

Základy praktické Bioinformatiky

PETRA MATOUŠKOVÁ

2023/2024

10/10

Základy praktické bioinformatiky

Téma 10/10

Opakování

Cíle:

Student bude dobře připraven na zkoušku 😊

Zkoušková písemka:

2 úlohy, každá 5 bodů

Hodnocení: max celkem 10 bodů

9-10b: 1

7-8b: 2

5-6b: 3

Verze 4.5.2021

A: Stáhněte si sekvenci NM_002085.5

- 1) Co tato sekvence kóduje?
- 2) Nalezněte isoformy příslušného enzymu a porovnejte je.
- 3) Navrhňte primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy BamHI a EcoR.I
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, zapište mutaci této AMK na alanin.

B: Pracujte s následující sekvencí peptidu:

```
EKHSGGFI FPPSYV FVVMSELT DQMT FMERVK NMIYVLY FDFWFE I FDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLI RNSWNFQFPY PLLPNVDFV GGLHCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRFDGNKPDITLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI
```

- 1) Identifikujte příslušný protein, zapište přístupový kód referenční sekvence.
- 2) Jaká je molekulová hmotnost tohoto peptidu?
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky?
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

Zkouškové otázky 2023


A) Pracujte s touto neznámou sekvencí:

```
181 elafqtiqei rhlkemyrln lsqdeikqev edylpsfckw agdfmhknts tdfpqmqsls  
241 psdnskdnst cnievvpmd ieesiwsprf glkgkidvtv gvkihrgykt kykimplelk  
301 tgkesnsieh rsqvvlytll sqerradpea gllylktgq mypvpanhld krellklrnq  
361 mafslfhris ksatrqtql aslpqiiee ktckycsqig ncalysrave qqmdcssvpi  
421 vmlpkieeet qhlkqthley fslwclmltl esqskdnkn hqniwlmpas emeksgscig
```

1. Sekvenci přepište do fasta formátu.
2. Zjistěte, co kóduje a z jakého organismu pochází.
3. Zjistěte, jaký bude nejdelší fragment po štěpení celého identifikovaného proteinu proteinázou kaspáza 1.
4. Jaká je 20 aminokyselina celého proteinu?
5. Existuje nějaký lidský homolog tohoto proteinu?

A) Pracujte s touto neznámou sekvencí:

```
181 elafqtiqei rhlkemyrln lsqdeikqev edylpsfckw agdfmhknts tdfpqmqlsl
241 psdnskdnst cnievvkpmd ieesiwsprf glkgkidvtv gvkihrgykt kykimplelk
301 tgkesnsieh rsqvvlytll sqerradpea gllylktgq mypvpanhld krellklrnq
361 mafslfhris ksatrqtql aslpqiiee ktckycsqig ncalysrave qqmdcssvpi
421 vmlpkieeet qhlkqthley fslwclmltl esqskdnkn hqniwlmpas emeksgscig
```

1. Sekvenci přepište do fasta formátu.  Filter Protein (SMS)
2. Zjistěte, co kóduje a z jakého organismu pochází.
3. Zjistěte, jaký bude nejdelší fragment po štěpení celého identifikovaného proteinu proteinázou kaspáza 1.
4. Jaká je 20 aminokyselina celého proteinu?
5. Existuje nějaký lidský homolog tohoto proteinu?



Sequence Manipulation Suite:

Filter Protein

Filter Protein removes non-protein characters from text. Use this program when you wish to remove digits and blank spaces from

Format Conversion

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Split FASTA
- Three to One
- Window Extractor DNA
- Window Extractor Protein

Sequence Analysis

- Codon Plot
- Codon Usage
- CpG Islands
- DNA Molecular Weight
- DNA Pattern Find
- DNA Stats
- Fuzzy Search DNA
- Fuzzy Search Protein
- Ident and Sim
- Multi Rev Trans
- Mutate for Digest
- ORF Finder
- Pairwise Align Codons
- Pairwise Align DNA
- Pairwise Align Protein
- PCR Primer Stats
- PCR Products
- Protein GRAVY
- Protein Isoelectric Point
- Protein Molecular Weight
- Protein Pattern Find

Paste the text into the text area below. Input limit is 500,000,000 characters.

```
361 mafslfhris ksatrqktql aslpqiiee ktckycsqig ncalysrave
qqmdcssvpi
421 vmlpkieeet qhlkathley fslwclmltl esgskdnkkn hqniwlmpas
emeksgscig
```

Submit Clear Reset

- remove non 'ACDEFGHIKLMNPQRSTVWY' characters
- replace removed characters with nothing
- don't change the case of remaining characters

*This page requires JavaScript. See [this page](#) for more information.
*You can mirror this page or use it offline.

Sun 14 Jun 00:36:59 2020
Valid XHTML 1.0; Valid CSS

Sequence Manipulation Suite - Pracovní - Microsoft Edge

about:blank

Filter Protein results

```
>filtered protein sequence consisting of 300 residues.
elafqtiqeirhlkemyrlnlsqdeikqevedylopsfckwagdfmhkntstfdpqqmqlsl
psdnskdnstcnievvkpmdieesiwsprfglkgkidvtvgvkihrgykkykimplelk
tgkesnsiehrrsqvvlytllsqerradpeagllylktgqmyvpvpanhldkrellklrnq
mafslfhrisksatrqktqlaslpqiieeektckycsqigncalysraveqqmdcssvpi
vmlpkieeetqhlkqthleyfslwclmltlesgskdnkknhqniwlmpasemeksgscig
```

A) Pracujte s touto neznámou sekvencí:

```
181 elafqtiqei rhlkemyrln lsqdeikqev edylpsfckw agdfmhknts tdfpqmqsls  
241 psdnskdnst cnievvkpmd ieesiwsprf glkgkidvtv gvkihrgykt kykimplelk  
301 tgkesnsieh rsqvvlytll sqerradpea gllylktgq mypvpanhld krellklrnq  
361 mafslfhris ksatrqtql aslpqiiee ktckycsqig ncalysrave qqmdcssvpi  
421 vmlpkieeet qhlkqthley fslwclmltl esqskdnkn hqniwlmpas emeksgscig
```


1. Sekvenci přepište do fasta formátu.
2. Zjistěte, co kóduje a z jakého organismu pochází. → BLASTp
3. Zjistěte, jaký bude nejdelší fragment po štěpení celého identifikovaného proteinu proteinázou kaspáza 1.
4. Jaká je 20 aminokyselina celého proteinu?
5. Existuje nějaký lidský homolog tohoto proteinu?

