



STRUCTURA BIOLOGICĂ.
§2.5. UTILIZAREA PROGRAMELOR
DEDICATE: EXEMPLE
ALINIAREA SECVENȚELOR

Sorana D. BOLBOACĂ

PROBLEMA: AMINOACIZI

- <http://www.uniprot.org/uniprot/P81364>
- Denumirea proteinei: Concanavalin-Ma
- Denumirea organismului: Canavalia rosea (Beach bean) (Canavalia maritima)
- Lungimea secvenței: 236 aa
- Funcție: Glucose/D-mannose specific lectin



PROBLEMA: AMINOACIZI

- <http://www.uniprot.org/uniprot/P55915>
- Denumirea proteinei:
Concanavalin-Br
- Denumirea organismului: *Canavalia brasiliensis* (Brazilian jack bean)
- Lungimea secvenței:
237 aa
- Funcție: Glucose/D-mannose specific lectin



IDENTIFICAREA SECVENȚEI

- Concanavalin-Ma:

Perez G, Perez C, Sousa-Cavada B, Moreira R, Richardson M. Comparison of the amino acid sequences of the lectins from seeds of *Dioclea lehmanni* and *Canavalia maritima*. *Phytochemistry* 1991;30:2619-2621.

- Concanavalin-Br:

Grangeiro TB. Thesis (1996). Universidade Federale do Ceara, Brazil.

SALVAREA ÎN FORMAT FASTA

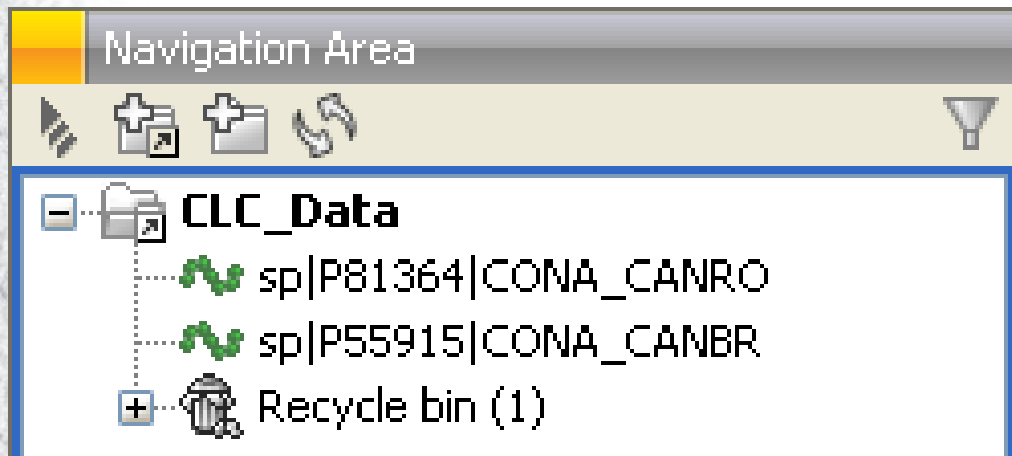
Concanavalin-Ma.fasta

```
>sp|P81364|CONA_CANRO Concanavalin-Ma OS=Canavalia rosea PE=1 SV=1
ADTIVAVELDTYPNTDVGDPSPHXXXXXXXXSVRXXTAKWNMQNGKVGTAHISYNSVGKRL
SAVVSYPNGDSATVSYDVLDLNVLPWVVRVGLSASTGLYKETNTILSWSFTSKLKSNSTH
ETNALHFMFNQFTKDQKDLILQSDATTGTDGNLXXTRVSSNGPSQGSTVGRALFYAPVHI
WESSATVAGFDATFXCLIKSPDSHPADGIAFFISNIDSSIPSGSTGRLGLFPDAN
```

