

Algoritmo para el cálculo de fragmentos de proteínas en los organismos secuenciados

Felipe Nicolás Araya Barrera

Objetivos

Obtener de una base de datos de proteínas lo siguiente:

- a) La cantidad de diferentes fragmentos de proteínas, de largo k entre 1 hasta 50
- b) Identificar cuáles son los fragmentos que más se repiten

4 bases de datos usadas:

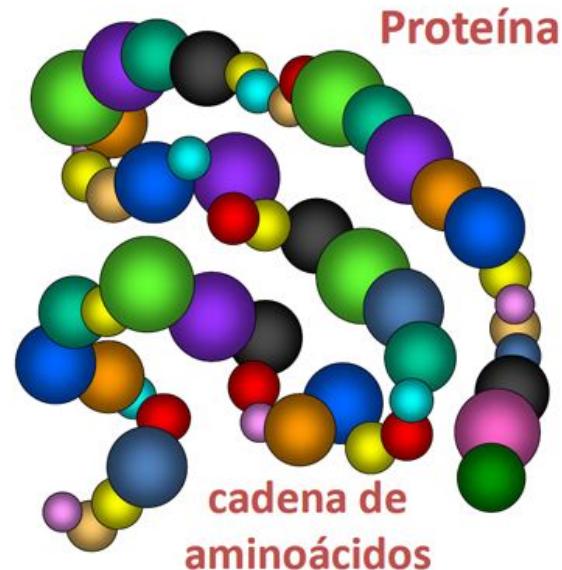
- UniProt-SwissProt
- UniProt-TrEMBL
- EROP-Moscow
- Homosapiens (derivada de UniProt-SwissProt)

Proteínas

Pertenecientes al grupo de biomoléculas orgánicas

Poseen la mayor diversidad de funciones que realizan en los seres vivos; prácticamente todos los procesos biológicos dependen de su presencia y/o actividad

Su unidad base es el aminoácido



Aminoácido

Molécula compuesta por un:

- a) Grupo amino (-NH₂)
- b) Grupo carboxilo (-COOH)
- c) Grupo Radical (R)

Grupo radical determina la identidad de aminoácido



Los amínoácidos conocidos y abreviaturas

Alanina - A	Cisteína - C	Ácido aspártico - D	Ácido glutámico - E
Fenilalalina - F	Glicina - G	Histidina - H	Isoleucina - I
Lisina - K	Leucina - L	Metionina - M	Aspargarina - N
Prolina - P	Glutamina - Q	Arginina - R	Serina - S
Treonina - T	Valina - V	Triptófano - W	Tirosina - Y

Unión de aminoácidos - macromoléculas

Tamaño	Tipo de estructura
Entre 2 y 14 aminoácidos	Oligopéptido
Entre 15 y 50 aminoácidos	Polipéptido
Mayor de 50 aminoácidos	Proteina

Formato FASTA

Texto utilizado para representar secuencias de péptidos, y en el que los aminoácidos se representan usando códigos de una única letra

```
>sp|Q9W7K1|3S32_PSETE Short neurotoxin 2 OS=Pseudonaja textilis PE=1 SV=1  
MKTLLLTLVMVTIMCLDLGYTLTCYKGYHDTVVKPHETICRYLIPATHGNAIPARGCG  
TSCPGGNHPVCCSTDLCNK
```

Formato FASTA - múltiples péptidos

```
>sp|Q9BV40|VAMP8_HUMAN Vesicle-associated membrane protein 8 OS=Homo sapiens GN=VAMP8 PE=1 SV=1  
MEEASEGGGNDRVRNLQSEVEGVKNIMTQNVERILARGENLEHLRNKTEDLEATSEHFKT  
TSQKVARKFWWKNVKMIVLICVIVFIIILFIVLFATGAFS  
>sp|Q9NZ42|PEN2_HUMAN Gamma-secretase subunit PEN-2 OS=Homo sapiens GN=PSENEN PE=1 SV=1  
MNLERVSNEEKLNLCRKYYLGGFAFLPFLWLVNIFWFFREAFLVPAYTEQSQIKGYVWRS  
AVGFLFWVIVLTSWITIFQIYRPRWGALGDYLSFTIPLGTP  
>sp|P31151|S10A7_HUMAN Protein S100-A7 OS=Homo sapiens GN=S100A7 PE=1 SV=4  
MSNTQAERSIIGMIDMFHKYTRRDDKIEKPSLLTMMKENFPNFLSACDKKGTNYLADVFE  
KKDKNEDKKIDFSEFLSLLGDIATDYHKQSHGAAPCSGGSQ  
>sp|Q15836|VAMP3_HUMAN Vesicle-associated membrane protein 3 OS=Homo sapiens GN=VAMP3 PE=1 SV=3  
MSTGPTAACGSNRLQQTQNQVDEVVDIMRVNVDKVLERDQKLSELDDRADALQAGASQF  
ETSAAKLKRKYWWKNCKMWAIGITVLVIFIIIIIVVVSS  
>sp|070404|VAMP8_MOUSE Vesicle-associated membrane protein 8 OS=Mus musculus GN=Vamp8 PE=1 SV=1  
MEEASGSAGNDRVRNLQSEVEGVKNIMTQNVERILSRGENLDHLRNKTEDLEATSEHFKT  
TSQKVARKFWWKNVKMIVIICVIVLIVILIILFATGTIP
```

Base de datos de proteínas UniProt

a) UniProt-SwissProt

b) UniProt-TrEMBL

c) Homosapiens (extraída de
UniProt-SwissProt)

The screenshot shows the UniProt homepage with the following sections:

- UniProtKB**: UniProt Knowledgebase. Sub-sections: Swiss-Prot (555,594) - Manually annotated and reviewed; TrEMBL (90,050,711) - Automatically annotated and not reviewed.
- UniRef**: Sequence clusters.
- UniParc**: Sequence archive.
- Proteomes**: Subcellular locations, Keywords.
- Supporting data**: Literature citations, Cross-ref. databases, Taxonomy, Diseases.

Getting started

Text search

Our basic text search allows you to search all the resources available



UniProt data

Download latest release

Get the UniProt data

Base de datos de proteínas EROP-Moscow

[Home Page](#) [Site Map](#) [Contact us](#) [Help](#)

EROP - Moscow
C
Knowledgebase
H

EROP-Moscow

Endogenous Regulatory OligoPeptide knowledgebase
EROP-Moscow: release of 19-Nov-2016, 14511 entries.

Project was performed by
Computer biochemistry group in [A. N. Bach Institute of Biochemistry, Russian Academy of Sciences](#), 33 Leninsky prosp., Moscow, 119071 Russian Federation
and supported by
The Russian Foundation for Basic Research (RFBR), Grant 02-07-90175 and the Program "Molecular and Cellular Biology" of the Russian Academy of Sciences,
Grant 10P.
EROP-Moscow is a curated oligopeptide (2-50 amino acid residues) sequence database which strives to provide a high level of annotations
(such as descriptions of the structure of an oligopeptide, its source and function, post-translational modifications, etc.).
First reference: Zamyatnin A.A. EROP-Moscow: Specialized data bank for endogenous regulatory oligopeptides, Protein Sequence and Data
Analysis, 4(1), 49-52, 1991.
Last reference: Zamyatnin A.A., Borchikov A.S., Vladimirov M.G., Voronina O.L. The EROP-Moscow oligopeptide database. Nucleic Acids
Research, 34(Database issue), D261-D266, 2006.

EROP-news

19-Nov-2016. 596 new entries.
01-Sep-2016. 855 new entries.
08-Aug-2016. 69 new entries.
25-Jul-2016. 2152 new entries.
07-Jul-2015. 351 new entries.
17-Feb-2014. 346 new entries.
01-Oct-2012. EROP publicationr: Zamyatnin A.A. Theoretical analysis of the structure

[Query EROP](#) [EROP Statistics](#)

SpToGull

EROP-Moscow author: Alexander A.Zamyatnin,
EROP-Moscow design and programs: Alexander S.Borchikov, Mikhail G.Vladimirov
<Legal disclaimer and copyright notice. Version 1.1, Aug 2003. Last updated on 19-Nov-2016>

[UniProt](#) [PIR](#) [PDB](#) [PubMed](#)

Bases de datos usadas (datasets)

Bases de datos	Cantidad de Proteínas	Tamaño (MB)
SwissProt	555.426	268,3
TrEMBL	89.396.316	40.900
EROP-Moscow	14.785	1,5
Proteínas humanas	86.298	47,2

Combinaciones posibles de aminoácidos

$$P = A^N$$

P = Combinaciones posibles

A = Aminoácidos totales (20)

N = Largo péptido

Tamaño péptido (N)	Combinaciones posibles (20^N)
2 aa.	400
3 aa.	8.000
4 aa.	160.000
5 aa.	3.200.000
10 aa.	$1,024 \times 10^{13}$
20 aa.	$1,049 \times 10^{26}$
50 aa.	$1,126 \times 10^{65}$

Algunas técnicas usadas

Algoritmo de fuerza bruta (búsqueda exhaustiva):

Texto: a n a l i s i s d e a l g o r i t m o s

x

Patrón: a | l | g | o

✓ ✓ x

a | l | g | o

• • •

✓ ✓ ✓ ✓

a | l | g | o

Algunas técnicas usadas (2)

Algoritmos de búsqueda de strings:

a) Algoritmo de Knuth-Morris-Pratt

T:	a	a	a	a	b	a	a	b	a	a	a	b
P1:	a	a	b									
P2:	a	a	b									
P3:		a	a	b	a	a	a					
P4:								a	a	a	b	

b) Algoritmo de Boyer-Moore

T:	a	a	a	a	b	a	a	b	a	a	a	b
P1:	a	a	b	a	a	a						
P2:		a	a	b	a	a	a					
P3:								a	a	a	a	a

Búsqueda de datos indexados (Indexed Text Searching)

Idea clave: Indexar palabra o cadena de strings

Palabra ejemplo: MISSISSIPPI\$

0	1	2	3	4	5	6	7	8	9	10	11
M	I	S	S	I	S	S	I	P	P	I	\$

Concepto clave: Sufijo

Sufijos

Sector de la palabra que comienza desde un punto medio (que puede ser incluso el comienzo de la palabra) hasta el final

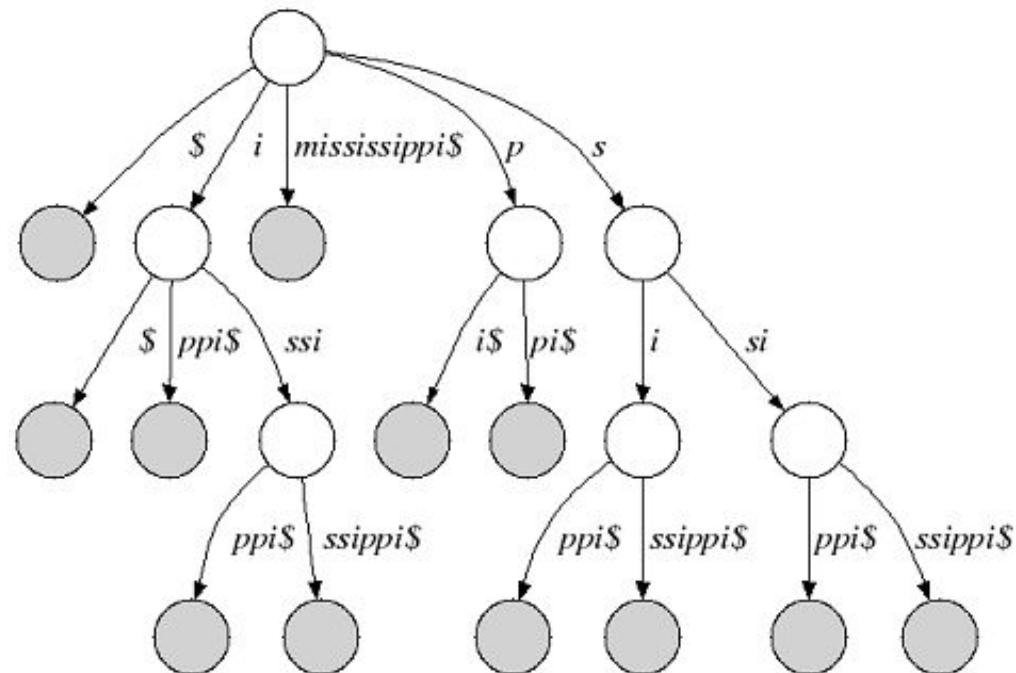
0	MISSISSIPPI\$
1	ISSISSIPPI\$
2	SSISSIPPI\$
3	SISSIPPI\$
4	ISSIPPI\$
5	SSIPPI\$
6	SIPPI\$
7	IPPI\$
8	PPI\$
9	PI\$
10	I\$
11	\$

Árbol de sufijos

Almacenar una cadena de caracteres con “información pre-procesada” sobre su estructura interna.

Ideal para problemas como hallar substrings en común

Requerimiento implementación: $20n$ bytes



Arreglo de sufijos

Similar al árbol de sufijos

Considera una cadena de texto T

$T[i,j]$ = subcadena que va desde i hasta j

Arreglo de sufijos SA[i] cumple que:

$$1 < i \leq n : T[SA[i - 1], n] < T[SA[i], n]$$

0	MISSISSIPPI\$	11	\$
1	ISSISSIPPI\$	10	I\$
2	SSISSIPPI\$	7	IPPI\$
3	SISSIPPI\$	4	ISSIPPI\$
4	ISSIPPI\$	1	ISSISSIPPI\$
5	SSIPPI\$	0	MISSISSIPPI\$
6	SIPPI\$	9	PI\$
7	IPPI\$	8	PPI\$
8	PPI\$	6	SIPPI\$
9	PI\$	3	SISSIPPI\$
10	I\$	5	SSIPPI\$
11	\$	2	SSISSIPPI\$

Arreglo de sufijos (2)

- Tamaño del arreglo: $4n$ bytes
- Cantidad de espacio utilizada es mucho menor que con el árbol de sufijos

i	0	1	2	3	4	5	6	7	8	9	10	11
$SA[i]$	11	10	7	4	1	0	9	8	6	3	5	2

Arreglo LCP

LCP: Prefijo común más largo

Estructura auxiliar al arreglo de sufijos, pero que en sus posiciones guarda las longitudes de los prefijos comunes más largos entre todos los pares de sufijos consecutivos correspondientes al arreglo de sufijos.

i	0	1	2	3	4	5	6	7	8	9	10	11
T[i]	m	i	s	s	i	s	s	i	p	p	i	\$
SA[i]	11	10	7	4	1	0	9	8	6	3	5	2
0	\$	i	i	i	i	m	p	p	s	s	s	s
1		\$	p	s	s	i	i	p	i	i	s	s
2			p	s	s	s	\$	i	p	s	i	i
3			i	i	i	s		\$	p	s	p	s
4			\$	p	s	i			i	i	p	s
5			p	s	s				\$	p	i	i
6			i	i	s					p	\$	p
7				\$	p	i				i		p
8					p	p				\$		i
9					i	p						\$
10						\$	i					
11							\$					

Arreglo LCP

- Se tiene arreglo de sufijos SA y texto T
- $T[i, j]$: substring de T que va desde la posición i hasta j
- $lcp(v, w)$: largo del prefijo común más largo entre 2 strings v y w

Se cumple que $LCP[0]$ está indefinido y que:

$$LCP[i] = lcp(T[SA[i - 1], n], T[SA[i], n]) \text{ para todo } 1 < i \leq n$$

i	0	1	2	3	4	5	6	7	8	9	10	11
T[i]	m	i	s	s	i	s	s	i	p	p	i	\$
SA[i]	11	10	7	4	1	0	9	8	6	3	5	2
LCP[i]	Ø	0	1	1	4	0	0	1	0	2	1	3

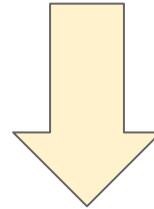
Implementación

Se siguen 3 etapas:

1. Extracción de cadena de proteínas
2. Obtención del arreglo de sufijos y arreglo LCP
3. Implementación principal

Extracción de proteínas

```
>SEQUENCE_1
MTEITAAMVKELRESTGAGMMDCKNALSETNGDFDKAVQLLREKGLGKAACKADRLAAEG
LVSVKVSDDFTIAAMRPSYLSYEDLDMTFVENEYKALVAELEKENEEERRRLKDPNKPEHK
IPQFASRKQLSDAILKEAEEKIKEELKAQGKPEKIWDNIIPGKMNSFIADNSQLDSKLTL
MGQFYVMDDKKTVEQVIAEKEKEFGGKIKIVEFICFEVGEGLEKKTEDFAAEVAAQL
>SEQUENCE_2
SATVSEINSETDFVAKNDQFIALTKDTTAHIQSNSLQSVEELHSSTINGVKFEEYLKSQI
ATIGENLVVRRFATLKAGANGVVNGYIHTNGRVGVVIAAACDSAEVASKSRDLLRQICMH
```



...EKKTEDFAAEVAAQL\$SATVSEINSETDFVA...

