

Introduction to applied bioinformatics

PETRA MATOUŠKOVÁ

2025/2026

10/10

Exam

Thursday 23.4.2026 12:20

Moodle: files with the version
open „task“ for submission (like for homeworks).

Usual Exam test:

2 parts, each five points

Maximum 10 points

8.5-10p: 1

7-8p: 2

5-6.5p: 3

Version 1

Find human protein sequence **FGF3**

- What is the accession number and function of this protein?
- How many valines does the sequence contain?
- How long is the longest fragment after trypsin digestion?
- How many human similar reference proteins are in the databases?
- Does the sequence contain any transmembrane regions?

Download sequence **NM_001757.3**

- What does this sequence encode ?
- Is there a complementary sequence to primer R1?

```
>R1  
GCTCTGACGCTCATGATGC
```

- In which exon is the complementary sequence for this primer?
- Is the primer suitable for PCR with Ta=60°C?
- Design suitable F primer for this experiment.

Last year Exam

A: Download the sequence of the accession number NM_003998.4

1. What does this sequence encode?
2. Download the seventh exon of this sequence in FASTA format, how long is it?
3. Is there a restriction enzyme that would cut the sequence of the seventh exon exactly once?
4. Find corresponding protein, does it contain any transmembrane helices?
5. How many threonines (T) does the respective protein contain?

B: Work with following peptide sequence:

```
181 mdvitstsfg vnidslnnpq dpfventkkl lrfdfldpff lsitvfpfli pilevlnicv
241 fprevtnflr ksvkrmkesr ledtqkhrvd flqlmidsqn sketeshkal sdlelvaqsi
301 ififagyett ssvlsfimye lathpdvqqk lgeeidavlp nkapptydtv lqmeylmrvv
361 netlrlfpia mrlervckkd veingmfipk gvvvmipsya lhrdpkywte pekflperfs
```

- 1) Rewrite this sequence in FASTA format.
- 2) What is the molecular weight of this peptide?
- 3) To which human protein this peptide probably belongs (find reference sequence NP_...)?
- 4) Find corresponding transcript sequence (mRNA), what is its accession number?
- 5) Manually design primers for PCR amplification of the **CDS**, so their T_m is not higher than 65°C.

A: Download the sequence of the accession number NM_003998.4

1. What does this sequence encode?
2. Download the seventh exon of this sequence in FASTA format, how long is it?
3. Is there a restriction enzyme that would cut the sequence of the seventh exon exactly once?
4. Find corresponding protein, does it contain any transmembrane helices?
5. How many threonins (T) does the respective protein contain?

Hints

A: Download the sequence of the accession number NM_003998.4

1. What does this sequence encode? → Search the NCBI with given accession number
 2. Download the seventh exon of this sequence in FASTA format, how long is it? → Check the Graphic in the NCBI
 3. Is there a restriction enzyme that would cut the sequence of the seventh exon exactly once? → Restriction summary
 4. Find corresponding protein, does it contain any transmembrane helices? → Find protein: TMHMM, Topcons...
 5. How many threonins (T) does the respective protein contain? → Find protein: Protein Stats
- Or through „Highlight sequence features“ on the left jump to exon 7 (out of 24)

GenBank ▾

Send to: ▾

Homo sapiens nuclear factor kappa B subunit 1 (NFKB1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_003998.4

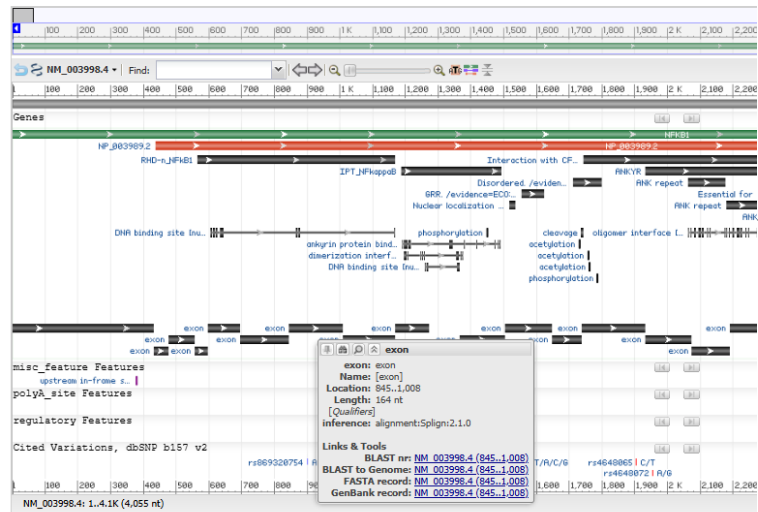
[FASTA](#) [Graphics](#)

[Go to:](#) ☐

LOCUS	NM_003998	4055 bp	mRNA	linear	PRI 20-NOV-2025
DEFINITION	Homo sapiens nuclear factor kappa B subunit 1 (NFKB1), transcript variant 1, mRNA.				
ACCESSION	NM_003998				

Homo sapiens nuclear factor kappa B subunit 1 (NFKB1), transcript variant 1, mRNA

NCBI Reference Sequence: NM_003998.4

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

```

721 aggttattgt tcagttggtc acaaatggaa aaaatatcca cctgcatgcc cacagcctgg
781 tgggaaaaca ctgtgaggat gggatctgca ctgtaactgc tggacccaag gacatgtgtg
841 tcggcttcgc aaacctgggt atacttcagt tgacaaagaa aaaagtattt gaacactctg
901 aagcacgaat gacagaggcg tgtataaggg gctataatcc tggactcttg gtgcaccctg
961 accttgccca ttgcaagca gaaggtggag gggaccggca gctgggagat cgggaaaaag
1021 agctaattcc ccaagcagct ctgcagcaga ccaaggagat ggacctcagc gtgtgtcggc
1081 tcatgtttac agctttttct cggatagca ctggcagctt cacaaggcgc ctggaacccg
1141 tggatcaga gccatctat gacagtaaag ccccaatgc atccaacttg aaaattgtaa
1201 gaattgacag gacagctgga tgtgtgactg gaggggagga aatttatctt ctttgtgaca
1261 aagttcagaa agatgacatc cagattcgat tttatgaaga ggaagaaaat ggtggagtct
1321 gggaaaggatt tggagatttt tccccacag atgttcatag acaatttgcc attgtcttca
1381 aaactccaaa gtataaagat attaatatta caaaaccagg ctctgtgttt gtccagcttc
1441 ggaggaaatc tgacttggaa actagtgaac caaaaccttt cctctactat cctgaaatca
1501 aagataaaga agaagtgcag aggaacacgtc agaagctcat gcccaatttt tcggatagtt
1561 tcggcgggtg tagtggtgct ggagctggag gaggaggcat gtttggtagt ggcggtgag
1621 gaggggcac tggaaagtaca ggtccagggt atagctttcc acactatgga tttcctactt
1681 atggtgggat tactttccat cctggaacta ctaaatctaa tgctgggatg aagcatggaa
1741 catggcacac tgaatctaaa aaggaccctg aaggttgtag caaaagtgtg gacaaaaaca
1801 ctgtaaacct ctttgggaaa gttattgaaa ccacagagca agatcaggag cccagcgagg
1861 ccacccteez gaateetaz etcactctaa cetatecaac aaaaacaaaa gaaagaetaz

```

```

845..1008
/gene="NFKB1"
/gene_synonym="CVID12; EBP-1; KB1; NF-kappa-B1;
NF-kappaB; NF-kappabeta; NF-kB; NF-kB1; NFkappaB;
NFKB-p105; NFKB-p50"
/inference="alignment:Splign:2.1.0"

```

[Details ▾](#)Display: [FASTA](#) [GenBank](#) [Help](#) [X](#)

1008-845+1=164nt

(write or show the number, not just the span!)