[← Edit Search](#)[Save Search](#)[Search Summary ▾](#)[? How to read this report?](#)[▶ BLAST Help Videos](#)[↶ Back to Traditional Results Page](#)**Job Title** filtered protein sequence consisting of 300...**RID** [2D24X59C013](#) Search expires on 04-23 23:56 pm [Download All ▾](#)**Program** BLASTP [?](#) [Citation ▾](#)**Database** **nr** [See details ▾](#)**Query ID** lc|Query_6258989**Description** filtered protein sequence consisting of 300 residues.**Molecule type** amino acid**Query Length** 300**Other reports** [Distance tree of results](#) [Multiple alignment](#) [MSA viewer](#) [?](#)**Filter Results****Organism** only top 20 will appear exclude[+ Add organism](#)**Percent Identity** to **E value** to **Query Coverage** to [Filter](#)[Reset](#)Compare these results against the new Clustered nr database [?](#)[BLAST](#)**Descriptions**[Graphic Summary](#)[Alignments](#)[Taxonomy](#)**Sequences producing significant alignments**[Download ▾](#)[Select columns ▾](#)Show [?](#) select all 100 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 replication helicase/nuclease 2 [Homo sapiens]	Homo sapiens	629	629	100%	0.0	100.00%	867	KAI2555973.1
<input checked="" type="checkbox"/>	DNA replication helicase/nuclease 2 [Homo sapiens]	Homo sapiens	629	629	100%	0.0	100.00%	867	KAI4076159.1
<input checked="" type="checkbox"/>	DNA replication ATP-dependent helicase/nuclease DNA2 isoform X1 [Homo sapiens]	Homo sapiens	629	629	100%	0.0	100.00%	1090	XP_006717743.1
<input checked="" type="checkbox"/>	DNA2 protein [Homo sapiens]	Homo sapiens	628	628	100%	0.0	100.00%	829	AAH63664.1
<input checked="" type="checkbox"/>	DNA2 protein [Homo sapiens]	Homo sapiens	628	628	100%	0.0	100.00%	687	AAI11741.1

A) Pracujte s touto neznámou sekvencí:

```
181 elafqtiqei rhlkemyrln lsqdeikqev edylpsfckw agdfmhknts tdfpqmqlsl
241 psdnskdnst cnievvpmd ieesiwsprf glkgkidvtv gvkihrgykt kykimplelk
301 tgkesnsieh rsqvvlytll sqerradpea gllylylktgq mypvpanhld krellklrnq
361 mafslfhris ksatrqktql aslpqiiee ktckycsqig ncalysrave qqmdcssvpi
421 vmlpkieeet qhlkqthley fslwclmltl esqskdnkn hqniwlmpas emeksgscig
```

1. Sekvenci přepište do fasta formátu.
2. Zjistěte, co kóduje a z jakého organismu pochází.
3. Zjistěte, jaký bude nejdelší fragment po štěpení celého identifikovaného proteinu proteinázou kaspáza 1.  PeptideCutter (+zaškrtnout navíc tabulku)
4. Jaká je 20 aminokyselina celého proteinu?
5. Existuje nějaký lidský homolog tohoto proteinu?

PeptideCutter

```
eiagtqieirhikemyrlnisqdeikqevdyips+ckwagdfmhkntsttdp  
mg151  
psdnskdnstcnievkvpmdieesiwspfrfalkgkidvtvgvkihrgyktykim  
pleik  
tgesnsiehrsqvlytllsqerradpeag1llylktgqmpvpanhidkrell  
klrnq  
mafslfhrisksatraktalaslogiieeektckycsaigncalysraveqmdc  
ssvp1  
vm1pkieeetqhlkathleyfslwclmltlesqskdnknhniwmpasemek  
gscig
```

the cleavage of the protein. the fields.

Please, select

- all available enzymes and chemicals
 only the following selection of **enzymes and chemicals**

- | | |
|--|---|
| <input type="checkbox"/> Arg-C proteinase | <input type="checkbox"/> Asp-N endopeptidase |
| <input type="checkbox"/> BNPS-Skatole | <input checked="" type="checkbox"/> Caspase 1 |
| <input type="checkbox"/> Caspase 3 | <input type="checkbox"/> Caspase 4 |
| <input type="checkbox"/> Caspase 6 | <input type="checkbox"/> Caspase 7 |
| <input type="checkbox"/> Caspase 9 | <input type="checkbox"/> Caspase 10 |
| <input type="checkbox"/> Chymotrypsin-high specificity (C-term to [FYW], not before P) | <input type="checkbox"/> Chymotrypsin-low specificity (C-term to [FYWML], n |
| <input type="checkbox"/> Clostripain (Clostridiopeptidase B) | <input type="checkbox"/> CNBr |
| <input type="checkbox"/> Factor Xa | <input type="checkbox"/> Formic acid |
| <input type="checkbox"/> Granzyme B | <input type="checkbox"/> Hydroxylamine |
| <input type="checkbox"/> LysC | <input type="checkbox"/> LysN |
| <input type="checkbox"/> Neutrophil elastase | |
| <input type="checkbox"/> Pepsin (pH1.3) | <input type="checkbox"/> Pepsin (pH>2) |
| <input type="checkbox"/> Proteinase K | <input type="checkbox"/> Staphylococcal peptidase I |
| <input type="checkbox"/> Thermolysin | <input type="checkbox"/> Thrombin |

- for the following enzymes an additional, more **sophisticated model** can be applied that attributes a probability of cleavage to each site :

