

## Lab sheet#7

### Protein sequence alignment and ExPASy Tools

**BLASTP** programs search protein databases using a protein query.

- Performing local alignment using **BLASTP**.



**Enter Query Sequence**

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

Query subrange [?](#)

From

To

Or, upload file  no file selected [?](#)

Job Title

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

Align two or more sequences [?](#)

**Standard**

Database  [?](#)

Organism [Optional](#)

exclude [Add organism](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude [Optional](#)

Models (XM/XP)  Non-redundant RefSeq proteins (WP)  Uncultured/environmental sample sequences

**Program Selection**

Algorithm

Quick BLASTP (Accelerated protein-protein BLAST)

**blastp (protein-protein BLAST)**

PSI-BLAST (Position-Specific Iterated BLAST)

PHI-BLAST (Pattern Hit Initiated BLAST)

DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm [?](#)

Search database **nr** using **Blastp (protein-protein BLAST)**

Show results in a new window

**Descriptions** | Graphic Summary | Alignments | Taxonomy

Sequences producing significant alignments Download ▾ Select columns ▾ Show 100 [?](#)

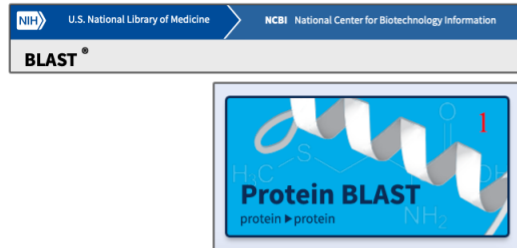
select all 30 sequences selected [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#) [MSA Viewer](#)

Description	Scientific Name	Max Score	Total Score	Query Cover	E value	Per. Ident	Acc. Len	Accession
<input checked="" type="checkbox"/> DNA repair protein XRCC1 isoform 1 [Mus musculus]	Mus musculus	1024	1024	100%	0.0	85.67%	631	NP_033558.3
<input checked="" type="checkbox"/> DNA repair protein XRCC1 isoform X1 [Mus caroli]	Mus caroli	1019	1019	100%	0.0	85.83%	631	XP_021023969.1
<input checked="" type="checkbox"/> DNA repair protein XRCC1 isoform 2 [Mus musculus]	Mus musculus	1018	1018	100%	0.0	85.51%	630	NP_001347097.1

**Multiple Sequence Alignment (MSA)** is a way of arranging three or more biological sequences (protein or nucleic acid) to identify regions of similarity that may be a consequence of functional or structural relationships between the sequences.

**Clustal Omega** is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences.

Multiple protein sequence alignment using **Clustal Omega**



The screenshot shows the NCBI BLAST search interface. Key elements include:
 

- Enter Query Sequence:** A text box containing "NP\_065826.3".
- Database:** A dropdown menu set to "Non-redundant protein sequences (nr)".
- Program Selection:** The "blastp (protein-protein BLAST)" algorithm is selected.
- BLAST button:** A red box highlights the "BLAST" button at the bottom.

Descriptions	Graphic Summary	Alignments	<b>Taxonomy</b>
<b>Reports</b>   Lineage   Organism   Taxonomy			
Pan troglodytes (chimpanzee) [primates]			
<a href="#">endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pan troglodytes]</a>	2096	0.0	<a href="#">XP_016812837</a>
<a href="#">endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pan troglodytes]</a>	2096	0.0	<a href="#">XP_016812901</a>
<a href="#">KIAA1324 isoform 4 [Pan troglodytes]</a>	2096	0.0	<a href="#">PNI54335</a>
<a href="#">endosome/lysosome-associated apoptosis and autophagy regulator 1 [Pan troglodytes]</a>	2090	0.0	<a href="#">NP_001267173</a>
Pongo pygmaeus (Bornean orangutan) [primates]			
<a href="#">endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pongo pygmaeus]</a>	2084	0.0	<a href="#">XP_054348983</a>
<a href="#">endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pongo pygmaeus]</a>	2084	0.0	<a href="#">XP_054356132</a>
<a href="#">endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X2 [Pongo pygmaeus]</a>	2042	0.0	<a href="#">XP_054295098</a>
<input type="text" value="Protein"/> <input type="text" value="NP_065826.3, XP_016812837.1, XP_054348983.1"/> <input type="button" value="Search"/>			
Advanced			

Items: 3

Filters activated: human. Clear all

endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform 1 [Homo sapiens]

1. 1013 aa protein  
 Accession: NP\_065826.3 GI: 1519473601  
 BioProject Nucleotide PubMed Taxonomy  
 GenPept Identical Proteins FASTA Graphics

endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pan troglodytes]