>NM_003998.4:845-1008 Homo sapiens nuclear factor kappa B subunit 1 (NFKB1), transcript variant 1, mRNA
CTTCGCAAACCTGGGTATACTTCATGTGACAAAGAAAAAGTATTTGAAACACTGGAAGCACGAATGACA
GAGGGCGTGTATAAGGGGCTATAATCCTGGACTCGCAGAAG

Sequence Manipulation Suite:

Restriction Summary

Restriction Summary accepts a DNA sequence and returns the number and positions of comm
Use this program if you wish to quickly determine whether or not an enzyme cuts a particular s

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit

>NM_003998.4:845-1008 Homo sapiens nuclear factor kappa B subunit 1 (NFKB1), transcript variant 1, mRNA
CTTCGCAAACCTGGGTATACTTCATGTGACAAAGAAAAAGTATTTGAAACACTGGAAGCACG
AATGACA
GAGGGCGTGTATAAGGGGCTATAATCCTGGACTCGCAGAAG

SubmitClearReset

Treat sequences as

linear

*This page requires JavaScript. See b

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

Sun 1 Oct 12:00:07 2023
Valid XHTML 1.0; Valid CSS

Sequence Manipulation Suite – Škola – Microsoft Edge

about:blank

AhII a ctagt	none
Alw441 g tgcac	108
AluI ag ct	158
Aor51HI agc gct	none
Apal gggcc c	none
ApaLI g tgcac	108
Ascl aalcacacc	none

1321 aggaaggatt tggagatttt tccccacag atgttcatag acaatttccc attgtcttca
1381 aaacttcaaa gtataaagat attaatatta caaaaccagc ctctgtgttt gtccagcttc
1441 agaggaaatc tgacttggaa actagtgaac caaaaccttt cctctactat cctgaaatca
1501 agataaaga agaagtgcag aggaacgctc agaagctcat gcccaatttt tggatagtt
1561 tccagctgag tagtggctct agagctgaag gccaagccat gtttggtagt agcagtgaa
1621 gaggggcac tggagtaca ggtccagggt atagcttccc acatatgga ttcttactt
1681 atgtggagat tactttccat cctggaaacta ctaaatctaa tgcgggagtg aagcatgaa
1741 ccatggacac tgaattcaaa aaggaccctg aaggttctga caaaagtgat gacaaaaaca
1801 ctataaacct ctttggaaa gttattgaaa ccacagagca agatcaggag cccagcagag
1861 ccaccgttgg gaatggtag gtcactctaa ctatgcaac aggaacaaaa gaaggaagtg
1921 ctggagttca gataacctc ttctagaga aggtatgca ccttgcacaa aggcatacca
1981 atgacctttt cgaactacgc gtcacagga acgtgaagat gctgtggcc gtccagcgcc
2041 atctcactgc tgtcaggat gagaatggg acagtgtctt acattagca atcatccacc
2101 ttacttcca acttggagg gattactag aagtcacatc tggtttgatt tctgatgaca
2161 tatcaacat gagaatgat ctgtaccaga cgccttgcg cttggcagtg atcataagc
2221 aggaagatgt gttgaaggat ttgtcagagg ctggggcgca cctgaacctt ctggaccgct
2281 tggtaactc tgtttgcac ctagtccca aagaaggaca tgataaagtl ctacgtatct
2341 tactaagca caaaaaggca cgaactactt ttgaccaccc caacgggagc ggtctgaag
2401 cttactcatc agccatgatg agcaatagcc tgccatgttt gctgctgctg gtggcctctg
2461 aggtctgact caatgtcag gacgagaagt cggggcgac agcactgcac ctgctgtga
2521 agcacgaca catctcattg gcaggtgccc tgcctctgga ggtgatgcc catgtgaca
2581 gtactacctc cgtatgaacc acacccctgc atatagcagc tggagagagg tccaccaagc
2641 tggcagctct tctcaagca gcaggagcag atccccgtgt ggagaacttt gacgtctct
2701 atgacttga tgactcttgg gaaaatgcag gagagatga aggatgttg cctggaacca
2761 gacctctaga tatggcacc agctgcagg tatttgacat attaatggg aaacatatg
2821 ggcagagatt tacatctgat gatttactag cacaaggaga catgaacag ctggctgaag
2881 atgtgaagct gcagctgtat aagttactag aaattcctga tccagacaaa aactggacta
2941 ctctggcaca gaaattaggt ctggagatc ttataatgac cticcggctg agtctgctc
3001 ctccaaaac acttatggac aactatgag tctctggagg tacactcaga gacttgatg
3061 gggccctgag acaaatggc tacaccgaag caatigaagt gatcaggca gctccagct
3121 cagtgaagca cactctcag gccactcgc tgcctctctc gctgcctcc acaaggcagc
3181 aatatagaga gctccgagc agtgacagtg tctgcagac ggcgtggag acatcttcc
3241 gcaactcag ctttaccgag tcttgacca gtgtgctc actgttaact ctcaacaaa
3301 tccccatga ttatggcag gaggacctc tgaaggcga aattatgct gctgaactt
3361 tcccacaccg tgtaaccaa agccataaa ttccactgct tgtgtcacaa gacagaagct
3421 gaagtgcac caaaggtgct cagagagcgc gccgcctga atcattctg atttaactg

CDS Feature 1 of 1 NM_003998 : 1 segment

438..3347
/gene="NFKB1"
/gene_synonym="CVID12; EBP-1; KBF1; NF-kappa-B1;
NF-kappaB; NF-kappabeta; NF-kB; NF-kB1; NFkappaB;
NFKB-p105; NFKB-p50"
/note="isoform 1 is encoded by transcript variant 1;
nuclear factor kappa-B DNA binding subunit; nuclear factor
NF-kappa-B p50 subunit; nuclear factor NF-kappa-B p105
subunit; DNA-binding factor KBF1; nuclear factor of kappa
light polypeptide gene enhancer in B-cells 1"
/codon_start=1
/product="nuclear factor NF-kappa-B p105 subunit isoform
1"
/protein_id="NP_003998.2"
/db_xref="CCDS: CCDS3657.1"
/db_xref="GeneID: 4798"
/db_xref="MIM: 164941"
/db_xref="MIM: 164941"
/translation="MAEDDPYLGRPEQHFLDPSLTHTFNPEVFQHQALPTADGPY
LQIALGKQGRFRFRVYCEGPGSHGLPGASSEKNKSYQKICIVYGPANVQLVLT
EKNIHLHAHSLVGHKDEGICTVTAGPDVNVGFANLGIHLVTKKVFLEARHTE
ACIRSYNPGLLVHPDLAYLQAEGGDRQLGRKELTRQALQGTKEKVDLSVRLMFT
AFLPDSTGSTRRLRPVSDAIVDSKAPNNSLKIIVNRDRTACVTTGGEIVLLCDKV
QKDDIQIRFYEIEENGWIEGDFSPDVRHQFAIVFKPKYVDINTTKPASVFLVQL
RKVSDLETSEPKFLYPPEIKDKEEVQRKRLKSNVSDSGGGSGAGAGGGGPFQSG
GGGGTGTGPGYSFPHYGFPTYGTFHPGTKSNAGWKGHTDESKVDPGCKDS
QKIVTVNLFGKIVTIEQQQEPSEATVNGEVLTYATGTKEESAGQDILFEKAMQ
LKKHANALFDYAVTGDVKNLLAVQRHLTAQDENGDSVLHAIHLHSLPLVDLLEV
TSGVTSDDINWRNDLQTPHLAVITKQVEDVLDLAGADLSLLDLRLGSLVHLAA
KEGKGLSTILLXHKKAALLDHPNGDGLNATHLWNSLPLLLVAAGADVHLAE
QKSGRTALNVEHONISLAGCLLLEGDAHVDTTYDGTTPHIAAGRGSRLAALLK
ANGADPLVENFVQLDOSHEVNADEGVVPGTTPDMATSHQVFDLNLNGPYEPPEF
TSDLLAQGDMQLAEKLLQYLKLETPDPDKMATLQKLGLGLINAFRLSPAPS
KTLMDNVEYSGGVRELVEALONGYTEAIEVQAASSPVKTTTSAHSLPLSPASTRQ
QIDELRDSDSVCSGVETSFRKLSFLESLTSGASLLTNKMPHYDQGEPLGKI"