Tamaño cadena base de datos

Bases de datos	Cantidad de Proteínas	Tamaño (MB)	Tamaño Cadena (MB)
SwissProt	555.426	268,3	199,5
TrEMBL	89.396.316	40.900	30.200
EROP-Moscow	14.785	1,5	0,3526
Proteínas humanas	86.298	47,2	38,2

Obtención de arreglos - 2 métodos

1. Guardar texto, SA y LCP en memoria RAM (Algoritmo Memoria Interna).

Total guardable en RAM: $9n$ bytes

Aplicable a bases de datos:

- SwissProt
- EROP-Moscow
- Proteínas Humanas

No aplicable a base de datos:

- TrEMBL (30 GB de capacidad)

2. Guardar texto en memoria RAM, pero SA y LCP se guardan en disco duro (Algoritmo Memoria Externa).

Total guardable en RAM: n bytes

Aplicable a bases de datos:

- SwissProt
- TrEMBL
- EROP-Moscow
- Proteínas Humanas

Implementación principal

Lógica: Aumentar la cantidad de diferentes substrings desde 0 considerando determinados largos de LCPs consecutivos

Restricciones:

- a) Si el largo del sufijo es mayor o igual a K, entonces el arreglo LCP puede ser analizado, en caso contrario se omite y se continua al siguiente valor del arreglo LCP.
- b) Si el prefijo del sufijo revisado solamente esté compuesto por los 20 aminoácidos conocidos.

Palabra ejemplo: BANANA\$

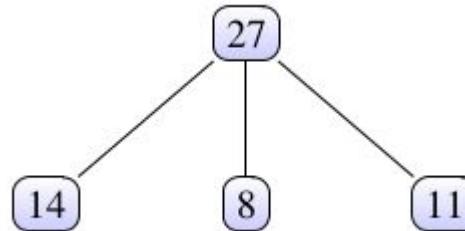
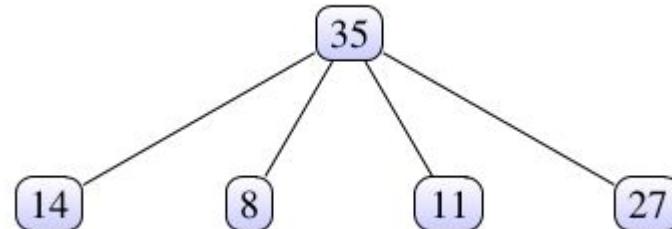
<i>SA[]</i>	<i>LCP[]</i>	<i>sufijo</i>
6	0	\$
5	1	A\$
3	3	ANA\$
1	0	ANANA\$
0	0	BANANA\$
4	2	NA\$
2	0	NANA\$

Variables a considerar - implementación

- a) activador: Esta variable es usada para verificar si se están sumando repeticiones de un determinado fragmento o no. Toma valores entre 0 y 1, donde 0 significa que se están buscando substrings y 1 significa que se están agregando repeticiones de determinado substring encontrado.
- b) contador: Variable usada para guardar la cantidad de repeticiones de un determinado fragmento.
- c) ds: Variable que guarda la cantidad de diferentes substrings encontrados de largo K.

Cola de prioridad (*priority queue*)

Es un tipo de estructura contenedora similar a una lista, vector o arreglo, con su característica principal que al único elemento que se puede acceder es aquel que sí o solo si tenga la prioridad o valor más alto que los demás elementos



Diferentes fragmentos de largo 3 (parte 1)

<i>SA[]</i>	<i>LCP[]</i>	<i>sufijo</i>
6	0	\$
5	1	A\$
3	3	ANA\$
1	0	ANANA\$
0	0	BANANA\$
4	2	NA\$
2	0	NANA\$

Inicializar con DS, activador y contador = 0.

Variable prohibida: \$

$LCP[0] = 0 < 3$ y activador = 0

=> Se mantiene DS = 0 ya que el largo del sufijo es menor a 3 ($T[SA[0]] = T[6] = \$$).

Diferentes fragmentos de largo 3 (parte 2)

$SA[]$	$LCP[]$	sufijo
6	0	\$
5	1	A\$
3	3	ANA\$
1	0	ANANA\$
0	0	BANANA\$
4	2	NA\$
2	0	NANA\$

$LCP[1] = 1 < 3$ y activador = 0

=> Se mantiene DS = 0 ya que el largo del sufijo es menor a 3 ($T[SA[1]] = T[5]=A\$$).

Diferentes fragmentos de largo 3 (parte 3)

SA[] LCP[] sufijo

6 0 \$

5 1 A\$

3 3 ANA\$

1 0 ANANA\$

0 0 BANANA\$

4 2 NA\$

2 0 NANA\$

LCP[2] = 3 = 3 y activador = 0

SA[2] = 3 => T[3,5] = ANA (no posee caracteres prohibidos)

=> contador = 2;

=> activador = 1;

Diferentes fragmentos de largo 3 (parte 4)

<i>SA[]</i>	<i>LCP[]</i>	<i>sufijo</i>	
6	0	\$	$LCP[3] = 0 < 3$ y activador = 1
5	1	A\$	$\Rightarrow DS = 0 + 1 = 1$
3	3	ANA\$	
1	0	ANANA\$	$\Rightarrow (ANA, 2)$ se guarda en cola de prioridad
0	0	BANANA\$	\Rightarrow contador = 0;
4	2	NA\$	
2	0	NANA\$	\Rightarrow activador = 0;

Diferentes fragmentos de largo 3 (parte 5)

SA[] LCP[] sufijo

6 0 \$

LCP[4] = 0 < 3 y activador = 0

5 1 A\$

3 3 ANA\$

SA[4] = 0 => T[0,2] = BAN (no posee caracteres prohibidos)

1 0 ANANA\$

=> DS = 1 + 1 = 2

0 0 BANANA\$

=> (BAN, 1) se guarda en cola de prioridad

4 2 NA\$

2 0 NANA\$

Diferentes fragmentos de largo 3 (parte 6)

SA[] LCP[] sufijo

6 0 \$

LCP[5] = 2 < 3 y activador = 0

5 1 A\$

SA[5] = 4 => T[4,6] = NA\$ (posee
caracteres prohibidos)

3 3 ANA\$

1 0 ANANA\$

0 0 BANANA\$

=> Avanzar hasta la siguiente iteración

4 2 NA\$

2 0 NANA\$

Diferentes fragmentos de largo 3 (parte 7)

<i>SA[]</i>	<i>LCP[]</i>	<i>sufijo</i>	
6	0	\$	$LCP[6] = 0 < 3$ y activador = 0
5	1	A\$	
3	3	ANA\$	$SA[6] = 2 \Rightarrow T[2,4] = NAN$ (no posee caracteres prohibidos)
1	0	ANANA\$	
0	0	BANANA\$	$\Rightarrow DS = 2 + 1 = 3$
4	2	NA\$	$\Rightarrow (NAN, 1)$ se guarda en cola de prioridad
2	0	NANA\$	

Diferentes fragmentos de largo 3 (parte 8)

- Diferentes fragmentos de largo 3 obtenidos en BANANA\$ = 3
- ANA se repite 2 veces, BAN y NAN se repiten una vez

Formato salida archivos

PNVGKS 1827

Tiempo utilizado: 36.0557 segundos

Diferentes substrings para 7 es 84118859

NNNNNNN	27091
QQQQQQQ	16272
SSSSSSS	6927
EEEEEEE	5342
AAAAAAA	4966
PPPPPPP	4189
GGGGGGG	4140
SDSDSDS	4106
DSDSDSD	4059
HTGEKPY	3252
TTTTTTT	2881
IHTGEKP	2029
ATVITNL	1679
LPWGQMS	1660
HVDHGKT	1659
YVLPWGQ	1651
GYVLPWG	1651
VLPWGQM	1648
TPGHVDF	1648
DTPGHVD	1648

Tiempo utilizado: 39.4517 segundos

Diferentes substrings para 8 es 100896814

NNNNNNNN	23879
QQQQQQQQ	13508

Resultados obtenidos

Sistema utilizado

Componente	Descripción
Nombre CPU	Intel(R) Xeon(R) CPU E5-2630 @2.30GHz
CPU (s)	12
Caché (capacidad)	15 MB
Memoria RAM (capacidad)	60 GB
Sistema Operativo	Ubuntu Server 16.04 (Linux versión 4.4.0-51)

Tiempos de ejecución Algoritmo Memoria Interna

Base de Datos	Tiempo SA (s)	Tiempo LCP (s)	Tiempo programa (s)
SwissProt	2.637,18	35,09	2.758,82
EROP-Moscow	1,92	0,03	3,58
Proteínas humanas	422,49	9,06	407,58

Tiempos de ejecución Algoritmo Memoria Externa (20 GB)

Base de Datos	Tiempo SA (s)	Tiempo LCP (s)	Tiempo guardado elementos (s)	Tiempo programa (s)
SwissProt	17,98	57,23	59,39	4.533,09
TrEMBL	7.981,08	10.785,03	7.722,16	8,5 días
EROP-Moscow	0,23	0,19	0,11	7,07
Proteínas humanas	4,10	9,40	9,81	834,60

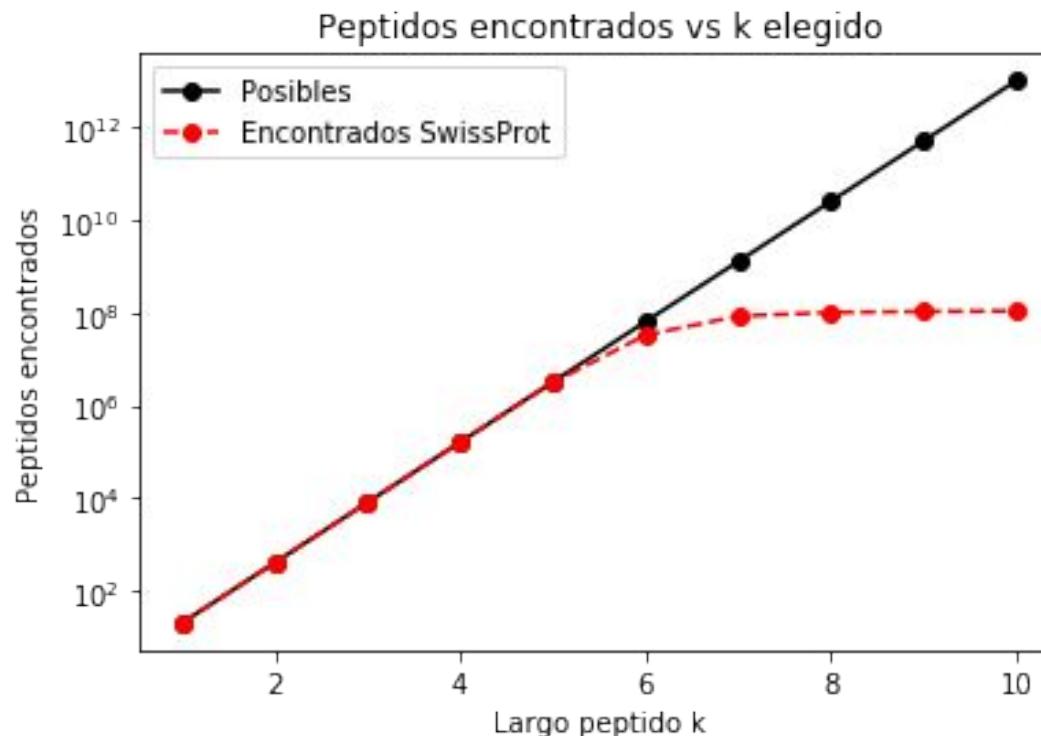
Tiempos totales

Base de Datos	Algoritmo Memoria Interna (s)	Algoritmo Memoria Externa (s)
SwissProt	5.431,09	4.667,69
TrEMBL	-	8,8 días
EROP-Moscow	9,11	7,6
Proteínas humanas	839,13	857,91

Diferentes fragmentos BD SwissProt (1)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00	20	100,00
2	400	400	100,00	400	100,00
3	8.000	8.000	100,00	8.000	100,00
4	160.000	159.999	99,99	159.999	99,99
5	3.200.000	3.113.509	97,29	3.113.509	97,29
6	64.000.000	32.921.109	51,43	32.921.109	51,43
7	1.280.000.000	84.118.859	6,57	84.118.859	6,57
8	25.600.000.000	100.896.814	0,39	100.896.814	0,39
9	512.000.000.000	105.834.330	0,020	105.834.330	0,020
10	$1,024 \times 10^{13}$	108.976.567	0,0010	108.976.567	0,0010

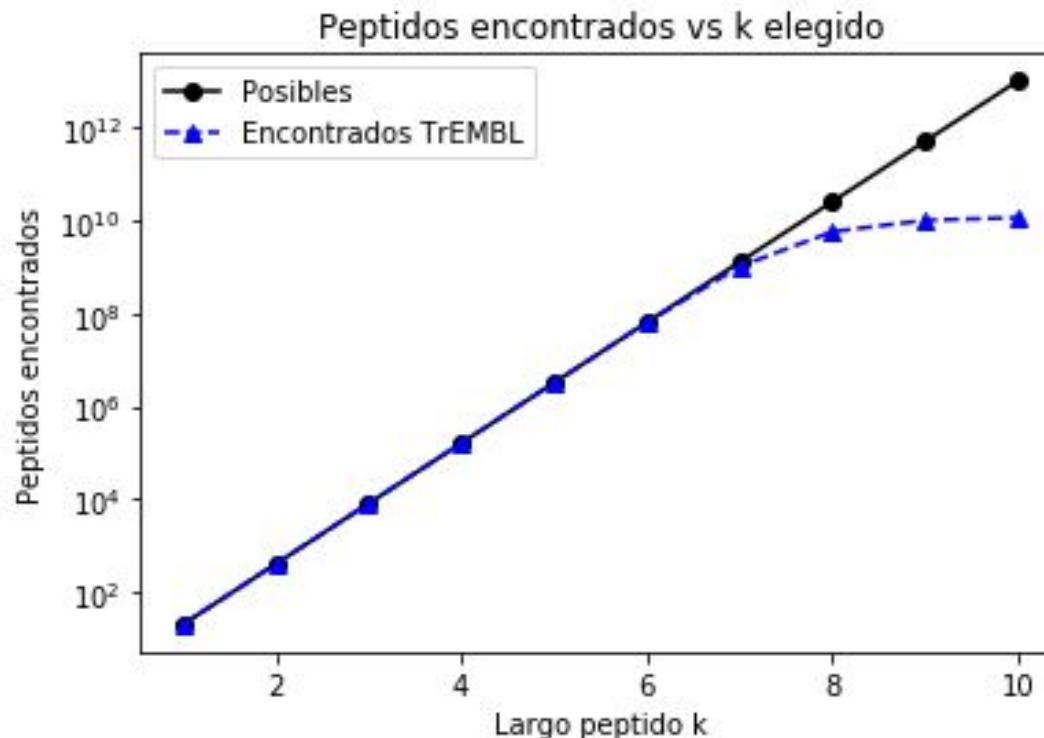
Diferentes fragmentos BD SwissProt (1)



