

Biological Database Search and Retrieval of Protein Sequence

Dr. Pradeep Kumar
Associate Professor
Department of Biotechnology
VBS Purvanchal University, Jaunpur

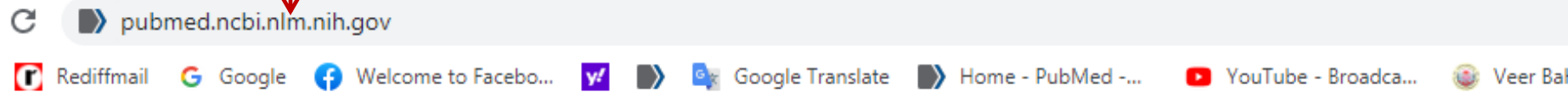
NCBI Protein

<https://www.ncbi.nlm.nih.gov/protein>

- The National Center for Biotechnology Information Protein database has sequences from
- GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>)
- RefSeq (<https://www.ncbi.nlm.nih.gov/refseq/>)
- TPA (<https://www.ncbi.nlm.nih.gov/genbank/tpa/>)
- SwissProt(<https://www.uniprot.org/uniprot/?query=swiss+prot&sort=score>),
- PIR(<https://proteininformationresource.org/>),
- PRF(<https://www.prf.or.jp/aboutdb-e.html>)
- PDB (www.rcsb.org/pdb).

Step 1

Open any web browser and type/enter pubmed.ncbi.nlm.nih.gov in the address bar and Click enter ↵



The Google logo, consisting of the word "Google" in its characteristic multi-colored font (blue, red, yellow, blue, green, red).

Search Google or type a URL



pubmed.ncbi.nlm.nih.gov homepage will open

The screenshot shows the NCBI homepage with a prominent red banner at the top containing COVID-19 information. Below the banner is a navigation menu on the left and a main content area with a 'Welcome to NCBI' message and six service tiles: Submit, Download, Learn, Develop, Analyze, and Research. A 'Popular Resources' list is on the right, and an 'NCBI News & Blog' section is at the bottom right.

ncbi.nlm.nih.gov

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

All Databases [Search]

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

NCBI Home
Resource List (A-Z)
All Resources
Chemicals & Bioassays
Data & Software
DNA & RNA
Domains & Structures
Genes & Expression
Genetics & Medicine
Genomes & Maps
Homology
Literature
Proteins
Sequence Analysis
Taxonomy
Training & Tutorials
Variation

Welcome to NCBI
The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

Analyze
Identify an NCBI tool for your data analysis task

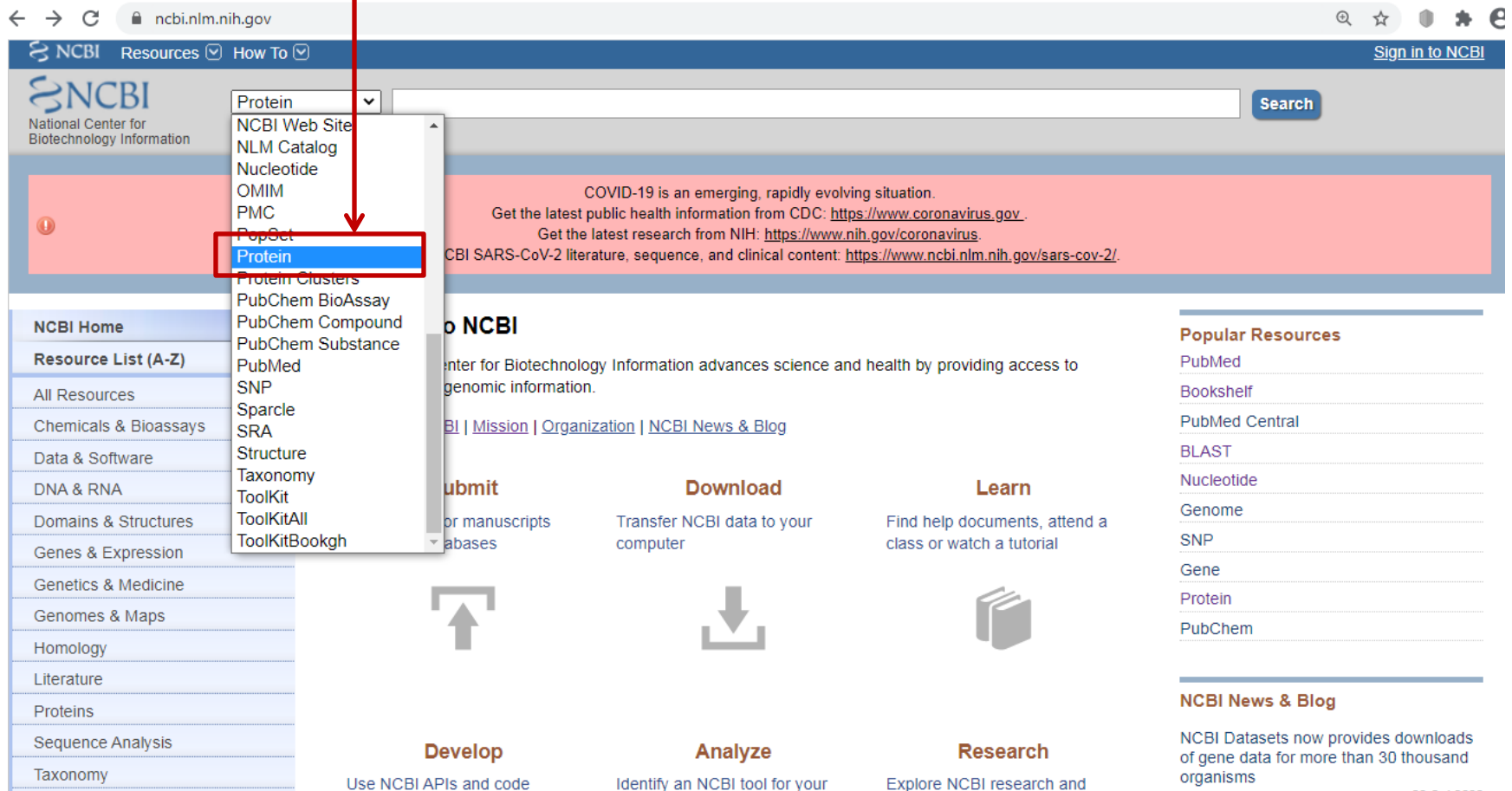
Research
Explore NCBI research and collaborative projects

