



# UNIVERSIDAD NACIONAL DE SAN AGUSTÍN

FACULTAD DE INGENIERÍA DE PRODUCCIÓN Y SERVICIOS  
CIENCIA DE LA COMPUTACIÓN

---

## **Práctica 03: Alineamiento de pares de secuencias**

---

**CURSO:**

BIOINFORMÁTICA

**ALUMNO**

KELVIN PAUL PUCHO ZEVALLOS

**DOCENTE:**

GUADALUPE DEL ROSARIO QUISPE SAJI

21 de Mayo del 2022

1. **Obtener secuencias de proteínas humanas HBA e HBB.**
  - a. **Realice el alineamiento global y local usando alguna de las herramientas propuestas en las tablas 1 y 2.**

## ALINEAMIENTO GLOBAL

Se usó Stretcher Del Instituto Pasteur; alineamiento global (EMBOSS)  
<http://www.bioinformatics.nl/cgi-bin/emboss/stretcher>

## CAPTURA DE LA HERRAMIENTA STRETCHER

En este caso ingrese la secuencia de proteínas humanas HBA e HBB

### stretcher

Needleman-Wunsch rapid global alignment of two sequences ([read the manual](#))

Only required fields are visible. ([show optional fields](#))

Input section

Select an input sequence. Use one of the following three fields:

1. To access a sequence from a database, enter the USA here:

2. To upload a sequence from your local computer, select it here: 

Seleccionar archivo Sin archivos seleccionados

3. To enter the sequence data manually, type here:

MVLSPADKTN VKAANGKVG A HAGEYGAEAL ERMFLSFPTT KTYFPHFELS  
HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK  
LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR

Select an input sequence. Use one of the following three fields:

1. To access a sequence from a database, enter the USA here:

2. To upload a sequence from your local computer, select it here: 

Seleccionar archivo Sin archivos seleccionados

3. To enter the sequence data manually, type here:

MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS  
TPDAVMGNPK VKAHGKKVLG AFSDGLAHLN NLKGTFTATLS ELHCDKLHVD  
PENFRLLGNV LVCVLAHHFG KEFTTPPVQAA YQKVVAGVAN ALAHKYH

Additional section

Output section

Run section

Email address:

If you are submitting a long job and would like to be informed by email when it finishes, enter your email address here.

Run stretcher

Restablecer

## CAPTURA DE LOS RESULTADOS DE LA HERRAMIENTA STRETCHER

```
# Length: 149
# Identity: 65/149 (43.6%)
# Similarity: 90/149 (60.4%)
# Gaps: 9/149 ( 6.0%)
# Score: 277
#
```

## OUTPUT FILE [outfile](#)

```
#####
# Program: stretcher
# Rundate: Sun 22 May 2022 00:31:09
# Commandline: stretcher
#   -auto
#   -asequence /var/lib/emboss-explorer/output/887899/.asequence
#   -bsequence /var/lib/emboss-explorer/output/887899/.bsequence
#   -outfile outfile
# Align_format: markx0
# Report_file: outfile
#####
```

```
#=====
#
# Aligned_sequences: 2
# 1:
# 2:
# Matrix: EBL0SUM62
# Gap_penalty: 12
# Extend_penalty: 2
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    90/149 (60.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 277
#
#
#=====
```

```

      10      20      30      40
MV-LSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-D
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD
      10      20      30      40

      50      60      70      80      90
LSH-----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLNKLKGTfATLSELHCDKLH
      50      60      70      80      90

      100     110     120     130     140
VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
:: :: :: :: :: :: :: :: :: :: :: :: :: :: :: ::
VDPENFRLLGNVLVLCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH
100      110      120      130      140
```

```
#-----
#-----
```

# CAPTURA DE LA HERRAMIENTA VIRGINIA ALINEAMIENTO LOCAL

PRSS De la Universidad de Virginia (B.Pearson)

[https://fasta.bioch.virginia.edu/fasta\\_www2/fasta\\_www.cgi](https://fasta.bioch.virginia.edu/fasta_www2/fasta_www.cgi)

**Compare Two Sequences**  
[Search Databases with FASTA](#)  
[Statistical Significance from Shuffles](#)  
[Find Internal Duplications \(lalign/plalign\)](#)

**Align two sequences** aligns two sequences using the indicated algorithm, and calculates the statistical significance using shuffled sequences.  
**New:** Annotation features available for Uniprot/SwissProt/PIR1 library searches.