Diferentes fragmentos BD TrEMBL (1)

Largo péptido (N)	N-péptidos posibles (20^N)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00
2	400	400	100,00
3	8.000	8.000	100,00
4	160.000	160.000	100,00
5	3.200.000	3.200.000	100,00
6	64.000.000	63.817.907	99,71
7	1.280.000.000	1.024.659.629	80,05
8	25.600.000.000	5.743.538.889	22,43
9	512.000.000.000	10.114.868.387	1,97
10	$1,024 \times 10^{13}$	11.524.607.918	0,11

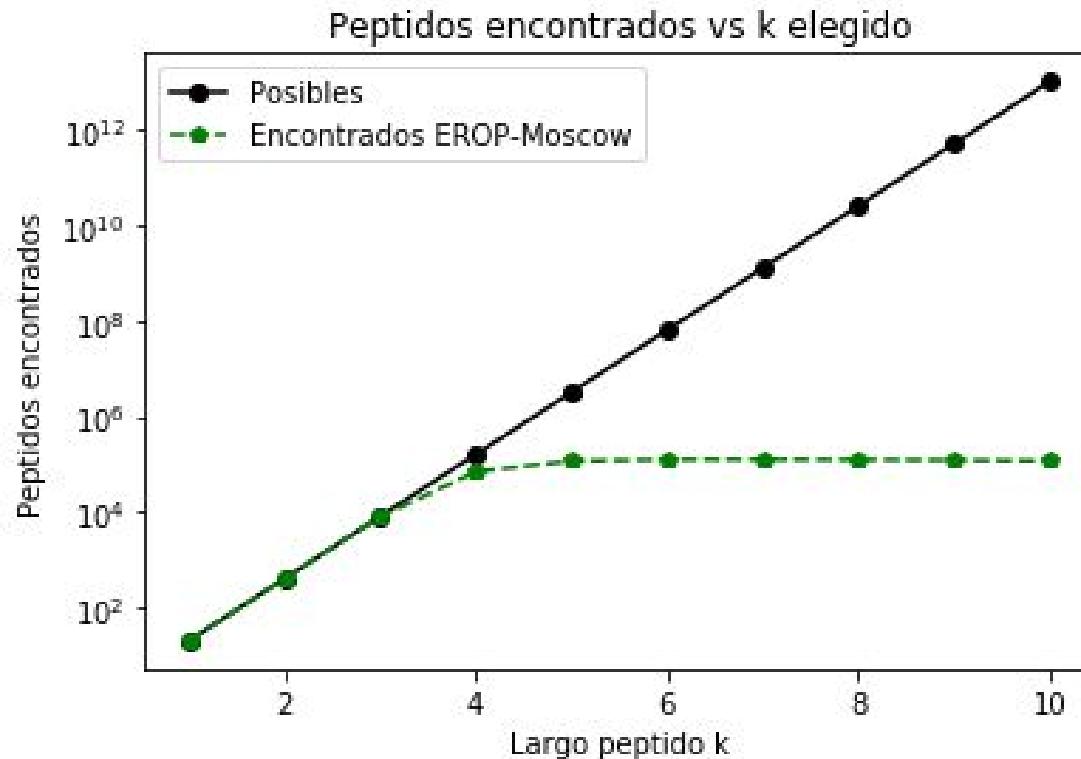
Diferentes fragmentos BD TrEMBL (2)



Diferentes fragmentos BD EROP-Moscow (1)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00	20	100,00
2	400	400	100,00	400	100,00
3	8.000	7.907	98,83	7.907	98,83
4	160.000	69.218	43,26	69.218	43,26
5	3.200.000	116.454	3,63	116.454	3,63
6	64.000.000	125.036	0,19	125.036	0,19
7	1.280.000.000	125.750	0,0098	125.750	0,0098
8	25.600.000.000	124.054	$4,84 \times 10^{-4}$	124.054	$4,84 \times 10^{-4}$
9	512.000.000.000	121.103	$2,36 \times 10^{-5}$	121.103	$2,36 \times 10^{-5}$
10	$1,024 \times 10^{13}$	117.457	$1,14 \times 10^{-6}$	117.457	$1,14 \times 10^{-6}$

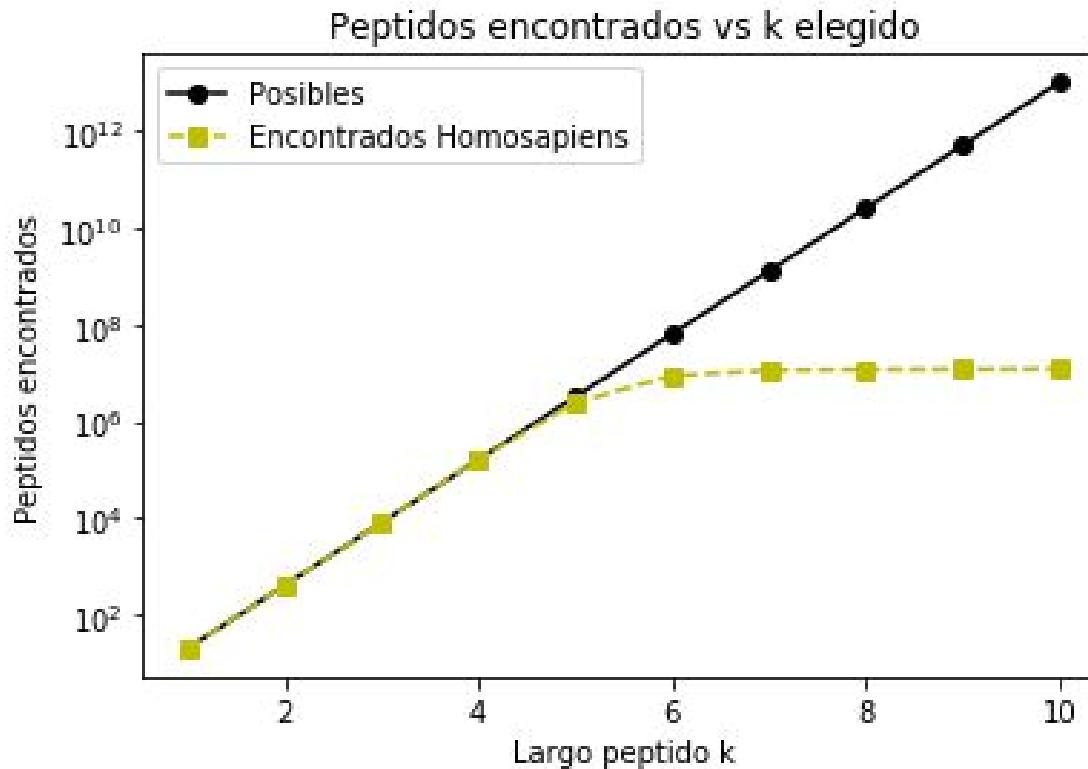
Diferentes fragmentos BD EROP-Moscow (2)



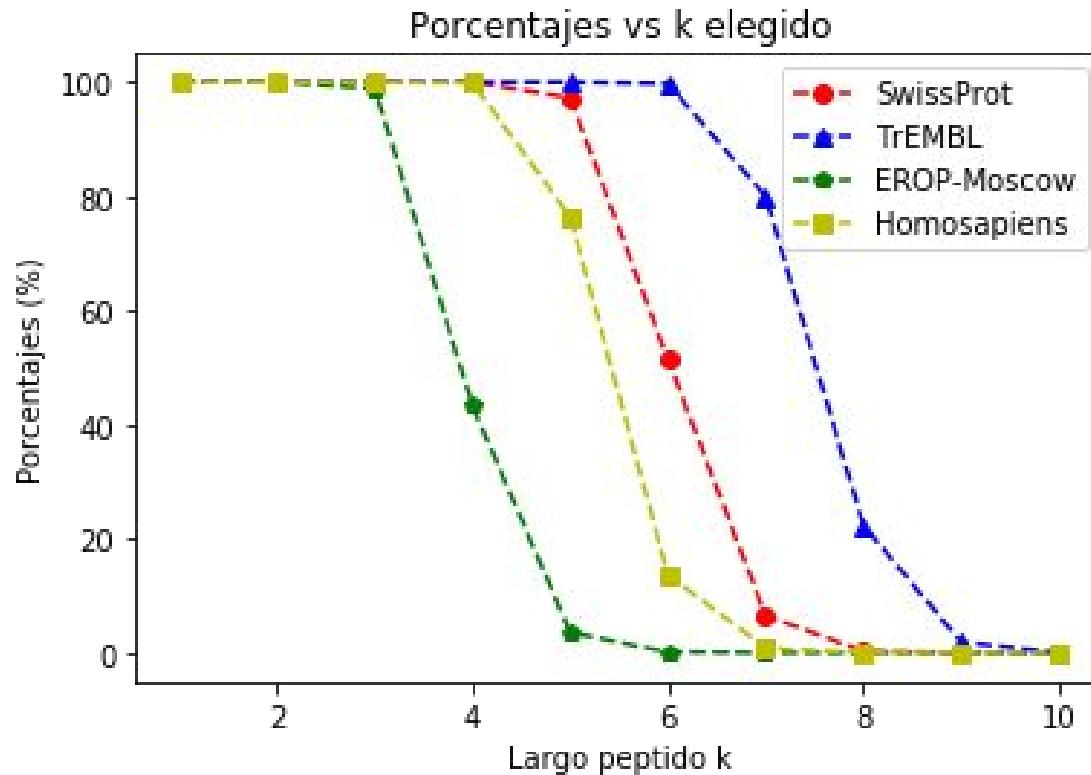
Diferentes fragmentos BD Homosapiens (1)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00	20	100,00
2	400	400	100,00	400	100,00
3	8.000	8.000	100,00	8.000	100,00
4	160.000	159.835	99,89	159.835	99,89
5	3.200.000	2.445.409	76,42	2.445.409	76,42
6	64.000.000	8.823.762	13,78	8.823.762	13,78
7	1.280.000.000	11.334.257	0,89	11.334.257	0,89
8	25.600.000.000	11.777.746	0,046	11.777.746	0,046
9	512.000.000.000	11.930.386	0,0023	11.930.386	0,0023
10	$1,024 \times 10^{13}$	12.032.248	$1,175 \times 10^{-4}$	12.032.248	$1,175 \times 10^{-4}$

Diferentes fragmentos BD Homosapiens (2)



Porcentaje de fragmentos encontrados



Fragmentos más repetidos BD SwissProt

k	Algoritmo Memoria Interna	Algoritmo Memoria Externa	k	Algoritmo Memoria Interna	Algoritmo Memoria Externa
1	L - 19.206.802	L - 19.206.802	5	NNNNN - 36.977	NNNNN - 36.977
	A - 16.435.230	A - 16.435.229		QQQQQ - 26.232	QQQQQ - 26.232
	G - 14.086.821	G - 14.086.821		SSSSS - 17.652	SSSSS - 17.652
	V - 13.664.059	V - 13.664.059		AAAAA - 17.020	AAAAA - 17.020
	E - 13.406.543	E - 13.406.543		EEEEEE - 13.572	EEEEEE - 13.572
2	LL - 1.886.886	LL - 1.886.886	6	NNNNNN - 31.269	NNNNNN - 31.269
	AA - 1.714.607	AA - 1.714.607		QQQQQQ - 20.197	QQQQQQ - 20.197
	AL - 1.683.204	AL - 1.683.204		SSSSSS - 10.379	SSSSSS - 10.379
	LA - 1.673.336	LA - 1.673.336		AAAAAA - 8.392	AAAAAA - 8.392
	LS - 1.337.091	LS - 1.337.091		EEEEEE - 8.077	EEEEEE - 8.077
3	AAA - 232.439	AAA - 232.439	7	NNNNNNNN - 27.091	NNNNNNNN - 27.091
	ALA - 185.750	ALA - 185.750		QQQQQQQQ - 16.272	QQQQQQQQ - 16.272
	LLL - 180.299	LLL - 180.299		SSSSSSSS - 6.927	SSSSSSSS - 6.927
	LAA - 179.704	LAA - 179.704		EEEEEEEE - 5.342	EEEEEEEE - 5.342
	AAL - 177.751	AAL - 177.751		AAAAAAA - 4.966	AAAAAAA - 4.966
4	AAAA - 47.284	AAAA - 47.284	8	NNNNNNNNN - 23.879	NNNNNNNNN - 23.879
	NNNN - 46.321	NNNN - 46.321		QQQQQQQQQ - 13.508	QQQQQQQQQ - 13.508
	SSSS - 39.923	SSSS - 39.923		SSSSSSSS - 5.006	SSSSSSSS - 5.006
	QQQQ - 36.956	QQQQ - 36.956		SDSDSDSD - 3.905	SDSDSDSD - 3.905
	EEEE - 29.213	EEEE - 29.213		DSDSDSDS - 3.875	DSDSDSDS - 3.875

Fragmentos más repetidos BD TrEMBL

k	Algoritmo Memoria Externa	k	Algoritmo Memoria Externa
1	L - 2.966.351.173 A - 2.715.689.443 G - 2.175.272.504 V - 2.065.665.475 S - 2.030.276.782	6	QQQQQQ - 2.355.101 AAAAAA - 1.496.140 SSSSSS - 1.411.002 GGGGGG - 898.024 NNNNNN - 861.564
2	AA - 313.028.096 LL - 306.619.288 LA - 285.314.569 AL - 281.904.750 AG - 212.968.976	7	QQQQQQQ - 1.798.936 SSSSSS - 922.266 AAAAAAA - 880.843 WTVYPPL - 815.253 GWTVYPP - 814.768
3	AAA - 45.269.369 LAA - 34.321.839 ALA - 34.225.481 AAL - 33.850.558 LLL - 32.856.884	8	QQQQQQQQ - 1.408.675 GWTVYPPL - 813.630 TGWTVYPP - 798.282 GTGWTVYP - 798.215 MIFFMVMP - 771.616
4	AAAA - 9.197.193 SSSS - 5.754.913 ALAA - 5.176.104 LLLL - 5.111.818 AALA - 5.085.892	9	QQQQQQQQQ - 1.124.827 GTGWTVYPP - 797.778 TGWTVYPPL - 797.178 LLLLSLPVL - 756.549 LLLSLPVLA - 751.235
5	QQQQ - 3.210.070 AAAAA - 3.108.174 SSSSS - 2.473.516 GGGGG - 1.679.232 PPPPP - 1.473.476	10	QQQQQQQQQQ - 911.438 GTGWTVYPPL - 796.679 LLLLSLPVLA - 749.816 PDMAFPRMNN - 667.479 FPRMNNMSFW - 661.041