2. 1013 aa protein  
 Accession: XP\_016812837.1 GI: 1034066223  
 BioProject Nucleotide PubMed Taxonomy  
 GenPept Identical Proteins FASTA Graphics

endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pongo pygmaeus]

48983.1 GI: 2463385750  
 BioProject Nucleotide PubMed Taxonomy  
 GenPept Identical Proteins FASTA Graphics

Format  
 Summary  
 GenPept  
 GenPept (full)  
 FASTA  
 FASTA (text)  
 ASN 1  
 Revision History  
 Accession List  
 GI List

Summary - Sort by Default order - Send to: -

```
>NP_065826.3 endosome/lysosome-associated apoptosis and autophagy regulator 1
isoform 1 [Homo sapiens]
MAEPGSHHLSARVRGRTERRIPRLRLLWAGTAFQVQTQGTGPELHACKESYHYEYACDSTGSRWRV
AVPHTPGLCTSLPDPKIGTECSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSAN
MELDDSAEASTGNCTSSKWWPRGDYIASNTDECTATLMYAVNKLQSGTVNFEYYPDSSIIFEPFVQNDQ
CQPNADDSRWMTTEKGWEPHVELNRGNVLYWRITAFVWTKVPKVLRNIAITGVAYTSECFCKP
GTYADKQGSFCKLCPANSYSNKGETSCHQCDPKYSEKSSSCNVRPACTDKDYFHTACDANGETQV
MYKWKPKICSEDELEGAVKLPASGVKTHCPCPNPFFFTNNSTCQPCPYGYSYNGSDCTRCAGTEPAVG
FEYKWWNTLPTNMETTSLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTUVVPGFRPPQSMADTE
NKEVARITFVFETLCTVNCLELYFMVGVNSRNTNPPVETWKGSKGQSYTYIIEENTTTSFTWAFQRTTFHE
ASRKYTNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSCTSCPAGYYIDRDSGTCCHSCPNTILKAH
QPYGVQACVPCPGPKTKNNKIHSLCYNDCTFSRNTPTRTFFNYFNSALANTVTLACGVSFTSKGLKYPHFHT
LSLCCGNQGRKMSVCTDNDVTLRIPEGESGFSKSIYAVVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVT
TDMTLDGITSPAELPHLESIGIPDVIFFRSNDVTSQSSGRSTTIRVRCSPQKTVPGSLLLPCTSDGT
CDGCNFHFLWESAAACPLCSVADYHAIYSSCVAGIQKTTYVWREPKLCGGISLPEQRVITCKTIDFWLK
VGI SAGTCTA ILLTVLTCYFWKKNQKLEYKYSKLVMNATLKDCLPAADSCAIMEGEDVEDDLIFTSKKS
LFGKIKSFTSKRTPDGFDSVPLKTS SGGLDMDL

>XP_016812837.1 endosome/lysosome-associated apoptosis and autophagy regulator 1
isoform X1 [Pan troglodytes]
MAEPGSHHLSARVRGRTERRIPRLRLLWAGTAFQVQTQGTGPELHACKESYHYEYACDSTGSRWRV
AVPHTPGLCTSLPDPKIGTECSFSCNAGEFLDMKDQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSAN
MELDDSAEASTGNCTSSKWWPRGDYIASNTDECTATLMYAVNKLQSGTVNFEYYPDSSIIFEPFVQNDQ
CQPNADDSRWMTTEKGWEPHVELNRGNVLYWRITAFVWTKVPKVLRNIAITGVAYTSECFCKP
GTYADKQGSFCKLCPANSYSNKGETSCHQCDPKYSEKSSSCNVRPACTDKDYFHTACDANGETQV
MYKWKPKICSEDELEGAVKLPASGVKTHCPCPNPFFFTNNSTCQPCPYGYSYNGSDCTRCAGTEPAVG
FEYKWWNTLPTNMETTSLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTUVVPGFRPPQSMADTE
NKEVARITFVFETLCTVNCLELYFMVGVNSRNTNPPVETWKGSKGQSYTYIIEENTTTSFTWAFQRTTFHE
ASRKYTNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSCTSCPAGYYIDRDSGTCCHSCPNTILKAH
QPYGVQACVPCPGPKTKNNKIHSLCYNDCTFSRNTPTRTFFNYFNSALANTVTLACGVSFTSKGLKYPHFHT
LSLCCGNQGRKMSVCTDNDVTLRIPEGESGFSKSIYAVVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVT
TDMTLDGITSPAELPHLESIGIPDVIFFRSNDVTSQSSGRSTTIRVRCNPQKTVPGSLLLPCTSDGT
CDGCNFHFLWESAAACPLCSVADYHAIYSSCVAGIQKTTYVWREPKLCGGISLPEQRVITCKTIDFWLK
VGI SAGTCTA ILLTVLTCYFWKKNQKLEYKYSKLVMNATLKDCLPAADSCAIMEGEDVEDDLIFTSKKS
LFGKIKSFTSKRTPDGFDSVPLKTS SGGPDMDL

>XP_054348983.1 endosome/lysosome-associated apoptosis and autophagy regulator 1
isoform X1 [Pongo pygmaeus]
MAEPGSHHLSARVRGRTERRIPRLRLLWAGTAFQVQTQGTGPELHACKESYHYEYACDSTGSRWRV
AVPHTPGLCTSLPDPKIGTECSFSCNAGEFLDMKNQCKPCAEGRYSLGTGIRFDEWDELPHGFASLSAN
MELDESAAEASTGNCTSSKWWPRGDYIASNTDECTATLMYAVNKLQSGTVNFEYYPDSSIIFEPFVQNDQ
CQPNADDSRWMTTEKGWEPHVELNRGNVLYWRITAFVWTKVSKVLRNIAITGVAYTSECFCKP
GTYADKQGSFCKLCPANSYSNKGETSCHQCDPKYSEKSSSCNVRPACTDKDYFHTACDANGETQV
MYKWKPKICSEDELEGAVKLPASGVKTHCPCPNPFFFTNNSTCQPCPYGYSYNGSDCTRCAGTEPAVG
FEYKWWNTLPTNMETTSLSGINFEYKGMTGWEVAGDHIYTAAGASDNDFMILTUVVPGFRPPQSMADTE
NKEVARITFVFETLCTVNCLELYFMVGVNSRNTNPPVETWKGSKGQSYTYIIEENTTTSFTWAFQRTTFHE
ASRKYTNDVAKIYSINVTNVMNGVASYCRPCALEASDVGSCTSCPAGYYIDRDSGTCCHSCPNTILKAH
QPYGVQACVPCPGPKTKNNKIHSLCYNDCTFSRNTPTRTFFNYFNSALANTVTLACGVSFTSKGLKYPHFHT
LSLCCGNQGRKMSVCTDNDVTLRIPEGESGFSKSIYAVVQAVIIPPEVTGYKAGVSSQPVSLADRLIGVT
TDMTLDGITSPAELPHLESIGIPDVIFFRSNDVTSQSSGRSTTIRVRCSPQKTVPGSLLLPCTSDGT
CDGCNFHFLWESAAACPLCSVADYHAIYSSCVAGIQKTTYVWREPKLCGGISLPEQRVITCKTIDFWLK
VGI SAGTCTA ILLTVLTCYFWKKNQKLEYKYSKLVMNATLKDCLPAADSCAIMEGEDVEDDLIFTSKKS
LFGKIKSFTSKRTPDGFDSVPLKTS SGGPDMDL
```