**Choose: (A) program and (B, C) sequences to compare:**

**(A) Program:** SSEARCH: local protein:protein

**(B.1) Enter first (query) sequence:** Accession/GI number Subset range:

HBA\_HUMAN

Annotate Query Sequence (SwissProt accessions)  
No annotation  
[Upload annotation file:](#) Seleccionar archivo Sin archivos seleccionados  
[Entrez protein](#) / [Entrez DNA](#) sequence browser  
[Uniprot sequence browser](#)

**(B.2) Or upload sequence from file:** Seleccionar archivo Sin archivos seleccionados  
☒ Protein ☐ DNA (both-strands) ☐ DNA (forward only) ☐ DNA (rev-comp only) [Use PSSM:](#) ☐

**(C.1) Enter the second sequence:** Accession/GI number Subset range:

HBB\_HUMAN

Annotate Target Sequence (SwissProt accessions)  
Uniprot Domains/Uniprot func sites  
[Upload annotation file:](#) Seleccionar archivo Sin archivos seleccionados  
[Compare Sequences](#) [Reset Form](#)

**(C.2) Or choose file of sequences/accessions:** Seleccionar archivo Sin archivos seleccionados

**Other comparison options:**  
**Scoring matrix:** Blosum50 (25%) **open:** -10 **ext:** -2 **Ktup:** ktup = 2

**Output limits:**  
**E():** 10  
Highlight ☒ similarities ☐ differences.

[FASTA program information](#) | [Download FASTA](#) | [About the Author](#)

Copyright © 1996, 1997, 1998, 1999, 2002, 2014, 2015 by William R. Pearson and The Rector & Visitors of the University of Virginia  
The FASTA package is open source software, licensed under the Apache License, Version 2.0 (the "License"); you may not copy this software except in compliance with the License. You may obtain a copy of the License at:  
<http://www.apache.org/licenses/LICENSE-2.0>

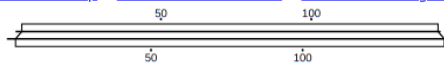
## CAPTURA DE LOS RESULTADOS DE LA HERRAMIENTA VIRGINIA

```
>>sp|P68871|HBB HUMAN Hemoglobin subunit beta; Beta-globin; Hemoglobin beta (147 aa)
s-w opt: 381 Z-score: 470.8 bits: 93.1 E(10000): 2e-20
Smith-Waterman score: 381; 43.4% identity (74.5% similar) in 145 aa overlap (3-141:4-146)
Sequence Lookup General re-search Pairwise alignment
```

Query: TMP.q  
1>>>sp|P69905|HBA\_HUMAN Hemoglobin subunit alpha; Alpha-globin; Hemoglobin alpha chain - 142 aa  
Library: TMP.q2  
147 residues in 1 sequences  
Statistics: (shuffled [500]) MLE statistics: Lambda= 0.1594; K=0.02337  
statistics sampled from 1 (1) to 500 sequences  
Algorithm: Smith-Waterman (SSE2, Michael Farrar 2006) (7.2 Nov 2010/SIMDe Nov 2020)  
Parameters: BL50 matrix (15:-5), open/ext: -10/-2  
Scan time: 0.010  
The best scores are:  
sp|P68871|HBB\_HUMAN Hemoglobin subunit beta; Beta-globin ( 147) 381 93.1 2e-20 0.434 0.745 145 align

>>>sp|P69905|HBA\_HUMAN, 142 aa vs TMP.q2 library

I  
>>sp|P68871|HBB\_HUMAN Hemoglobin subunit beta; Beta-globin; Hemoglobin beta (147 aa)  
s-w opt: 381 Z-score: 470.8 bits: 93.1 E(10000): 2e-20  
Smith-Waterman score: 381; 43.4% identity (74.5% similar) in 145 aa overlap (3-141:4-146)  
[Sequence Lookup](#) [General re-search](#) [Pairwise alignment](#)



[\[Domains\]](#)

[\[alignment\]](#)