Fragmentos más repetidos

BD

EROP-Moscow

k	Algoritmo Memoria Interna	Algoritmo Memoria Externa	k	Algoritmo Memoria Interna	Algoritmo Memoria Externa
1	G - 30.373	G - 30.373	6	RRRRRR - 160	RRRRRR - 160
	L - 23.867	L - 23.867		FDEIDR - 127	FDEIDR - 127
	K - 21.956	K - 21.956		NFDEID - 114	NFDEID - 114
	S - 21.486	S - 21.486		CGLSGL - 98	CGLSGL - 98
	C - 21.365	C - 21.365		GLSGLC - 94	GLSGLC - 94
2	GL - 3.307	GL - 3.307	7	NFDEIDR - 102	NFDEIDR - 102
	RR - 2.904	RR - 2.904		CGLSGLC - 88	CGLSGLC - 88
	GG - 2.872	GG - 2.872		VCGLSGL - 79	VCGLSGL - 79
	GK - 2.558	GK - 2.558		MEHFRWG - 76	MEHFRWG - 76
	AA - 2.493	AA - 2.493		FDEIDRS - 72	FDEIDRS - 72
3	RRR - 1.253	RRR - 1.253	8	VCGLSGLC - 71	VCGLSGLC - 71
	FGL - 483	FGL - 483		EHFRWGKP - 68	EHFRWGKP - 68
	GGG - 479	GGG - 479		MEHFRWGK - 68	MEHFRWGK - 68
	LSG - 469	LSG - 469		SMEHFRWG - 65	SMEHFRWG - 65
	AAK - 451	AAK - 451		YSMEHFRW - 62	YSMEHFRW - 62
4	RRRR - 712	RRRR - 712	9	MEHFRWGKP - 68	MEHFRWGKP - 68
	GGGG - 183	GGGG - 183		SMEHFRWGK - 65	SMEHFRWGK - 65
	GCSC - 172	GCSC - 172		YSMEHFRWG - 61	YSMEHFRWG - 61
	CCSG - 167	CCSG - 167		EHFRWGKPV - 59	EHFRWGKPV - 59
	LSGL - 162	LSGL - 162		GGTCNTPGC - 59	GGTCNTPGC - 59
5	RRRRR - 373	RRRRR - 373	10	SMEHFRWGKP - 65	SMEHFRWGKP - 65
	FDEID - 144	FDEID - 144		YSMEHFRWGK - 61	YSMEHFRWGK - 61
	DEIDR - 139	DEIDR - 139		MEHFRWGKPV - 59	MEHFRWGKPV - 59
	CGLSG - 128	CGLSG - 128		SYSMEHFRWG - 53	SYSMEHFRWG - 53
	GGGGG - 121	GGGGG - 121		EHFRWGKPVG - 46	EHFRWGKPVG - 46

Fragmentos más repetidos

BD

Homosapiens

k	Algoritmo Memoria Interna	Algoritmo Memoria Externa
1	L - 3.979.937	L - 3.979.937
	S - 3.123.633	S - 3.123.633
	A - 2.631.167	A - 2.631.167
	E - 2.467.325	E - 2.467.325
	G - 2.448.081	G - 2.448.081
	P - 2.336.530	P - 2.336.530
	T - 2.248.068	T - 2.248.068
	V - 2.229.816	V - 2.229.816
	K - 2.082.746	K - 2.082.746
	R - 2.012.813	R - 2.012.813
	I - 1.863.090	I - 1.863.090
	D - 1.723.766	D - 1.723.766
	Q - 1.715.384	Q - 1.715.384
	F - 1.448.281	F - 1.448.281
	N - 1.448.115	N - 1.448.115
	Y - 1.078.310	Y - 1.078.310
	H - 997.696	H - 997.696
	M - 922.926	M - 922.926
	C - 805.709	C - 805.709
	W - 527.194	W - 527.194

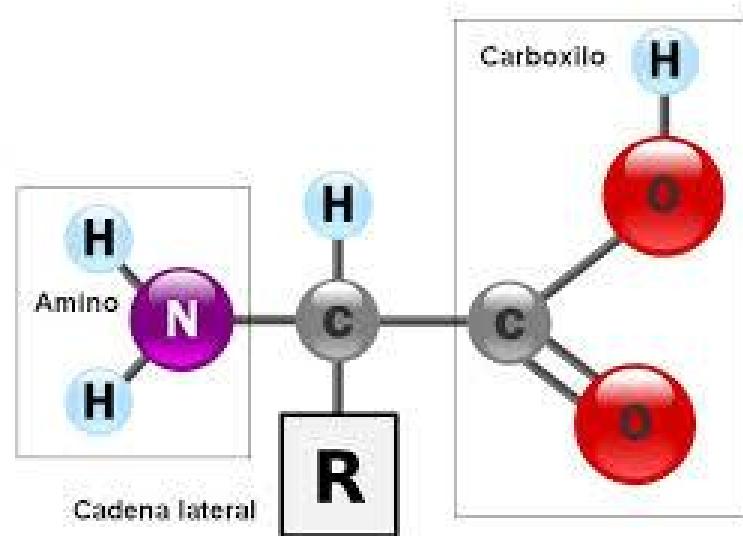
Análisis (enfoque algorítmico)

- Resultados similares para ambos algoritmos
- Uso del priority queue

Análisis (enfoque biológico)

Base de datos UniProt: Los más repetidos son L (Leucina), A (Alanina), G (Glicina), V (Valina). Son residuos hidrofóbicos

Base de datos EROP-Moscow: Los más repetidos son G, L, K (Lisina - aminoácido que posee carga positiva en su grupo radical R), S (Serina) y C (Cisteína). Son oligopéptidos reguladores, moléculas flexibles ya que poseen pocas interacciones intramoleculares débiles.



Conclusiones generales

- Resultados obtenidos
- Extensión del algoritmo a otros temas
- Posible trabajo a futuro

**"Si te caes y te levantas, no caiste,
solamente tomaste un impulso"**

Gracias por su atención

Apéndice

Fórmulas matemáticas

Considerando una proteína compuesta por n aminoácidos:

Máximo número posible de fragmentos de tamaño k (considerando repeticiones de fragmentos):

$$N_k^{teorica} = n - k + 1$$

Máximo número posible de fragmentos (con repeticiones) que se pueden obtener para una molécula de tamaño n , partiendo desde $k = 2$ hasta $k = n - 1$:

$$N_{max}^{teorica} = \sum_2^{n-1} \frac{k(k-1)}{2} - 1$$

Número de estructuras idénticas:

$$R_k = \sum_1^m (i - 1)$$

Número máximo de diferentes fragmentos de largo k :

$$N_k^{diff} = N_k^{teorica} - R_k$$

Número máximo de diferentes fragmentos totales:

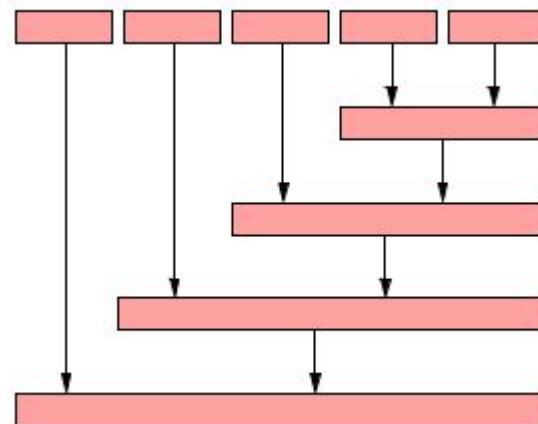
$$N_{max}^{diff} = \left[\sum_2^{n-1} \frac{k(k-1)}{2} - 1 \right] - \sum_2^{n-1} \left[\sum_1^m (i - 1) \right]$$

Algoritmo Memoria Interna

- Arreglo de sufijos: Kasai (ordenamiento de los primeros 2^k caracteres).
- Arreglo LCP: se crea un arreglo de sufijos inverso como auxiliar para encontrar el arreglo LCP.
- Ambos arreglos se guardan en memoria RAM.

Algoritmo Memoria Externa (1)

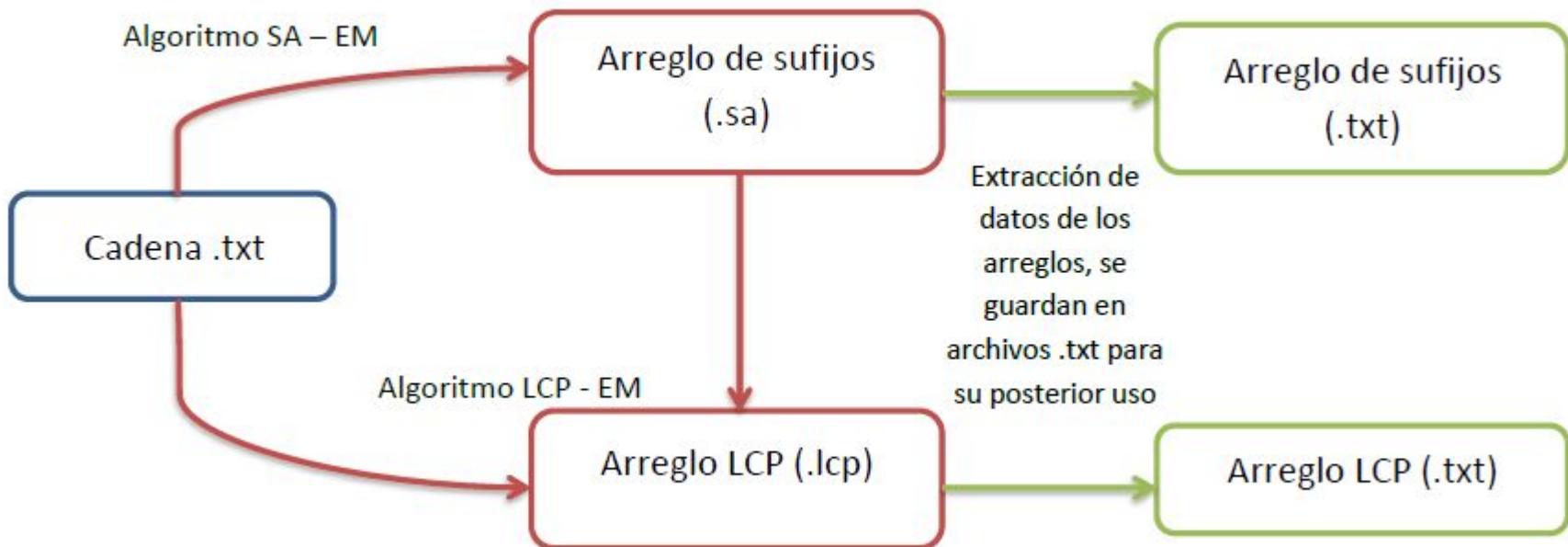
- External Memory Suffix Array (Karkkainen y Kempa)



Algoritmo Memoria Externa (2)

- External Memory LCP Array (Karkkainen y Kempa): Se obtiene en base al texto y el arreglo de sufijos obtenido en memoria externa.
 - Usa varios arreglos permutados como métodos auxiliares.
-
- Tanto el arreglo de sufijos EM como el arreglo LCP EM guardan sus valores utilizando una codificación especial denominada V-byte (archivo guardado pesa $5n$ bytes)

Conversión a archivo .txt manipulable

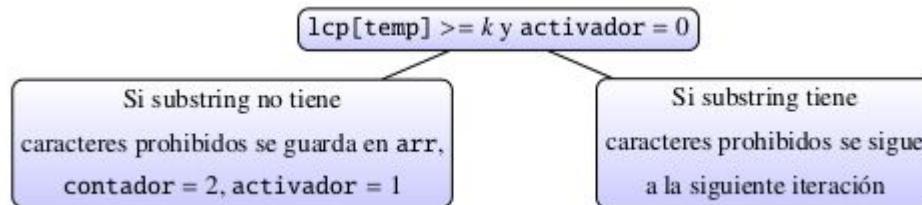


Pesos archivos transformados

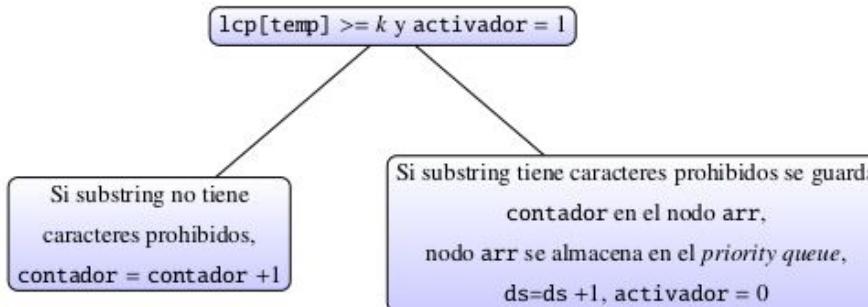
Base de Datos	Archivo .sa	Archivo .lcp	Archivo accesible SA	Archivo accesible LCP	Total
SwissProt	997,5 MB	997,5 MB	1,9 GB	524,8 MB	4,41 GB
TrEMBL	151 GB	151 GB	351,3 GB	85,3 GB	738,6 GB
EROP-Moscow	1,8 MB	1,8 MB	2,4 MB	807,9 KB	6,80 MB
Proteínas humanas	191 MB	191 MB	332,6 MB	120,8 MB	835,4 MB

Casos aplicados Implementación principal

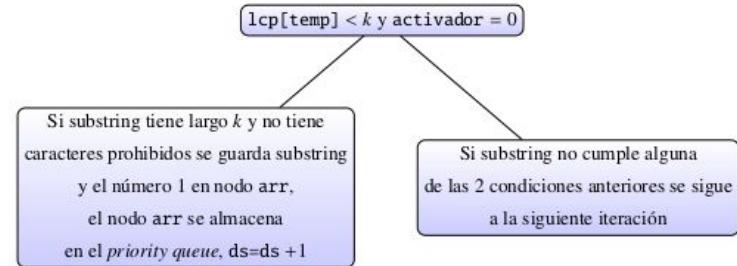
Caso a):



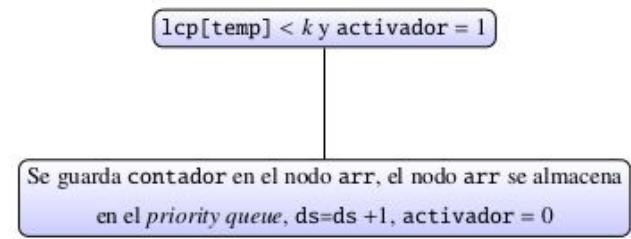
Caso b):



Caso c):



Caso d):



Diferentes fragmentos BD SwissProt (K entre 1 a 25)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00	20	100,00
2	400	400	100,00	400	100,00
3	8.000	8.000	100,00	8.000	100,00
4	160.000	159.999	99,99	159.999	99,99
5	3.200.000	3.113.509	97,29	159.999	97,29
6	64.000.000	32.921.109	51,43	32.921.109	51,43
7	1.280.000.000	84.118.859	6,57	84.118.859	6,57
8	25.600.000.000	100.896.814	0,39	100.896.814	0,39
9	512.000.000.000	105.834.330	0,020	105.834.330	0,020
10	1.024×10^{13}	108.976.567	$1,00 \times 10^{-3}$	108.976.567	0,0010
11	2.048×10^{14}	111.551.595	$5,44 \times 10^{-5}$	111.551.595	$5,44 \times 10^{-5}$
12	4.096×10^{15}	113.751.287	$2,77 \times 10^{-6}$	113.751.287	$2,77 \times 10^{-6}$
13	8.192×10^{16}	115.660.117	$1,41 \times 10^{-7}$	115.660.117	$1,41 \times 10^{-7}$
14	1.638×10^{18}	117.330.023	$7,16 \times 10^{-9}$	117.330.023	$7,16 \times 10^{-9}$
15	3.276×10^{19}	118.797.220	$3,62 \times 10^{-10}$	118.797.220	$3,62 \times 10^{-10}$
16	6.553×10^{20}	120.090.802	$1,83 \times 10^{-11}$	120.090.802	$1,83 \times 10^{-11}$
17	1.310×10^{22}	121.236.766	$9,24 \times 10^{-13}$	121.236.766	$9,24 \times 10^{-13}$
18	2.621×10^{23}	122.255.013	$4,66 \times 10^{-14}$	122.255.013	$4,66 \times 10^{-14}$
19	5.242×10^{24}	123.160.391	$2,34 \times 10^{-15}$	123.160.391	$2,34 \times 10^{-15}$
20	1.048×10^{26}	123.966.286	$1,18 \times 10^{-16}$	123.966.286	$1,18 \times 10^{-16}$
21	2.097×10^{27}	124.685.091	$5,94 \times 10^{-18}$	124.685.091	$5,94 \times 10^{-18}$
22	4.194×10^{28}	125.325.735	$2,98 \times 10^{-19}$	125.325.735	$2,98 \times 10^{-19}$
23	8.388×10^{29}	125.896.961	$1,50 \times 10^{-20}$	125.896.961	$1,50 \times 10^{-20}$
24	1.677×10^{31}	126.406.287	$7,53 \times 10^{-22}$	126.406.287	$7,53 \times 10^{-22}$
25	3.355×10^{32}	126.859.712	$3,78 \times 10^{-23}$	126.859.712	$3,78 \times 10^{-23}$