Details Display: FASTA GenBank Help

TMHMM result

WEBSEQUENCE Length: 969
WEBSEQUENCE Number of predicted TMHs: 0
WEBSEQUENCE Exp number of AAs in TMHs: 0.0355600000000001
WEBSEQUENCE Exp number, first 60 AAs: 0
WEBSEQUENCE Total prob of N-in: 0.00028
WEBSEQUENCE TMH*912.0 outside 1 969

No THs:

TMHMM posterior probabilities for WEBSEQUENCE

probability

transmembrane inside outside

Sequence Manipulation Suite:

Protein Stats

Protein Stats returns the number of occurrences of each residue in the sequence you enter. Per and for certain groups of residues, allowing you to quickly compare the results obtained for diffe

61 Threonins

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

MAEDDPYLGRPEQMFHLDPSLTHTIFNPEVFQPQMALPTADGPY
LQILEQPKQRGFRFRYVCEGSPSHGGLPGASSEKNKKSYPQVKICNYVGPAKVIVQLVT
NGKNIHLHAHSLVGKHCEGICTVTAGPKDMVVGAFANLGHVTKKKVFETLEARMTE
ACIRGYNPGLLVHPDLAYLQAEGGGDF
AFLPDSTGSFTRRLEPVVSDAIYDSKA
QKDDIQIRFYEEEEENGWEGFGDFSF

SubmitClearReset

*This page requires JavaScript

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

Sun 1 Oct 12:00:07 2023
Valid XHTML 1.0; Valid CSS

Q	41	4.23
R	35	3.61
S	60	6.19
T	61	6.30
V	60	6.19
W	4	0.41

B: Work with following peptide sequence:

```
181 mdvitstsfg vnidslnnpq dpfventkkl lrfdfldpff lsitvfpfli pilevlnicv
241 fprevtnflr ksvkrmkesr ledtqkhrvd flqlmidsqn sketeshkal sdlelvaqsi
301 ififagyett ssvlsfimye lathpdvqqk lqeeidavlp nkapptydtv lqmeyldmvv
361 netlrlfpia mrlervckkd veingmfipk gvvvmipsya lhrdpkywte pekflperfs
```

- 1) Rewrite this sequence in FASTA format.
- 2) What is the molecular weight of this peptide?
- 3) To which human protein this peptide probably belongs (find reference sequence NP_..)?
- 4) Find corresponding transcript sequence (mRNA), what is its accession number?
- 5) Manually design primers for PCR amplification of the **CDS**, so their T_m is not higher than 65°C.

Hints

B: Work with following peptide sequence:

```
181 mdvitstsfg vnidslnnpq dpfventkkl lrfdfldpff lsitvfpfli pilevlnicv
241 fprevtnflr ksvkrmkesr ledtqkhrvd flqlmidsqn sketeshkal sdlelvaqsi
301 ififagyett ssvlsfimye lathpdvqqk lqeeidavlp nkapptydtv lqmeylmrvv
361 netlrlfpia mrlervckkd veingmfipk gvvvmipsya lhrdpkywte pekflperfs
```

- 1) Rewrite this sequence in FASTA format. —→ SMS: Filter protein
- 2) What is the molecular weight of this peptide? —→ SMS: Molecular weight
- 3) To which human protein this peptide probably belongs (find reference sequence NP_...)? —→ BLASTp – Reference sequence database
- 4) Find corresponding transcript sequence (mRNA), what is its accession number? —→ Find DNA: NM_
- 5) Manually design primers for PCR amplification of the **CDS**, so their T_m is not higher than 65°C.
—→ Find CDS: F primer-check by oligocalc
R primer- reverse complement - oligocalc

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 to make it suitable for other applications.

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

```
sdlelvaqsi
301 ififagyett ssvlsfimye lathpdvqqk lgeeidavlp nkapptydtv
lqmeylmrvv
361 n
pekfl
```

Subn

Filter Protein results

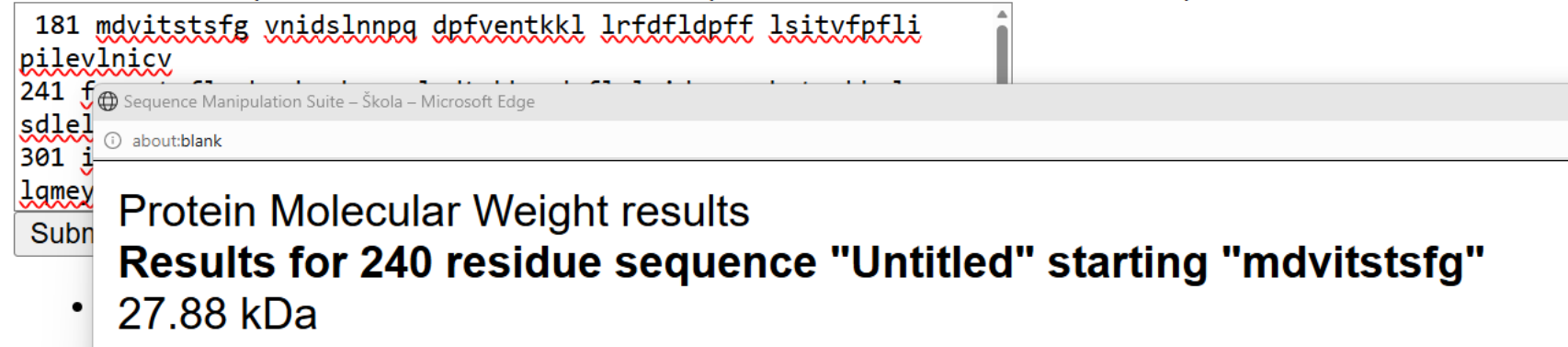
- >filtered protein sequence consisting of 240 residues.
 - mdvitstsfgvnidslnnpqdpfventkkllrfdfldpfflsitvfpflipilevlnicv
- *This
*You c
- ```
fprevtnflrksvkrmkesrledtqkhrvdfllqmidsqnsketeshkalsdlelvaqsi
ififagyettssvlsfimyelathpdvqqklqeeidavlpnkapptydtvlgmeylmrvv
netlrlfpiamrlervckkdveingmfipkgvvvmipsyalhrdpkywtepekflperfs
```

## Sequence Manipulation Suite:

### Protein Molecular Weight

Protein Molecular Weight accepts one or more protein sequences and calculates molecular weight. You can append epitopes and fusion proteins using the supplied list. Use Protein Molecular Weight when you wish to predict the location on a gel in relation to a set of protein standards.

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 200,000,000 characters.



The screenshot shows a web browser window with the title "Sequence Manipulation Suite - Škola - Microsoft Edge". The address bar shows "about:blank". The main content area displays a protein sequence input field with the following text: "181 mdvitstsfg vnidslnnpq dpfventkk1 lrfdfldpff lsitvfpfli pilevlnicy". Below the input field, the results are displayed: "Protein Molecular Weight results" and "Results for 240 residue sequence 'Untitled' starting 'mdvitstsfg'". A single bullet point indicates the molecular weight: "• 27.88 kDa".

