accession.version! Read more...

Species
Animals (2,002)
Plants (1,609)
Fungi (2,342)
Protists (2,846)
Bacteria (160,923)
Archaea (1,558)
Viruses (42)
Customize ...

Molecule types
genomic DNA/RNA (168,981)
mRNA (2,256)
Customize ...

Source databases
INSDC (GenBank) (85,520)
RefSeq (85,814)
Customize ...

Genetic compartments

Summary 20 per page Sort by Default order Send to:

Items: 1 to 20 of 171348

<< First < Prev Page 1 of 8568 Next > Last >>

Found 173215 nucleotide sequences. Nucleotide (171348) EST (1834) GSS (33)

1. Trypanosoma grayi **glutamate dehydrogenase** partial mRNA
   3,000 bp linear mRNA
   Accession: XM_009315222.1 GI: 686642102
   GenBank FASTA Graphics

2. Human **glutamate dehydrogenase** (GDH) mRNA, complete cds
   2,950 bp linear mRNA
   GenBank FASTA Graphics

3. Escherichia coli **glutamate dehydrogenase** (gdhA) gene, complete cds
   1,779 bp linear DNA

Secuencia nucleotídica

Escherichia coli glutamate dehydrogenase (gdhA) gene, complete cds

GenBank: J01615.1

FASTA Graphics **Formato**

Go to:

```
LOCUS       ECOGDHA                 1779 bp    DNA     linear   BCT 10-FEB-2004
DEFINITION  Escherichia coli glutamate dehydrogenase (gdhA) gene, complete cds.
ACCESSION   J01615 K00565 M23171 X00988
VERSION     J01615.1  GI:146123
KEYWORDS    NADP-specific glutamate dehydrogenase; dehydrogenase; gdhA gene;
            glutamate dehydrogenase.
SOURCE      Escherichia coli
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1  (bases 361 to 714)
  AUTHORS   Mattaj,I.W., McPherson,M.J. and Wootton,J.C.
  TITLE     Localisation of a strongly conserved section of coding sequence in
            glutamate dehydrogenase genes
  JOURNAL   FEBS Lett. 147 (1), 21-25 (1982)
   PUBMED   6754449
REFERENCE   2  (bases 1 to 498)
  AUTHORS   Valle,F., Sanvicente,E., Seeburg,P., Covarrubias,A., Rodriguez,R.L.
            and Bolivar,F.
  TITLE     Nucleotide sequence of the promoter and amino-terminal coding
            region of the glutamate dehydrogenase structural gene of
            Escherichia coli
  JOURNAL   Gene 23 (2), 199-209 (1983)
   PUBMED   6225701
REFERENCE   3  (bases 121 to 1779)
```

Formato FASTA

Escherichia coli glutamate dehydrogenase (gdhA) gene, c

GenBank: J01615.1

GenBank Graphics

```
>gi|146123|gb|J01615.1|ECOGDHA Escherichia coli glutamate dehydrogenase (gdhA) gene,
complete cds
CCGGGTGGCAAAACTTTAGCGTCTGAGGTTATCGCATTTGGTTATGAGATTACTCTCGTTATTAATTTGC
TTTCCTGGGTCATTTTTTTCTTGCTTACCGTCACATTCTTGATGGTATAGTCGAAAACTGCAAAAGCACA
TGACATAAACAACATAAGCACAATCGTATTAATATATAAGGGTTTTATATCTATGGATCAGACATATTCT
CTGGAGTCATTCCTCAACCATGTCCAAAAGCGCGACCCGAATCAAACCGAGTTCGCGCAAGCCGTTCGTG
AAGTAATGACCACACTCTGGCCTTTTCTTGAACAAAATCCAAAATATCGCCAGATGTCATTACTGGAGCG
TCTGGTTGAACCGGAGCGCGTGATCCAGTTTCGCGTGGTATGGGTTGATGATCGCAACCAGATACAGGTC
AACCGTGCATGGCGTGTGCAGTTCAGCTCTGCCATCGGCCCGTACAAAGGCGGTATGCGCTTCCATCCGT
CAGTTAACCTTTCCATTCTCAAATTCCTCGGCTTTGAACAAACCTTCAAAAATGCCCTGACTACTCTGCC
GATGGGCGGTGGTAAAGGCGGCAGCGATTTCGATCCGAAAGGAAAAAGCGAAGGTGAAGTGATGCGTTTT
TGCCAGGCGCTGATGACTGAACTGTATCGCCACCTGGGCGCGGATACCGACGTTCCGGCAGGTGATATCG
GGGTTGGTGGTCGTGAAGTCGGCTTTATGGCGGGGATGATGAAAAAGCTCTCCAACAATACCGCCTGCGT
CTTCACCGGTAAGGGCCTTTCATTTGGCGGCAGTCTTATTCGCCCGGAAGCTACCGGCTACGGTCTGGTT
TATTTCACAGAAGCAATGCTAAAACGCCACGGTATGGGTTTTGAAGGGATGCGCGTTTCCGTTTCTGGCT
CCGGCAACGTCGCCCAGTACGCTATCGAAAAAGCGATGGAATTTGGTGCTCGTGTGATCACTGCGTCAGA
CTCCAGCGGCACTGTAGTTGATGAAAGCGGATTCACGAAAGAGAAACTGGCACGTCTTATCGAAATCAAA
GCCAGCCGCGATGGTCGAGTGGCAGATTACGCCAAAGAATTTGGTCTGGTCTATCTCGAAGGCCAACAGC
CGTGGTCTCTACCGGTTGATATCGCCCTGCCTTGCGCCACCCAGAATGAACTGGATGTTGACGCCGCGCA
TCAGCTTATCGCTAATGGCGTTAAAGCCGTCGCCGAAGGGGCAAATATGCCGACCACCATCGAAGCGACT
GAACTGTTCCAGCAGGCAGGCGTACTATTTGCACCGGGTAAAGCGGCTAATGCTGGTGGCGTCGCTACAT
CGGGCCTGGAAATGGCACAAAACGCTGCGCGCCTGGGCTGGAAAGCCGAGAAAGTTGACGCACGTTTGCA
TCACATCATGCTGGATATCCACCATGCCTGTGTTGAGCATGGTGGTGAAGGTGAGCAAACCAACTACGTG
CAGGGCGCGAACATTGCCGGTTTTGTGAAGGTTGCCGATGCGATGCTGGCGCAGGGTGTGATTTAAGTTG
TAAATGCCTGATGGCGCTACGCTTATCAGGCCTACAAATGGGCACAATTCATTGCAGTTACGCTCTAATG
TAGGCCGGGCAAGCGCAGCGCCCCCGGCAAAATTTCAGGCGTTTATGAGTATTTAACGGATGATGCTCCC
CACGGAACATTTCTTATGGGCCAACGGCATTTCTTACTGTAGTGCTCCCAAAACTGCTTGTCGTAACGAT
AACACGCTTCAAGTTCAGCATCCGTTAAC
```