Diferentes fragmentos BD SwissProt (K entre 26 a 50)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
26	$6,710 \times 10^{33}$	127.262.721	$1,89 \times 10^{-24}$	127.262.721	$1,89 \times 10^{-24}$
27	$1,342 \times 10^{35}$	127.619.577	$9,50 \times 10^{-26}$	127.619.577	$9,50 \times 10^{-26}$
28	$2,684 \times 10^{36}$	127.934.582	$4,76 \times 10^{-27}$	127.934.582	$4,76 \times 10^{-27}$
29	$5,368 \times 10^{37}$	128.211.241	$2,38 \times 10^{-28}$	128.211.241	$2,38 \times 10^{-28}$
30	$1,073 \times 10^{39}$	128.452.495	$1,19 \times 10^{-29}$	128.452.495	$1,19 \times 10^{-29}$
31	$2,147 \times 10^{40}$	128.661.754	$5,99 \times 10^{-31}$	128.661.754	$5,99 \times 10^{-31}$
32	$4,294 \times 10^{41}$	128.841.734	$2,99 \times 10^{-32}$	128.841.734	$2,99 \times 10^{-32}$
33	$8,589 \times 10^{42}$	128.994.975	$1,50 \times 10^{-33}$	128.994.975	$1,50 \times 10^{-33}$
34	$1,717 \times 10^{44}$	129.123.214	$7,51 \times 10^{-35}$	129.123.214	$7,51 \times 10^{-35}$
35	$3,435 \times 10^{45}$	129.227.999	$3,76 \times 10^{-36}$	129.227.999	$3,76 \times 10^{-36}$
36	$6,871 \times 10^{46}$	129.311.151	$1,88 \times 10^{-37}$	129.311.151	$1,88 \times 10^{-37}$
37	$1,374 \times 10^{48}$	129.374.023	$9,41 \times 10^{-39}$	129.373.023	$9,41 \times 10^{-39}$
38	$2,748 \times 10^{49}$	129.418.106	$4,70 \times 10^{-40}$	129.418.106	$4,70 \times 10^{-40}$
39	$5,497 \times 10^{50}$	129.444.962	$2,35 \times 10^{-41}$	129.444.962	$2,35 \times 10^{-41}$
40	$1,099 \times 10^{52}$	129.455.435	$1,17 \times 10^{-42}$	129.455.435	$1,17 \times 10^{-42}$
41	$2,199 \times 10^{53}$	129.450.779	$5,88 \times 10^{-44}$	129.450.779	$5,88 \times 10^{-44}$
42	$4,398 \times 10^{54}$	129.431.855	$2,94 \times 10^{-45}$	129.431.855	$2,94 \times 10^{-45}$
43	$8,796 \times 10^{55}$	129.399.352	$1,47 \times 10^{-46}$	129.399.352	$1,47 \times 10^{-46}$
44	$1,759 \times 10^{57}$	129.354.043	$7,35 \times 10^{-48}$	129.354.043	$7,35 \times 10^{-48}$
45	$3,518 \times 10^{58}$	129.296.497	$3,67 \times 10^{-49}$	129.296.497	$3,67 \times 10^{-49}$
46	$7,036 \times 10^{59}$	129.227.572	$1,83 \times 10^{-50}$	129.227.572	$1,83 \times 10^{-50}$
47	$1,407 \times 10^{61}$	129.147.725	$9,17 \times 10^{-52}$	129.147.725	$9,17 \times 10^{-52}$
48	$2,814 \times 10^{62}$	129.057.687	$4,58 \times 10^{-53}$	129.057.687	$4,58 \times 10^{-53}$
49	$5,629 \times 10^{63}$	128.957.792	$2,29 \times 10^{-54}$	128.957.792	$2,29 \times 10^{-54}$
50	$1,125 \times 10^{65}$	128.848.422	$1,14 \times 10^{-55}$	128.848.422	$1,14 \times 10^{-55}$

Diferentes fragmentos BD TrEMBL (K entre 1 a 25)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00
2	400	400	100,00
3	8.000	8.000	100,00
4	160.000	160.000	100,00
5	3.200.000	3.200.000	100,00
6	64.000.000	63.817.907	99,71
7	1.280.000.000	1.024.659.629	80,05
8	25.600.000.000	5.743.538.889	22,43
9	512.000.000.000	10.114.868.387	1,97
10	1.024×10^{13}	11.524.607.918	0,11
11	2.048×10^{14}	12.097.210.719	$5,90 \times 10^{-3}$
12	4.096000×10^{15}	12.496.100.196	$3,05 \times 10^{-4}$
13	8.192000×10^{16}	12.830.754.238	$1,56 \times 10^{-5}$
14	1.638400×10^{18}	13.123.976.715	$8,01 \times 10^{-7}$
15	3.276800×10^{19}	13.384.096.861	$4,08 \times 10^{-8}$
16	6.553600×10^{20}	13.616.603.068	$2,07 \times 10^{-9}$
17	1.310720×10^{22}	13.825.879.858	$1,05 \times 10^{-10}$
18	2.621440×10^{23}	14.015.107.413	$5,34 \times 10^{-12}$
19	5.242880×10^{24}	14.186.650.854	$2,70 \times 10^{-13}$
20	1.048576×10^{26}	14.342.664.055	$1,36 \times 10^{-14}$
21	2.097152×10^{27}	14.484.928.487	$6,90 \times 10^{-16}$
22	4.194304×10^{28}	14.614.871.775	$3,48 \times 10^{-17}$
23	8.388608×10^{29}	14.733.705.124	$1,75 \times 10^{-18}$
24	1.677722×10^{31}	14.842.531.635	$8,84 \times 10^{-20}$
25	3.355443×10^{32}	14.942.295.108	$4,45 \times 10^{-21}$

Diferentes fragmentos BD TrEMBL (K entre 26 a 50)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AME”	Porcentaje encontrado (%)
26	6.710886×10^{33}	15.033.804.724	$2,24 \times 10^{-22}$
27	1.342177×10^{35}	15.117.770.220	$1,12 \times 10^{-23}$
28	2.684355×10^{36}	15.194.832.726	$5,66 \times 10^{-25}$
29	5.368709×10^{37}	15.265.555.484	$2,84 \times 10^{-26}$
30	1.073742×10^{39}	15.330.434.357	$1,42 \times 10^{-27}$
31	2.147484×10^{40}	15.389.925.372	$7,16 \times 10^{-29}$
32	4.294967×10^{41}	15.444.437.953	$3,59 \times 10^{-30}$
33	8.589935×10^{42}	15.494.326.223	$1,80 \times 10^{-31}$
34	1.717987×10^{44}	15.539.916.767	$9,04 \times 10^{-33}$
35	3.435974×10^{45}	15.581.490.496	$4,53 \times 10^{-34}$
36	6.871948×10^{46}	15.619.332.864	$2,27 \times 10^{-35}$
37	1.374390×10^{48}	15.653.677.726	$1,13 \times 10^{-36}$
38	2.748779×10^{49}	15.684.729.946	$5,70 \times 10^{-38}$
39	5.497558×10^{50}	15.712.714.045	$2,85 \times 10^{-39}$
40	1.099512×10^{52}	15.737.807.161	$1,43 \times 10^{-40}$
41	2.199023×10^{53}	15.760.177.009	$7,16 \times 10^{-42}$
42	4.398047×10^{54}	15.779.975.924	$3,58 \times 10^{-43}$
43	8.796093×10^{55}	15.797.327.916	$1,79 \times 10^{-44}$
44	1.759219×10^{57}	15.812.367.378	$8,98 \times 10^{-46}$
45	3.518437×10^{58}	15.825.209.981	$4,49 \times 10^{-47}$
46	7.036874×10^{59}	15.835.967.244	$2,25 \times 10^{-48}$
47	1.407375×10^{61}	15.844.731.164	$1,12 \times 10^{-49}$
48	2.814750×10^{62}	15.851.604.529	$5,63 \times 10^{-51}$
49	5.629500×10^{63}	15.856.668.836	$2,81 \times 10^{-52}$
50	1.125900×10^{65}	15.860.001.319	$1,40 \times 10^{-53}$

Diferentes fragmentos BD EROP-Moscow (K entre 1 a 25)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00	20	100,00
2	400	400	100,00	400	100,00
3	8.000	7.907	98,83	7.907	98,83
4	160.000	69.218	43,26	69.218	43,26
5	3.200.000	116.454	3,63	116.454	3,63
6	64.000.000	125.036	0,19	125.036	0,19
7	1.280.000.000	125.750	0,0098	125.750	0,0098
8	25.600.000.000	124.054	$4,84 \times 10^{-4}$	124.054	$4,84 \times 10^{-4}$
9	512.000.000.000	121.103	$2,36 \times 10^{-5}$	121.103	$2,36 \times 10^{-5}$
10	$1,024 \times 10^{13}$	117.457	$1,14 \times 10^{-6}$	117.457	$1,14 \times 10^{-6}$
11	$2,048 \times 10^{14}$	113.572	$5,54 \times 10^{-8}$	113.572	$5,54 \times 10^{-8}$
12	$4,096 \times 10^{15}$	109.595	$2,67 \times 10^{-9}$	109.595	$2,67 \times 10^{-9}$
13	$8,192 \times 10^{16}$	105.433	$1,28 \times 10^{-10}$	105.433	$1,28 \times 10^{-10}$
14	$1,638 \times 10^{18}$	101.061	$6,16 \times 10^{-12}$	101.061	$6,16 \times 10^{-12}$
15	$3,276 \times 10^{19}$	96.751	$2,95 \times 10^{-13}$	96.751	$2,95 \times 10^{-13}$
16	$6,553 \times 10^{20}$	92.402	$1,40 \times 10^{-14}$	92.402	$1,40 \times 10^{-14}$
17	$1,310 \times 10^{22}$	87.989	$6,71 \times 10^{-16}$	87.989	$6,71 \times 10^{-16}$
18	$2,621 \times 10^{23}$	83.546	$3,18 \times 10^{-17}$	83.546	$3,18 \times 10^{-17}$
19	$5,242 \times 10^{24}$	79.152	$1,50 \times 10^{-18}$	79.152	$1,50 \times 10^{-18}$
20	$1,048 \times 10^{26}$	74.746	$7,12 \times 10^{-20}$	74.746	$7,12 \times 10^{-20}$
21	$2,097 \times 10^{27}$	70.365	$3,35 \times 10^{-21}$	70.365	$3,35 \times 10^{-21}$
22	$4,194 \times 10^{28}$	66.016	$1,57 \times 10^{-22}$	66.016	$1,57 \times 10^{-22}$
23	$8,388 \times 10^{29}$	61.820	$7,36 \times 10^{-24}$	61.820	$7,36 \times 10^{-24}$
24	$1,677 \times 10^{31}$	57.634	$3,43 \times 10^{-25}$	57.634	$3,43 \times 10^{-25}$
25	$3,355 \times 10^{32}$	53.497	$1,59 \times 10^{-26}$	53.497	$1,59 \times 10^{-26}$

Diferentes fragmentos BD EROP-Moscow (K entre 26 a 50)

Largo péptido (N)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
26	$6,710 \times 10^{33}$	49.611	$7,39 \times 10^{-28}$	49.611	$7,39 \times 10^{-28}$
27	$1,342 \times 10^{35}$	45.777	$3,41 \times 10^{-29}$	45.777	$3,41 \times 10^{-29}$
28	$2,684 \times 10^{36}$	41.972	$1,56 \times 10^{-30}$	41.972	$1,56 \times 10^{-30}$
29	$5,368 \times 10^{37}$	38.273	$7,12 \times 10^{-32}$	38.273	$7,12 \times 10^{-32}$
30	$1,073 \times 10^{39}$	34.679	$3,22 \times 10^{-33}$	34.679	$3,22 \times 10^{-33}$
31	$2,147 \times 10^{40}$	31.284	$1,45 \times 10^{-34}$	31.284	$1,45 \times 10^{-34}$
32	$4,294 \times 10^{41}$	28.052	$6,53 \times 10^{-36}$	28.052	$6,53 \times 10^{-36}$
33	$8,589 \times 10^{42}$	24.990	$2,90 \times 10^{-37}$	24.990	$2,90 \times 10^{-37}$
34	$1,717 \times 10^{44}$	22.069	$1,28 \times 10^{-38}$	22.069	$1,28 \times 10^{-38}$
35	$3,435 \times 10^{45}$	19.402	$5,64 \times 10^{-40}$	19.402	$5,64 \times 10^{-40}$
36	$6,871 \times 10^{46}$	16.859	$2,45 \times 10^{-41}$	16.859	$2,45 \times 10^{-41}$
37	$1,374 \times 10^{48}$	14.497	$1,05 \times 10^{-42}$	14.497	$1,05 \times 10^{-42}$
38	$2,748 \times 10^{49}$	12.410	$4,51 \times 10^{-44}$	12.410	$4,51 \times 10^{-44}$
39	$5,497 \times 10^{50}$	10.629	$1,93 \times 10^{-45}$	10.629	$1,93 \times 10^{-45}$
40	$1,099 \times 10^{52}$	9.006	$8,19 \times 10^{-47}$	9.006	$8,19 \times 10^{-47}$
41	$2,199 \times 10^{53}$	7.559	$3,43 \times 10^{-48}$	7.559	$3,43 \times 10^{-48}$
42	$4,398 \times 10^{54}$	6.230	$1,41 \times 10^{-49}$	6.230	$1,41 \times 10^{-49}$
43	$8,796 \times 10^{55}$	5.004	$5,68 \times 10^{-51}$	5.004	$5,68 \times 10^{-51}$
44	$1,759 \times 10^{57}$	3.878	$2,20 \times 10^{-52}$	3.878	$2,20 \times 10^{-52}$
45	$3,518 \times 10^{58}$	2.897	$8,23 \times 10^{-54}$	2.897	$8,23 \times 10^{-54}$
46	$7,036 \times 10^{59}$	2.031	$2,88 \times 10^{-55}$	2.031	$2,88 \times 10^{-55}$
47	$1,407 \times 10^{61}$	1.295	$9,20 \times 10^{-57}$	1.295	$9,20 \times 10^{-57}$
48	$2,814 \times 10^{62}$	758	$2,69 \times 10^{-58}$	758	$2,69 \times 10^{-58}$
49	$5,629 \times 10^{63}$	386	$6,85 \times 10^{-60}$	386	$6,85 \times 10^{-60}$
50	$1,125 \times 10^{65}$	116	$1,03 \times 10^{-61}$	116	$1,03 \times 10^{-61}$

Diferentes fragmentos BD Homosapiens (K entre 1 a 25)

