

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)

Query subrange

Or, upload file

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Standard

Database

Organism exclude

Optional
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown

Exclude 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

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**

Pan troglodytes (chimpanzee) [primates]			
endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pan troglodytes]	2096	0.0	XP_016812837
endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pan troglodytes]	2096	0.0	XP_016812901
KIAA1324 isoform 4 [Pan troglodytes]	2096	0.0	PN154335
endosome/lysosome-associated apoptosis and autophagy regulator 1 [Pan troglodytes]	2090	0.0	NP_001267173

Pongo pygmaeus (Bornean orangutan) [primates]			
endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pongo pygmaeus]	2084	0.0	XP_054348983
endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X1 [Pongo pygmaeus]	2084	0.0	XP_054356132
endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform X2 [Pongo pygmaeus]	2042	0.0	XP_054295098

Protein: NP_065826.3, XP_016812837.1, XP_054348983.1

Items: 3

Filters activated: human. [Clear all](#)

[endosome/lysosome-associated apoptosis and autophagy regulator 1 isoform 1 \[Homo sapiens\]](#)

- 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
[ide](#) [Taxonomy](#)

Format

Summary
 GenPept
 GenPept (full)
 FASTA
 FASTA (text)

ASN.1
 Annotation History
 Accession List
 GI List

[Proteins](#) [FASTA](#) [Graphics](#)

48983.1 GI: 2463385750
[ide](#) [Taxonomy](#)

[Proteins](#) [FASTA](#) [Graphics](#)