181 mdvitstsfg vnidslnnpq dpfventkk1 lrfdfldpff lsitvfpfli  
pilevlnicy  
241 f  
sdlel  
301 i  
lqmev  
Subn

Protein Molecular Weight results  
Results for 240 residue sequence "Untitled" starting "mdvitstsfg"

- 27.88 kDa



[← Edit Search](#)[Save Search](#)[Search Summary ▼](#)[? How to read this report?](#)[▶ BLAST Help Videos](#)[↶ Back to Traditional Results Page](#)

|               |                                                                                                                          |
|---------------|--------------------------------------------------------------------------------------------------------------------------|
| Job Title     | 181:kappa-casein precursor                                                                                               |
| RID           | <a href="#">YE71BPRT014</a> Search expires on 04-22 14:41 pm <a href="#">Download All</a> ▼                              |
| Program       | BLASTP <a href="#">?</a> <a href="#">Citation</a> ▼                                                                      |
| Database      | refseq_protein <a href="#">See details</a> ▼                                                                             |
| Query ID      | lcl Query_10941476                                                                                                       |
| Description   | unnamed protein product                                                                                                  |
| Molecule type | amino acid                                                                                                               |
| Query Length  | 240                                                                                                                      |
| Other reports | <a href="#">Distance tree of results</a> <a href="#">Multiple alignment</a> <a href="#">MSA viewer</a> <a href="#">?</a> |

## Filter Results

Organism only top 20 will appear

☐ exclude[+ 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 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"/> | <a href="#">cytochrome P450 3A4 isoform 1 [Homo sapiens]</a>                    | <a href="#">Homo sapiens</a>    | 491         | 491           | 100%          | 6e-171    | 100.00%      | 503        | <a href="#">NP_059488.2</a>    |
| <input checked="" type="checkbox"/> | <a href="#">cytochrome P450 3A4 [Pan paniscus]</a>                              | <a href="#">Pan paniscus</a>    | 487         | 487           | 100%          | 1e-169    | 99.17%       | 503        | <a href="#">XP_057158705.1</a> |
| <input checked="" type="checkbox"/> | <a href="#">cytochrome P450 family 3 subfamily A member 4 [Pan troglodytes]</a> | <a href="#">Pan troglodytes</a> | 487         | 487           | 100%          | 1e-169    | 99.17%       | 503        | <a href="#">NP_001116247.1</a> |

# NM\_017460.6

Nucleotide

Nucleotide

cyp3a4

Search

Create alert

Advanced

Help

Species

Animals (467)

Plants (4)

Fungi (37)

Protists (2)

Customize ...

Molecule types

genomic DNA/RNA (1,310)

mRNA (178)

Customize ...

Source databases

INSDC (GenBank) (1,571)

RefSeq (134)

Customize ...

Sequence Type

Nucleotide (1,067)

EST (38)

GSS (2)

Sequence length

Custom range...

Release date

Custom range...

Revision date

Custom range...

Clear all

Show additional filters

Summary

20 per page

Sort by Default order

Send to:

Filters: [Manage Filters](#)

GENE

Was this helpful?

👍

👎

[CYP3A4 – cytochrome P450 family 3 subfamily A member 4](#)

[Homo sapiens \(human\)](#)

Also known as: CP33, CP34, CYP3A, CYP3A3, CYP3A3, CYP3A4, HLP, NF-25, P450C3, P450PCN1, VDDR3

Gene ID: 1576

RefSeq products

Orthologs

Genome Data Viewer

New - Visualize gene across multiple species

RefSeq Sequences

Showing 2 of 2 (by status, accession number)

| Transcript                     | nt    | Protein                        | aa  | Isoform | Status      |
|--------------------------------|-------|--------------------------------|-----|---------|-------------|
| <a href="#">NM_017460.6</a>    | 2,781 | <a href="#">NP_059488.2</a>    | 503 | 1       | MANE Select |
| <a href="#">NM_001202855.3</a> | 2,778 | <a href="#">NP_001189784.1</a> | 502 | 2       | curated     |

View full table

Results by taxon

Top Organisms [\[Tree\]](#)

synthetic construct (747)

Homo sapiens (194)

Macaca mulatta (30)

unidentified (30)

Pan troglodytes (21)

All other taxa (685)

More...

Find related data

Database: [Select](#)

Find items

Search details

cyp3a4[All Fields]

Search

See more...

Recent activity

Turn Off

Clear

Q cyp3a4 (1707)

Nucleotide

cytochrome P450 3A4 isoform 1 [Homo sapiens]

Protein

Homo sapiens nuclear factor kappa B subunit 1 (NFKB1), transcript variant 1

Nucleotide

amyloid-beta precursor protein isoform a precursor [Homo sapiens]

Protein

F:ATG GCT CTC ATC CCA GAC TT  
R:TCA GGC TCC ACT TAC GGT G

### Oligonucleotide Properties Calculator

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

[Nucleotide base codes](#)  
ATG GCT CTC ATC CCA GAC TT

Reverse Complement Strand(5' to 3') is:  
AAG TCT GGG ATG AGA GCC AT

5' modification (if any)

3' modification (if any)

Select molecule  
ssDNA

50 nM Primer

1 Measured Absorbance at 260 nanometers

50 mM Salt (Na<sup>+</sup>)

Calculate

Swap Strands

BLAST

mfold

Physical Constants

Melting Temperature (T<sub>M</sub>) Calculations

Length: 20 Molecular Weight: 6028<sup>4</sup> GC content: 50%

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 5.087 microMolar<sup>5</sup> and contains 30.7 micrograms.

Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.

RlnK 33.404 cal/(°K\*mol)

deltaH 158 Kcal/mol

deltaG 25.2 Kcal/mol

deltaS 412 cal/(°K\*mol)

Deprecated Hairpin/self dimerization 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 calculator supports the IUPAC alphabet, and sequence if it contains an

Paste the raw sequence or

TGCATTG  
GCATGAGGTTTGCTCTCATGAAC  
TCAAA  
TTGTA  
CGTTG  
CTAAA  
Sub

### Reverse Con

\*This  
\*You c  
>NM\_017460  
TCAGGCTCCF  
TTGAAGAAGT  
GGAGAAGTTC  
GCCAATGCAC  
CTTGTTCTTC

Sun 1 O  
Valid XH

### Oligonucleotide Properties Calculator

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

[Nucleotide base codes](#)  
TCA GGC TCC ACT TAC GGT G

Reverse Complement Strand(5' to 3') is:  
CAC CGT AAG TGG AGC CTG A

5' modification (if any)

3' modification (if any)

Select molecule  
ssDNA

50 nM Primer

1 Measured Absorbance at 260 nanometers

50 mM Salt (Na<sup>+</sup>)

Calculate

Swap Strands

BLAST

mfold

Physical Constants

Melting Temperature (T<sub>M</sub>) Calculations

Length: 19 Molecular Weight: 5779.8<sup>4</sup> GC content: 58%

1 ml of a sol'n with an Absorbance of 1 at 260 nm is 5.265 microMolar<sup>5</sup> and contains 30.4 micrograms.

Thermodynamic Constants Conditions: 1 M NaCl at 25°C at pH 7.

RlnK 33.404 cal/(°K\*mol)

deltaH 159.8 Kcal/mol

deltaG 25.6 Kcal/mol

deltaS 416.6 cal/(°K\*mol)

Deprecated Hairpin/self dimerization calculations

5 (Minimum base pairs required for single primer self-dimerization)

4 (Minimum base pairs required for a hairpin)

Check Self-Complementarity

# 2024 Exam

---

# Work with the following sequence:

---

MNWE LLWLLVLCALLLLLVQLLRFLRADGDLTLLWAEWQGRRPEWELTDMVWVTGASS  
GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPLDLTDTGSHEA  
ATKAVLQEFGRIDILVNNGGMSQRS LCMDTSLDVYRKLIELNYLGTVSLTKCVLPHMIER  
KQ GKIVTVNSILGII SVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN  
IVENSLAGEVT KTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY  
MPTWAWWITNKMGGKKRIENFKSGVDADSSYFKIFKTKHD

- Identify the sequence.
- Does this protein have any transmembrane regions?
- How many isoforms this protein has? Align them.
- How many cysteins does the sequence have?
- Design primers to amplify the CDS of the respective gene.