Largo péptido (<i>N</i>)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
1	20	20	100,00	20	100,00
2	400	400	100,00	400	100,00
3	8.000	8.000	100,00	8.000	100,00
4	160.000	159.835	99,89	159.835	99,89
5	3.200.000	2.445.409	76,42	2.445.409	76,42
6	64.000.000	8.823.762	13,78	8.823.762	13,78
7	1.280.000.000	11.334.257	0,89	11.334.257	0,89
8	25.600.000.000	11.777.746	0,046	11.777.746	0,046
9	512.000.000.000	11.930.386	0,0023	11.930.386	0,0023
10	$1,024 \times 10^{13}$	12.032.248	$1,18 \times 10^{-4}$	12.032.248	$1,18 \times 10^{-4}$
11	$2,048 \times 10^{14}$	12.113.988	$5,91 \times 10^{-6}$	12.113.988	$5,91 \times 10^{-6}$
12	$4,096 \times 10^{15}$	12.182.561	$2,97 \times 10^{-7}$	12.182.561	$2,97 \times 10^{-7}$
13	$8,192 \times 10^{16}$	12.241.962	$1,49 \times 10^{-8}$	12.241.962	$1,49 \times 10^{-8}$
14	$1,638 \times 10^{18}$	12.294.014	$7,50 \times 10^{-10}$	12.294.014	$7,50 \times 10^{-10}$
15	$3,276 \times 10^{19}$	12.340.141	$3,76 \times 10^{-11}$	12.340.141	$3,76 \times 10^{-11}$
16	$6,553 \times 10^{20}$	12.380.952	$1,89 \times 10^{-12}$	12.380.952	$1,89 \times 10^{-12}$
17	$1,310 \times 10^{22}$	12.417.964	$9,47 \times 10^{-14}$	12.417.964	$9,47 \times 10^{-14}$
18	$2,621 \times 10^{23}$	12.451.529	$4,75 \times 10^{-15}$	12.451.529	$4,75 \times 10^{-15}$
19	$5,242 \times 10^{24}$	12.482.187	$2,38 \times 10^{-16}$	12.482.187	$2,38 \times 10^{-16}$
20	$1,048 \times 10^{26}$	12.510.252	$1,19 \times 10^{-17}$	12.510.252	$1,19 \times 10^{-17}$
21	$2,097 \times 10^{27}$	12.536.181	$5,98 \times 10^{-19}$	12.536.181	$5,98 \times 10^{-19}$
22	$4,194 \times 10^{28}$	12.560.224	$2,99 \times 10^{-20}$	12.560.224	$2,99 \times 10^{-20}$
23	$8,388 \times 10^{29}$	12.582.575	$1,50 \times 10^{-21}$	12.582.575	$1,50 \times 10^{-21}$
24	$1,677 \times 10^{31}$	12.603.460	$7,51 \times 10^{-22}$	12.603.460	$7,51 \times 10^{-22}$
25	$3,355 \times 10^{32}$	12.622.974	$3,76 \times 10^{-24}$	12.622.974	$3,76 \times 10^{-24}$

Diferentes fragmentos BD Homosapiens (K entre 26 a 50)

Largo péptido (N)	N-péptidos posibles (20^N)	Encontrados “AMI”	Porcentaje encontrado (%)	Encontrados “AME”	Porcentaje encontrado (%)
26	$6,710 \times 10^{33}$	12.641.270	$1,88 \times 10^{-25}$	12.641.270	$1,88 \times 10^{-25}$
27	$1,342 \times 10^{35}$	12.658.508	$9,43 \times 10^{-27}$	12.658.508	$9,43 \times 10^{-27}$
28	$2,684 \times 10^{36}$	12.674.724	$4,72 \times 10^{-28}$	12.674.724	$4,72 \times 10^{-28}$
29	$5,368 \times 10^{37}$	12.689.960	$2,36 \times 10^{-29}$	12.689.960	$2,36 \times 10^{-29}$
30	$1,073 \times 10^{39}$	12.704.363	$1,18 \times 10^{-30}$	12.704.363	$1,18 \times 10^{-30}$
31	$2,147 \times 10^{40}$	12.717.937	$5,92 \times 10^{-32}$	12.717.937	$5,92 \times 10^{-32}$
32	$4,294 \times 10^{41}$	12.730.708	$2,96 \times 10^{-33}$	12.730.708	$2,96 \times 10^{-33}$
33	$8,589 \times 10^{42}$	12.742.719	$1,48 \times 10^{-34}$	12.742.719	$1,48 \times 10^{-34}$
34	$1,717 \times 10^{44}$	12.754.002	$7,42 \times 10^{-36}$	12.754.002	$7,42 \times 10^{-36}$
35	$3,435 \times 10^{45}$	12.764.580	$3,72 \times 10^{-37}$	12.764.580	$3,72 \times 10^{-37}$
36	$6,871 \times 10^{46}$	12.774.485	$1,86 \times 10^{-38}$	12.774.485	$1,86 \times 10^{-38}$
37	$1,374 \times 10^{48}$	12.783.835	$9,30 \times 10^{-40}$	12.783.835	$9,30 \times 10^{-40}$
38	$2,748 \times 10^{49}$	12.792.665	$4,65 \times 10^{-41}$	12.792.665	$4,65 \times 10^{-41}$
39	$5,497 \times 10^{50}$	12.800.978	$2,33 \times 10^{-42}$	12.800.978	$2,33 \times 10^{-42}$
40	$1,099 \times 10^{52}$	12.808.759	$1,17 \times 10^{-43}$	12.808.759	$1,17 \times 10^{-43}$
41	$2,199 \times 10^{53}$	12.816.054	$5,83 \times 10^{-45}$	12.816.054	$5,83 \times 10^{-45}$
42	$4,398 \times 10^{54}$	12.822.954	$2,92 \times 10^{-46}$	12.822.954	$2,92 \times 10^{-46}$
43	$8,796 \times 10^{55}$	12.829.470	$1,46 \times 10^{-47}$	12.829.470	$1,46 \times 10^{-47}$
44	$1,759 \times 10^{57}$	12.835.544	$7,30 \times 10^{-49}$	12.835.544	$7,30 \times 10^{-49}$
45	$3,518 \times 10^{58}$	12.841.211	$3,65 \times 10^{-50}$	12.841.211	$3,65 \times 10^{-50}$
46	$7,036 \times 10^{59}$	12.846.521	$1,83 \times 10^{-51}$	12.846.521	$1,83 \times 10^{-51}$
47	$1,407 \times 10^{61}$	12.851.470	$9,13 \times 10^{-53}$	12.851.470	$9,13 \times 10^{-53}$
48	$2,814 \times 10^{62}$	12.856.127	$4,57 \times 10^{-54}$	12.856.127	$4,57 \times 10^{-54}$
49	$5,629 \times 10^{63}$	12.860.412	$2,29 \times 10^{-55}$	12.860.412	$2,29 \times 10^{-55}$
50	$1,125 \times 10^{65}$	12.864.382	$1,14 \times 10^{-56}$	12.864.382	$1,14 \times 10^{-56}$

Fragmentos más repetidos BD SwissProt (K entre 11 a 20)

k	Algoritmo Memoria Interna	Algoritmo Memoria Externa
11	NNNNNNNNNNNN - 17.610 QQQQQQQQQQQQQ - 8.638 SDSDSDSDSDS - 3.549	NNNNNNNNNNNN - 17.610 QQQQQQQQQQQQQ - 8.638 SDSDSDSDSDS - 3.549
12	NNNNNNNNNNNNN - 16.128 QQQQQQQQQQQQQ - 7.613 SDSDSDSDSDSD - 3.454	NNNNNNNNNNNNN - 16.128 QQQQQQQQQQQQQ - 7.613 SDSDSDSDSDSD - 3.454
13	NNNNNNNNNNNNN - 14.789 QQQQQQQQQQQQQ - 6.764 SDSDSDSDSDSDS - 3.387	NNNNNNNNNNNNN - 14.789 QQQQQQQQQQQQQ - 6.764 SDSDSDSDSDSDS - 3.387
14	NNNNNNNNNNNNNN - 13.570 QQQQQQQQQQQQQQ - 6.035 SDSDSDSDSDSDSD - 3.304	NNNNNNNNNNNNNN - 13.570 QQQQQQQQQQQQQQ - 6.035 SDSDSDSDSDSDSD - 3.304
15	NNNNNNNNNNNNNN - 12.445 QQQQQQQQQQQQQQQ - 5.406 SDSDSDSDSDSDSDS - 3.243	NNNNNNNNNNNNNN - 12.445 QQQQQQQQQQQQQQQ - 5.406 SDSDSDSDSDSDSDS - 3.243
16	NNNNNNNNNNNNNNN - 11.404 QQQQQQQQQQQQQQQ - 4.850 SDSDSDSDSDSDSDSD - 3.163	NNNNNNNNNNNNNNN - 11.404 QQQQQQQQQQQQQQQ - 4.850 SDSDSDSDSDSDSDSD - 3.163
17	NNNNNNNNNNNNNNNN - 10.436 QQQQQQQQQQQQQQQQ - 4.353 SDSDSDSDSDSDSDS - 3.105	NNNNNNNNNNNNNNNN - 10.436 QQQQQQQQQQQQQQQQ - 4.353 SDSDSDSDSDSDSDS - 3.105
18	NNNNNNNNNNNNNNNN - 9.537 QQQQQQQQQQQQQQQQQ - 3.915 SDSDSDSDSDSDSDSD - 3.028	NNNNNNNNNNNNNNNN - 9.537 QQQQQQQQQQQQQQQQQ - 3.915 SDSDSDSDSDSDSDSD - 3.028
19	NNNNNNNNNNNNNNNNN - 8.701 QQQQQQQQQQQQQQQQQ - 3.524 SDSDSDSDSDSDSDSDS - 2.980	NNNNNNNNNNNNNNNNN - 8.701 QQQQQQQQQQQQQQQQQ - 3.524 SDSDSDSDSDSDSDSDS - 2.980
20	NNNNNNNNNNNNNNNNN - 7.921 QQQQQQQQQQQQQQQQQ - 3.165 SDSDSDSDSDSDSDSDSD - 2.919	NNNNNNNNNNNNNNNNN - 7.921 QQQQQQQQQQQQQQQQQ - 3.165 SDSDSDSDSDSDSDSDSD - 2.919

Fragmentos más repetidos BD SwissProt (K entre 21 a 30)

k	Algoritmo Memoria Interna	Algoritmo Memoria Externa
21	NNNNNNNNNNNNNNNNNNNN - 7.195 SDSDSDSDSDSDSDSDSDS - 2.872 DSDSDSDSDSDSDSDSDSD - 2.842	NNNNNNNNNNNNNNNNNNNN - 7.195 SDSDSDSDSDSDSDSDSDS - 2.872 DSDSDSDSDSDSDSDSDSD - 2.842
22	NNNNNNNNNNNNNNNNNNNN - 6.519 SDSDSDSDSDSDSDSDSDS - 2.813 DSDSDSDSDSDSDSDSDSDS - 2.795	NNNNNNNNNNNNNNNNNNNN - 6.519 SDSDSDSDSDSDSDSDSDS - 2.813 DSDSDSDSDSDSDSDSDSDS - 2.795
23	NNNNNNNNNNNNNNNNNNNN - 5.882 SDSDSDSDSDSDSDSDSDS - 2.767 DSDSDSDSDSDSDSDSDSD - 2.737	NNNNNNNNNNNNNNNNNNNN - 5.882 SDSDSDSDSDSDSDSDSDS - 2.767 DSDSDSDSDSDSDSDSDSD - 2.737
24	NNNNNNNNNNNNNNNNNNNN - 5.298 SDSDSDSDSDSDSDSDSD - 2.709 DSDSDSDSDSDSDSDSDSDS - 2.691	NNNNNNNNNNNNNNNNNNNN - 5.298 SDSDSDSDSDSDSDSDSD - 2.709 DSDSDSDSDSDSDSDSDSDS - 2.691
25	NNNNNNNNNNNNNNNNNNNN - 4.762 SDSDSDSDSDSDSDSDSDS - 2.663 DSDSDSDSDSDSDSDSDSD - 2.640	NNNNNNNNNNNNNNNNNNNN - 4.762 SDSDSDSDSDSDSDSDSDS - 2.663 DSDSDSDSDSDSDSDSDSD - 2.640
26	NNNNNNNNNNNNNNNNNNNN - 4.266 SDSDSDSDSDSDSDSDSDS - 2.612 DSDSDSDSDSDSDSDSDSDS - 2.594	NNNNNNNNNNNNNNNNNNNN - 4.266 SDSDSDSDSDSDSDSDSDS - 2.612 DSDSDSDSDSDSDSDSDSDS - 2.594
27	NNNNNNNNNNNNNNNNNNNN - 3.812 SDSDSDSDSDSDSDSDSDS - 2.566 DSDSDSDSDSDSDSDSDSD - 2.545	NNNNNNNNNNNNNNNNNNNN - 3.812 SDSDSDSDSDSDSDSDSDS - 2.566 DSDSDSDSDSDSDSDSDSD - 2.545
28	NNNNNNNNNNNNNNNNNNNN - 3.408 SDSDSDSDSDSDSDSDSDS - 2.517 DSDSDSDSDSDSDSDSDSDS - 2.499	NNNNNNNNNNNNNNNNNNNN - 3.408 SDSDSDSDSDSDSDSDSDS - 2.517 DSDSDSDSDSDSDSDSDSDS - 2.499
29	NNNNNNNNNNNNNNNNNNNN - 3.037 SDSDSDSDSDSDSDSDSDS - 2.471 DSDSDSDSDSDSDSDSDSD - 2.451	NNNNNNNNNNNNNNNNNNNN - 3.037 SDSDSDSDSDSDSDSDSDS - 2.471 DSDSDSDSDSDSDSDSDSD - 2.451
30	NNNNNNNNNNNNNNNNNNNN - 2.690 SDSDSDSDSDSDSDSDSDS - 2.423 DSDSDSDSDSDSDSDSDSDS - 2.406	NNNNNNNNNNNNNNNNNNNN - 2.690 SDSDSDSDSDSDSDSDSDS - 2.423 DSDSDSDSDSDSDSDSDSDS - 2.406

Kentre 31 a 40

k	Algoritmo Memoria Interna
41	SD - 2.017 DSD - 2.012 NN - 646
42	SD - 1.994 DSD - 1.978 NNN - 560
43	SD - 1.960 DSD - 1.955 NN - 486
44	SD - 1.937 DSD - 1.921 LGDPDNYTPANPLNTPPHIKPEWYFLFAYAILRSIPNKLGGVLA - 443
45	SD - 1.905 DSD - 1.898 LGDPDNYTPANPLNTPPHIKPEWYFLFAYAILRSIPNKLGGVLA - 441
46	SD - 1.883 DSD - 1.866 LGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWR - 428
47	SD - 1.851 DSD - 1.844 GCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDR - 422
48	SD - 1.829 DSD - 1.812 LGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDR - 421
49	SD - 1.797 DSD - 1.790 LGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRF - 418
50	SD - 1.775 DSD - 1.758 LLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRF - 411

k	Algoritmo Memoria Externa
41	SD - 2.017 DSD - 2.012 NN - 646
42	SD - 1.994 DSD - 1.978 NN - 560
43	SD - 1.960 DSD - 1.955 NN - 486
44	SD - 1.937 DSD - 1.921 LGDPDNYTPANPLNTPPHIKPEWYFLFAYAILRSIPNKLGGVLA - 443
45	SD - 1.905 DSD - 1.898 LGDPDNYTPANPLNTPPHIKPEWYFLFAYAILRSIPNKLGGVLA - 441
46	SD - 1.883 DSD - 1.866 LGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWR - 428
47	SD - 1.851 DSD - 1.844 GCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDR - 422
48	SD - 1.829 DSD - 1.812 LGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDR - 421
49	SD - 1.797 DSD - 1.790 LGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRF - 418
50	SD - 1.775 DSD - 1.758 LLGCTIKPKLGLSAKNYGRAVYECLRGGLDFTKDDENVNSQPFMRWRDRF - 411