Chymotrypsin
Trypsin

Please enter the lowest cleavage probability that you would like to be displayed: %

Please indicate the way you would like the cleavage sites to be displayed

- Map of cleavage sites. Please select the number of amino acid within one block:
- Table of sites, sorted alphabetically by enzyme and chemical name
- Table of sites, sorted sequentially by amino acid number

Expasy

PeptideCutter

PeptideCutter

The sequence to investigate:

```
10 20 30 40 50 60  
ELAFQTIQEI RHLKEMYRLN LSQDEIKQEV EDYLPSCFKI AGDFMHNKTS TDFPQMLLSL  
70 80 90 100 110 120  
PSDNSKDNST CNIEVVKPYD IEESIWSPRF GLKGIKIDTV GVKIHRGYKT KYKIMPLELK  
130 140 150 160 170 180  
TGKESNSIEH RSQVLYTLL SQERRADPEA GLLLYLKTGQ MYPVPANHLD KRELLKLRNQ  
190 200 210 220 230 240  
MAFSLFHRIS KSATRQKTQL ASLPQIIIEE KTCKYCSQIG NCALYSRAVE QQMDCSSVPI  
250 260 270 280 290 300  
VMLPKIEEET QHLKQTHLEY FSLWCLMLTL ESQSKDNKNI HQNIWLMMPAS EMEKSGSCTIG
```

The sequence is 300 amino acids long.

Available enzymes

The enzyme(s) that you have chosen:

- Caspase 1

You have chosen to display all possible cleaving enzymes.

None of the enzymes cut.

Špatně – vložen jen neznámý peptid

PeptideCutter

```
eiagtqiearhikemyrlnisqdeikqvedyips+ckwagdfmhnkststfpg  
mg1s1  
psdnskdnstcnievkvpmdieesiwspfrfalkgkidvtvgkikhrgytkykim  
pleik  
fgkesnsiehrsqvlytllsqerradpeaglllylktgqmpvpanhdkrell  
klrnq  
mafslfhrisksatraktalaslogiieeektckycsaigncalysraveagmdc  
ssvp1  
vmipkieetqhlkathleyfslwclmlflesqskdnknhqnihwmpasemeks  
gscig
```

the cleavage of the protein. the fields.

Please, select

- all available enzymes and chemicals
 only the following selection of **enzymes and chemicals**

- | | |
|--|---|
| <input type="checkbox"/> Arg-C proteinase | <input type="checkbox"/> Asp-N endopeptidase |
| <input type="checkbox"/> BNPS-Skatole | <input checked="" type="checkbox"/> Caspase1 |
| <input type="checkbox"/> Caspase3 | <input type="checkbox"/> Caspase4 |
| <input type="checkbox"/> Caspase6 | <input type="checkbox"/> Caspase7 |
| <input type="checkbox"/> Caspase9 | <input type="checkbox"/> Caspase10 |
| <input type="checkbox"/> Chymotrypsin-high specificity (C-term to [FYW], not before P) | <input type="checkbox"/> Chymotrypsin-low specificity (C-term to [FYWML], not |
| <input type="checkbox"/> Clostripain (Clostridiopeptidase B) | <input type="checkbox"/> CNBr |
| <input type="checkbox"/> Factor Xa | <input type="checkbox"/> Formic acid |
| <input type="checkbox"/> GranzymeB | <input type="checkbox"/> Hydroxylamine |
| <input type="checkbox"/> LysC | <input type="checkbox"/> LysN |
| <input type="checkbox"/> Neutrophil elastase | |
| <input type="checkbox"/> Pepsin (pH1.3) | <input type="checkbox"/> Pepsin (pH>2) |
| <input type="checkbox"/> Proteinase K | <input type="checkbox"/> Staphylococcal peptidase I |
| <input type="checkbox"/> Thermolysin | <input type="checkbox"/> Thrombin |

- for the following enzymes an additional, more **sophisticated model** can be applied that attributes a probability of cleavage to

Chymotrypsin
Trypsin

Please enter the lowest cleavage probability that you would like to be displayed: %

Please indicate the way you would like the cleavage sites to be displayed

- Map of cleavage sites. Please select the number of amino acid within one block:
- Table of sites, sorted alphabetically by enzyme and chemical name
- Table of sites, sorted sequentially by amino acid number

```
          97          98          99          100          101          102  
PAPEQVEKGG VSNVTEAKLI VFLTSIFVKA GCSPSDIGII APYRQLKII NDLLARSIGM  
VEVNTVDKYQ GRDKSIVLVS FVRSNKDGTV GELLKDWRL NVAITRAKHK LILLGCVPSL  
  
          103          104          105          106  
NCYPPLEKLL NHLNSEKLII DLPSREHESL CHILGDFQRE
```

The sequence is 1060 amino acids long.

Available enzymes

The enzyme(s) that you have chosen:

- Caspase1

You have chosen to display all possible cleaving enzymes.

These enzymes cleave the sequence:

Name of enzyme	No. of cleavages	Positions of cleavage sites
Caspase1	2	614 875


At these positions the following enzymes cleave:

- Please note that the size of the peptides are calculated as if **all chosen enzymes** were used.
- Please be aware of the fact that the present version of the PeptideCutter program do not calculate the mass of the peptides occurring amino acid residues, and giving peptide masses as [M]. If you want to select

Position of cleavage site	Name of cleaving enzyme(s)	Peptide length [aa]	Peptide mass [Da]
614	Caspase1	614	70406.081
875	Caspase1	261	29334.597
1060	end of sequence	185	20710.061

A) Pracujte s touto neznámou sekvencí:

```
181 elafqtiqei rhlkemyrln lsqdeikqev edylpsfckw agdfmhknts tdfpqmqsls  
241 psdnskdnst cnievvpmd ieesiwsprf glkgkidvtv gvkihrgykt kykimplelk  
301 tgkesnsieh rsqvvlytll sqerradpea gllylktgq mypvpanhld krellklrnq  
361 mafslfhris ksatrqtql aslpqiiee ktckycsqig ncalysrave qqmdcssvpi  
421 vmlpkieeet qhlkqthley fslwclmltl esqskdnkn hqniwlmpas emeksgscig
```

1. Sekvenci přepište do fasta formátu.
2. Zjistěte, co kóduje a z jakého organismu pochází.
3. Zjistěte, jaký bude nejdelší fragment po štěpení celého identifikovaného proteinu proteinázou kaspáza 1.
4. Jaká je 20 aminokyselina celého proteinu?  Protein Range Extractor (SMS)
5. Existuje nějaký lidský homolog tohoto proteinu?