	10	20	30	40	50	60	70	80	90	100	110
sp P69	MVLSPADKTNVKA	AWGKVG	AHAGEYGA	EALERMFL	SFPTTKTY	FPHF-DLS----	HGSAQVKG	HGKKVAD	ALTNAVAH	VDDMPNALS	SALSDLHA
	..	..	..	..	..	..	..	..	..	..	..
sp P68	MVHLTPEEKSA	VTALWGKV-	NVDEVGGE	ALGRLLV	VPWTQR	FFESFGDL	STPDVAV	MGNPKV	KAHGKKV	LGA	FS
	..	..	..	..	..	..	..	..	..	..	..
	120	130	140								
sp P69	LPAEFTPAVHAS	LDKFLAS	VSTVLT	SKYR							
	..	..	..	..	..	..	..	..	..	..	..
sp P68	FGKEFTPPVQA	AYQKV	VAGVAN	ALAHKYH							
	120	130	140								

142 residues in 1 query sequences  
147 residues in 1 library sequences  
Tcomplib [36.3.8i Sept, 2021] (48 proc in memory [0G])  
start: Sat May 21 19:05:05 2022 done: Sat May 21 19:05:05 2022  
Total Scan time: 0.010 Total Display time: 0.000

Function used was SSEARCH [36.3.8i Sept, 2021]

**b. Realizar el alineamiento de pares usando el Blast do site NCBI.**

BLAST<sup>®</sup> » blastp suite

Align Sequences

blastnblastpblastxtblastntblastx

BLASTP programs search protein sul

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

P69905

Query subrange [?](#)

From

To

Or, upload file

Seleccionar archivo Sin archivos seleccionados [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

☒ Align two or more sequences [?](#)

Enter Subject Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#) [Clear](#)

P68871

Subject subrange [?](#)

From

To

Or, upload file

Seleccionar archivo Sin archivos seleccionados [?](#)

Program Selection

Algorithm

☒ blastp (protein-protein BLAST)

Choose a BLAST algorithm [?](#)

BLAST

Search protein sequence using Blastp (protein-protein BLAST)

☐ Show results in a new window

Note: Parameter values that differ from the default are highlighted in yellow

+ Algorithm parameters

NIH National Library of Medicine  
National Center for Biotechnology Information

Log in

BLAST<sup>®</sup> » blastp suite-2sequences » results for RID-8JWPMV4N114

Home Recent Results Saved Strategies Help

[← Edit Search](#)

[Save Search](#)

[Search Summary ▾](#)

[How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

Job Title

sp|P69905|

RID

[8JWPMV4N114](#) Search expires on 05-23 07:16 am [Download All ▾](#)

Program

Blast 2 sequences [Citation ▾](#)

Query ID

[P69905.2](#) (amino acid)

Query Descr

RecName: Full=Hemoglobin subunit alpha; AltName: Full= ...

Query Length

142

Subject ID

[P68871.2](#) (amino acid)

Subject Descr

RecName: Full=Hemoglobin subunit beta; AltName: Full=E ...

Subject Length

147

Other reports

[Multiple alignment](#)

[MSA viewer](#) [?](#)

Filter Results

Percent Identity

to

E value

to

Query Coverage

to

[Filter](#)

[Reset](#)

Descriptions

Graphic Summary

Alignments

Dot Plot

hover to see the title click to show alignments

Alignment Scores ☒ < 40 ☐ 40 - 50 ☐ 50 - 80 ☐ 80 - 200 ☐ >= 200 [?](#)

1 sequences selected [?](#)

Distribution of the top 1 Blast Hits on 1 subject sequences



[← Edit Search](#)

[Save Search](#)

[Search Summary ▾](#)

[? How to read this report?](#)

[▶ BLAST Help Videos](#)

[↶ Back to Traditional Results Page](#)

Job Title **sp|P69905|**  
RID [8JWPMV4N114](#) Search expires on 05-23 07:16 am [Download All ▾](#)  
Program Blast 2 sequences [Citation ▾](#)  
Query ID [P69905.2](#) (amino acid)  
Query Descr RecName: Full=Hemoglobin subunit alpha; AltName: Full= ...  
Query Length 142  
Subject ID [P68871.2](#) (amino acid)  
Subject Descr RecName: Full=Hemoglobin subunit beta; AltName: Full=E ...  
Subject 147  
Length

Other reports [Multiple alignment](#) [MSA viewer](#) [?](#)

#### Filter Results

Percent Identity

 to 

E value

 to 

Query Coverage

 to 

[Filter](#)