Fragmentos más repetidos BD TrEMBL (K entre 11 a 30)

k	Algoritmo Memoria Externa
11	QQQQQQQQQQQ - 747.275 LLLLSLPVLAG - 651.789 GTGWTVYPPLS - 641.850
12	MAFPRMNNMSFW - 635.480 DMAFPRMNNMSF - 635.036 PDMAFPRMNNMS - 626.591
13	DMAFPRMNNMSFW - 634235 PDMAFPRMNNMSF - 624989 LLLLSLPVLAGAI - 609167
14	PDMAFPRMNNMSFW - 624.205 LLLLSLPVLAGAIT - 602.580 LLLSLPVLAGAITM - 594.451
15	LLLLSLPVLAGAITM - 593.277 APDMAFPRMNNMSFW - 587.410 GAPDMAFPRMNNMSF - 573.789
16	GAPDMAFPRMNNMSFW - 573.072 LLLLSLPVLAGAITML - 560.280 LLL.SLPVLAGAITML - 553.123
17	LLLLSLPVLAGAITML - 552.018 DMAFPRMNNMSFWLLPP - 517.441 PDMAFPRMNNMSFWLLP - 513.511
18	PDMAFPRMNNMSFWLLPP - 510.800 LLLLSLPVLAGAITMLLT - 509.597 MIFFMVMPIMIGGFGNWL - 502.774
19	LLLLSLPVLAGAITMLTD - 494.639 APDMAFPRMNNMSFWLLPP - 482.567 GAPDMAFPRMNNMSFWLLP - 475.434
20	GAPDMAFPRMNNMSFWLLPP - 473.240 LLLLSLPVLAGAITMLTDR - 447.133 APDMAFPRMNNMSFWLLPPS - 440.860

k	Algoritmo Memoria Externa
21	GAPDMAFPRMNNMSFWLLPPS - 432.279 LLLLSLPVLAGAITMLTDRN - 419.728 GGFGNWLVPLMLGAPDMAFPR - 396.661
22	MIFFMVMPIMIGGFGNWLVPLM - 378.931 IGGFGNWLVPLMLGAPDMAFPR - 377.153 FGNWLVPLMLGAPDMAFPRMNN - 359.982
23	GFGNWLVPLMLGAPDMAFPRMNN - 359.640 GGFGNWLVPLMLGAPDMAFPRMN - 359.610 WLVPLMLGAPDMAFPRMNNMSFW - 344.612
24	GFGNWLVPLMLGAPDMAFPRMNN - 359.443 NWLVPLMLGAPDMAFPRMNNMSFW - 344.262 GNWLVPLMLGAPDMAFPRMNNMSF - 343.907
25	GNWLVPLMLGAPDMAFPRMNNMSFW - 343.736 GFGNWLVPLMLGAPDMAFPRMNNMS - 343.581 GGFGNWLVPLMLGAPDMAFPRMNNNM - 343.560
26	GFGGNWLVPLMLGAPDMAFPRMNNMS - 343.421 FGNWLVPLMLGAPDMAFPRMNNMSFW - 343.361 GFGNWLVPLMLGAPDMAFPRMNNMSF - 343.280
27	GFGGNWLVPLMLGAPDMAFPRMNNMSF - 343.121 GFGNWLVPLMLGAPDMAFPRMNNMSFW - 343.114 IGGFGNWLVPLMLGAPDMAFPRMNNMS - 325.259
28	GFGGNWLVPLMLGAPDMAFPRMNNMSFW - 342.955 IGGFGNWLVPLMLGAPDMAFPRMNNMSF - 324.970 MIFFMVMPIMIGGFGNWLVPLMLGAPDMA - 280.394
29	IGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 324.816 MIFFMVMPIMIGGFGNWLVPLMLGAPDMA - 280.009 FFMVMPIIGGFGNWLVPLMLGAPDMAFP - 279.921
30	IFFMVMPIIGGFGNWLVPLMLGAPDMAFP - 279.703 MIFFMVMPIMIGGFGNWLVPLMLGAPDMAF - 279.654 FFMVMPIIGGFGNWLVPLMLGAPDMAFP - 279.556

k	Algoritmo Memoria Externa
31	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFP - 279.595 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPR - 279.343 GGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 276.883
32	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPR - 279.236 GGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 276.219 IGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 272.788
33	IGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 272.129 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNN - 261.538 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMN - 261.458
34	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMN - 261.358 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNN - 261.348 VMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 259.364
35	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNN - 261.249 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMS - 258.297 MVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 258.227
36	IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMS - 258.112 MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNNM - 258.097 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSF - 258.094
37	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMS - 258.015 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 257.967 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSF - 257.910
38	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSF - 257.813 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 257.783 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWL - 213.300
39	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 257.688 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWL - 213.149 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLL - 212.672
40	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWL - 213.071 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLL - 212.521 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 211.173

k	Algoritmo Memoria Externa
41	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLL - 212.443 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 211.022 FFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPP - 210.576
42	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 210.944 IFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPP - 210.428 AFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 190.888
43	MIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPP - 210.350 HAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 190.613 AHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSF - 185.499
44	AHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 185.410 TAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSF - 185.297 VTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMS - 184.986
45	TAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 185.209 VTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSF - 184.849 IVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMS - 177.134
46	VTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 184.761 IVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSF - 176.999 AFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 166.132
47	IVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFW - 176.912 HAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 165.913 TAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLL - 161.308
48	AHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPP - 161.121 TAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 161.059 VTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLL - 160.969
49	TAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPP - 160.978 VTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLP - 160.720 IVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLL - 155.384
50	VTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPP - 160.640 IVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPP - 155.328 TAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPRMNNMSFWLLPPS - 134.371

Fragmentos más repetidos BD EROP-Moscow (K entre 11 a 20)

k	Algoritmo “Memoria Interna”	Algoritmo “Memoria Externa”
11	YSMEHFRWGKP - 61 SMEHFRWGKPV - 56 SYSMEHFRWGK - 53	YSMEHFRWGKP - 61 SMEHFRWGKPV - 56 SYSMEHFRWGK - 53
12	YSMEHFRWGKPV - 54 SYSMEHFRWGKP - 53 SMEHFRWGKPVG - 44	YSMEHFRWGKPV - 54 SYSMEHFRWGKP - 53 SMEHFRWGKPVG - 44
13	SYSMEHFRWGKPV - 47 YSMEHFRWGKPVG - 44 ALRHYINLITRQR - 35	SYSMEHFRWGKPV - 47 YSMEHFRWGKPVG - 44 ALRHYINLITRQR - 35
14	SYSMEHFRWGKPVG - 41 ALRHYINLITRQRY - 34 KLKKTETQEKNPLP - 32	SYSMEHFRWGKPVG - 41 ALRHYINLITRQRY - 34 KLKKTETQEKNPLP - 32
15	GTVKDLLIGAGKSAA - 30 GKEVGMDVIRTGIDV - 27 CGETCFGTCNTPGC - 27	GTVKDLLIGAGKSAA - 30 GKEVGMDVIRTGIDV - 27 CGETCFGTCNTPGC - 27
16	GKEVGMDVIRTGIDVA - 26 IGKEVGMDVIRTGIDV - 26 KEVGMDVIRTGIDVAG - 26	GKEVGMDVIRTGIDVA - 26 IGKEVGMDVIRTGIDV - 26 KEVGMDVIRTGIDVAG - 26
17	GKEVGMDVIRTGIDVAG - 26 IGKEVGMDVIRTGIDVA - 25 YYSLRHYINLITRQRY - 25	GKEVGMDVIRTGIDVAG - 26 IGKEVGMDVIRTGIDVA - 25 YYSLRHYINLITRQRY - 25
18	IGKEVGMDVIRTGIDVAG - 25 GKEVGMDVIRTGIDVAGC - 23 CKKKCKGGWKCKFNMCV - 23	IGKEVGMDVIRTGIDVAG - 25 GKEVGMDVIRTGIDVAGC - 23 CKKKCKGGWKCKFNMCV - 23
19	GKEVGMDVIRTGIDVAGCK - 23 CKKKCKGGWKCKFNMCVK - 22 GTVKDLLIGAGKSAAQSVL - 22	GKEVGMDVIRTGIDVAGCK - 23 CKKKCKGGWKCKFNMCVK - 22 GTVKDLLIGAGKSAAQSVL - 22
20	GKEVGMDVIRTGIDVAGCKI - 22 IGKEVGMDVIRTGIDVAGCK - 22 EVGMDVIRTGIDVAGCKIKG - 21	GKEVGMDVIRTGIDVAGCKI - 22 IGKEVGMDVIRTGIDVAGCK - 22 EVGMDVIRTGIDVAGCKIKG - 21

Fragmentos más repetidos BD EROP-Moscow (K entre 21 a 30)

k	Algoritmo Memoria Interna
21	GKEVGMDVIRTGIDVAGCKIK - 21 IGKEVGMDVIRTGIDVAGCKI - 21 KEVGMDVIRTGIDVAGCKIKG - 21
22	GKEVGMDVIRTGIDVAGCKIKG - 21 IGKEVGMDVIRTGIDVAGCKIK - 20 KEVGMDVIRTGIDVAGCKIKGE - 19
23	IGKEVGMDVIRTGIDVAGCKIKG - 20 GKEVGMDVIRTGIDVAGCKIKGE - 19 KEVGMDVIRTGIDVAGCKIKGEC - 19
24	GKEVGMDVIRTGIDVAGCKIKGEC - 19 IGKEVGMDVIRTGIDVAGCKIKGE - 18 KGVKHIGKEVGMDVIRTGIDVAGC - 16
25	IGKEVGMDVIRTGIDVAGCKIKGEC - 18 KGVKHIGKEVGMDVIRTGIDVAGCK - 16 KHIGKEVGMDVIRTGIDVAGCKIKG - 15
26	KGVKHIGKEVGMDVIRTGIDVAGCKI - 15 GVKHIGKEVGMDVIRTGIDVAGCKIK - 14 VKhIGKEVGMDVIRTGIDVAGCKIKG - 14
27	GVKHIGKEVGMDVIRTGIDVAGCKIKG - 14 KGVKHIGKEVGMDVIRTGIDVAGCKIK - 14 KHIGKEVGMDVIRTGIDVAGCKIKGEC - 13
28	KGVKHIGKEVGMDVIRTGIDVAGCKIKG - 14 GVKHIGKEVGMDVIRTGIDVAGCKIKGE - 12 VKhIGKEVGMDVIRTGIDVAGCKIKGEC - 12
29	GVKHIGKEVGMDVIRTGIDVAGCKIKGEC - 12 KGVKHIGKEVGMDVIRTGIDVAGCKIKGE - 12 SCEIEKEGNKDCKKKCKGGWKCKFNMCV - 11
30	KGVKHIGKEVGMDVIRTGIDVAGCKIKGEC - 12 EPPISLDLTFHLLREVLEMARAEQLAQQA - 10 EPPISLDLTFHLLREVLEMARAEQLAQQAH - 10

k	Algoritmo Memoria Externa
21	GKEVGMDVIRTGIDVAGCKIK - 21 IGKEVGMDVIRTGIDVAGCKI - 21 KEVGMDVIRTGIDVAGCKIKG - 21
22	GKEVGMDVIRTGIDVAGCKIKG - 21 IGKEVGMDVIRTGIDVAGCKIK - 20 KEVGMDVIRTGIDVAGCKIKGE - 19
23	IGKEVGMDVIRTGIDVAGCKIKG - 20 GKEVGMDVIRTGIDVAGCKIKGE - 19 KEVGMDVIRTGIDVAGCKIKGEC - 19
24	GKEVGMDVIRTGIDVAGCKIKGEC - 19 IGKEVGMDVIRTGIDVAGCKIKGE - 18 KGVKHIGKEVGMDVIRTGIDVAGC - 16
25	IGKEVGMDVIRTGIDVAGCKIKGEC - 18 KGVKHIGKEVGMDVIRTGIDVAGCK - 16 KHIGKEVGMDVIRTGIDVAGCKIKG - 15
26	KGVKHIGKEVGMDVIRTGIDVAGCKI - 15 GVKHIGKEVGMDVIRTGIDVAGCKIK - 14 VKhIGKEVGMDVIRTGIDVAGCKIKG - 14
27	GVKHIGKEVGMDVIRTGIDVAGCKIKG - 14 KGVKHIGKEVGMDVIRTGIDVAGCKIK - 14 KHIGKEVGMDVIRTGIDVAGCKIKGEC - 13
28	KGVKHIGKEVGMDVIRTGIDVAGCKIKG - 14 GVKHIGKEVGMDVIRTGIDVAGCKIKGE - 12 VKhIGKEVGMDVIRTGIDVAGCKIKGEC - 12
29	GVKHIGKEVGMDVIRTGIDVAGCKIKGEC - 12 KGVKHIGKEVGMDVIRTGIDVAGCKIKGE - 12 SCEIEKEGNKDCKKKCKGGWKCKFNMCV - 11
30	KGVKHIGKEVGMDVIRTGIDVAGCKIKGEC - 12 EPPISLDLTFHLLREVLEMARAEQLAQQA - 10 EPPISLDLTFHLLREVLEMARAEQLAQQAH - 10

K entre 31 a 40

k	Algoritmo Memoria Interna
31	EPPISLDLTFHLLREVLEMARAELAQQAH - 10 EPPISLDLTFHLLREVLEMARAELAQQAH - 10 ISLDLTFHLLREVLEMARAELAQQAHNSRK - 10
32	EPPISLDLTFHLLREVLEMARAELAQQAH - 10 EPPISLDLTFHLLREVLEMARAELAQQAHSN - 10 PISLDLTFHLLREVLEMARAELAQQAHNSRK - 10
33	EPPISLDLTFHLLREVLEMARAELAQQAHSN - 10 EPPISLDLTFHLLREVLEMARAELAQQAHNSR - 10 PPISLDLTFHLLREVLEMARAELAQQAHNSRK - 10
34	EPPISLDLTFHLLREVLEMARAELAQQAHNSR - 10 EPPISLDLTFHLLREVLEMARAELAQQAHNSRK - 10 SEEPPISDLTFHLLREVLEMARAELAQQAHSN - 10
35	EPPISLDLTFHLLREVLEMARAELAQQAHNSRK - 10 SEEPPISDLTFHLLREVLEMARAELAQQAHNSR - 10 PISLDLTFHLLREVLEMARAELAQQAHNSRKLME - 9
36	SEEPPISDLTFHLLREVLEMARAELAQQAHNSRK - 10 EPPISLDLTFHLLREVLEMARAELAQQAHNSRKL - 9 EPPISLDLTFHLLREVLEMARAELAQQAHNSRKL - 9
37	EPPISLDLTFHLLREVLEMARAELAQQAHNSRKL - 9 EPPISLDLTFHLLREVLEMARAELAQQAHNSRKLME - 9 SEEPPISDLTFHLLREVLEMARAELAQQAHNSRKL - 9
38	EPPISLDLTFHLLREVLEMARAELAQQAHNSRKLME - 9 SEEPPISDLTFHLLREVLEMARAELAQQAHNSRKL - 9 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 7
39	SEEPPISDLTFHLLREVLEMARAELAQQAHNSRKLME - 9 PTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 DMAIEKFDKSLLKKTETQEKNPLPSKETIEQEKGAGES - 6
40	FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 PTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDM - 6