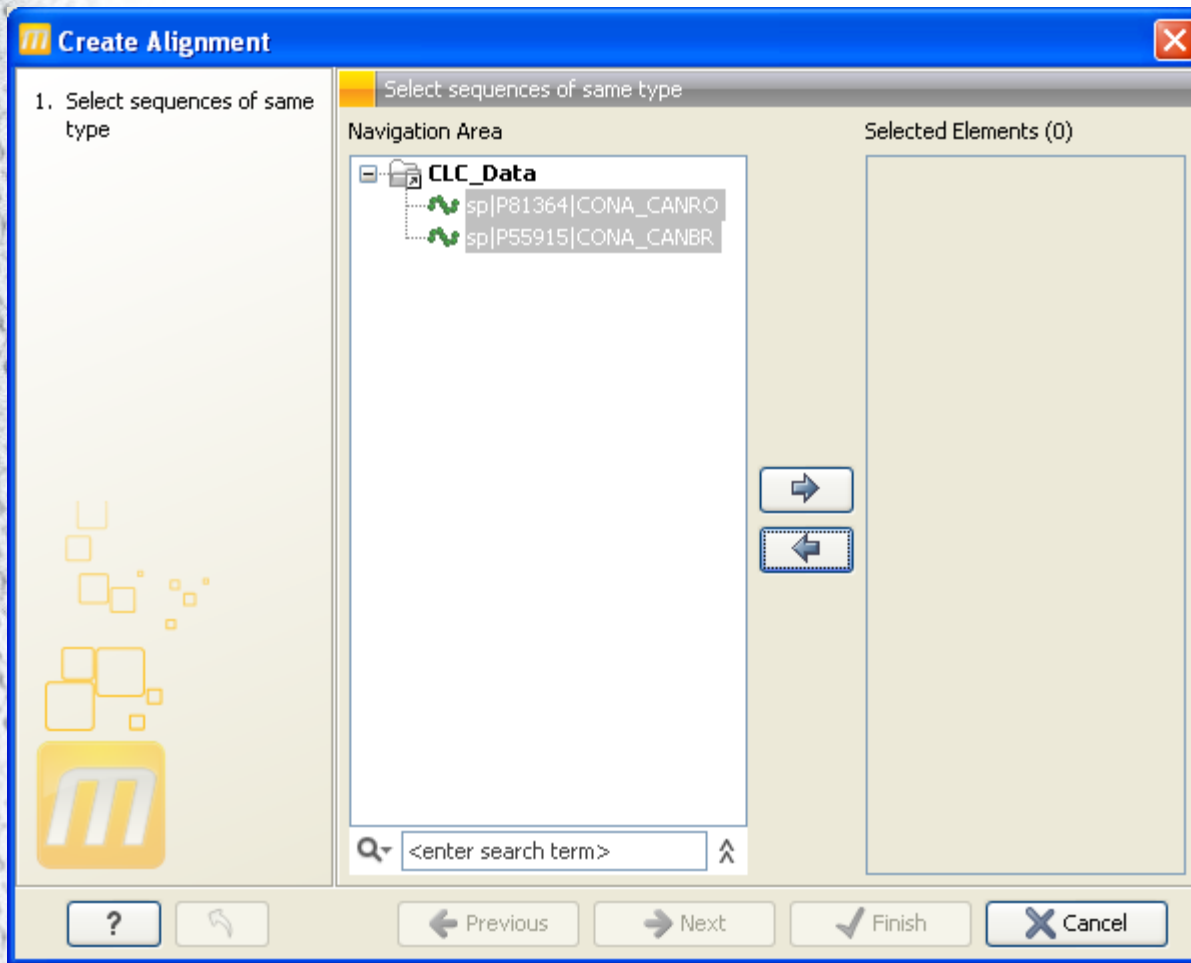
Concanavalin-Br.fasta

```
>sp|P55915|CONA_CANBR Concanavalin-Br OS=Canavalia brasiliensis PE=1 SV=1
ADTIVAVELDTYPNTDIGDPSPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVGKR
LSAVVSYPNGDSATVSYDVLDLNVLPWVVRVGLSASTGLYKETNTILSWSFTSKLKSNST
HETNALHFMFNQFSKDQKDLILQGDATTGTEGNLRLTRVSSNGSPQGSVGRALFYAPVH
IWESSAVVASFEATFTFLIKSPDSHPADGIAFFISNIDSSIPSGSTGRLGLFPDAN
```