[Reset](#)

[Descriptions](#)

[Graphic Summary](#)

[Alignments](#)

[Dot Plot](#)

Alignment view [Pairwise](#) ▾

[? Restore defaults](#)

[Download ▾](#)

1 sequences selected [?](#)

[Download ▾](#) [GenPept](#) [Graphics](#)

[▾ Next](#) [▲ Previous](#) [◀ Descriptions](#)

#### hemoglobin subunit beta [Homo sapiens]

Sequence ID: [NP\\_000509.1](#) Length: 147 Number of Matches: 1

[See 30 more title\(s\) ▾](#) [See all Identical Proteins\(IPG\)](#)

Range 1: 4 to 146 [GenPept](#) [Graphics](#)

[▾ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
114 bits(286)	2e-38	Compositional matrix adjust.	63/145(43%)	88/145(60%)	8/145(5%)
Query 3	LSPADKTNVKAANGKVG	GAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSAQV			56
	L+P +K+ V A WGV	+ E G EAL R+ + +P T+ +F F DLS			G+ +V
Sbjct 4	LTPEEKSAVTALWGKV--	NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKV			61
Query 57	KGHGKKVADALTNVAHV	DMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPA			116
	K HGKKV A ++ +AH+D++	+ LS+LH KL VDP NF+LL + L+ LA H			
Sbjct 62	KAHGKKVLGAFSDGLAHL	DNLKGTFAITLSEHCCKLHVDPENFRLLGNVLVCVLAHHFGK			121
Query 117	EFTPAVHASLDKFLASV	STVLTSKY 141			
	EFTP V A+ K +A V+ L KY				
Sbjct 122	EFTPPVQAAYQKVVAGV	ANALAHKY 146			

#### Related Information

[Gene](#) - associated gene details

[Genome Data Viewer](#) - aligned

genomic context

[Identical Proteins](#) - identical proteins  
to NP\_000509.1

- c. Utilice una herramienta de comparación del site EBI. Varie la matriz de puntuación (por ejemplo, pruebe diferentes matrices PAM (2) y BLOSUM (2)) y registrar los efectos en la puntuación, el número de brechas, la identidad porcentual y la longitud de la región alineada.

<https://www.ebi.ac.uk/Tools/emboss/>

## CAPTURA DE LA HERRAMIENTA EMOSS NEEDLE

# EMBOSS Needle

[Input form](#)[Web services](#)[Help & Documentation](#)[Bioinformatics Tools FAQ](#)[Feedback](#)

Tools > Pairwise Sequence Alignment > EMBOSS Needle

## Pairwise Sequence Alignment

EMBOSS Needle reads two input sequences and writes their optimal global sequence alignment to file.

STEP 1 - Enter your protein sequences

Enter a pair of

PROTEIN

sequences. Enter or paste your first **protein** sequence in any supported format:

MVLSPADKTN VKAAWGKVGA HAGEYGAEL ERMFLSFPTT KTYFPHFDLS HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK  
LLSHCLLVTL AAHLPAEFTP AVHASLDFKL ASVSTVLTSK YR

Or, upload a file: [Seleccionar archivo](#) Sin archivos seleccionados

Use a [example sequence](#) | [Clear sequence](#) | [See more example inputs](#)

AND

Enter or paste your second **protein** sequence in any supported format:

MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS TPDVVMGNPK VKAHGKKVLG AFSDGLAHLN NLKGTFTALS ELHCDKLHVD  
PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVAVGVAN ALAHKYH