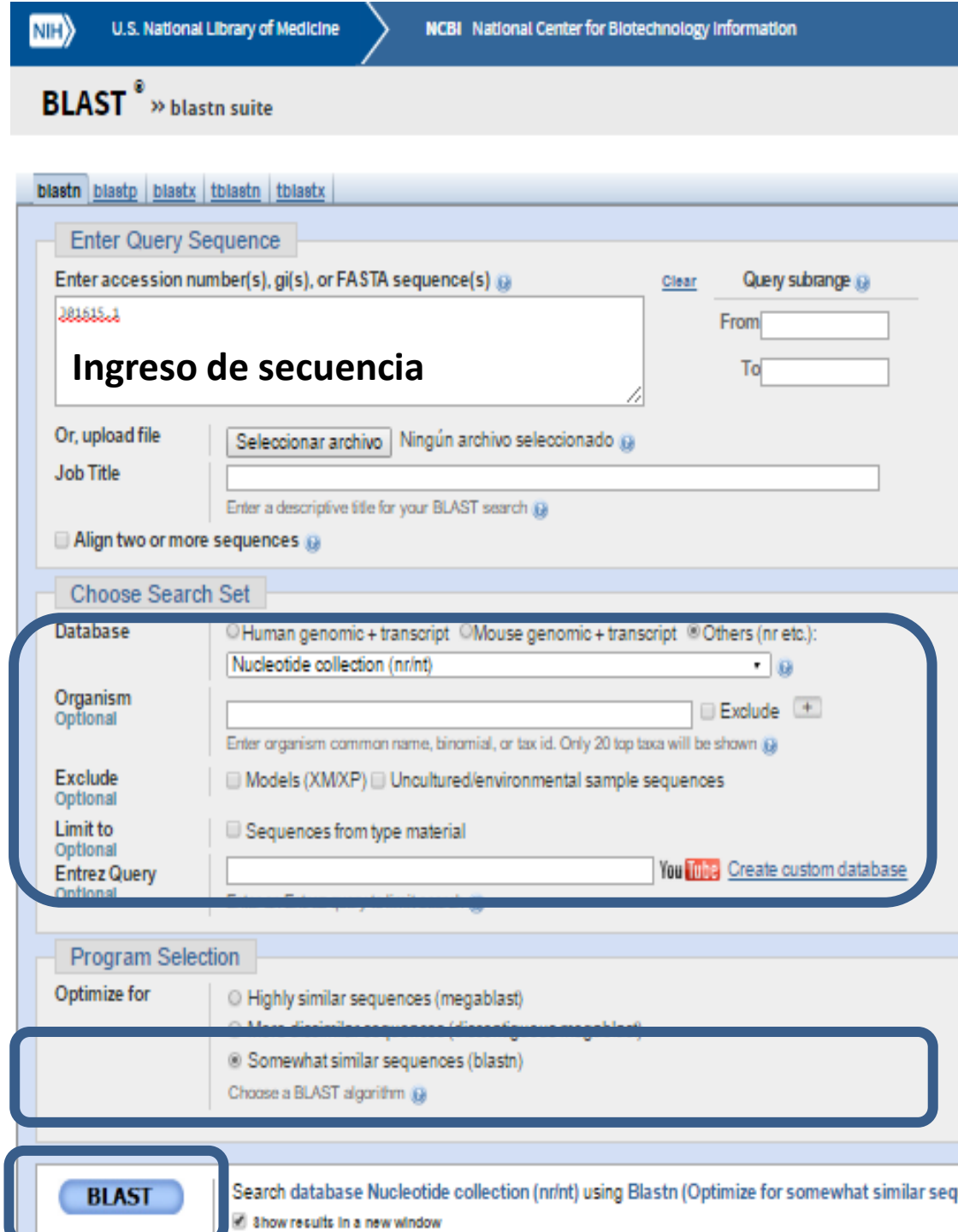
Herramientas de comparación: BLAST n y BLAST p