Format Conversion

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Split FASTA
- Three to One
- Window Extractor DNA
- Window Extractor Protein

Sequence Analysis

- Codon Plot
- Codon Usage
- CpG Islands
- DNA Molecular Weight
- DNA Pattern Find
- DNA Stats
- Fuzzy Search DNA
- Fuzzy Search Protein
- Ident and Sim
- Multi Rev Trans
- Mutate for Digest
- ORF Finder
- Pairwise Align Codons
- Pairwise Align DNA
- Pairwise Align Protein
- PCR Primer Stats
- PCR Products
- Protein GRAVY
- Protein Isoelectric Point

Sequence Manipulation Suite:

Range Extractor Protein

Range Extractor Protein accepts a protein sequence along with a set of positions or ranges. The residues corresponding to the position: or as lowercase text. Use Range Extractor Protein to obtain subsequences using position information.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 500,000,000 characters.

```
APEQVEKGG
VSNVTEAKLIVFLTSIFVKAGCSPSDIGIIPYRQQLKIINDLLARSIGMVEVNTVDKYQG
RDKSIVLVS
FVRSNKDGTGVELLKDWRRLNVAITRAKHKLILLGCVPSLNCYPPLEKLLNHLNSEKLIID
LPSREHESL
CHILGDFQRE
```

Enter the residue positions or ranges to be extracted. Use "." to represent a range, and use a comma to separate entries. The words 'start' and 'length' of the sequence. Arithmetic expressions can be included in the ranges. For example, to obtain the last three residues of a sequence along with the center residue, the ranges '(center - 30)..(center - 1), center, (center + 1)..(center + 30)' can be used.

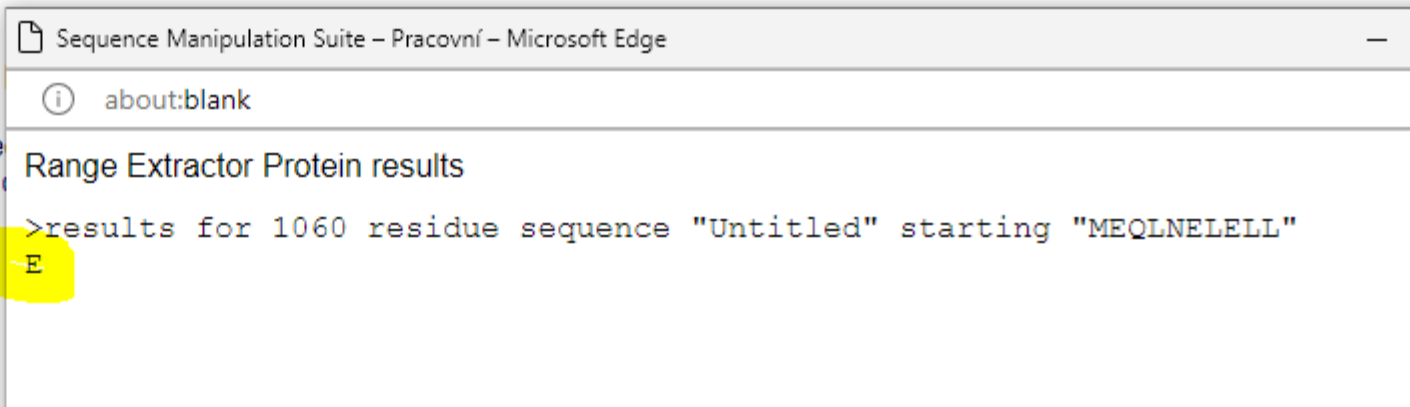
- Sequence segments should

*This page requires JavaScript. See

*You can [mirror this page](#) or use it

Sun 14 Jun 00:37:00 2020

Valid XHTML 1.0; Valid CSS



A) Pracujte s touto neznámou sekvencí:

```
181 elafqtiqei rhlkemyrln lsqdeikqev edylpsfckw agdfmhknts tdfpqmqlsl
241 psdnskdnst cnievvpmd ieesiwsprf glkgkidvtv gvkihrgykt kykimplelk
301 tgkesnsieh rsqvvlytll sqerradpea gllylktgq mypvpanhld krellklrnq
361 mafslfhris ksatrqtql aslpqiiee ktckycsqig ncalysrave qqmdcssvpi
421 vmlpkieeet qhlkqthley fslwclmltl esqskdnkn hqniwlmpas emeksgscig
```

1. Sekvenci přepište do fasta formátu.
2. Zjistěte, co kóduje a z jakého organismu pochází.
3. Zjistěte, jaký bude nejdelší fragment po štěpení celého identifikovaného proteinu proteinázou kaspáza 1.
4. Jaká je 20 aminokyselina celého proteinu?
5. Existuje nějaký lidský homolog tohoto proteinu? → BLASTp (Refseq, organismus=Homo sapiens)

[← Edit Search](#)[Save Search](#)[Search Summary ▾](#)[? How to read this report?](#)[▶ BLAST Help Videos](#)[↶ Back to Traditional Results Page](#)**i** Your search is limited to records that include: **Homo sapeins (taxid:9606)**

Job Title	filtered protein sequence consisting of 300...
RID	2D2SCC4B013 <small>Search expires on 04-24 00:06 am</small> Download All ▾
Program	BLASTP ? Citation ▾
Database	refseq_protein See details ▾
Query ID	lcl Query_2928689
Description	filtered protein sequence consisting of 300 residues.
Molecule type	amino acid
Query Length	300
Other reports	Distance tree of results Multiple alignment MSA viewer ?