Send to: ▾

Summary ▾ Sort by Default order ▾

```
>NP_065826.3 endosome/lysosome-associated apoptosis and autophagy regulator 1
isoform 1 [Homo sapiens]
MAEPGHSHHLSARVRGRTERRIPRLRLLWAGTAQVTQGTGPELHACKESEYHETYACDSTGSRWRV
AVPHTPGCLTSPLDPVKGTECSFSNCAGEFLDMKQDSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSAN
MELDDSAESTGCNTSSKWVPRGDYIASNTDECATLMYAVNLKQSGTVNFYEEYYDSSIIEFFVQNDQ
CQPNAADSRWMKTTKEGWEFHSVELNRGNNVLYWRTTAFSVWTKVPKPVLVRNIAITGVAYTSECFCPKP
GTYADKQGSFCKLCPANSYSNKGETSCHQCDPDYSEKGSSSCNVRPACTKDVFYTHTACDANEGETQL
MYWKAKPKICSEDLEGAVKLPAASGVKTHCPCPNGFFKTNNSTCOPCPYGSYSNGSDCTRCPAGTEPAVG
FEYKWWNTLPTNMETTTLVSGINFEYKGMGTGEVAGDHITYTAAGASNDMFILTLVVPGRFPQSVVMADTE
NKEVARITFVFEITLCTVNCELYFMVGVNSRTNTPVETWKGSKGKQSYTYIIIEENTTSTWAFQRTTFHE
ASRKYTNDVAKIYSINVNTVMNGVASYCRPCALEASDVGSCTSCPAGYIIDRDSTCHSCPTNTILKAH
QFYGVQACVPCGPTKNNKIHSCLYNDCTSFSRNTTRTFNYNFSALANTVTLAGGPFSTSKGKLYFHHFT
LSLCGNQGRKMSVCTDNVTDLRIPEGESGSKSITAIVCQAVIIPFEMTGYKAGVSSQPVSLADRIGVVT
TDMTLDGITSPAELPHPESLGPIDPVIFFYRSNDVTQSCSSGRSTTIRRCSPQKTPVGSLLLPGTCSDGT
CDCGNFHFLWESAAACPLCSVADYHAIVSCVAGIQKTTTVWREPKLCSGGISLPEQRVTICKTIDFWLK
VGISAGTCTAILLTWLTCYFWKKNQKLEYKYSKLVNMNATLKDCLPAADSCAIMEGEDVEDDLIFTSKKS
LFGKIKSFTSKRTPDGFDSVPLKTSSGGPMDML
```

```
>XP_016812837.1 endosome/lysosome-associated apoptosis and autophagy regulator 1
isoform X1 [Pan troglodytes]
MAEPGHSHHLSARVRGRTERRIPRLRLLWAGTAQVTQGTGPELHACKESEYHETYACDSTGSRWRV
AVPHTPGCLTSPLDPVKGTECSFSNCAGEFLDMKQDSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSAN
MELDDSAESTGCNTSSKWVPRGDYIASNTDECATLMYAVNLKQSGTVNFYEEYYDSSIIEFFVQNDQ
CQPNAADSRWMKTTKEGWEFHSVELNRGNNVLYWRTTAFSVWTKVPKPVLVRNIAITGVAYTSECFCPKP
GTYADKQGSFCKLCPANSYSNKGETSCHQCDPDYSEKGSSSCNVRPACTKDVFYTHTACDANEGETQL
MYWKAKPKICSEDLEGAVKLPAASGVKTHCPCPNGFFKTNNSTCOPCPYGSYSNGSDCTRCPAGTEPAVG
FEYKWWNTLPTNMETTTLVSGINFEYKGMGTGEVAGDHITYTAAGASNDMFILTLVVPGRFPQSVVMADTE
NKEVARITFVFEITLCTVNCELYFMVGVNSRTNTPVETWKGSKGKQSYTYIIIEENTTSTWAFQRTTFHE
ASRKYTNDVAKIYSINVNTVMNGVASYCRPCALEASDVGSCTSCPAGYIIDRDSTCHSCPSNTILKAH
QFYGVQACVPCGPTKNNKIHSCLYNDCTSFSRNTTRTFNYNFSALANTVTLAGGPFSTSKGKLYFHHFT
LSLCGNQGRKMSVCTDNVTDLRIPEGESGSKSITAIVCQAVIIPFEMTGYKAGVSSQPVSLADRIGVVT
TDMTLDGITSPAELPHPESLGPIDPVIFFYRSNDVTQSCSSGRSTTIRRCSPQKTPVGSLLLPGTCSDGT
CDCGNFHFLWESAAACPLCSVADYHAIVSCVAGIQKTTTVWREPKLCSGGISLPEQRVTICKTIDFWLK
VGISAGTCTAILLTWLTCYFWKKNQKLEYKYSKLVNMNATLKDCLPAADSCAIMEGEDVEDDLIFTSKKS
LFGKIKSFTSKRTPDGFDSVPLKTSSGGPMDML
```

```
>XP_054348983.1 endosome/lysosome-associated apoptosis and autophagy regulator 1
isoform X1 [Pongo pygmaeus]
MAEPGHSHHLSARGRGRTERRIPRLRLLWAGTAQVTQGTGPELHACKESEYHETYACDSTGSRWRV
AVPHTPGCLTSPLDPVKGTECSFSNCAGEFLDMKQDSCKPCAEGRYSLGTGIRFDEWDELPHGFASLSAN
MELDDSAESTGCNTSSKWVPGDYIASNTDECATLMYAVNLKQSGTVNFYEEYYDSSIIEFFVQNDQ
CQPNAADSRWMKTTKEGWEFHSVELNRGNNVLYWRTTAFSVWTKVPKPVLVRNIAITGVAYTSECFCPKP
GTYADKQGSFCKLCPANSYSNKGETSCHQCDPDYSEKGSSSCNVRPACTKDVFYTHTACDANEGETQL
MYWKAKPKICSEDLEGAVKLPAASGVKTHCPCPNGFFKTNNSTCOPCPYGSYSNGSDCTRCPAGTEPAVG
FEYKWWNTLPTNMETTTLVSGINFEYKGMGTGEVAGDHITYTAAGASNDMFILTLVVPGRFPQSVVMADTE
NKEVARITFVFEITLCTVNCELYFMVGVNSRTNTPVETWKGSKGKQSYTYIIIEENTTSTWAFQRTTFHE
ASRKYTNDVAKIYSINVNTVMNGVASYCRPCALEASDVGSCTSCPAGYIIDRDSTCHSCPTNTILKAH
QFYGVQACVPCGPTKNNKIHSCLYNDCTSFSRNTTRTFNYNFSALANTVTLAGGPFSTSKGKLYFHHFT
LSLCGNQGRKMSVCTDNVTDLRIPEGESGSKSITAIVCQAVIIPFEMTGYKAGVSSQPVSLADRIGVVT
TDMTLDGITSPAELPHPESLGPIDPVIFFYRSNDVTQSCSSGRSTTIRRCSPQKTPVGSLLLPGTCSDGT
CDCGNFHFLWESAAACPLCSVADYHAIVSCVAGIQKTTTVWREPKLCSGGISLPEQRVTICKTIDFWLK
VGISAGTCTAILLTWLTCYFWKKNQKLEYKYSKLVNMNATLKDCLPAADSCAIMEGEDVEDDLIFTSKKS
LFGKIKSFTSKRTPDGFDSVPLKTSSGGPMDML
```