BLAST n

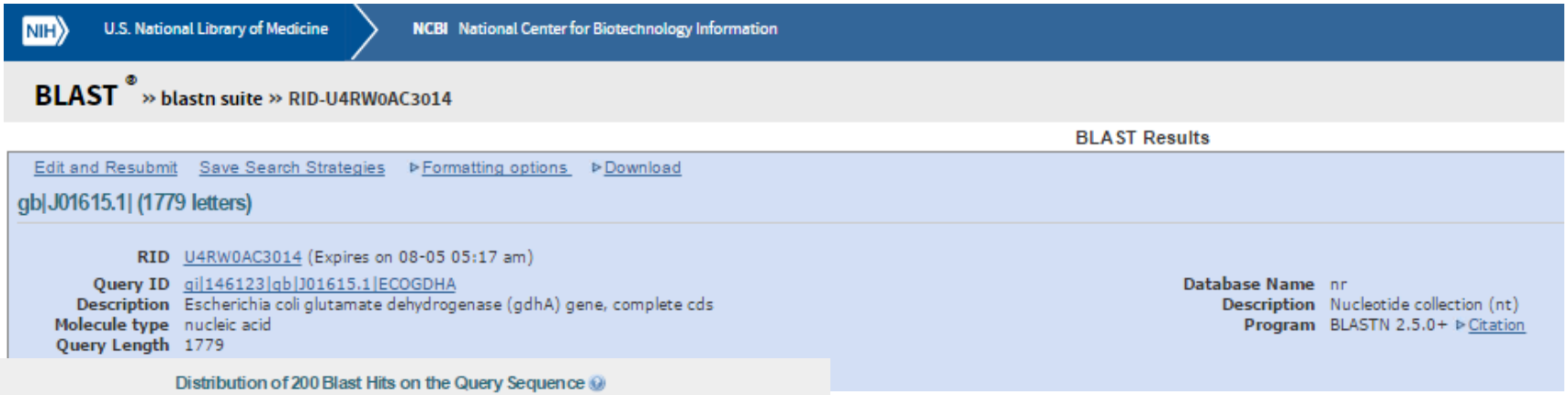
Caso de estudio

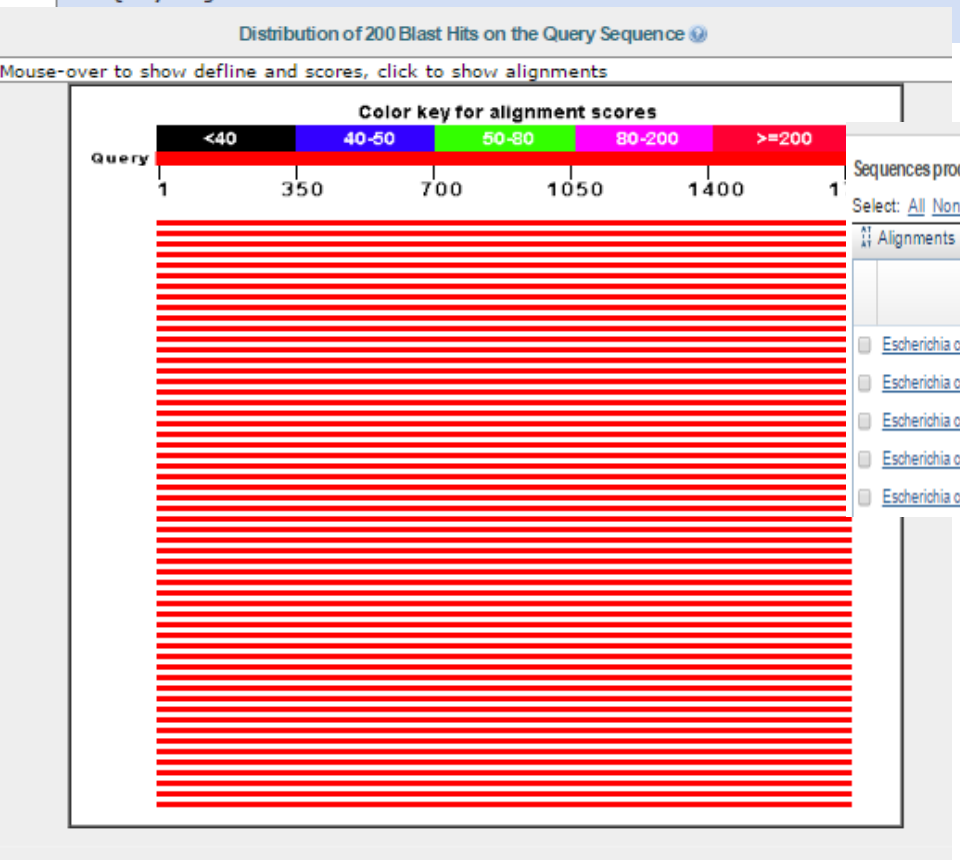
Tengo una secuencia aminoacídica de un gen XX