Popular Resources
PubMed
Bookshelf
PubMed Central
BLAST
Nucleotide
Genome
SNP
Gene
Protein
PubChem

NCBI News & Blog
NCBI Datasets now provides downloads of gene data for more than 30 thousand organisms
08 Oct 2020
NCBI Datasets now offers Gene tables:
Recent enhancements in Genome Workbench version 3.5.0
08 Oct 2020
New Features Version 3.5.0 of Genome Workbench NCBI's sequence annotation

Step 2

Click on dropdown menu(All databases) and select protein database by clicking on Protein



The screenshot shows the NCBI website interface. A dropdown menu is open, displaying a list of databases. The 'Protein' option is highlighted with a red box, and a red arrow points to it from the text above. The website header includes the NCBI logo, navigation links, and a search bar. A COVID-19 information banner is visible in the background. The main content area features sections for 'Develop', 'Analyze', and 'Research', along with a 'Popular Resources' sidebar on the right.

NCBI Resources How To Sign in to NCBI

Protein

- NCBI Web Site
- NLM Catalog
- Nucleotide
- OMIM
- PMC
- PopSet
- Protein**
- Protein Clusters
- PubChem BioAssay
- PubChem Compound
- PubChem Substance
- PubMed
- SNP
- Sparcle
- SRA
- Structure
- Taxonomy
- ToolKit
- ToolKitAll
- ToolKitBookgh

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Develop
Use NCBI APIs and code

Analyze
Identify an NCBI tool for your

Research
Explore NCBI research and

Popular Resources

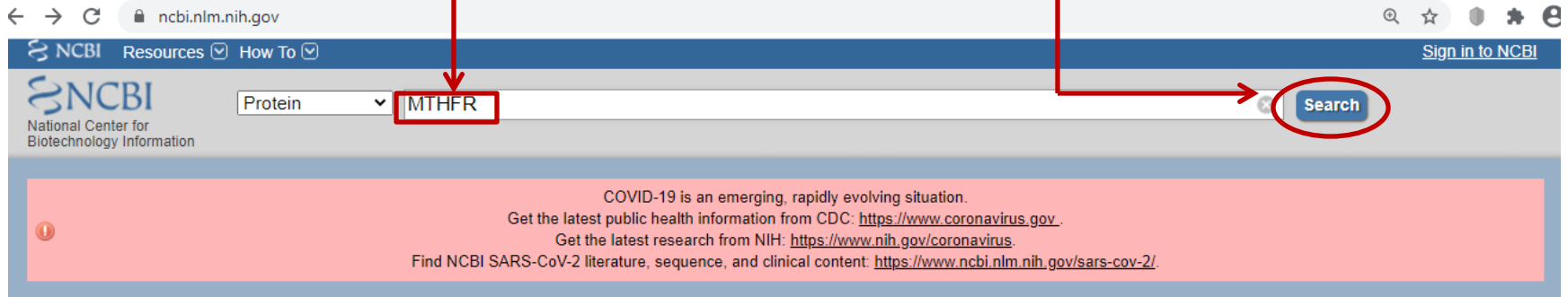
- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog