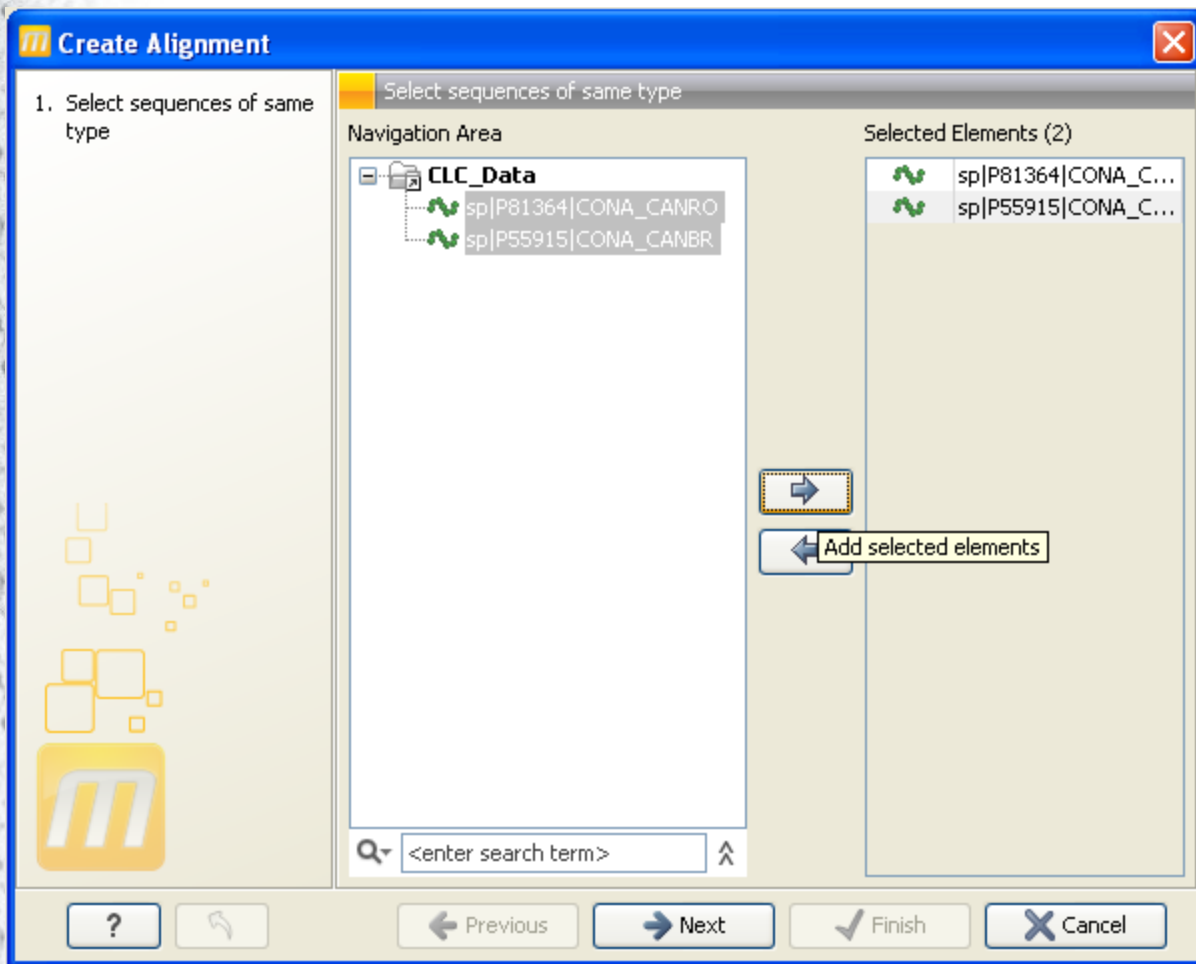
IMPORTUL FIȘIERULUI FASTA ÎN CLC



CLC: ALINIEREA SECVENȚELOR



CLC: ALINIAREA SECVENȚELOR



CLC: ALINIAREA SECVENȚELOR

Create Alignment

1. Select sequences of same type
2. Set parameters

Set parameters

Gap settings

Gap open cost:

Gap extension cost:

End gap cost:

Alignment

Less accurate (fast)