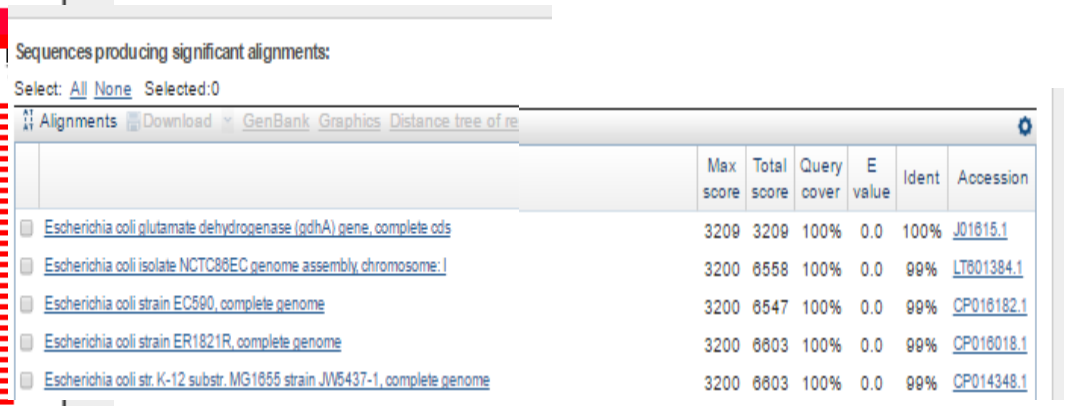
¿Cómo puedo determinar la identidad del gen o similitud con otros ?



Resultados







- Colores
- E-value
- %identidad
- %Cobertura

BLAST p

Caso de estudio

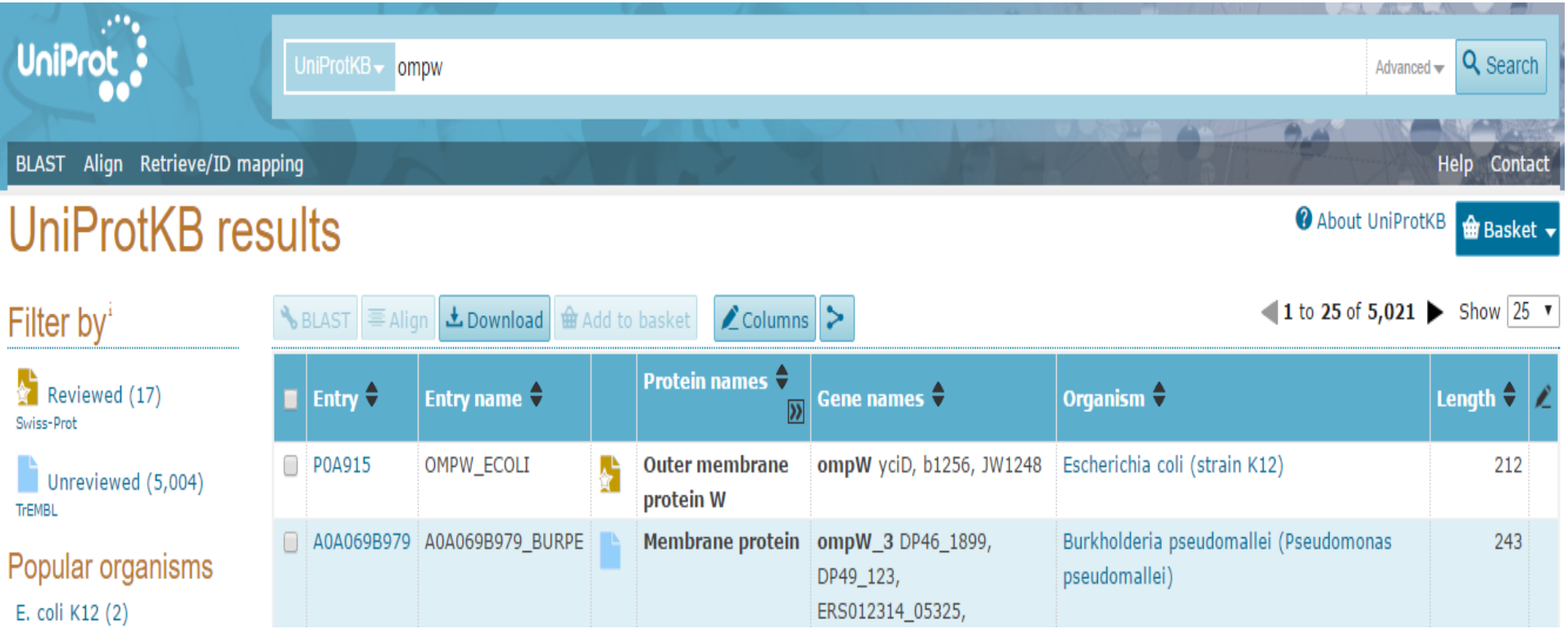
Tengo una secuencia aminoacídica de una proteína

¿Cómo puedo determinar la identidad de la proteína?