Clustal Omega

Input form
Web services
Help & Documentation
Bioinformatics Tools FAQ

STEP 1 - Enter your input sequences

Enter or paste a set of  
 PROTEIN

sequences in any supported format:
 

```

                >NP_065826.3 endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform 1 [Homo sapiens]
                MAEPGSHHLSARVRGRTERRIPLRLLWAGTAFQVTQGTGPELHACKESEYHYEYACDSTGSRWRV
                AVPHTPGLCTSLPDIKTECSFSCNAGEFLDMKDQSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSAN
                MELDSSAAESTGNCTSSKWVPRGDIASNTDECTATLMYAVNLKQSGTVNFEYYPDSIIFFVQNDQ
                CQPNADDSRWMTTEKGEWFEHSEVLRGNNVLYWRTTAFSVWTKVPKPLVRNIAITGVAYTSECFFCKP
                GTYADKQSSFFCKLCPANSYSNKGETSCHQCDPPDYKSEKSSSCNVRPACTDKDYFYHTACDANGETQL
                MYKWAKPKICSEDLGAVKLPASGVKTHCPCPNPFFKTNNSTCQPCPYGYSNGSDCTRCAPAGTEPAVG
                FEVGMWNTLDTNMETTULSCINEFVKMTQMEVAGDHYTAAGASNDDEMILTUVBCERDPSVMADE
            
```

1

Or, upload a file: Choose File no file selected
 Use an example sequence | Clear sequence | See more example inputs