Or, upload a file: [Seleccionar archivo](#) Sin archivos seleccionados

STEP 2 - Set your pairwise alignment options

OUTPUT FORMAT

pair



## CAPTURA DE LOS RESULTADOS USANDO LA MATRIZ PAM 10

```
#####
# Program: needle
# Runday: Sun 22 May 2022 00:23:31
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20220522-003606-0740-44446359-plm.asequence
#   -bsequence emboss_needle-I20220522-003606-0740-44446359-plm.bsequence
#   -datafile EPAM10
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM10
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 203
# Identity:      61/203 (30.0%)
# Similarity:    61/203 (30.0%)
# Gaps:          117/203 (57.6%)
# Score: 136.0
#
#
#=====

EMBOSS_001      1 MV-LSPADKTNVKAANGKV-----GAHAGEYGAEALERM-----F      34
                  |||.|.|.|.|.|||      |      |||.|.      |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKVNVDVGG-----EALGRLLVYPWTQRF      42

EMBOSS_001      35 LSFPTTKTYFPHF----DLSHGSAQ-----VKGHGKKV--A--DA      66
                  |      ||      |||.|||      |      |.
EMBOSS_001      43 -----FESFGDLS-----TPDAVMGNPKVKAHGKKVLGAFSDG      75

EMBOSS_001      67 LTNAVAHVDDMPN----ALSALSDLHAHKLRVDPVNFKLLSH---CLLV      108
                  |      |||.      |      |.|||.|.|||.|||.|||.      |.
EMBOSS_001      76 L----AHLN---NLKGTFA--TLSELHCDKLHVDPENFRLLGNVLCVLA-      115

EMBOSS_001      109 TLAHLPA----EFTPAVHASLDKFLASVSTVLTISKYR-----      142
                  ||      |||.|.      |      |.
EMBOSS_001      116 ---AH---HFGKEFTPPVQA-----A-----YQKVVAGVANALAH      144

EMBOSS_001      143 ---      142
EMBOSS_001      145 KYH      147

#-----
#-----
```

## CAPTURA DE LOS RESULTADOS USANDO LA MATRIZ PAM 20

```
#####
# Program: needle
# Rundate: Sun 22 May 2022 00:37:45
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20220522-004724-0258-80798525-plm.asequence
#   -bsequence emboss_needle-I20220522-004724-0258-80798525-plm.bsequence
#   -datafile EPAM20
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align format: pair
# Report file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM20
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 173
# Identity:      66/173 (38.2%)
# Similarity:    69/173 (39.9%)
# Gaps:          57/173 (32.9%)
# Score: 187.0
#
#
#=====

EMBOSS_001      1 MV-LSPADKTNVKAAGKVGAGHAGEYGAEAL-----ERMFLSFPT      39
                  |||.|.:.|..|.|.||||| ..|.|.||||| ..|.|.|.
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFG-      47

EMBOSS_001     40 TKTYFPHFDLSH-----GSAQVKGHGKKV---ADALTNAVAHVDDMPNA      80
                  |||. ..|.|.||||| ..|. |||.|:
EMBOSS_001     48 -----DLSTPDVAMGNPKVKAHGKKVLGAFSDGL----AHLDN----      81

EMBOSS_001     81 L----SALSDLHAHKLRVDPVNFKLLSH---CLLVTLAAHLPA----EFT      119
                  | ..||:|..|.|||.|||.||.. |. | ||
EMBOSS_001     82 LKGTfATLSELHCDKLHVDPENFRLLGNVLCVL----AH---HFGKEFT      124

EMBOSS_001    120 PAVHASLDKFLASVSTVLTSKYR      142
                  |.|.|...|.|.|.|.|||.
EMBOSS_001    125 PPVQAAYQKVVAGVANALAHKYH      147

#-----
#-----
```

## CAPTURA DE LOS RESULTADOS USANDO LA MATRIZ BLOSUM30

```
#####
# Program: needle
# Rundate: Sun 22 May 2022 00:45:10
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20220522-005808-0204-77193821-plm.asequence
#   -bsequence emboss_needle-I20220522-005808-0204-77193821-plm.bsequence
#   -datafile EBL0SUM30
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMB0SS_001
# 2: EMB0SS_001
# Matrix: EBL0SUM30
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    102/149 (68.5%)
# Gaps:          9/149 ( 6.0%)
# Score: 401.5
#
#=====

EMB0SS_001      1 MV-LSPADKTNVKAAWGKVGGAHAGEYGAELERMFLSFPTTKTYFPHF-D      48
                  || |:|.:|:|.|.||| |. :|:|.|||.||:|.:.|.:|. |
EMB0SS_001      1 MVHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMB0SS_001     49 LS-----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLR      93
                  || :|..| |.|||||.||:|.:.|:|:|:|:|:|:|:|:|..| |.
EMB0SS_001     49 LSTPDVAMGNPKVKAHGKKVLGAFSDGLAHLNLTGTFATLSELHCDKLH      98

EMB0SS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR      142
                  |||.||:| |..|:|.|||.||:|.:.|:|:|:|:|:|:|:|:|..| |.
EMB0SS_001     99 VDPENFRLLGNVLVLCVLAHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147

#-----
#-----
```