NCBI Datasets now provides downloads of gene data for more than 30 thousand organisms

Step 3

Enter the name of the protein in the search bar (MTHFR in this example) and click on search



- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)

Submit

Deposit data or manuscripts into NCBI databases



Develop

Use NCBI APIs and code

Download

Transfer NCBI data to your computer



Analyze

Identify an NCBI tool for your

Learn

Find help documents, attend a class or watch a tutorial



Research

Explore NCBI research and

Popular Resources

[PubMed](#)

[Bookshelf](#)

[PubMed Central](#)

[BLAST](#)

[Nucleotide](#)

[Genome](#)

[SNP](#)

[Gene](#)

[Protein](#)

[PubChem](#)

NCBI News & Blog

NCBI Datasets now provides downloads of gene data for more than 30 thousand organisms

<https://www.ncbi.nlm.nih.gov/protein/?term=MTHFR> page will open
scroll down the page to see your protein of interest

NCBI Resources How To Sign in to NCBI

Protein Protein MTHFR Search

Create alert Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

Species Summary 20 per page Sort by Default order Send to Filters: Manage Filters

Animals (1,019)
Plants (311)
Fungi (3,102)
Protists (350)
Bacteria (56,850)
Archaea (1,001)
Viruses (10)
Customize ...

Source databases
PDB (42)
RefSeq (15,751)
UniProtKB / Swiss-Prot (15)
Customize ...

Genetic compartments
Mitochondrion (1)
Plasmid (101)

Sequence length
Custom range...

Molecular weight
Custom range...

Release date
Custom range...

Revision date
Custom range...

Clear all

GENE Was this helpful?

[MTHFR – methylenetetrahydrofolate reductase](#)

[Homo sapiens \(human\)](#)
GeneID: 4524

[RefSeq transcripts \(5\)](#) [RefSeq proteins \(5\)](#) [RefSeqGene \(1\)](#) [PubMed \(3,549\)](#)

[Orthologs](#) [Genome Browser](#) [BLAST](#) [Download](#)

RefSeq Sequences +

Items: 1 to 20 of 63429

[MTHFR \(Cervus elaphus hippelaphus\)](#)

1. 610 aa protein
Accession: OWK08248.1 GI: 1207830732
[BioProject](#) [Nucleotide](#) [Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

<< First < Prev Page 1 of 3172 Next > Last >>

Results by taxon
Top Organisms [Tree](#)
[Acinetobacter baumannii \(6362\)](#)
[Salmonella enterica \(2896\)](#)
[Escherichia coli \(2761\)](#)
[Clostridioides difficile \(1847\)](#)
[Klebsiella pneumoniae \(942\)](#)
All other taxa (48621)
More...

Find related data
Database: Select
Find items

Search details
MTHFR[All Fields]
Search See more...

Recent activity
Turn Off Clear

Total number of items available in the protein database related to our search i.e. "MTHFR"

<https://www.ncbi.nlm.nih.gov/protein/?term=MTHFR>

(No. 12 is our protein of interest in this example)

Step 4

Click on the FASTA to obtain the Protein sequence of the *methylenetetrahydrofolate reductase isoform 2[Homo sapiens]* (Sequence No.12)

[methylenetetrahydrofolate reductase isoform 2 \[Homo sapiens\]](#)

12. 656 aa protein

Accession: NP_005948.3 GI: 87240000

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

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

[methylenetetrahydrofolate reductase isoform 1 \[Homo sapiens\]](#)

13. 697 aa protein

Accession: NP_001317287.1 GI: 1060099355

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

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

[methylenetetrahydrofolate reductase isoform b \[Mus musculus\]](#)

14. 654 aa protein

Accession: NP_034970.2 GI: 31543271

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

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

[methylenetetrahydrofolate reductase isoform a \[Mus musculus\]](#)