STEP 2 - Set your parameters

OUTPUT FORMAT  
 ClustalW with character counts

2

The default settings will fulfill the needs of most users.  
More options... *(Click here, if you want to view or change the default settings.)*

STEP 3 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email)*

Submit

3

Results for job clustalo-l20201226-153502-0424-69399751-p1m

Alignments
Result Summary
Guide Tree
Phylogenetic Tree
Results Viewers
Submission Details

Download Alignment File
Show Colors

4

CLUSTAL 0(1.2.4) multiple sequence alignment
 

```

                NP_001270610.1 MAEPGH-----RLSAFQVTQGTGPELHACKESEYHYEYTA 35
                NP_065826.3 MAEPGSHHLSARVRGRTERRIPLRLLWAGTAFQVTQGTGPELHACKESEYHYEYTA 60
                NP_001267173.1 MAEPGSHHLSARVRGRTERRIPLRLLWAGTAFQVTQGTGPELHACKESEYHYEYTA 60
                ***** :*****
            
```

a

Alignments
Result Summary
Guide Tree
Phylogenetic Tree
Results Viewers
Submission Details

Download Alignment File
Hide Colors

b

CLUSTAL 0(1.2.4) multiple sequence alignment
 

```

                NP_001270610.1 MAEPGH-----RLSAFQVTQGTGPELHACKESEYHYEYTA 35
                NP_065826.3 MAEPGSHHLSARVRGRTERRIPLRLLWAGTAFQVTQGTGPELHACKESEYHYEYTA 60
                NP_001267173.1 MAEPGSHHLSARVRGRTERRIPLRLLWAGTAFQVTQGTGPELHACKESEYHYEYTA 60
                ***** :*****
            
```

**Consensus Symbols:**

- An \* (asterisk) indicates positions which have a single, fully conserved residue.
- A : (colon) indicates conservation between groups of strongly similar properties
- A . (period) indicates conservation between groups of weakly similar properties

What do the **colors** mean when I show them on protein alignments?

This protein-only option colors the residues according to their physicochemical properties:

Residue	Colour	Property
AVFPMILW	RED	Small (small+ hydrophobic (incl.aromatic -Y))
DE	BLUE	Acidic
RK	MAGENTA	Basic - H
STYHCNGQ	GREEN	Hydroxyl + sulfhydryl + amine + G
Others	Grey	Unusual amino/imino acids etc

### ExPasy tools

It is an extensible and integrative portal which provides access to over 160 databases and software tools, developed by SIB Groups and supporting a range of life science and clinical research domains, from genomics, proteomics and structural biology, systems biology and medical chemistry.

### ExPasy translate tool

Translate is a tool which allows the translation of a nucleotide (DNA/RNA) sequence to a protein sequence.

**Open reading frames (ORFs)** are defined as spans of DNA sequence between start and stop codons.



- Protein predicted sequence using **ExPasy translation tool**.

- Protein primary **structure analysis** using **ProtParam** on ExPasy.

**ProtParam** is a tool which allows the computation of various physical and chemical parameters for a given protein stored in Swiss-Prot or TrEMBL or for a user entered protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY)

Advanced

**XRCC1 [Homo sapiens]**

GenBank: CAG33009.1

[GenPept](#) [Identical Proteins](#) [Graphics](#)

>CAG33009.1 XRCC1 [Homo sapiens]

```

MPEIRLRHVVCSQSDSTHCAENLLKADTYRKRRAAKAGKETSIVVLQLEKEEQIHSVDIGNDGSFAFV
LVGSSAGGAGEQDYEVLLVTSFMSPESESRGSPNRRVRFPGDKLVRAAAEKRWDRVKIVCSQPYSKDS
PFGLSFVRFHSPDKDEAEAPSQKVTYTKLQGFVKEEDESANSLRPGALFFSRINKTSPVTASDPAGPS
YAAATLQASSAASASPVSRRAIGSTKPKQESPKGRKLDLNOEKKTPSKPPADLSPSVKRPKLPAPTR
TPATAPVPAQAVTGKPRGEGTEPRRPAQGEELCKLQGVVVLSGFNFRSELRKALELGAKYR
PDWTRDSTHLICAFANTPKYSQVVLGGRIVRKEWLDCHRRRRLPSQRYLMAGPGSSSEDEASHSGG
SGDEAPKLPKQPKTKPTQAAGPSSPOKPPTEETKAASPVLQEDIDIEGVSEGGQNGAEDSGDTE
ELRRVAEQEHLRPPGQENGEDPYAGSDENTDSEEHQEPDLPVPELPOFFQGHFFLYGEFPGDERR
KLIRYVTAQFNGELEDYMSDRVQFVITAQEWDPSEFEEALMNPFLAFVRRWVWYSCNEKQLLPHQLYGVV
PQA
                    
```

Please note that you may only fill out **one** of the following fields at a time.