## CAPTURA DE LOS RESULTADOS USANDO LA MATRIZ BLOSUM35

```
#####
# Program: needle
# Rundate: Sun 22 May 2022 00:55:12
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20220522-010204-0578-10321633-p2m.asequence
#   -bsequence emboss_needle-I20220522-010204-0578-10321633-p2m.bsequence
#   -datafile EBL0SUM35
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMB0SS_001
# 2: EMB0SS_001
# Matrix: EBL0SUM35
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    97/149 (65.1%)
# Gaps:          9/149 ( 6.0%)
# Score: 398.5
#
#
#=====

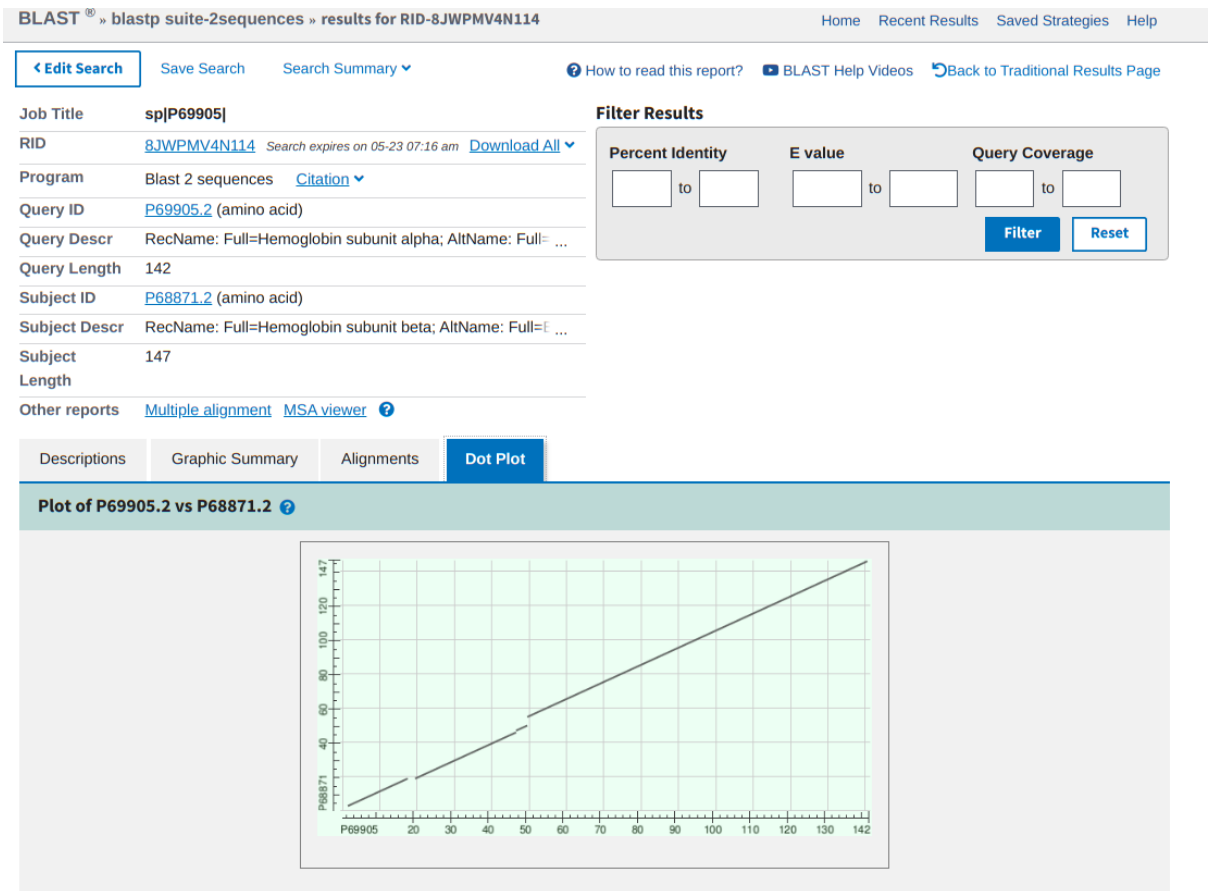
EMB0SS_001      1 MV-LSPADKTNVKAAGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                || |:|.:.|.|.|.||||:.. |.|.||||.||||:|.|.:.|.|. |
EMB0SS_001      1 MVHLTPEEKSAVTALWGKVNVDD--EVGGEALGRLLVVYPWTQRFFESFGD      48

EMB0SS_001      49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDDLHAKLR      93
                || :|...||.|||||.||:|.:.|:|.:.|.:.|.:.|.:.|.|.
EMB0SS_001      49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFFATLSELHCDKLH      98

EMB0SS_001      94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR      142
                |||.||:|:|.:.|.:.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMB0SS_001      99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAVQKVAGVANALAHKYH      147

#-----
#-----
```