# Work with the following sequence:

---

MNWELLWLLVLCALLLLLVQLLRFLRADGDLTLLWAEWQGRRPEWELTDMVWVTGASS  
GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPLDLTDTGSHEA  
ATKAVLQEFGRIDILVNNGGMSQRSLCMDTSLDVYRKLIELNYLGTVSLTKCVLPHMIER  
KQGKIVTVNSILGII SVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN  
IVENSLAGEVTKTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY  
MPTWAWWITNKMGGKKRIENFKSGVDADSSYFKIFKTKHD

- Identify the sequence. → **BLASTp**
- Does this protein have any transmembrane regions? → **TMHMM, TopCons, Phobius...  
....not Protter (just figure)**
- How many isoforms this protein has? Align them. → **Uniprot for isoforms, multalin for alignment**
- How many cysteins does the sequence have? → **SMS: Protein Stats**
- Design primers to amplify the CDS of the respective gene. → **Manual design (OligoCalc for check-similar Tm)  
SMS:Reverse complement for R primer**

# Work with the following sequence:

- Identify the sequence.

DHRS7

BLAST<sup>®</sup> » blastp suite » results for RID-0Y787PHZ016 [Home](#) [Recent Results](#) [Saved Strategies](#) [Help](#)

[← Edit Search](#) [Save Search](#) [Search Summary ▾](#) [? How to read this report?](#) [▶ BLAST Help Videos](#) [↶ Back to Traditional Results Page](#)

**Job Title** Protein Sequence

**RID** [0Y787PHZ016](#) Search expires on 04-29 14:15 pm [Download All ▾](#)

**Program** BLASTP [?](#) [Citation ▾](#)

**Database** nr [See details ▾](#)

**Query ID** lcl|Query\_6710324

**Description** unnamed protein product

**Molecule type** amino acid

**Query Length** 339

**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**  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](#)

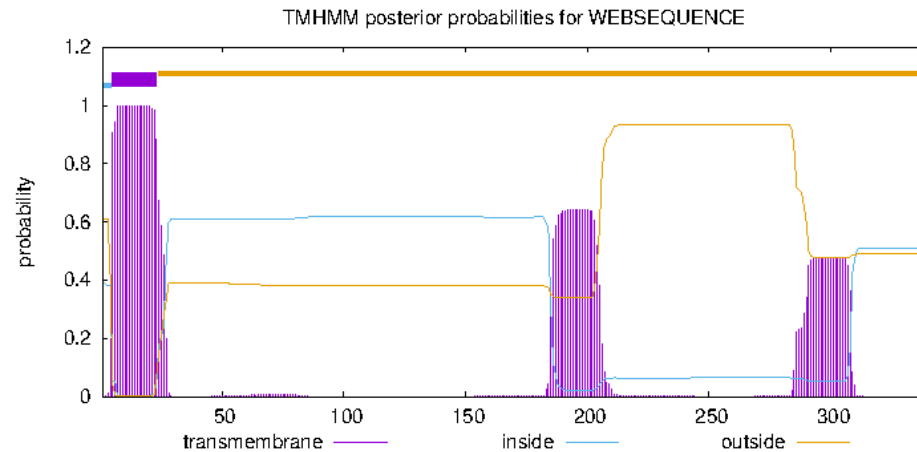
|                                     | Description                                                                                    | Scientific Name              | Max Score | Total Score | Query Cover | E value | Per. Ident | Acc. Len | Accession                      |
|-------------------------------------|------------------------------------------------------------------------------------------------|------------------------------|-----------|-------------|-------------|---------|------------|----------|--------------------------------|
| <input checked="" type="checkbox"/> | <a href="#">dehydrogenase/reductase SDR family member 7 isoform 1 precursor (Homo sapiens)</a> | <a href="#">Homo sapiens</a> | 693       | 693         | 100%        | 0.0     | 100.00%    | 339      | <a href="#">NP_057113.1</a>    |
| <input checked="" type="checkbox"/> | <a href="#">dehydrogenase/reductase SDR family member 7 isoform X1 (Pan paniscus)</a>          | <a href="#">Pan paniscus</a> | 691       | 691         | 100%        | 0.0     | 99.71%     | 339      | <a href="#">XP_008955244.1</a> |

# Work with the following sequence:

- Does this protein have any transmembrane regions?

Yes 1: **TMHMM result**

```
WEBSEQUENCE Length: 339
WEBSEQUENCE Number of predicted TMHs: 1
WEBSEQUENCE Exp number of AAs in TMHs: 44.121
WEBSEQUENCE Exp number, first 60 AAs: 20.90194
WEBSEQUENCE Total prob of N-in: 0.39052
WEBSEQUENCE POSSIBLE N-term signal sequence
WEBSEQUENCE TMHMM2.0 inside 1 4
WEBSEQUENCE TMHMM2.0 TMhelix 5 23
WEBSEQUENCE TMHMM2.0 outside 24 339
```





# Work with the following sequence:

- How many isoforms this protein has? Align them.

## 2 isoforms in uniprot

### Sequence & Isoform<sup>i</sup>

Align isoforms (2) [Add isoforms](#)

Sequence status<sup>i</sup> | Complete

This entry describes 2 isoforms<sup>i</sup> produced by **Alternative splicing**.

### Q9Y394-1

This isoform has been chosen as the **canonical** sequence. All positional informati

Name 1

Tools

[Download](#)

[Add](#)

[Highlight](#)

[Copy sequence](#)

Length 339

Mass (Da) 38,299

10182030405060708090100110120130

MINWELLWLVLVLCALLLLVQLRLFRADGDLTLLEHAGRPPEMELTDMYVWYTGSSGDEELAYQLSKLGYSVLSSRRPHELEKPKRCLLENGHKEKEDILVLPDLTDTGSHEHATKAVLQDFG

150160170180190

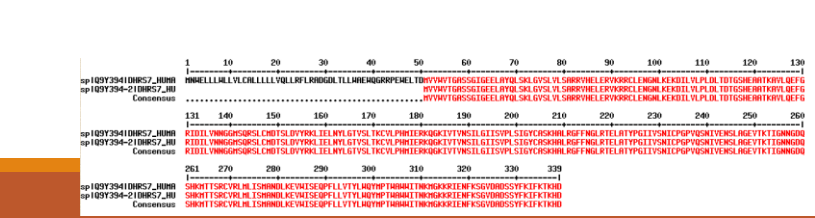
MSQRSLSMDTSLDVYRKLIENLYLGTVSLTKCVLPHEIERKQGIKIVTVNSILK

200210220230240250260

LKEVWISQEPFLLVTVLWQYMPMTAWMIITNKMGGKRIENFKSGVDADSSYFK

### Q9Y394-2

Name 2  
See also sequence in UniParc or sequence clusters in UniRef



## 4 isoforms in NCBI

### DHRS7 – dehydrogenase/reductase 7

*Homo sapiens* (human)

Also known as: CGI-86, SDR34C1, retDSR4, retSDR4

Gene ID: 51635

[RefSeq products](#)

[Orthologs](#)

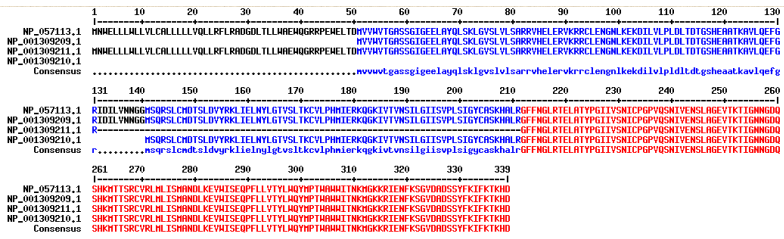
[Genome Data Viewer](#)

New - Visualize gene across multiple species

### RefSeq Sequences

Showing 4 of 4 (by status, accession number)

| Transcript     | nt    | Protein        | aa  | Isoform | Status      |
|----------------|-------|----------------|-----|---------|-------------|
| NM_016029.4    | 1,956 | NP_057113.1    | 339 | 1       | MANE Select |
| NM_001322280.2 | 1,927 | NP_001309209.1 | 289 | 2       | curated     |
| NM_001322281.2 | 1,963 | NP_001309210.1 | 199 | 3       | curated     |
| NM_001322282.2 | 1,716 | NP_001309211.1 | 259 | 4       | curated     |



# Work with the following sequence:

- How many cysteins does the sequence have?