Propiedades de las proteínas

Base de datos con información sobre proteínas

http://www.uniprot.org/



Copiar código UniProt

UniProtKB - P0A915 (OMPW_ECOLI)

Display

Entry

Feature viewer

Feature table

None

- [x] Function
- [x] Names & Taxonomy
- [x] Subcellular location
- [ ] Pathology & Biotech
- [x] PTM / Processing
- [ ] Expression
- [x] Interaction

BLAST | Align | Format | Add to basket | History

Protein: **Outer membrane protein W**

Gene: **ompW**

Organism: *Escherichia coli (strain K12)*

Status: Reviewed - Annotation score: ●●●○○ - Experimental evidence at protein level

Function

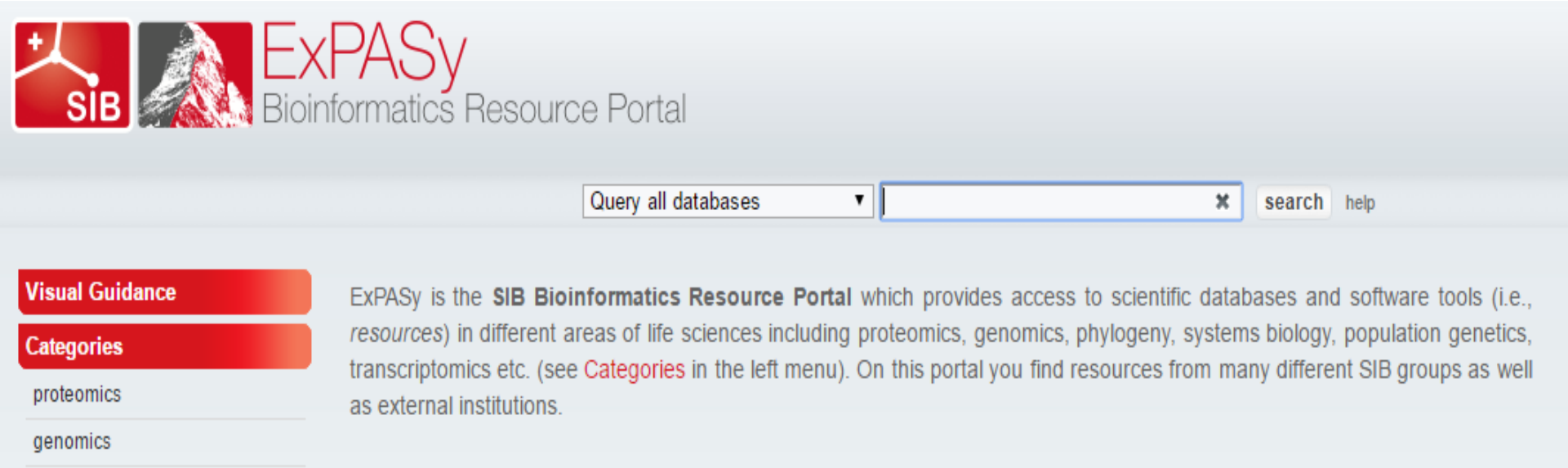
Acts as a receptor for colicin S4. 1 Publication

Enzyme and pathway databases

BioCyc	EcoCyc:EG11124-MONOMER. ECOL316407:JW1248-MONOMER.

Base de datos con información sobre proteínas

http://www.expasy.org/

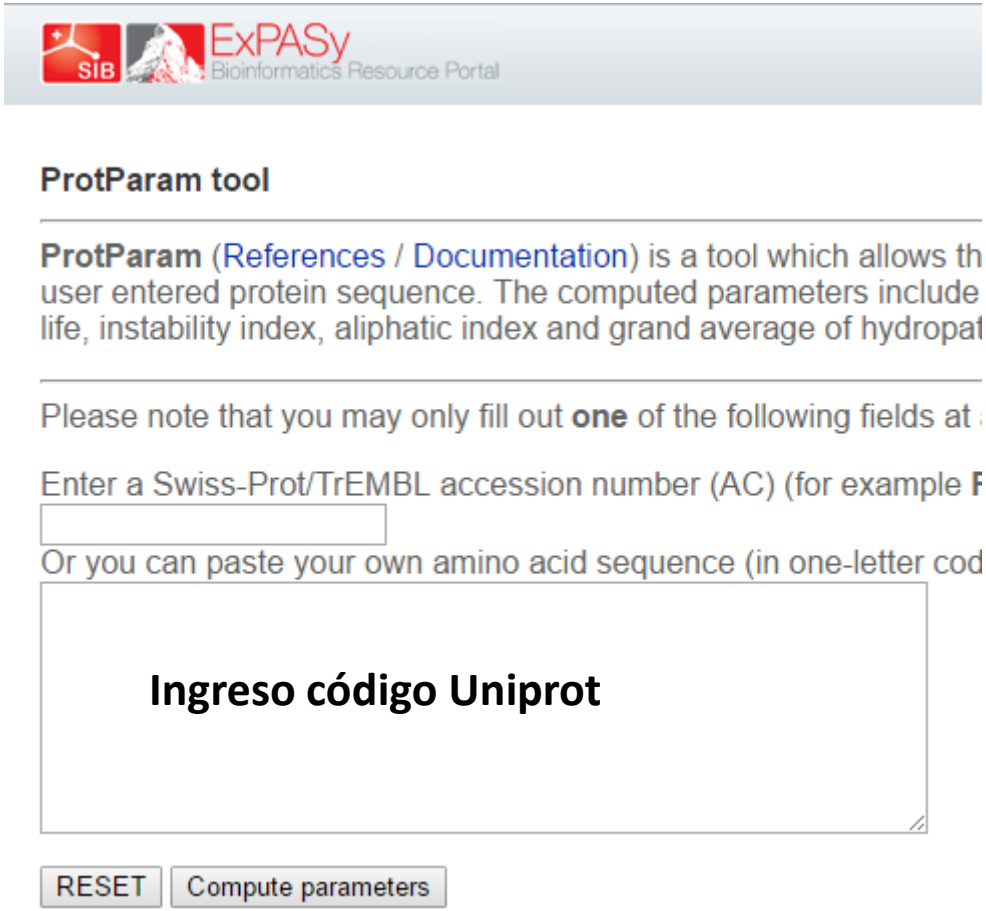


Peso molecular y propiedades fisicoquímicas

Ruta:

- Página inicio ExPASy
- Resources A..Z

  Compute ProtParam tool

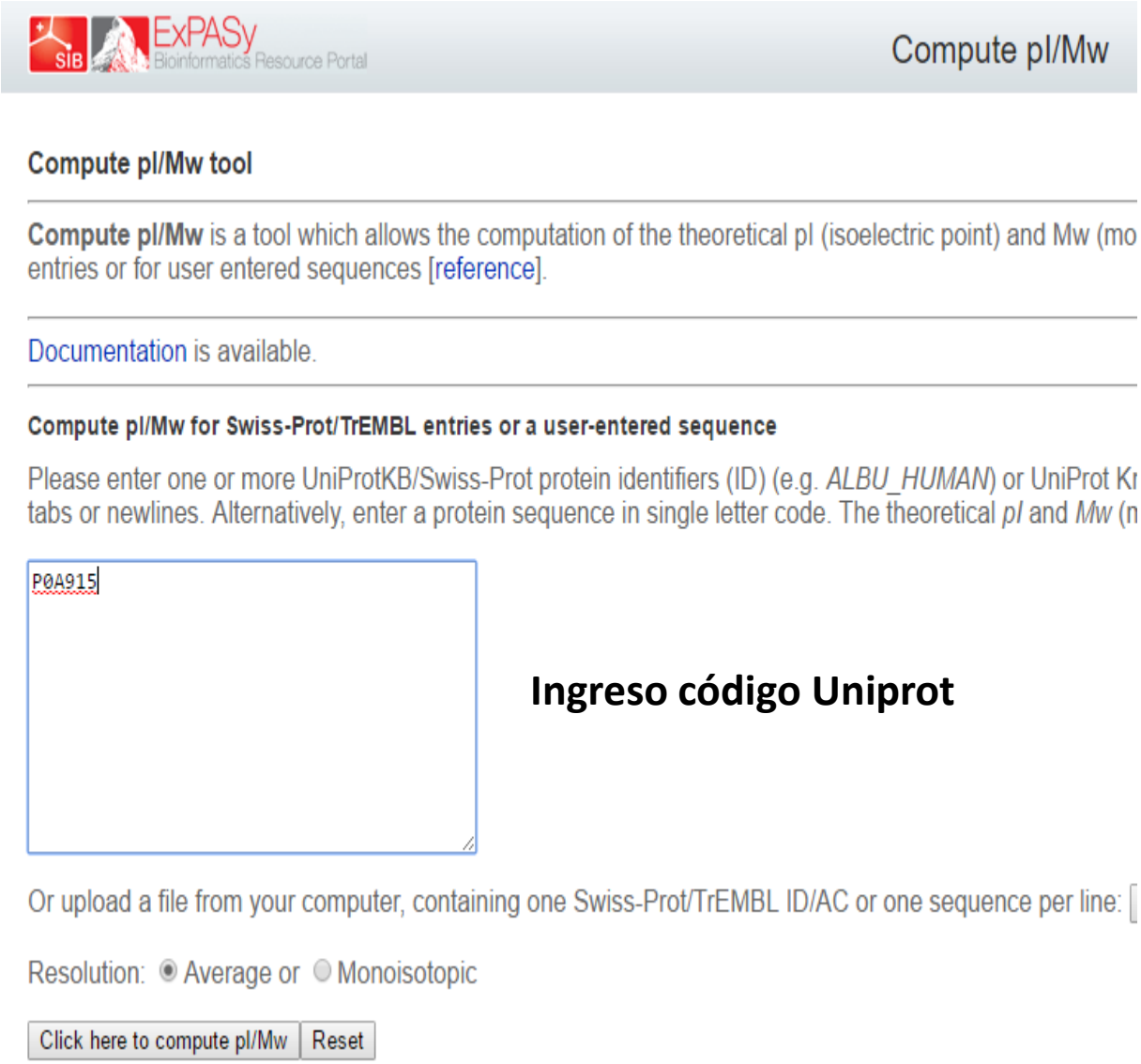


Punto isoeléctrico

Ruta:

- Página inicio ExPASy
- Resources A..Z

  Compute pI/Mw

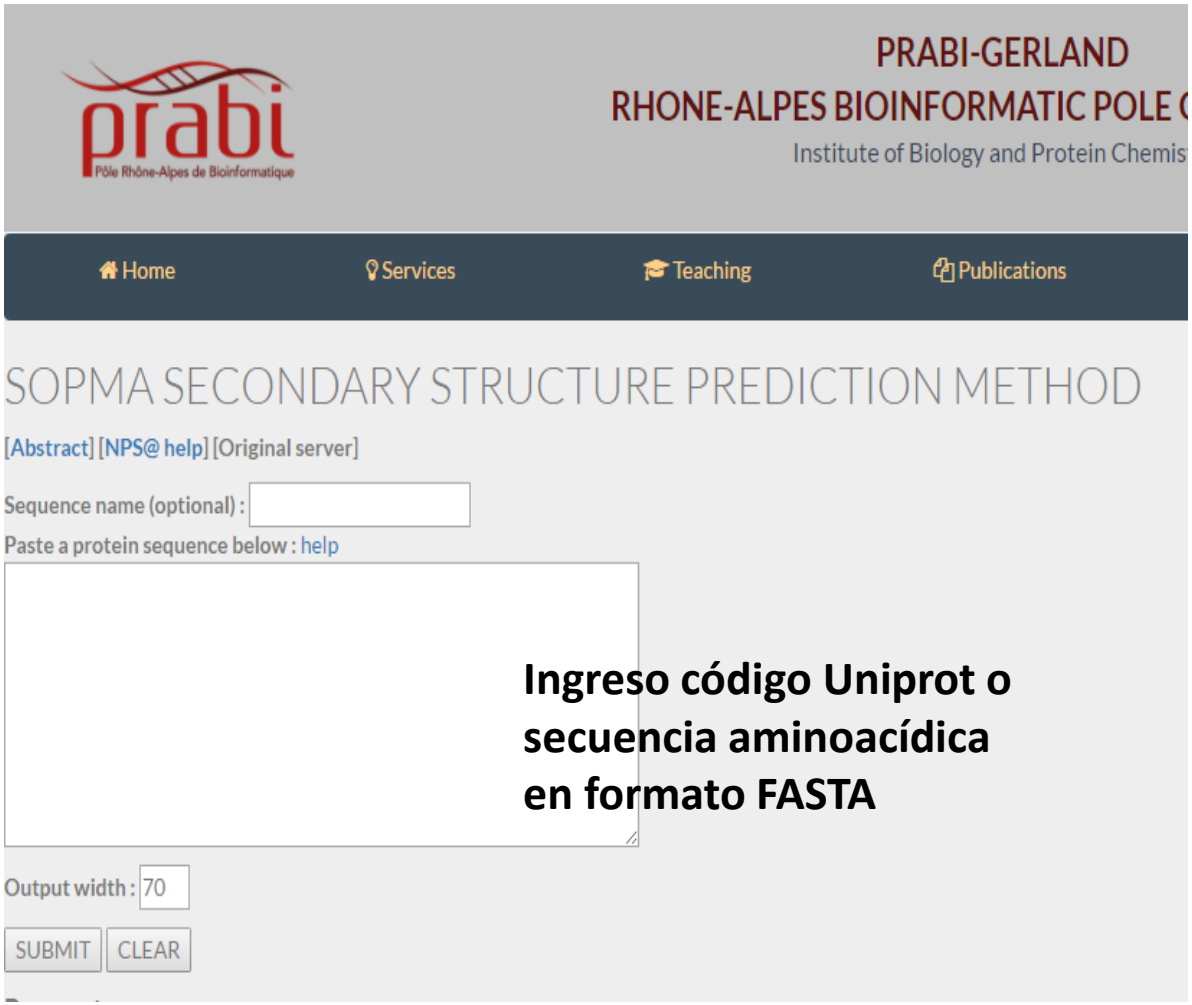


Estructura secundaria

Ruta:

- Página inicio ExPASy
- Resources A..Z

  SOPMA

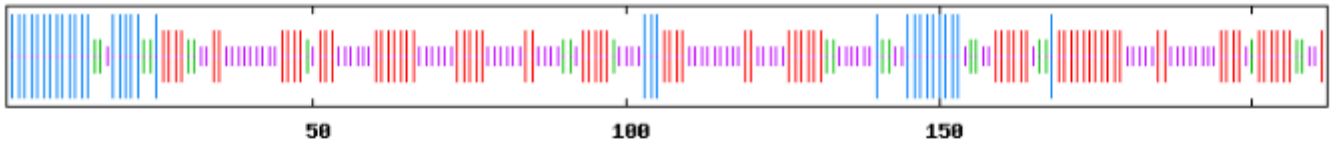


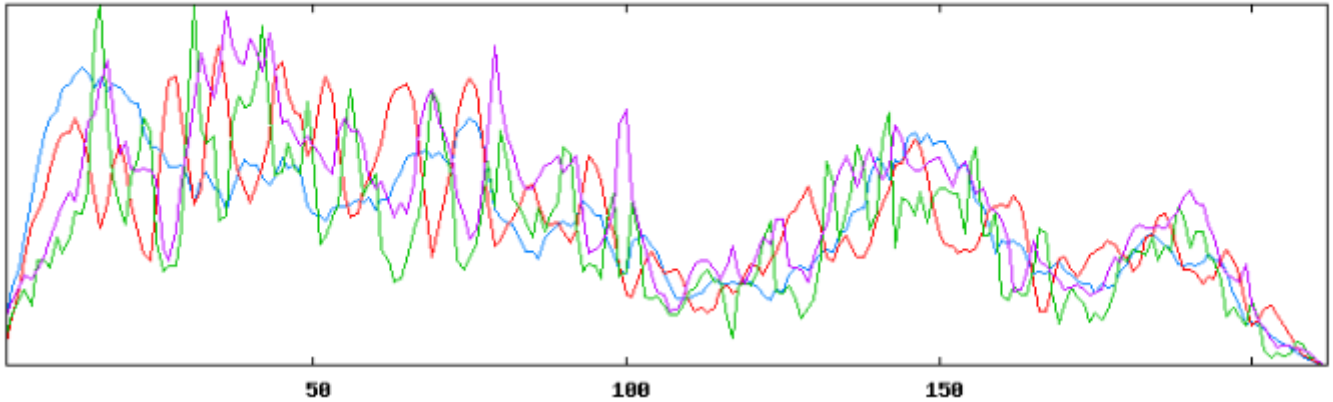
Ingreso código Uniprot o secuencia aminoacídica en formato FASTA