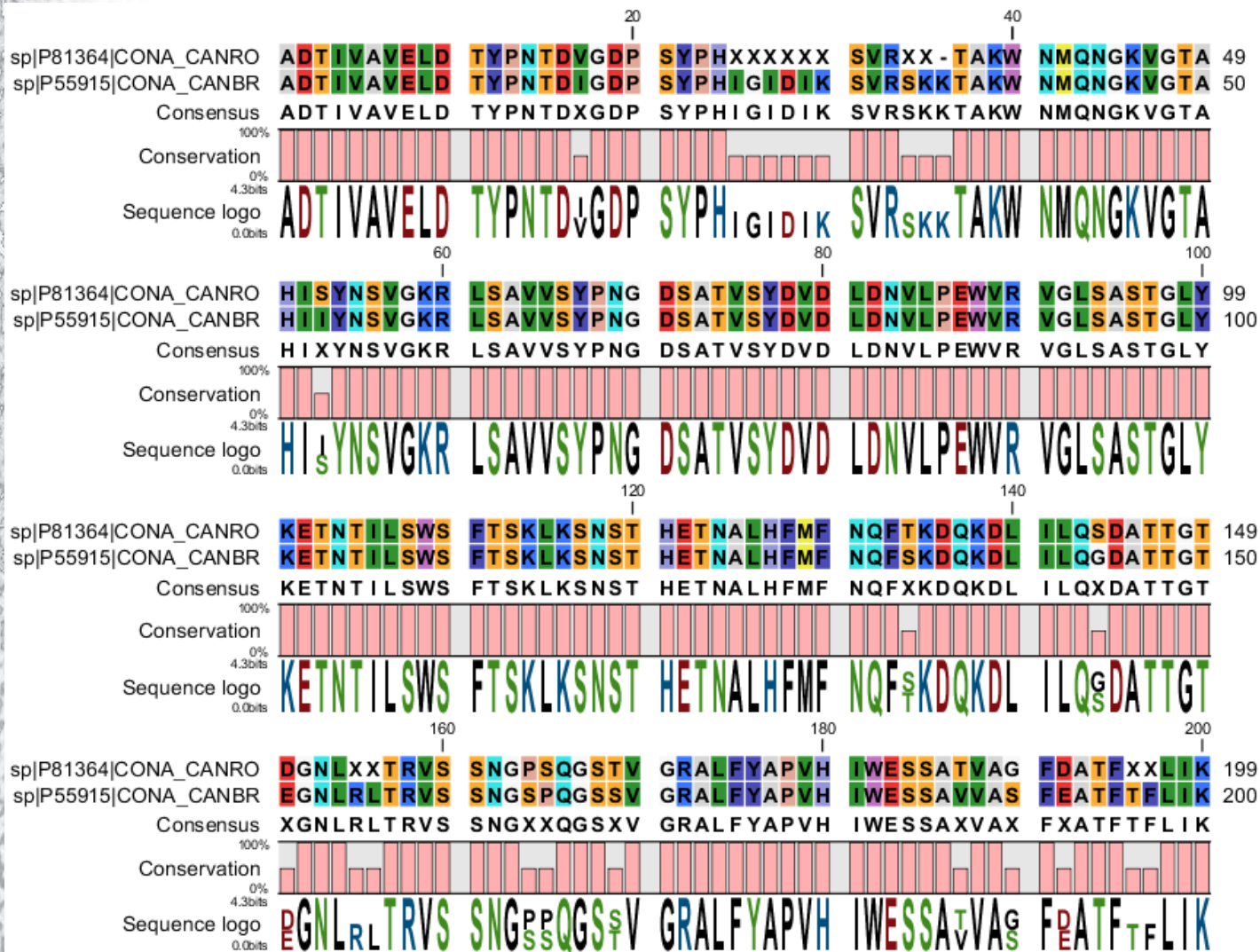
Very accurate (slow)

Redo alignments

Use fixpoints

?

CLC: REZULTATE ALINIERE



ALINIEREA UNIPROT

The screenshot shows a web browser window with the UniProt website. The address bar displays the URL <http://www.uniprot.org/uniprot/P81364?tab=align>. The browser's address bar also shows a search for "Concanavalin-Ma - Canavalia rosea (...)". The UniProt logo and "UniProtKB" are visible in the top navigation bar. Below the navigation bar, there are five tabs: "Search", "Blast *", "Align", "Retrieve", and "ID Mapping *". The "Align" tab is selected. The main content area is titled "Sequences or UniProt identifiers" and contains a text input field with the following text:

```
P81364  
P55915
```

To the right of the input field are two buttons: "Align" and "Clear".