Filter Results**Organism** only top 20 will appear exclude

Type common name, binomial, taxid or group name

[+ Add organism](#)**Percent Identity****E value****Query Coverage** to to to **Filter****Reset****Descriptions**[Graphic Summary](#)[Alignments](#)[Taxonomy](#)**Sequences producing significant alignments**[Download ▾](#)[Select columns ▾](#)

Show

[?](#) select all **4 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 replication ATP-dependent helicase/nuclease DNA2 isoform X1 [Homo sapiens]	Homo sapiens	629	629	100%	0.0	100.00%	1090	XP_006717743.1
<input checked="" type="checkbox"/>	DNA replication ATP-dependent helicase/nuclease DNA2 [Homo sapiens]	Homo sapiens	628	628	100%	0.0	100.00%	1060	NP_001073918.2
<input checked="" type="checkbox"/>	DNA replication ATP-dependent helicase/nuclease DNA2 isoform X2 [Homo sapiens]	Homo sapiens	535	535	85%	0.0	100.00%	836	XP_016871288.1
<input checked="" type="checkbox"/>	DNA replication ATP-dependent helicase/nuclease DNA2 isoform X3 [Homo sapiens]	Homo sapiens	254	254	40%	2e-78	100.00%	700	XP_011537719.1

Zkoušková písemka:

2 úlohy, každá 5 bodů

Verze 4.5.2021

Hodnocení: max celkem 10 bodů

9-10b: 1

7-8b: 2

5-6b: 3

A: Stáhněte si sekvenci NM_002085.5

- 1) Co tato sekvence kóduje?
- 2) Naleznete isoformy příslušného enzymu a porovnejte je.
- 3) Navrhnete primery, tak aby bylo kódující sekvenci tohoto genu možné vložit do plazmidu mezi restrikční endonukleázy BamHI a EcoR.I
- 4) Ověřte, zda jsou tyto RE vhodné pro klonování tohoto genu.
- 5) Přeložte tuto sekvenci do proteinu a identifikujte 100. aminokyselinu, запиште mutaci této AMK na alanin.

B: Pracujte s následující sekvencí peptidu:


```
EKHSGGFI FPPSYV FVVMSELT DQMT FMERVK NMIYVLY FDFWFE I FDMKKWDQFYSEVL  
GRPTTLSETMGKADVWLI RNSWNFQFPY PLLPNVDFV GGLHCKPAKPLPKEMEDFVQSSG  
ENGVVVFSLGSMVSNMTEERANVIASALAQIPQKVLWRFDGNKPDITLGLNTRLYKWIPQN  
DLLGHPKTRAFITHGGANGIYEAI
```

- 1) Identifikujte příslušný protein, запиште přístupový kód referenční sekvence.
- 2) Jaká je molekulová hmotnost tohoto peptidu?
- 3) Obsahuje celý identifikovaný protein signální peptid nebo transmembránové úseky?
- 4) Kde je tento protein v buňce lokalizován?
- 5) Ve kterých tkáních je nejvíce exprimován příslušný gen?

B) Stáhněte si sekvenci NM_000946.3

1. Co tato sekvence kóduje?
2. Jak dlouhá je kódující oblast pro příslušný protein a jak dlouhý bude přeložený protein?
3. Navrhněte manuálně primery, které by bylo možné použít pro amplifikaci pátého exonu.
4. Porovnejte vytvořené primery se sekvencí mRNA a příslušného exonu.
5. Existuje nějaký restriční enzym, který by sekvenci třetího exonu štěpil právě jednou?

B) Stáhněte si sekvenci NM_000946.3

1. Co tato sekvence kóduje?  NCBI
2. Jak dlouhá je kódující oblast pro příslušný protein a jak dlouhý bude přeložený protein?
3. Navrhněte manuálně primery, které by bylo možné použít pro amplifikaci pátého exonu.
4. Porovnejte vytvořené primery se sekvencí mRNA a příslušného exonu.
5. Existuje nějaký restriční enzym, který by sekvenci třetího exonu štěpil právě jednou?



Nucleotide

Nucleotide ▾

Advanced

GenBank ▾

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA


NCBI Reference Sequence: NM_000946.3

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

Go to:

LOCUS NM_000946 1423 bp mRNA linear PRI 03-APR-2024
DEFINITION Homo sapiens DNA primase subunit 1 (PRIM1), mRNA.
ACCESSION NM_000946
VERSION NM_000946.3
KEYWORDS RefSeq; MANE Select.
SOURCE Homo sapiens (human)
ORGANISM [Homo sapiens](#)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Haplorrhini;
Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Schimmel,J., Munoz-Subirana,N., Kool,H., van Schendel,R. and
Tijsterman,M.
TITLE Small tandem DNA duplications result from CST-guided Pol
alpha-primase action at DNA break termini
JOURNAL Nat Commun 12 (1), 4843 (2021)
PUBMED [34376693](#)
REMARK GeneRIF: Small tandem DNA duplications result from CST-guided Pol
alpha-primase action at DNA break termini.
Publication Status: Online-Only

B) Stáhněte si sekvenci NM_000946.3

1. Co tato sekvence kóduje?
2. Jak dlouhá je kódující oblast pro příslušný protein a jak dlouhý bude přeložený protein?  NCBI (Graphic)
3. Navrhněte manuálně primery, které by bylo možné použít pro amplifikaci pátého exonu.
4. Porovnejte vytvořené primery se sekvencí mRNA a příslušného exonu.
5. Existuje nějaký restriční enzym, který by sekvenci třetího exonu štěpil právě jednou?