- d. Para o programa BLASTP NCBI observe que la salida de un alineamiento de pares incluye una visualización de la matriz de puntos.



- e. Realizar el alineamiento de las proteínas usando el programa blast instalado localmente

- Primero instale el blast+ en mi entorno de mi ordenador
- Se crea una carpeta de **Programas**
- Segundo descargar los archivos fasta de proteínas de humanos hba e hbb y colocarlos en la carpeta **Programas**

Los archivos son los siguientes:

```
>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2
MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAMGNPK
VKAHGKKVLGAFSDGLAHLNLTGTFATLSSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH|
```

hba\_human.fa

```
>sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens OX=9606 GN=HBA PE=1 SV=2
MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAMGNPK
VKAHGKKVLGAFSDGLAHLNLTGTFATLSSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH|
```

hbb\_human.fa

- Tercero ubicamos los archivos y ejecutamos el siguiente comando

```
blastp -query hba_human.fa -subject hbb_human.fa
-out result.txt
```

- Los resultados se muestran en la terminal pero tambien se puede guardar en el archivo result.txt

```
judal@judal-OMEN-by-HP-Laptop-15-dc0xxx:~/Documentos/programas$ blastp -query hba_human.fa -subject hbb_human.fa
BLASTP 2.9.0+

Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A.
Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of
protein database search programs", Nucleic Acids Res. 25:3389-3402.

Reference for composition-based statistics: Alejandro A. Schaffer,
L. Aravind, Thomas L. Madden, Sergei Shavirin, John L. Spouge, Yuri
I. Wolf, Eugene V. Koonin, and Stephen F. Altschul (2001),
"Improving the accuracy of PSI-BLAST protein database searches with
composition-based statistics and other refinements", Nucleic Acids
Res. 29:2994-3005.

Database: User specified sequence set (Input: hbb_human.fa).
1 sequences; 142 total letters

Query= sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606
GN=HBB PE=1 SV=2
Length=147

Sequences producing significant alignments:

                                Score      E
                                (Bits)    Value
sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens OX=9... 114      2e-38

> sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens
OX=9606 GN=HBA1 PE=1 SV=2
Length=142

Score = 114 bits (286), Expect = 2e-38, Method: Compositional matrix adjust.
Identities = 63/145 (43%), Positives = 88/145 (61%), Gaps = 8/145 (6%)

Query    4    LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVVMGNPKV   61
          L+P +K+ V A WGKV + E G EAL R+ + +P T+ +F F DLS G+ +V
Sbjct    3    LSPADKTNVKAAGWKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS----HGSAQV   56

Query    62    KAHGKKVLGAFSDGLAHLNLTGKTFATLSSEHLCDKLHVDPENFRLLGNVLVCVLAHHFGK 121
          K HGKKV A ++ +AH+D++ + LS+LH KL VDP NF+LL + L+ LA H
Sbjct    57    KGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLKRVDPVNFKLLSHCLLVTLAAHLPA 116

Query    122   EFTPPVQAAYQKVVAGVANALAHKY 146
          EFTP V A+ K +A V+ L KY
Sbjct    117   EFTPAVHASLDKFLASVSTVLTSKY 141

Lambda      K      H      a      alpha
0.320      0.137    0.422    0.792    4.96

Gapped
Lambda      K      H      a      alpha      sigma
0.267      0.0410   0.140    1.90     42.6     43.6

Effective search space used: 16506

Database: User specified sequence set (Input: hbb_human.fa).
Posted date: Unknown
Number of letters in database: 142
Number of sequences in database: 1

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Neighboring words threshold: 11
Window for multiple hits: 40
```