ALINIAREA UNIPROT

UniProt

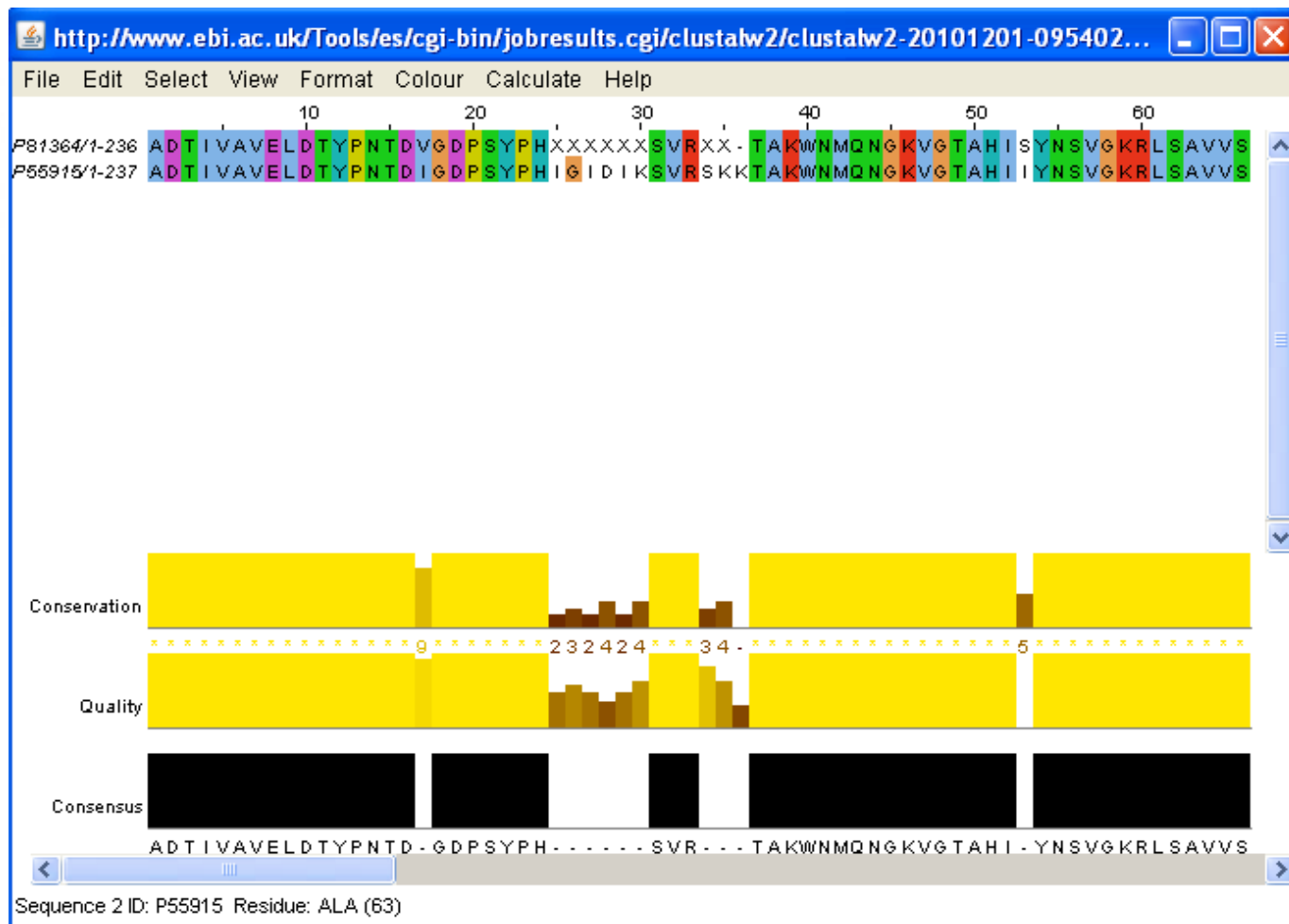
Search Blast **Align** Retrieve ID Mapping

Uniprot identifiers

```
>P81364
ADTIVAVELDTYPNTDVGDPSPHXXXXXXXXSVRXXTAKWNMONGKVGTAHISYNSVGKRL
SAVVSYPNGDSATVSYDVDLDNVLPWVRVGLSASTGLYKETNTILSWSFTSKLKSNSTH
ETNALHFMFNQFTKDQKDLILQSDATTGTDGNLXXTRVSSNGPSQGSTVGRALFYAPVHI
WESSATVAGFDATFXXLIKSPDSHPADGIAFFISNIDSSIPSGSTGRLLGLFPDAN
>P55915
ADTIVAVELDTYPNTDIGDPSPHIGIDIKSVRSKKTAKWNMONGKVGTAHIIYNSVGKR
```