Nucleotide

Nucleotide

Advanced

Graphics

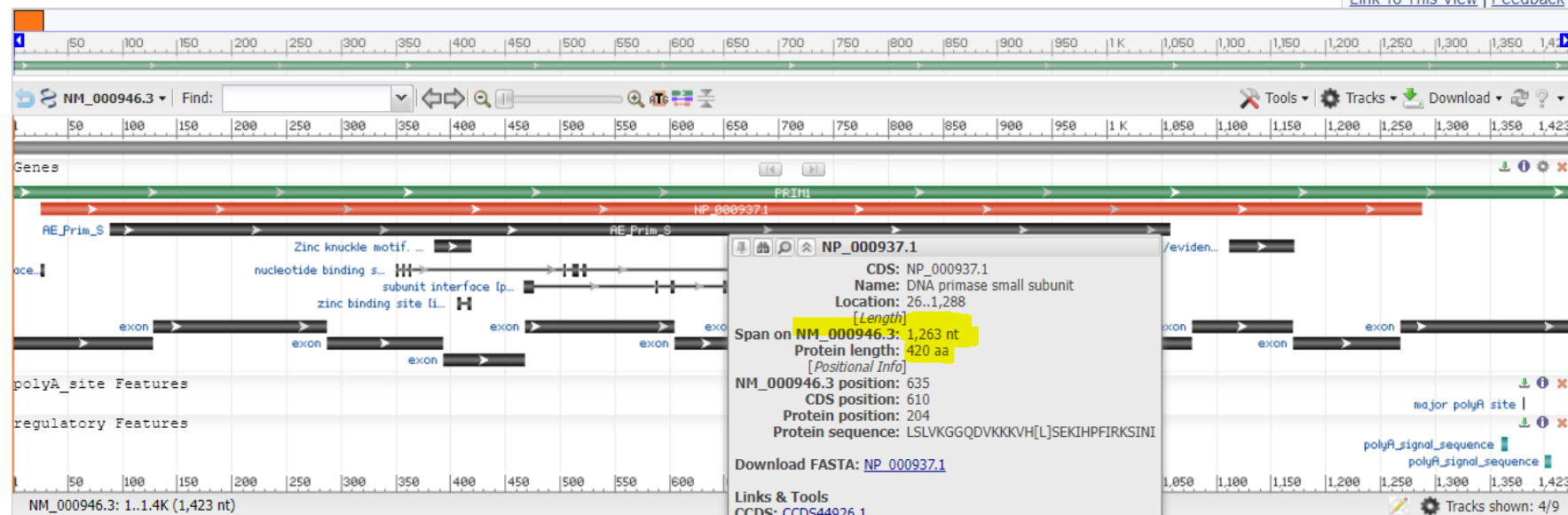
Send to:

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA


NCBI Reference Sequence: NM_000946.3

[GenBank](#) [FASTA](#)

[Link To This View](#) | [Feedback](#)



B) Stáhněte si sekvenci NM_000946.3

1. Co tato sekvence kóduje?
2. Jak dlouhá je kódující oblast pro příslušný protein a jak dlouhý bude přeložený protein?
3. Navrhněte manuálně primery, které by bylo možné použít pro amplifikaci pátého exonu.  Manuálně, OligoCalc kontrola
4. Porovnejte vytvořené primery se sekvencí mRNA a příslušného exonu.
5. Existuje nějaký restriční enzym, který by sekvenci třetího exonu štěpil právě jednou?

```

/ gene="PRIM1"
/ gene_synonym="p49; PDIL"
ORIGIN
1 cttaccgtgg cgagttccgc gctcaatgga gacgtttgac cccaccgagc tgcccagact
61 gcttaaacctt tattaccgga ggctctttcc ctactctcag tactactcgt ggctcaacta
121 cgggggagtg ataagaatt actttcaaca cgtgaattt tcattcacat tgaagatga
181 tatttacatt cgctaccaat ccttcaacaa ccagatgatg ctggaaaagg agatcgagaa
241 aatgaatcca tacaagattg atataggcgc agtatattct cacagacca atcaacacaa
301 tacagtgaag ctgggagctt tccaggctca gaaaaaagaa ctggtatttg acattgacat
361 gacagactat gacgatgtga ggagatgttg tagttctgca gacatattgc ctaagtgtctg
421 gaccctcatg acaatggcca tacgcatcat tgacagagca ttgaaggagg actttggatt
481 taagcatcgt cctcgggat attctgggaag gagaggtgtt cattgttggg tctgtgatga
541 atcagttaga aaactgtctt ctgcagtagc ttctgggata gttgagtatt tgagcctgtt
601 aaaaagggtgt caagactgta aaaaagaagt tcaccttaag gaaaaaattc acccttttat
661 cagaaaatct ataacataa taaaaaataa ctttgaagaa tatgccttgg ttaatcaaga
721 tattctcgaa aataaagaaa gctgggataa gattttagcc ctgttctcgt aaacaattca
781 tgatgaactt caacaaagct tccaaaagtc tcaacaattc cttcagcgtt ggggacactt
841 gaagaaagta gccagcagat atcagaataa catcaaaaat gacaatatg gaccctggct
901 ggagtgaggag attatgtccc agtactgttt tccacggctg gatatacagt tcagcaagg
961 aatcaatcat ctactgaaga gcccttttag tgttcacctt aaaaacagtc gcatacttgt
1021 gcctattgat ttgcagaagc tggaccagtt tgatccattt actgttccga ccataagcct
1081 catctgccgt gaattggatg ccatttccac taatgaagag gaaaaagagg agaatgaagc
1141 tgaatctgat gtcaaacata gaaccagaga ttataaagag accagtctag caccttatgt
1201 gaaagttttt gaacatttct ttgaaaatct ggataaatcc cgaaaaggag aactcttaa
1261 gaagagtgat ttacaaaaag atttctgaag acagagctcc tcaaacattt gtggatatct
1321 tctgccttca accacagatc aaatacttca agagccattt aataaatatg gcagaactat
1381 atatgtgtct taaacctcaa agtaaaatct ccttgagaaa taa
//

```

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

[GenBank](#) [Graphics](#)

```

>NM_000946.3:468-604 Homo sapiens DNA primase subunit 1 (PRIM1), mRNA
AGGACTTTGGATTTAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAGGTGTTTCATTGTTGGGTCTGTGA
TGAATCAGTTAGAAAACGTCTTCTGCAGTACGTTCTGGGATAGTTGAGTATTTGAGCCTTGTAAG