7 cysteins: [Sequence Manipulation Suite:](#)

## Protein Stats

Protein Stats returns the number of occurrences of each residue in the sequence you enter. Percentage totals are also given for each res

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

```
MNWE LLWLLVLCALLLLLVQLRFLRADGDLTLLWAEWQGRPEWELTDMVVWVTGASS
GIGEELAYQLSKLGVSLVLSARRVHELERVKRRCLENGNLKEKDILVLPDLTDTGSHEA
ATKAVLQEFGRIDLNVNNGGMSQRS LCM DTS LDVYRKLIELNYLGTVSLTKCVLPHMIER
KQ GKIVTVNSILGIISVPLSIGYCASKHALRGFFNGLRTELATYPGIIVSNICPGPVQSN
IVENSLAGEVTKTIGNNGDQSHKMTTSRCVRLMLISMANDLKEVWISEQPFLLVTYLWQY
MPTWAWWITNKMGGKRIENFKSGVDADSSYFKIFKTKHD
```

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

\*You c

Sequence Manipulation Suite – Škola – Microsoft Edge

about:blank

Protein Stats results

Results for 339 residue sequence "Untitled" starting "MNWE LLWLL"

| Pattern: | Times found: | Percentage: |
|----------|--------------|-------------|
| A        | 16           | 4.72        |
| B        | 0            | 0.00        |
| C        | 7            | 2.06        |
| D        | 14           | 4.13        |

# Work with the following sequence:

- Design primers to amplify the CDS of the respective gene.

## DHRS7 – dehydrogenase/reductase 7

*Homo sapiens* (human)

Also known as: CGI-86, SDR34C1, retDSR4, retSDR4

Gene ID: 51635

RefSeq products   Orthologs   Genome Data Viewer

New - Visualize gene across multiple species

### RefSeq Sequences

Showing 4 of 4 (by status, accession number)

| Transcript     | nt    | Protein        | aa  | Isoform | Status      |
|----------------|-------|----------------|-----|---------|-------------|
| NM_016029.4    | 1,956 | NP_057113.1    | 339 | 1       | MANE Select |
| NM_001322280.2 | 1,927 | NP_001309209.1 | 289 | 2       | curated     |
| NM_001322281.2 | 1,963 | NP_001309210.1 | 199 | 3       | curated     |
| NM_001322282.2 | 1,716 | NP_001309211.1 | 259 | 4       | curated     |

F: ATG AAC TGG GAG CTG CTG C

R: TCA GTC ATG TTT TGT CTT AAA GAT TTT

FASTA

### Homo sapiens dehydrogenase/reductase 7 (DHRS7),

NCBI Reference Sequence: NM\_016029.4

GenBank   Graphics

>NM\_016029.4:90-1109 Homo sapiens dehydrogenase/reductase 7 (DHRS7), transcribed mRNA

```
ATGAAC TGGAGCTGCTGCTGGCTGCTGGTGTGTCGCGCTGCTCCTGCTCTTGGTGCAGCTGCTGC
GCTTCTCTGAGGGCTGACGGCGACCTGACGCTACTATGGGCCGAGTGGCAGGGACGACGCCAGAATGGGA
GCTGACTGATATGGTGGTGGGTGACTGGAGCCTCGAGTGAATTGGTGAGGAGCTGGCTTACAGTTG
TCTAACTAGGAGTTTCTCTTGTGCTGCAGCAGAAGATGCATGAGCTGGAAGGGTGAAAGGAAGAT
GCCTAGAGAATGGCAATTTAAAGAAAAAGATATACTTGTTTTGCCCTTGACCTGACCGACACTGGTTTC
CCATGAAGCGGCTACCAAGCTGTTCTCAGGAGTTTGGTAGAATCGACATTCTGGTCAACATGGTGGA
ATGTCACAGCGTTCTCTGTGCATGGATACCAGCTTGGATGTCTACAGAAAGCTAATAGAGCTTAACTACT
TAGGGACGGTGTCTTGCAGAAATGTGTTCTGCTCATGATCGAGAGGAAGCAAGGAAAGATTGTTAC
TGTGAATAGCATCTGGGTATCATATCTGTACCTCTTCCATTGGATACTGTGCTAGCAAGCATGCTCTC
CGGGGTTTTTTAATGGCTTTCGAACAGAAGCTTGCCACATACCCAGGTATAATAGTTCTTAACATTTGCC
CAGGACCTGTGCAATCAAAATATTGGAGAATTCCCTAGCTGGAGAAGTCACAAAGACTATAGGCAATAA
TGGAGACCAGTCCCAAGATGACAACAGCTGTTGTGTGCGGCTGATGTTAATCAGCATGGCCAATGAT
TTGAAAGAAGTTTGGATCTCAGAACAACCTTTCTTGTAGTAGAACATATTTGTGGCAATACATGCCAACCT
GGGCTGGTGGATAACCAACAGATGGGGAAGAAAGGATTGAGAACTTTAAGAGTGGTGTGGATGCAGA
CTCTTCTATTTTAAATCTTTAAGACAAAACATGACTGA
```

### Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below  
GC and Molecular Weight calculations are for single-stranded DNA or RNA

Nucleotide base codes  
ATG AAC TGG GAG CTG CTG C

Reverse Complement Strand(5' to 3') is:  
GCA GCA GCT CCC AGT TCA T

5' modification (if any) 3' modification (if any) select ss/ds and DNA or RNA  
[5] (Minimum base pairs required for single primer self-dimerization)  
[4] (Minimum base pairs required for a hairpin)  
50 mM Primer  
50 mM Salt (Na<sup>+</sup>)

Calculate Swap Strands BLAST Check Self-Complementarity

Physical Constants  
Length: 19 bases  
GC content: 58 %  
Molecular Weight: 5868.9  
1 ml of a sol'n with an Absorbance of 1 at 260 nm is 4.839 microMolar and contains 28.4 micrograms

Melting Temperature (T<sub>m</sub>) Calculations  
1 53 °C (Basic)  
2 50 °C (Salt Adjusted)  
3 55 °C (Nearest Neighbor)

Thermodynamic Constants  
Conditions: 6 M NaCl at 36°C at pH 7

### Sequence Manipulation Suite:

#### Reverse Complement

Reverse Complement converts a DNA sequence into its reverse, complement, or 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 characters.

```
GCCAATGAT
TTGAAGAAGATTTGGATCTCAGAACCACTCTTCTGTAGTAACATATTTGGCAATACA
TGCCCAACT
GGGCTGGTGGATACCAACAAGATGGGGAAGAAAGATTGAGAACTTTAAGAGTGGTGT
GGATGCAGA
CTCTTCTATTTTAAATCTTTAAGACAAAACATGACTGA
```

Submit Clear Reset

Sequence Manipulation Suite - Skola - Microsoft Edge

about:blank

#### Reverse Complement results

>NM\_016029.4:90-1109 Homo sapiens dehydrogenase/reductase 7 (DHRS7), transcribed mRNA  
TCAGTCACTGTTTTGTCTTAAAGATTTTAAATAAGAGAAGTCTGCGTACCCACACCACTCTT  
AAGATCTCTCAATCCTTTTCTCCCACTCTGTTGGTTATCCACAGGCCAGTGTGGCAT  
GTATTGGCCACAAATATGTTACTAACAAAGAAAGTGTGTTCTGAGATCCCAACTTCTTCAA  
ATCATTGGCCATGCTGATTAAACATCAGCCGCACACACAGACTGGTTGTCTCTGTGGGA  
CTGGTCTCCATTATTGCTATAGTCTTTGTGACATCTCAGCTAGGGAAATCTCCACAAT  
ATTTGATTGCACAGTCTCTGGGCAATGTTAGAACTATTATACCTGGTATGTGGCAAG  
TCTCTTTGGAAGGCCATTAAGAAACCCGGAGAGATGCTCTCTAGCACAGTATCCAA  
GGAAAGAGTACAGATATGATACCCAGGATGCTATTACAGTAACCAATCTTCCCTGCT  
CCTCTGATCATGTGAGGCAGAACACATTTTGTCAAGGACACCGTCCCTAAGTAGTTAAG  
CTCTATTAGCTTCTGTAGACATCCCACTGGTATCCATGCACAGAGAACGCTGGGACAT  
TCCACATTGTTGACCAAGATGTGATTCTACCAAACTCCTGGAGAACAGCTTTGGTAGC  
CGCTTCATGAGAACAGTGTCTGTCAGTCAAGGGGCAAAAGATATATCTTTTCTTT  
TAAATGCGCATCTCTAGGCACTCTTTTCAACCTTTTCAGCTCAGCACTCTCTGGC  
TGACAGCACAAGAGAACTCTAGTTTAGACAACTGGTAAGCCAGCTCCTCCCAATGCC  
ACTCGAGGCTCCAGTCACCCACACACCATATCAGTCAGCTCCCATCTCTGGGCTCGTC  
CTGCCACTCGGCCATAGTAGCTGAGCTCAGCTCGCCCTCAGCCCTCAGGAAGCGCAGCAGCTG  
CACCAGAGCAGGACAGCGGCCACAGCAGCAGCCACAGCAGCAGCTCCAGTTCTAT