Align
Clear

ALINIAREA UNIPROT



ALINIAREA UNIPROT

ClustalW results

 Learn how to print this alignment in colour

```
P81364 ADTIVAVELDTYPNTDVGDPSYPHXXXXXXXXSVRXX-TAKWNMQNGKVGTAHISYNSVGKR 59 CONA_CANMR
P55915 ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNMQNGKVGTAHIIYNSVGKR 60 CONA_CANBR
*****:***** *** *****
P81364 LSAVVSYPNGDSATVSYDVLDLNVLPWVRVGLSASTGLYKETNTILSWSFTSKLKSNST 119 CONA_CANMR
P55915 LSAVVSYPNGDSATVSYDVLDLNVLPWVRVGLSASTGLYKETNTILSWSFTSKLKSNST 120 CONA_CANBR
*****:*****:*****:*** *****.***:*****
P81364 HETNALHFMFNQFTKDQKDLILQSDATTGTDGNLXXTRVSSNGPSQGSTVGRALFYAPVH 179 CONA_CANMR
P55915 HETNALHFMFNQFSKDQKDLILQGDATTGTEGNLRLTRVSSNGSPQGSVGRALFYAPVH 180 CONA_CANBR
*****:*****.*****:*** *****.***:*****
P81364 IWESSATVAGFDATFXXLIKSPDHPADGIAFFISNIDSSIPSGSTGRLLGLFPDAN 236 CONA_CANMR
P55915 IWESSAVVASFEATFTFLIKSPDHPADGIAFFISNIDSSIPSGSTGRLLGLFPDAN 237 CONA_CANBR
*****.***:*** *****
```

ALINIEREA UNIPROT

<input checked="" type="checkbox"/>	P55915	CONA_CANBR	★	Concanavalin-Br
<input checked="" type="checkbox"/>	P02866	CONA_CANEN	★	Concanavalin-A
<input checked="" type="checkbox"/>	P14894	CONA_CANGL	★	Concanavalin-A
<input checked="" type="checkbox"/>	P81460	CONA_CANLI	★	Concanavalin-A
<input checked="" type="checkbox"/>	P81364	CONA_CANRO	★	Concanavalin-Ma

Canavalia brasiliensis (Brazilian jack bean)	237
Canavalia ensiformis (Jack bean) (Horse bean)	290
Canavalia gladiata (Sword bean) (Dolichos gladiatus)	290
Canavalia lineata	237
Canavalia rosea (Beach bean) (Canavalia maritima)	236
Canavalia virosa	237

ALINIAREA UNIPROT

The screenshot displays a web browser window with the URL <http://www.ebi.ac.uk/Tools/es/cgi-bin/jobresults.cgi/clustalw2/clustalw2-20101201-100024...>. The browser window contains a sequence alignment tool interface with a menu bar (File, Edit, Select, View, Format, Colour, Calculate, Help) and a sequence alignment view. The alignment shows four protein sequences: P81460/1-237, P81461/1-237, P55915/1-237, and P81364/1-236. The sequences are color-coded by amino acid type. Below the alignment, there are three horizontal tracks: Conservation (a dotted line), Quality (a bar chart), and Consensus (a black bar). The consensus sequence is shown as ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNNMQNGKVGTAHI IYNSV GKRLSAVVS. The sequence ID P14894 is mentioned at the bottom of the window.

File Edit Select View Format Colour Calculate Help

10 20 30 40 50 60

P81460/1-237 ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNNMQNGKVGTAHI IYNSV GKRLSAVVS

P81461/1-237 ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNNMQNGKVGTAHI IYNSV GKRLSAVVS

P55915/1-237 ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNNMQNGKVGTAHI IYNSV GKRLSAVVS