```

FOLLOW NCBI

Connect with NLM

National Library of Medicine
8600 Rockville Pike

Web Policies
FOIA

Help
Accessibility

```

468..604
/ gene="PRIM1"
/ gene_synonym="p49; PDIL"
/inference="alignment:Splign:2.1.0"

```

exon Feature 5 of 13 NM_000946 : 1 segment

Details Display FASTA GenBank Help

F: AGG ACT TTG GAT TTA AGC ATC G

Oligo Calc: Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
OD calculations are for single-stranded DNA or RNA

Nucleotide base codes
AGG ACT TTG GAT TTA AGC ATC G

Reverse Complement Strand(s) to 3' is:
CGA TGC TTA AAT CCA AAG TCC T

5' modification (if any) 3' modification (if any) Select molecule
50 mM Primer 1 Measured Absorbance at 260 nanometers
50 mM Salt (Na⁺)

Calculate Swap Strands BLAST mfold

Physical Constants Melting Temperature (T_m) Calculations

Length: 22 Molecular Weight: 6785.5 GC content: 41% 1 51.1 °C (Basic)
1 ml of a stock with an Absorbance of 1 at 260 nm 2 58.2 °C (Salt Adjusted)
is 4.112 micromolar and contains 27.9 micrograms 3 52.68 °C (Nearest Neighbor)

Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7
DeltaH: 33.464 kcal/mol (-136.16 kJ/mol) deltaTm: 173.9 °C (Kcal/mol)
DeltaS: 26.9 kcal/mol deltaTm: 457.7 °C (KJ/mol)

Designated Hairpin-loop formation calculations
5 (Minimum base pairs required for single primer self-dimerization)
4 (Minimum base pairs required for a hairpin) Check Self-Complementarity

Sequence Manipulation Suite:

Reverse Complement

Reverse Complement converts a DNA sequence into its reverse, complement, or reverse-complement counterpart. The entire IUPAC DNA alphabet is : want to work with the reverse-complement of a sequence if it contains an ORF on the reverse strand.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 100,000,000 characters.

```
>NM_000946.3:468-604 Homo sapiens DNA primase subunit 1
(PRIM1), mRNA
AGGACTTTGGATTAAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAGGTGTTCAATTGTTGG
GTCTGTGA
TGAATCAGTTAGAAAACGTCTTCTGCAGTACGTTCTGGGATAGTTGAGTATTTGAGCCTTG
TAAAG
```

Submit Clear Reset

- reverse-complement

*This page requires JavaScript. See [browser compatibility](#).

*You can mirror this page or use it

Sun 14 Jun 00:37:00 2020
Valid XHTML 1.0; Valid CSS

Sequence Manipulation Suite – Pracovní – Microsoft Edge

about:blank

Reverse Complement results

```
>NM_000946.3:468-604 Homo sapiens DNA primase subunit 1 (PRII
CTTTACAAGGCTCAAATACTCAACTATCCCAGAACGTACTGCAGAAGACAGTTTCTAAC
TGATTCATCACAGACCCCAACAATGAACACCTCTCCTCCAGAATATACCCAGAGACGATG
CTTAAATCCAAAGTCTCT
```

R: CTT TAC AAG GCT CAA ATA CTC AA (C)

Oligo Calc: Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below
OD calculations are for single-stranded DNA or RNA

Nucleotide base codes
CTT TAC AAG GCT CAA ATA CTC AA

Reverse Complement Strand(5' to 3') is:
TTG AGT ATT TGA GCC TTG TAA AG


5' modification (if any) 3' modification (if any) Select molecule
50 nM Primer 1 Measured Absorbance at 260 nanometers
50 mM Salt (Na⁺)

Calculate Swap Strands BLAST mfold

Physical Constants		Melting Temperature (T _m) Calculations	
Length: 23	Molecular Weight: 6975.64	GC content: 35%	1 49.9 °C (Basic)
1 ml of a sol'n with an Absorbance of 1 at 260 nm			2 57.6 °C (Salt Adjusted)
is 3.944 microMolar and contains 27.5 micrograms.			3 51.14 °C (Nearest Neighbor)
Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.			
RinK: 33.404	cal/(K*mol)	deltaH: 175.6	Kcal/mol
deltaG: 26.4	Kcal/mol	deltaS: 464.8	cal/(K*mol)
Deprecated Hairpin/self dimerization calculations			
5 (Minimum base pairs required for single primer self-dimerization)		Check Self-Complementarity	
4 (Minimum base pairs required for a hairpin)			

Citation: Kibbe WA. 'OligoCalc: an online oligonucleotide properties calculator'. (2007)

B) Stáhněte si sekvenci NM_000946.3

1. Co tato sekvence kóduje?
2. Jak dlouhá je kódující oblast pro příslušný protein a jak dlouhý bude přeložený protein?
3. Navrhněte manuálně primery, které by bylo možné použít pro amplifikaci pátého exonu.
4. Porovnejte vytvořené primery se sekvencí mRNA a příslušného exonu.  Multalin
(R primer zpátky rc“)
5. Existuje nějaký restriční enzym, který by sekvenci třetího exonu štěpil právě jednou?