Enter a Swiss-Prot/TrEMBL accession number (AC) (for example **P05130**) or a sequence identifier (ID) (for example **KPC1\_DROME**):

Or you can paste your own amino acid sequence (in one-letter code) in the box below:

SGDEAPKLPKQPKTKPTQAAGPSSPOKPPTEETKAASPVLQEDIDIEGVSEGGQNGAEDSGDTE
 ELRRVAEQEHLRPPGQENGEDPYAGSDENTDSEEHQEPDLPVPELPOFFQGHFFLYGEFPGDERR
 KLIRYVTAQFNGELEDYMSDRVQFVITAQEWDPSEFEEALMNPFLAFVRRWVWYSCNEKQLLPHQLYGVV
 PQA

**ProtParam**

User-provided sequence: **a**

```

10      20      30      40      50      60
MPEIRLRHVV SCSSQSDSTHCAENLLKADTYRKRRAAKAGE KTISVVLQLE KEEQIHSVDI
70      80      90     100     110     120
GNDGSFAFVE LVGSSAGGAGE QDYEVLLVTSFMSPESESR GSGSNRRVRFH FPGDKLVRAA
130     140     150     160     170     180
AEKRWDRVKI VCSQPYSKDS PFGLSFVRFH SPDKDEAEA PSQKVTYTKL GQFRVKEEDE
190     200     210     220     230     240
SANSLRPGAL FFSRINKTSP VTASDPAGPS YAAATLQASS AASSASPVSR AIGSTSKPQE
250     260     270     280     290     300
SPKGRKLDL NQEEKTKPSK PPAQLSPSPV KRPKLPAPTR TPATAPVPAQ AQGAVTKKPR
310     320     330     340     350     360
GEGTEPRRPR AGPEELGKIL QGVVVLVSGF QNFRSELRD KALELGAKYR PDWTRDSTHL
370     380     390     400     410     420
ICAFANTPKY SQVVLGGRIVRKEWLDCH RRRRRLPSQR YLMAGPGSSS EDEASHSGG
430     440     450     460     470     480
SGDEAPKLPQ KQPKTKPT QAAGPSSPOK PPTPEETKAA SPVLQEDIDI EGVSEGGQNG
490     500     510     520     530     540
GAEDSGDTE ELRRVAEQE HRLPPGQEN GEDPYAGSD ENTDSEEHQE PPDLPVPELP
550     560     570     580     590     600
DFFQGHFFL YGEFPGDERR KLIRYVTAQ FNGELEDYMS DRVQFVITAQ EWDPSEFEEALMD
610     620     630
NPSLAFVRRP WIYSCNEKQK LLPHQLYGVV PQA
                    
```

**Number of amino acids:** 633 **b**

**Molecular weight:** 69497.53

**Theoretical pI:** 5.93

**Amino acid composition:**

Ala (A)	52	8.2%
Arg (R)	43	6.8%
Asn (N)	15	2.4%
Asp (D)	35	5.5%
Cys (C)	6	0.9%
Gln (Q)	33	5.2%
Glu (E)	56	8.8%
Gly (G)	45	7.1%

**Total number of negatively charged residues (Asp + Glu):** 91

**Total number of positively charged residues (Arg + Lys):** 83

**Atomic composition:** **c**

Carbon	C	3036
Hydrogen	H	4792
Nitrogen	N	878
Oxygen	O	968
Sulfur	S	13

**Formula:** C<sub>3036</sub>H<sub>4792</sub>N<sub>878</sub>O<sub>968</sub>S<sub>13</sub>

**Total number of atoms:** 9687

**Extinction coefficients:** **d**

Extinction coefficients are in units of M<sup>-1</sup> cm<sup>-1</sup>, at 280 nm measured in water.

Ext. coefficient 52745  
Abs 0.1% (=1 g/l) 0.759, assuming all pairs of Cys residues form cystines

Ext. coefficient 52370  
Abs 0.1% (=1 g/l) 0.754, assuming all Cys residues are reduced

**Estimated half-life:**

The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).  
>20 hours (yeast, in vivo).  
>10 hours (Escherichia coli, in vivo).

**Instability index:**

The instability index (II) is computed to be 63.35  
This classifies the protein as unstable.

**Aliphatic index:** 62.89

**Grand average of hydropathicity (GRAVY):** -0.780

Prepared by: Imtinan Alkhudair 2023