P81364/1-236 ADTIVAVELDTYPNTDVGDPSYPHXXXXXXXXSVRXX-TAKWNNMQNGKVGTAHISYNSV GKRLSAVVS

P02866/1-290

P14894/1-290

Conservation

Quality

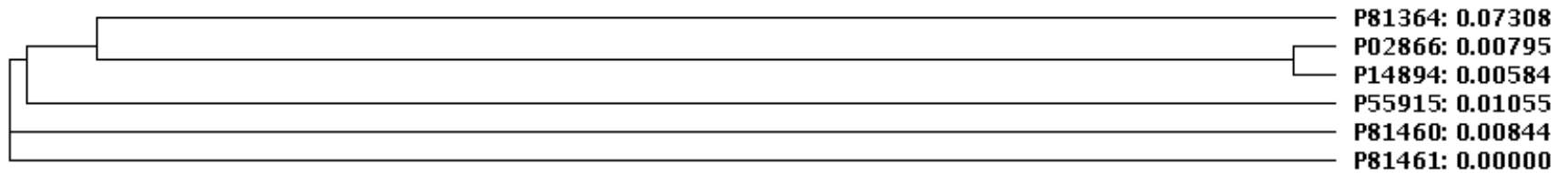
Consensus

ADTIVAVELDTYPNTDIGDPSYPHIGIDIKSVRSKKTAKWNNMQNGKVGTAHI IYNSV GKRLSAVVS

Sequence ID: P14894

ALINIAREA UNIProt

ClustalW tree



PROBLEMA: NUCLEOTIDE

AAV1gp1 nonstructural protein [*Adeno-associated virus - 1*]

Gene ID: 1450375, updated on 12-Mar-2010

Summary

Gene symbol	AAV1gp1
Gene description	nonstructural protein
Locus tag	AAV1gp1
Gene type	protein coding
RefSeq status	PROVISIONAL
Organism	Adeno-associated virus - 1 (old-name: adeno-associated virus type 1)
Lineage	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus

AAV1gp2 capsid protein [*Adeno-associated virus - 1*]

Gene ID: 1450374, updated on 12-Mar-2010

Summary

Gene symbol	AAV1gp2
Gene description	capsid protein
Locus tag	AAV1gp2
Gene type	protein coding
RefSeq status	PROVISIONAL
Organism	Adeno-associated virus - 1 (old-name: adeno-associated virus type 1)
Lineage	Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus

SALVAREA ÎN FORMAT FASTA

AAV1gp1.fasta

```
>gi|9632547:335-2206 Adeno-associated virus - 1, complete genome
ATGCCGGGCTTCTACGAGATCGTGATCAAGGTGCCGAGCGACCTGGACGAGCACCTGCCGGGCATTTCTG
ACTCGTTTGTGAGCTGGGTGGCCGAGAAGGAATGGGAGCTGCCCCGGATTCTGACATGGATCTGAATCT
GATTGAGCAGGCACCCCTGACCGTGGCCGAGAAGCTGCAGCGCGACTTCCTGGTCCAATGGCGCCGCGTG
AGTAAGGCCCCGGAGGCCCTCTTCTTTGTTTCAGTTCGAGAAGGGCGAGTCTACTTCCACCTCCATATTC
TGGTGGAGACCACGGGGGTCAAATCCATGGTGCTGGGCCGCTTCCTGAGTCAGATTAGGGACAAGCTGGT
```

AAV1gp2.fasta

```
>gi|9632547:2223-4433 Adeno-associated virus - 1, complete genome
ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCATTTCGCGAGTGGTGGG
ACTTGAAACCTGGAGCCCCGAAGCCCAAAGCCAACCAGCAAAGCAGGACGACGGCCGGGGTCTGGTGCT
TCCTGGCTACAAGTACCTCGGACCCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGGCGGACGCA
GCGGCCCTCGAGCACGACAAGGCTACGACCAGCAGCTCAAAGCGGGTGACAATCCGTACCTGCGGTATA
ACCACGCCGACGCCGAGTTTCAGGAGCGTCTGCAAGAAGATACGTCTTTTGGGGGCAACCTCGGGCGAGC
AGTCTTCCAGGCCAAGAAGCGGGTTCTCGAACCTCTCGGTCTGGTTGAGGAAGGCGCTAAGACGGCTCCT
```

IMPORTUL FIȘIERULUI FASTA ÎN CLC

Navigation Area

CLC_Data

- gi|9632547:335-2206
- gi|9632547:2223-4433
- Recycle bin (6)

Create Alignment

1. Select sequences of same type

Select sequences of same type

Navigation Area

CLC_Data

- gi|9632547:335-2206
- gi|9632547:2223-4433

Selected Elements (2)

gi 9632547:335-2206
gi 9632547:2223-4...

<enter search term>

? Refresh Previous Next Finish Cancel

ALINIEREA SECVENȚELOR

Create Alignment

1. Select sequences of same type
2. Set parameters

Set parameters

Gap settings

Gap open cost:

Gap extension cost:

End gap cost:

Alignment

Less accurate (fast)

Very accurate (slow)

Redo alignments

Use fixpoints

? ↶ ↷ ✓ ✕

Previous Next Finish Cancel

ALINIAREA SECVENȚELOR