### Oligonucleotide Properties Calculator

Enter Oligonucleotide Sequence Below  
GC and Molecular Weight calculations are for single-stranded DNA or RNA

Nucleotide base codes  
TCA GTC ATG TTT TGT CTT AAA GAT TTT AA

Reverse Complement Strand(5' to 3') is:  
TTA AAA TCT TTA AGA CAA AAC ATG ACT GA

5' modification (if any) 3' modification (if any) select ss/ds and DNA or RNA  
[5] (Minimum base pairs required for single primer self-dimerization)  
[4] (Minimum base pairs required for a hairpin)  
50 mM Primer  
50 mM Salt (Na<sup>+</sup>)

Calculate Swap Strands BLAST Check Self-Complementarity

Physical Constants  
Length: 29 bases  
GC content: 24 %  
Molecular Weight: 8886.9  
1 ml of a sol'n with an Absorbance of 1 at 260 nm is 4.839 microMolar and contains 28.4 micrograms

Melting Temperature (T<sub>m</sub>) Calculations  
1 52 °C (Basic)  
2 60 °C (Salt Adjusted)  
3 54 °C (Nearest Neighbor)

# Exam test:

## Maximum 10 points

8.5-10p: 1

7-8p: 2

5-6.5p: 3

### Version 1

Find human protein sequence **FGF3**

- What is the accession number and function of this protein?
- How many valines does the sequence contain?
- How long is the longest fragment after trypsin digestion?
- How many human similar reference proteins are in the databases?
- Does the sequence contain any transmembrane regions?

Download sequence **NM\_001757.3**

- What does this sequence encode ?
- Is there a complementary sequence to primer R1?

>R1  
GCTCTGACGCTCATGATGC

- In which exon is the complementary sequence for this primer?
- Is the primer suitable for PCR with Ta=60°C?
- Design suitable F primer for this experiment.

# Download from database the sequence NM\_001282692.1

---

- What does this mRNA encode?
- How long is the second exon?
- Will the CDS of this sequence be cleaved by the following enzymes: BamHI, EcoRI or Scal?
- What fragments are formed after cleavage by all these enzymes at once?
- Compare how similar is the corresponding protein to the respective mouse homologue?

# Download from database the sequence NM\_001282692.1

---

- What does this mRNA encode? → Finding the sequence in NCBI
- How long is the second exon? → NCBI: highlight sequence features or Graphic
- Will the CDS of this sequence be cleaved by the following enzymes: BamHI, EcoRI or Scal? → Get CDS only and SMS: Restriction summary
- What fragments are formed after cleavage by all these enzymes at once? → SMS: Restriction digest
- Compare how similar is the corresponding protein to the respective mouse homologue? → Get protein and mouse protein and Lalign or Needle pairwise comparison

# Download from database the sequence NM\_001282692.1

- What does this mRNA encode?

FMO1

GenBank ▾

## Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), transcript variant 1, mRNA

NCBI Reference Sequence: NM\_001282692.1

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

[Go to:](#) ☐

|            |                                                                                                    |         |      |        |                 |
|------------|----------------------------------------------------------------------------------------------------|---------|------|--------|-----------------|
| LOCUS      | NM_001282692                                                                                       | 2198 bp | mRNA | linear | PRI 09-FEB-2025 |
| DEFINITION | Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), transcript variant 1, mRNA. |         |      |        |                 |
| ACCESSION  | NM_001282692                                                                                       |         |      |        |                 |

Download from database the sequence NM\_001282692.1

- How long is the second exon?

$$298..486 : 486-298+1 = 189\text{nt}$$

GenBank

Send to: 

Change region shown

Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), transcript variant 1, mRNA

NCBI Reference Sequence: NM\_001282692.1

FASTA Graphics

Go to:

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

NM\_001282692

2198 bp

mRNA

linear

PRI 09-FEB-2005

Homo sapiens

Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), transcript variant 1, mRNA.

NM\_001282692

NM\_001282692.1

RefSeq.

Homo sapiens (human)

121

gcttgtaga

gcagccaagg

gtgggggga

122

gcttgtaga

aaaagcctg

cttcattct

ctcatcgag

agaacatg

gc caagcaggt

181

gcattctg

gagctggg

cagcgccct

gcctccat

catgctgt

ctt ggaagaa

241

ctggagcc

ctgcttga

gagagcag

gacctggg

ggctggag

attaccgaa

301

catgtga

agggcag

cagctctac

agctctg

gtgggtt

cttcaaac

ctgcaagg

361

atgtctgt

actcagct

tccattcca

gaagattac

caaatatg

tgc caaatct

421

caattctg

atactcaa

aagtatga

aaccattt

gaccttga

acacattca

481

ttcagcca

aagctgcg

tgtaacaaa

tgctcagt

ctgctgtc

tgccaatgg

541

gaggtgtc

ctatgcag

agagaagca

gagtcagca

ctttgtgc

tgctatggt

601

tgactcgt

ttcttctaa

tctctattg

ccactgg

ctcttcag

tattatgac

661

ttaaagcg

agtaactta

tggcgacaa

tataagac

cagatatat

taaggacaa

721

agagctct

tgattga

gggaattct

ggcacagac

ttgcttga

ggcagccac

781

ctggcgaa

aggtgttct

cagcacacc

ggagggag

gggtatcg

ccgaatttt

841

gactgggc

accatggg

catgtgtt

atgacagc

ttcagaac

gttgaaat

901

tccctccaa

cccaattgt

gactgttg

atggagcaa

agataaac

ctggctaac

961

catgcaaa

acggttaa

accagaag

aggactcag

tgaagagt

ttgtctaa

1021

gatgagct

caggagcat

catactgg

aaagtgtt

tcaggcca

cataaaag

1081

gtaaaagaa

actctgtcat

attacaat

acttcaag

aagccctat

tgactcat

1141

gtctttgca

ctggatacac

attgtcttc

cccttctg

atgagctgt

agtgaagt

1201

gaagatg

agccctact

gtacagat

atcttctg

catactga

aaagccaac

1261

ctggcatta

ttggctcat

caaaccttg

ggctcatga

tactacag

agaacacaa

1321

ctcgtggg

ctgttgagt

cttgaagat

gtaaatagt

taccacacc

aagtgctac

1381

atagagaaa

taagtacag

gaagaagaa

aagccaggt

gcttggtct

gtgtctacg

1441

aaagtttaa

atcagatt

tatcacat

atagatgac

tcttgacct

tataatgca

1501

aaaccacac

ttgtctcat

gtctcaag

gctcacac

tggtctgac

gctctcttt

1561

ggccactgc

cacatacca

gttctcgt

actggccag

gaaatggga

aggagccga

1621

aatgcatca

tgaccacgt

ggagccaac

ttcagggt

ctaaagct

agttgtaca

1681

gagcttcat

ctctcttga

aagtttct

aaagtctta

gctttctgc

ttgtctgtg

1741

gctattttt

tgattttct

ataagtaa

gactctcaa

atggagatg

cacagagtg

1801

attacaagt

ctccaattc

tctctacag

caatattgc

ttcacagtt

taaaactgat

1861

tcaaatagta

aagcgacc

tctgccttc

ctggctg

caggctac

catgtgtat

1921

cttgagctc

tccagcttc

acttctaag

ctagagaag

ataactaag

ctttgtgca

1981

tttgaaggt

gttgaaga

tacagttca

tttgaaga

aaagctgt

ctgacagac

2041

cttgagcat

catactctt

tccatata

actatttca

cagatctca

ctaaacacc

2101

ttaactcaa

aagtatgtg

ttgtctga

atggtgtct

tatagtact

gcttattaa

2 of 8

1 segment

298..486

/gene="FMO1"