>F

AGG ACT TTG GAT TTA AGC ATC G

>R_{rc}

TTG AGT ATT TGA GCC TTG TAA AG

>NM_000946.3:468-604 Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

AGGACTTTGGATTAAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAGGTGTTTCATTGTTGGGTCTGTGA

TGAATCAGTTAGAAAAGTGTCTTCTGCAGTACGTTCTGGGATAGTTGAGTATTTGAGCCTTGTAAG

	1	10	20	30	40	50	60	70	80	90	100	110	120	130
F	-----													
R _{rc}	AGGACTTTGGATTAAAGCATCG													
NM_000946.3:468-604	AGGACTTTGGATTAAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAGGTGTTTCATTGTTGGGTCTGTGATGATCAGTTAGAAAAGTGTCTTCTGCAGTACGTTCTGGGATAGTTGAGTATTTGAGCCT													
Consensus	aggactttggattaaagcatcg.....ttgagtattgagcct													
	131	137												
F	-----													
R _{rc}	TGTAAG													
NM_000946.3:468-604	TGTAAG													
Consensus	tgaag													

	261	270	280	290	300	310	320	330	340	350	360	370	380	390
F	-----													
R _{rc}	ATATAGGCGCAGTATATTCTCAGACCCCAATCAACACATACAGTGAAGCTGGGAGCTTCCAGGCTCAGGAAAAGAGACTGGTATTTGACATTGACATGACAGACTATGACGATGTGAGGAGATGTTG													
NM_000946.3:468-604	ATATAGGCGCAGTATATTCTCAGACCCCAATCAACACATACAGTGAAGCTGGGAGCTTCCAGGCTCAGGAAAAGAGACTGGTATTTGACATTGACATGACAGACTATGACGATGTGAGGAGATGTTG													
Consensus													
	391	400	410	420	430	440	450	460	470	480	490	500	510	520
F	-----													
R _{rc}	AGGACTTTGGATTAAAGCATCG													
NM_000946.3:468-604	AGGACTTTGGATTAAAGCATCGTCTCTGGGTATATTCTGGAAGGAGAGGTGTTTCATTGTTGGGTCTGTGATGATCAGTTAGAAAAGTGTCTTCTGCAGTACGTTCTGGGATAGTTGAGTATTTGAGCCTTGTAAG													
Consensus	aggactttggattaaagcatcg.....ttgagtattgagccttgtaag													
	521	530	540	550	560	570	580	590	600	610	620	630	640	650
F	-----													
R _{rc}	TTGAGTATTTGAGCCTTGTAAG													
NM_000946.3:468-604	CATTGTTGGGTCTGTGATGATCAGTTAGAAAAGTGTCTTCTGCAGTACGTTCTGGGATAGTTGAGTATTTGAGCCTTGTAAGGGTGGTCAAGACGTTAAAAGAAAGTTCACCTAAGTGA AAAAATTC													
Consensusttgagtattgagccttgtaag.....													
	651	660	670	680	690	700	710	720	730	740	750	760	770	780
F	-----													
R _{rc}	ACCCCTTTATCAGAAATCTATAACATATAAAAAAATCTTTGAGAAATATGCTTGGTGTATCAGATATTTCTGAAATTAAGAAAGCTGGGATAGATTTTAGCCCTTGTCTGAAACATTC													
NM_000946.3:468-604	ACCCCTTTATCAGAAATCTATAACATATAAAAAAATCTTTGAGAAATATGCTTGGTGTATCAGATATTTCTGAAATTAAGAAAGCTGGGATAGATTTTAGCCCTTGTCTGAAACATTC													
Consensus													

B) Stáhněte si sekvenci NM_000946.3

1. Co tato sekvence kóduje?
2. Jak dlouhá je kódující oblast pro příslušný protein a jak dlouhý bude přeložený protein?
3. Navrhněte manuálně primery, které by bylo možné použít pro amplifikaci pátého exonu.
4. Porovnejte vytvořené primery se sekvencí mRNA a příslušného exonu.
5. Existuje nějaký restriční enzym, který by sekvenci třetího exonu štěpil právě jednou?
→ Restriction summary (SMS)

Homo sapiens DNA primase subunit 1 (PRIM1), mRNA

NCBI Reference Sequence: NM_000946.3

[GenBank](#) [Graphics](#)

```
>NM_000946.3:287-393 Homo sapiens DNA primase subunit 1 (PRIM1), mRNA
CCCAATCAACACAATACAGTGAAGCTGGGAGCTTCCAGGCTCAGGAAAAAGAACTGGTATTTGACATTG
ACATGACAGACTATGACGATGTGAGGAGATGTTGTAG
```



Format Conversion

- Combine FASTA
- EMBL to FASTA
- EMBL Feature Extractor
- EMBL Trans Extractor
- Filter DNA
- Filter Protein
- GenBank to FASTA
- GenBank Feature Extractor
- GenBank Trans Extractor
- One to Three
- Range Extractor DNA
- Range Extractor Protein
- Reverse Complement
- Split Codons
- Split FASTA
- Three to One
- Window Extractor DNA
- Window Extractor Protein

Sequence Analysis

- Codon Plot
- Codon Usage
- CpG Islands
- DNA Molecular Weight
- DNA Pattern Find
- DNA Stats
- Fuzzy Search DNA
- Fuzzy Search Protein
- Ident and Sim
- Multi Rev Trans
- Mutate for Digest
- ORF Finder
- Pairwise Align Codons
- Pairwise Align DNA
- Pairwise Align Protein
- PCR Primer Stats
- PCR Products
- Protein GRAVY
- Protein Isoelectric Point
- Protein Molecular Weight
- Protein Pattern Find
- Protein Stats
- Restriction Digest
- Restriction Summary
- Reverse Translate
- Translate

Sequence Figures

Sequence Manipulation Suite:

Restriction Summary

Restriction Summary accepts a DNA sequence and returns the number and positions of commonly used restriction endonuclease cut sites. Use this program on a segment of DNA.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 100,000,000 characters.

```
>NM_000946.3:287-393 Homo sapiens DNA primase subunit 1
(PRIM1), mRNA
CCCAATCAACACAATACAGTGAAGCTGGGAGCTTCCAGGCTCAGGAAAAAGAACTGGTATT
TGACATTG
ACATGACAGACTATGACGATGTGAGGAGATGTTGTAG
```

- Treat sequences as molecules.

*This page requires JavaScript. See [browser compatibility](#).

*You can [mirror this page](#) or use it

Sun 14 Jun 00:37:00 2020
Valid XHTML 1.0; Valid CSS

Enzyme	Count
NaeI gcc ggc	none
NarI gg cgc	none
NcoI c catgg	none
NdeI ca tatg	none
NdeII gatc	none
NgoMIV g ccggc	none
NheI g ctagc	none
NlaIII catg	76
NotI gc ggccgc	none
NruI tcg cga	none