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]
MAEPGHSHHLSARVRGRTERRIPRLWRLLLWAGTAFQVTQGTGPELHACKESEYHYEYTA
AVPHTPGLCTSLLDPKIGTECSFCNAGEFLDMKDQSCKPCAEGRYSLGTRIFDEWDELPHGFASLSAN
MELEDDSAESTGNCTSSKWVPRGDYIASNTDECTALMYAVNLQSGTVNFEEYYPDSSIIFFFVQNDQ
COPNADDSSRWMKTTKGWEFHNSVELNRGNNVLYWRTTAFSVWTKVPKPVLVRNIAITGVAYTSECFCPKP
GTYADKQGSSFCKLCPANSYSNKGETSCHQCDPPDKYSEKGGSSCNVRPACTDKDVFYTHIACDANGETQL
MYKWAQPKICSEDLEGAVKLPLASGVKTHCPCPNPGFFKTNNSTCQPCPYGSYNGSDCTRCPAGTEPAVG
CEVYIMMINLPTNMETTVLCOINEEVVOMTQWEVACDQHIVTAACACDNDMILTVRCERPPDSVMDTE
```

Or, upload a file: Choose File no file selected

Use a example sequence | Clear sequence | See more example inputs

STEP 2 - Set your parameters

OUTPUT FORMAT
ClustalW with character counts

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)

Submit

a

b

c

d

e

f

g

h

i

j

k

l

m

n

o

p

q

r

s

t

u

v

w

x

y

z

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

kk

ll

mm

nn

oo

pp

qq

rr

ss

tt

uu

vv

ww

xx

yy

zz

aa

bb

cc

dd

ee

ff

gg

hh

ii

jj

<

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 + sulphydryl + 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.

The screenshot shows the ExPasy translation tool interface. At the top, there's a search bar with "Nucleotide" and "Search" buttons. Below it, the nucleotide sequence NM_006297.3 is shown, along with "FASTA" and "Graphics" options. The main area displays the protein sequence for "Homo sapiens X-ray repair cross complementing 1 (XRCC1), mRNA". The sequence starts with a methionine (M) and ends with a stop codon (TAA). The tool also includes options for translating reading frames, selecting strands, and choosing genetic codes.

The screenshot shows the ProtParam tool interface. It displays the protein sequence for "Homo sapiens X-ray repair cross complementing 1 (XRCC1)" with a start codon (M) and a stop codon (TAA). The sequence is presented in two frames: "5' Frame 1" and "5' Frame 2". The sequence is annotated with various parameters and features, such as hydrophobicity plots and secondary structure predictions.

- 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)

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:

```
SQGS8APKLPOKQPTKTPQAAGPSPDKPKTPPEETKAASPLVQLEDIEDEQVSEGDQQN
GAEDSGOTED
ELRVAQAEHLRPLQGQEENEDPAGYSTDNTDSEENQEPDPLVPELPDPFGQHIFL
YEVSEGDQQN
KLIRYVTAFNGELEDYMSDRVQFVITAEWOPSFEEALMDNPSPSLAFVRPRWYSCNEOK
LLPHOLYGVV
PQA
```

Number of amino acids: 633 b

Molecular weight: 69497.53

Theoretical pI: 5.93

Amino acid composition: [CSV format](#)

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: C₃₀₃₆H₄₇₉₂N₈₇₈O₉₆₈S₁₃
Total number of atoms: 9687

Extinction coefficients: d

Extinction coefficients are in units of M⁻¹ cm⁻¹, 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