k	Algoritmo Memoria Externa
31	EPPISLDLTFHLLREVLEMARAELAQQAH - 10 EPPISLDLTFHLLREVLEMARAELAQQAH - 10 ISLDLTFHLLREVLEMARAELAQQAHNSRK - 10
32	EPPISLDLTFHLLREVLEMARAELAQQAH - 10 EPPISLDLTFHLLREVLEMARAELAQQAHSN - 10 PISLDLTFHLLREVLEMARAELAQQAHNSRK - 10
33	EPPISLDLTFHLLREVLEMARAELAQQAHSN - 10 EPPISLDLTFHLLREVLEMARAELAQQAHNSR - 10 PPISLDLTFHLLREVLEMARAELAQQAHNSRK - 10
34	EPPISLDLTFHLLREVLEMARAELAQQAHNSR - 10 EPPISLDLTFHLLREVLEMARAELAQQAHNSRK - 10 SEEPPISDLTFHLLREVLEMARAELAQQAHSN - 10
35	EPPISLDLTFHLLREVLEMARAELAQQAHNSRK - 10 SEEPPISDLTFHLLREVLEMARAELAQQAHNSR - 10 PISLDLTFHLLREVLEMARAELAQQAHNSRKLME - 9
36	SEEPPISDLTFHLLREVLEMARAELAQQAHNSRK - 10 EPPISLDLTFHLLREVLEMARAELAQQAHNSRKL - 9 EPPISLDLTFHLLREVLEMARAELAQQAHNSRKL - 9
37	EPPISLDLTFHLLREVLEMARAELAQQAHNSRKL - 9 EPPISLDLTFHLLREVLEMARAELAQQAHNSRKLME - 9 SEEPPISDLTFHLLREVLEMARAELAQQAHNSRKL - 9
38	EPPISLDLTFHLLREVLEMARAELAQQAHNSRKLME - 9 SEEPPISDLTFHLLREVLEMARAELAQQAHNSRKL - 9 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 7
39	SEEPPISDLTFHLLREVLEMARAELAQQAHNSRKLME - 9 PTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 DMAIEKFDKSLLKKTETQEKNPLPSKETIEQEKGAGES - 6
40	FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 PTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDM - 6

k	Algoritmo Memoria Interna
41	FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDM - 6 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 KPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKGAGES - 5
42	SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDM - 6 DKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKGAGES - 5 FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDMP - 4
43	SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKGAGES - 4 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDMP - 4 ARYRCCRSQSRSRCCRPRRRCRRRRRSCRARRRATCCRRLRY - 3
44	ARYRCCRSQSRSRCCRPRRRCRRRRRSCRARRRATCCRRLRY - 3 EKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLKCL - 3 YEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLK - 3
45	ARYRCCRSQSRSRCCRPRRRCRRRRRSCRARRRATCCRRLRYRL - 3 DDESECIINTRDSPWGRCRTRMCMSCCPRNGCTVYHWRRGHG - 3 PIYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLK - 3
46	IYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLKCL - 3 PIYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLK - 3 ARYRCCRSRSRCCRPRRRCRRRRRCCRRRRVCCRRYTVRCRR - 2
47	PIYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLKCL - 3 ARYRCCRSRSRCCRPRRRCRRRRRCCRRRRVCCRRYTVRCRR - 2 ATTKYISY GALQRNTVPCSRGGASYYNCRPGAQANPYSRGCSAITRC - 2
48	ARYRCCRSQSRSRCCRPRRRCRRRRRCRARRRAMRCCRRYRLRCR - 2 ATTKYISY GALQRNTVPCSRGGASYYNCRPGAQANPYSRGCSAITRC - 2 CESGPCCDNCKFLKEGTICKMARGDNHDYCNGKTCDCPRNPYKGEHD 2
49	ARYRCCRSQSRSRCCRPRRRCRRRRRCRARRRAMRCCRRYRLRCR - 2 EFELDRICGYGTARCRKKCRSQEYRIGRCPTNYACCLRKWDESLLNRTK - 2 DCESGPCCDNCKFLKEGTICKMARGDNHDYCNGKTCDCPRNPYKGEHD - 2
50	ARYRCCRSQSRSRCCRPRRRCRRRRRCRARRRAMRCCRRYRLRCR - 2 EFELDRICGYGTARCRKKCRSQEYRIGRCPTNYACCLRKWDESLLNRTK - 2 ACGAEGSCHFGGGELCNDRKRCGPTGKYKRGACCGLKQAAPCGYYS - 1

k	Algoritmo Memoria Externa
41	FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDM - 6 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDD - 6 KPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKGAGES - 5
42	SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDM - 6 DKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKGAGES - 5 FPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDMP - 4
43	SDKPDMAEIEKFDKSKLKKTETQEKNPLPSKETIEQEKGAGES - 4 SFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNVAHVDDMP - 4 ARYRCCRSQSRSRCCRPRRRCRRRRRSCRARRRATCCRRLRY - 3
44	ARYRCCRSQSRSRCCRPRRRCRRRRRSCRARRRATCCRRLRY - 3 EKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLKCL - 3 YEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLK - 3
45	ARYRCCRSQSRSRCCRPRRRCRRRRRSCRARRRATCCRRLRYRL - 3 DDESECIINTRDSPWGRCRTRMCMSCCPRNGCTVYHWRRGHG - 3 PIYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLK - 3
46	IYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLKCL - 3 PIYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLK - 3 ARYRCCRSRSRCCRPRRRCRRRRRCCRRRRVCCRRYTVRCRR - 2
47	PIYEKKYGQVPMCDAGEQCAVRKGARIGKLDCPRTSCNSFLKCL - 3 ARYRCCRSRSRCCRPRRRCRRRRRCCRRRRVCCRRYTVRCRR - 2 ATTKYISY GALQRNTVPCSRGGASYYNCRPGAQANPYSRGCSAITRC - 2
48	ARYRCCRSQSRSRCCRPRRRCRRRRRCRARRRAMRCCRRYRLRCR - 2 ATTKYISY GALQRNTVPCSRGGASYYNCRPGAQANPYSRGCSAITRC - 2 CESGPCCDNCKFLKEGTICKMARGDNHDYCNGKTCDCPRNPYKGEHD 2
49	ARYRCCRSQSRSRCCRPRRRCRRRRRCRARRRAMRCCRRYRLRCR - 2 EFELDRICGYGTARCRKKCRSQEYRIGRCPTNYACCLRKWDESLLNRTK - 2 DCESGPCCDNCKFLKEGTICKMARGDNHDYCNGKTCDCPRNPYKGEHD - 2
50	ARYRCCRSQSRSRCCRPRRRCRRRRRCRARRRAMRCCRRYRLRCR - 2 EFELDRICGYGTARCRKKCRSQEYRIGRCPTNYACCLRKWDESLLNRTK - 2 ACGAEGSCHFGGGELCNDRKRCGPTGKYKRGACCGLKQAAPCGYYS - 1

Fragментos más repetidos BD Homosapiens (K entre 11 a 20)

k	Algoritmo “Memoria Interna”	Algoritmo “Memoria Externa”
11	QQQQQQQQQQQQ - 2.279 QDTELVETRPA - 1.549 TELVETRPAGD - 1.548	QQQQQQQQQQQQ - 2.279 QDTELVETRPA - 1.549 TELVETRPAGD - 1.548
12	QQQQQQQQQQQQ - 2.074 QDTELVETRPAG - 1.547 DTELVETRPAGD - 1.547	QQQQQQQQQQQQ - 2.074 QDTELVETRPAG - 1.547 DTELVETRPAGD - 1.547
13	QQQQQQQQQQQQQQ - 1.882 QDTELVETRPAGD - 1.547 TQDTELVETRPAG - 1.545	QQQQQQQQQQQQQQ - 1.882 QDTELVETRPAGD - 1.547 TQDTELVETRPAG - 1.545
14	QQQQQQQQQQQQQQ - 1.708 TQDTELVETRPAGD - 1.545 QTQDTELVETRPAG - 1.533	QQQQQQQQQQQQQQ - 1.708 TQDTELVETRPAGD - 1.545 QTQDTELVETRPAG - 1.533
15	QQQQQQQQQQQQQQ - 1.553 QTQDTELVETRPAGD - 1.533 DQTQDTELVETRPAG - 1.517	QQQQQQQQQQQQQQ - 1.553 QTQDTELVETRPAGD - 1.533 DQTQDTELVETRPAG - 1.517
16	DQTQDTELVETRPAGD - 1.517 DGEDQTQDTELVETRP - 1.515 EDQTQDTELVETRPAG - 1.514	DQTQDTELVETRPAGD - 1.517 DGEDQTQDTELVETRP - 1.515 EDQTQDTELVETRPAG - 1.514
17	DGEDQTQDTELVETRPA - 1.514 EDQTQDTELVETRPAGD - 1.514 GEDQTQDTELVETRPAG - 1.512	DGEDQTQDTELVETRPA - 1.514 EDQTQDTELVETRPAGD - 1.514 GEDQTQDTELVETRPAG - 1.512
18	DGEDQTQDTELVETRPAG - 1.512 GEDQTQDTELVETRPAGD - 1.512 PPHIKPEWYFLFAYTILR - 1.467	DGEDQTQDTELVETRPAG - 1.512 GEDQTQDTELVETRPAGD - 1.512 PPHIKPEWYFLFAYTILR - 1.467
19	DGEDQTQDTELVETRPAGD - 1.512 TPPHIKPEWYFLFAYTILR - 1.467 PPHIKPEWYFLFAYTILRS - 1.465	DGEDQTQDTELVETRPAGD - 1.512 TPPHIKPEWYFLFAYTILR - 1.467 PPHIKPEWYFLFAYTILRS - 1.465
20	TPPHIKPEWYFLFAYTILRS - 1.465 NTPPHIKPEWYFLFAYTILR - 1.461 LNTPPHIKPEWYFLFAYTIL - 1.460	TPPHIKPEWYFLFAYTILRS - 1.465 NTPPHIKPEWYFLFAYTILR - 1.461 LNTPPHIKPEWYFLFAYTIL - 1.460

Fragmentos más repetidos BD Homosapiens (K entre 21 a 30)

k	Algoritmo Memoria Interna
21	LNTPPHIKPEWYFLFAYTILR - 1.460 NTPPHIKPEWYFLFAYTILRS - 1.459 PLNTPPHIKPEWYFLFAYTIL - 1.459
22	PLNTPPHIKPEWYFLFAYTILR - 1.459 LNTPPHIKPEWYFLFAYTILRS - 1.458 NPLNTPPHIKPEWYFLFAYTIL - 1.454
23	PLNTPPHIKPEWYFLFAYTILRS - 1.457 NPLNTPPHIKPEWYFLFAYTILR - 1.454 IKPEWYFLFAYTILRSVPNKLGG - 1.452
24	NPLNTPPHIKPEWYFLFAYTILRS - 1.452 HIKPEWYFLFAYTILRSVPNKLGG - 1.449 ANPLNTPPHIKPEWYFLFAYTILR - 1.449
25	PPIHKPEWYFLFAYTILRSVPNKLG - 1.448 TPPHIKPEWYFLFAYTILRSVPNKL - 1.448 ANPLNTPPHIKPEWYFLFAYTILRS - 1.447
26	TPPHIKPEWYFLFAYTILRSVPNKLG - 1.448 PPIHKPEWYFLFAYTILRSVPNKLGG - 1.447 NTPPHIKPEWYFLFAYTILRSVPNKL - 1.442
27	TPPHIKPEWYFLFAYTILRSVPNKLGG - 1.447 NTPPHIKPEWYFLFAYTILRSVPNKLG - 1.442 LNTPPHIKPEWYFLFAYTILRSVPNKL - 1.441
28	LNTPPHIKPEWYFLFAYTILRSVPNKLG - 1.441 NTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.441 PLNTPPHIKPEWYFLFAYTILRSVPNKL - 1.440
29	LNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.440 PLNTPPHIKPEWYFLFAYTILRSVPNKLG - 1.440 NPLNTPPHIKPEWYFLFAYTILRSVPNKL - 1.435
30	PLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.439 NPLNTPPHIKPEWYFLFAYTILRSVPNKLG - 1.435 ANPLNTPPHIKPEWYFLFAYTILRSVPNKL - 1.430

k	Algoritmo Memoria Externa
21	LNTPPHIKPEWYFLFAYTILR - 1.460 NTPPHIKPEWYFLFAYTILRS - 1.459 PLNTPPHIKPEWYFLFAYTIL - 1.459
22	PLNTPPHIKPEWYFLFAYTILR - 1.459 LNTPPHIKPEWYFLFAYTILRS - 1.458 NPLNTPPHIKPEWYFLFAYTIL - 1.454
23	PLNTPPHIKPEWYFLFAYTILRS - 1.457 NPLNTPPHIKPEWYFLFAYTILR - 1.454 IKPEWYFLFAYTILRSVPNKLGG - 1.452
24	NPLNTPPHIKPEWYFLFAYTILRS - 1.452 HIKPEWYFLFAYTILRSVPNKLGG - 1.449 ANPLNTPPHIKPEWYFLFAYTILR - 1.449
25	PPIHKPEWYFLFAYTILRSVPNKLG - 1.448 TPPHIKPEWYFLFAYTILRSVPNKL - 1.448 ANPLNTPPHIKPEWYFLFAYTILRS - 1.447
26	TPPHIKPEWYFLFAYTILRSVPNKLG - 1.448 PPIHKPEWYFLFAYTILRSVPNKLGG - 1.447 NTPPHIKPEWYFLFAYTILRSVPNKL - 1.442
27	TPPHIKPEWYFLFAYTILRSVPNKLGG - 1.447 NTPPHIKPEWYFLFAYTILRSVPNKLG - 1.442 LNTPPHIKPEWYFLFAYTILRSVPNKL - 1.441
28	LNTPPHIKPEWYFLFAYTILRSVPNKLG - 1.441 NTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.441 PLNTPPHIKPEWYFLFAYTILRSVPNKL - 1.440
29	LNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.440 PLNTPPHIKPEWYFLFAYTILRSVPNKLG - 1.440 NPLNTPPHIKPEWYFLFAYTILRSVPNKL - 1.435
30	PLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.439 NPLNTPPHIKPEWYFLFAYTILRSVPNKLG - 1.435 ANPLNTPPHIKPEWYFLFAYTILRSVPNKL - 1.430

K entre 31 a 40

k	Algoritmo Memoria Interna
31	NPLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.434 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.430 TPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.427
32	ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.429 TPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.424 PPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.422
33	TPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.422 PPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.420 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.420
34	TPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.420 PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.419 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.417
35	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.416 TPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.416 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.415
36	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.414 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.413 TPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.412
37	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.412 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.409 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.409
38	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.408 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.407 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.405
39	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.404 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.403 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.402
40	NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.399 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.398 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.397

k	Algoritmo Memoria Externa
31	NPLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.434 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.430 TPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.427
32	ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGG - 1.429 TPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.424 PPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.422
33	TPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.422 PPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.420 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.420
34	TPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.420 PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.419 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.417
35	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.416 TPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.416 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.415
36	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.414 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.413 TPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.412
37	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.412 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.409 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.409
38	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.408 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.407 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.405
39	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.404 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.403 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLS - 1.402
40	NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.399 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSI - 1.398 LNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.397

k	Algoritmo Memoria Interna
41	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.396 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSIL - 1.394 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.391
42	NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.391 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.386 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.385
43	ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.386 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.382 YTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.381
44	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLL - 1.380 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLS - 1.378 YTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.377
45	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLS - 1.378 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.374 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLL - 1.374
46	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.374 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLS - 1.372 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.370
47	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.370 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.368 NYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.364
48	DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSIL - 1.364 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.362 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.362
49	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.362 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.356 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILILA - 1.346
50	DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.356 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILILA - 1.346 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILILAM - 1.341

k	Algoritmo Memoria Externa
41	PLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.396 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSIL - 1.394 NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.391
42	NPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.391 ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.386 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLAL - 1.385
43	ANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.386 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALL - 1.382 YTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.381
44	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLL - 1.380 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLS - 1.378 YTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.377
45	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLS - 1.378 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.374 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLL - 1.374
46	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.374 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLS - 1.372 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSIL - 1.370
47	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSIL - 1.370 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSI - 1.368 NYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.364
48	DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSIL - 1.364 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.362 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.362
49	PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.362 DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILI - 1.356 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSILILA - 1.346
50	DPDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLLSILIL - 1.356 PDNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSILILA - 1.346 DNYTLANPLNTPPHIKPEWYFLFAYTILRSVPNKLGGVLALLSILILAM - 1.341