Resultados

```
Sequence length :   212
```

```
SOPMA :
   Alpha helix      (Hh) :    35 is  16.51%
   3₁₀ helix        (Gg) :     0 is   0.00%
   Pi helix         (Ii) :     0 is   0.00%
   Beta bridge      (Bb) :     0 is   0.00%
   Extended strand  (Ee) :    74 is  34.91%
   Beta turn        (Tt) :    21 is   9.91%
   Bend region      (Ss) :     0 is   0.00%
   Random coil      (Cc) :    82 is  38.68%
   Ambiguous states (?)  :     0 is   0.00%
   Other states          :     0 is   0.00%
```





Dominios transmembrana

Stockholm Bioinformatics Centre SBC

Phobius

A combined transmembrane topology and signal peptide predictor

Ruta:

- Página inicio ExPASy
- Resources A..Z

  Phobius

Normal prediction | Constrained prediction | PolyPhobius | Instructions | Download | Mirror site at KU

Normal prediction

Paste your protein sequence here in Fasta format:

Ingreso código Uniprot o secuencia aminoacídica en formato FASTA

Or: Select the sequence file you wish to use [Seleccionar archivo] Ningún archivo seleccionado

Select output format:

- Short
- Long without Graphics
- Long with Graphics

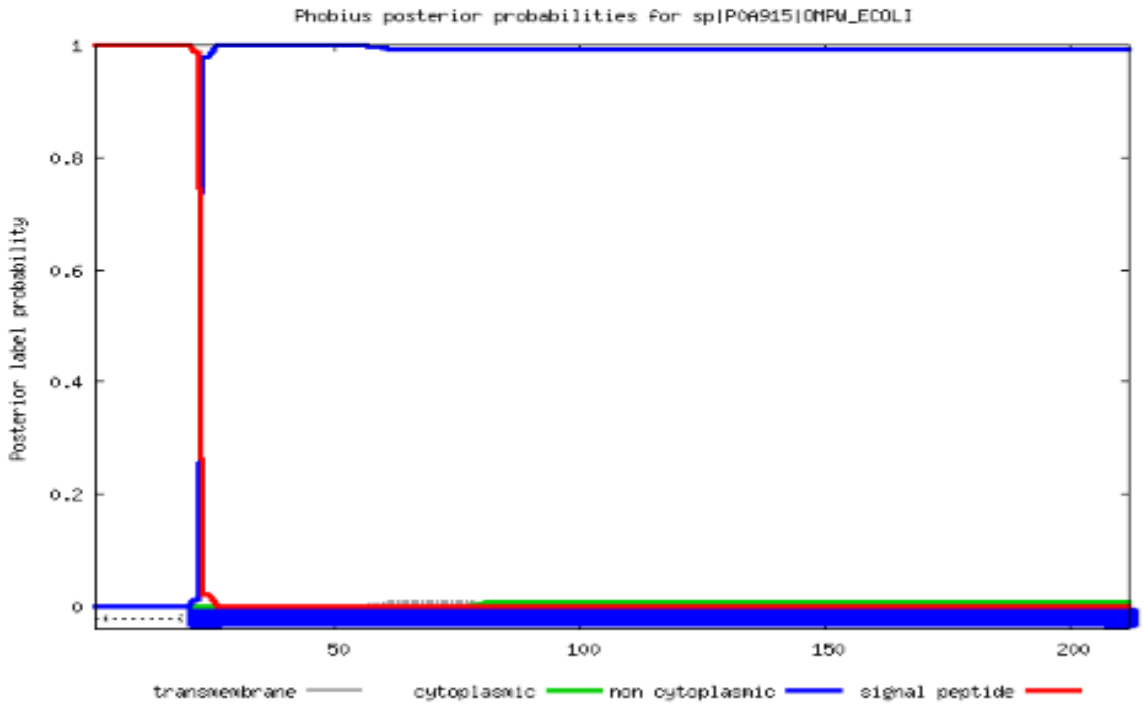
Resultados

¿Posee péptido señal o dominios transmembrana?

Phobius prediction

Prediction of sp|P0A915|OMPW_ECOLI

```
ID   sp|P0A915|OMPW_ECOLI
FT   SIGNAL        1     21
FT   REGION        1      3       N-REGION.
FT   REGION        4     15       H-REGION.
FT   REGION       16     21       C-REGION.
FT   TOPO_DOM     22    212       NON CYTOPLASMIC.
//
```



The probability data used in the plot is found here, and the gnuplot script is here.

¿Preguntas, sugerencias, comentarios?