/inference="alignment;Sp|gn:2

exon

Feature

Details

# Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), trar

NCBI Reference Sequence: NM\_001282692.1

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

Genes

NP\_001269621.1

FMO1

exon

misc\_feature Features

upstream in...

polyA\_site Features

regulatory Features

Cited Variations, dbSNP b157 v2

rs10912675 | T/C

rs1126692 | A/C/G

rs606396

BLAST nr: [NM\\_001282692.1 \(298..486\)](#)

BLAST to Genome: [NM\\_001282692.1 \(298..486\)](#)

FASTA record: [NM\\_001282692.1 \(298..486\)](#)



# Download from database the sequence NM\_001282692.1

- Will the CDS of this sequence be cleaved by the following enzymes: BamHI, EcoRI or Scal?

BamHI, Scal- cuts 1x

EcoRI –does not cut

|               |                   |
|---------------|-------------------|
| BamI tgg cca  | 17, 381, 1056, 11 |
| BamHI g gatcc | 1378              |
| EcoRI ggaattc | none              |
| EcoRI glaattc | none              |
| Scal agt act  | 521               |

FASTA

## Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), transcript variant 1, mRNA

NCBI Reference Sequence: NM\_001282692.1

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

```
>NM_001282692.1:154-1764 Homo sapiens flavin containing dimethylaniline monooxygenase 1 (FMO1), transcript variant 1, mRNA
ATGCAGGAGAACATGGCCAAGCGAGTTGCCATTGTGGGAGCTGGGGTCAGCGGCCTGGCCTCCATCAAGT
GCTGTCTGGAAGAAGGACTGGAGCCACCTGCTTTGAGAGGAGCGATGACCTTGGGGGCTGTGGAGATT
CACCGAACATGTTGAAGAAGGCAGAGCCAGTCTCTACAAGTCTGTGGTTTCCAACAGCTGCAAGGAGATG
TCTTGTTACTCAGACTTTCCATTCCAGAAGATTATCCAAACTATGTGCCAAATTCATTCCTCGGAAT
ATCTCAAAATGTATGCAAAACCACTTTGACCTTCTGAAACACATTCAATTCAAGACCAAAAGTCTGCAGTGT
AACAAAATGCTCAGATTCTGCTGTCTCTGGCCAATGGGAGGTGGTCACTATGCATGAAGAGAAGCAAGAG
TCAGCCATCTTTGATGCTGTCATGGTCTGCACTGGCTTTCTTACTAATCCTTATTTGCACTGGATTCTCT
TTCCAGGTATTAATGCTTTTAAAGGCGAGTACTTTCATAGCGGCGAATATAGCATTCAGATATATTTAA
```

Download from database the sequence NM\_001282692.1

- What fragments are formed after cleavage by all these enzymes at once?

Three fragments: 857 + 520 + 234 bp

### Sequence Manipulation Suite:

## Restriction Digest

Restriction Digest cleaves a DNA sequence in a virtual restriction digest, with one, two, or three restriction enzymes that produced them. You can digest linear or circular molecules, and even a mixture of molecules (by entering

Paste the raw sequence or one or more FASTA sequences into the text area below. Input limit is 100,000,00

GCCCAGAAAT  
 GCCATCATGACGCCAAGTGGGACCGAACATTCAAGGTCATCAAGGCTCGAGTTGTACAAGAGT  
 CTCATCATTC  
 CCTTTGGAAGATTTCTTAAAGCTTTAGCTTCTGCGCTTGCTGTGGCTATTTTCTGAT  
 TTTCCTATA  
 A

- Treat sequences as  molecules.
- Digest with  and  and

[illegible]

# Download from database the sequence NM\_001282692.1

- Compare how similar is the corresponding protein to the respective mouse homologue?

Input sequence ⓘ

Sequence type

☒ Protein ☐ DNA

Paste your first sequence here - or use the example sequence

```
>sp|Q01740|FMO1_HUMAN Flavin-containing monooxygenase 1 OS=Homo sapiens C
MAKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEGRASLYKSVVS
NSCKEMSCYSDFPFPEYDYPNYVPNSQFLEYLKMYANHFDLLKHIQFKTKVCSVTKCSDSA
VSGQWEVVTMHEEKQESAIFDAVMVCTGFLTNPYLPLDSFPGINAFKQYFHSRQYKHPD
IFKDKRVLVIGMNSGTDIAVEASHLAEKVFLSTTGGGWVISRIFDSGYPWDMVMFMTRFQ
NMLRNSLPTPIVTWLMERKINNWLHNHANYGLIPEDRTQLKEFVLNDELPGRIITGKVIFR
PSIKEVKENSIVFNNTSKEEPIDIIVFATGYTFAFPFLDES VVKVEDGQASLYKYIFPAH
LQSTLALQTLKPLQCMPTQCTQPMWVWVLYKQVWVLRDQVAFETWLRKVKQSMF
```

Zvolit soubor Nevybrán žádný soubor

Paste your second sequence here - or use the example sequence

```
>sp|P50285|FMO1_MOUSE Flavin-containing monooxygenase 1 OS=Mus musculus O
MVKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEGRASLYKSVVS
NSSREMSCYPDFPFPEYDYPNFVPSLFLLEYLKLYSTQFNLQRCIYFNTKVCISITKRPDFA
VSGQWEVVTVTNGKQNSAIFDAVMVCTGFLTNPHLPLDSFPGILTFKGEYFHSRQYKHPD
IFKDKRVLVVG MNSGTDIAVEASHLAKKVFLSTTGGAWVISRVFDSGYPWDMIFMTRFQ
NMLRNLLPTPIVSWLISKMNWFWNHVNYGVAPEDRTQLREPVLNDELPGRIITGKVFIK
PSIKEVKENSIVFNNTSKEEPIDIIVFATGYTFAFPFLDES VVKVEDGQASLYKYIFPAH
```

```

#
Aligned_sequences: 2
1: FMO1_HUMAN
2: FMO1_MOUSE
Matrix: EBL0SUM62
Gap_penalty: 10.0
Extend_penalty: 0.5
#
Length: 532
Identity: 443/532 (83.3%)
Similarity: 485/532 (91.2%)
Gaps: 0/532 (0.0%)
Score: 2396.0
#
#
#=====
```

|            |     |                                                     |     |
|------------|-----|-----------------------------------------------------|-----|
| FMO1_HUMAN | 1   | MAKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEG  | 50  |
| FMO1_MOUSE | 1   | MVKRVAIVGAGVSGLASIKCCLEEGLEPTCFERSDDLGGWLRFTEHVEEG  | 50  |
| FMO1_HUMAN | 51  | RASLYKSVVSNCKEMSCYSDFPFPEYDYPNYVPNSQFLEYLKMYANHFDL  | 100 |
| FMO1_MOUSE | 51  | RASLYKSVVSNSSREMSCYPDFPFPEYDYPNFVPSLFLLEYLKLYSTQFNL | 100 |
| FMO1_HUMAN | 101 | LKHIQFKTKVCSVTKCSDSAVSGQWEVVTMHEEKQESAIFDAVMVCTGFL  | 150 |
| FMO1_MOUSE | 101 | QRCIYFNTKVCISITKRPDFAVSGQWEVVTVTNGKQNSAIFDAVMVCTGFL | 150 |