15. 695 aa protein

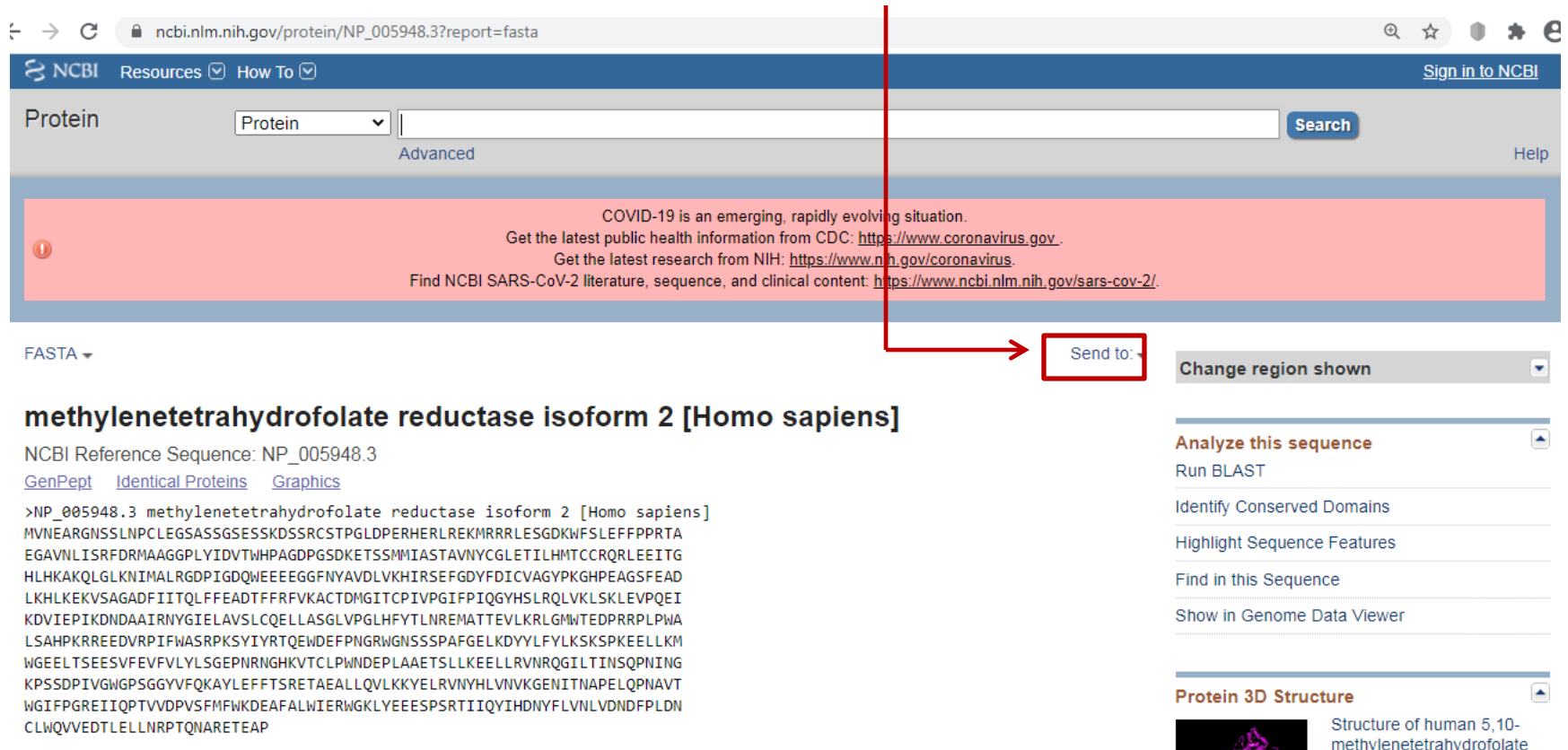
Accession: NP_001155270.1 GI: 239985494

[BioProject](#) [Nucleotide](#) [PubMed](#) [Taxonomy](#)

https://www.ncbi.nlm.nih.gov/protein/NP_005948.3?report=fasta will open
and the Protein sequence of
methylenetetrahydrofolate reductase isoform 2[Homo sapiens] will open.

Step 5

Click on send to



NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

FASTA

methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]

NCBI Reference Sequence: NP_005948.3

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

```
>NP_005948.3 methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]
MVNEARGNSSLNPCLEGSASSGSESSKDSRCSTPGLDPERHERLREKMRRLSEGDKWFSLFFPPRTA
EGAVNLIISRFDRMAAGGPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILHMTCCRQRLEEITG
HLHKAKQLGLKNIMALRGDPIGDQWEEEEEFGFNAYVDLVKHIRSEFGDYFDICVAGYPKGHPPEAGSFEAD
LKHLKEKVSAGADFIITQLFFEADTFFRFVKACTIONMGITCPVPGIFPIQGYHSRLQLVKLSKLEVPQEI
KDVIIEPIKDNDAAIRNYGIELAVSLCQELLASGLVPLGHFYTLNREMATTEVLKRLGMWTEDPRRPLPWA
LSAHPKRREEDVRPIFWASRPKSYIYRTQEWDFPNGRWGNSSSPAFGELKDYYLFYLSKSPKEELLKM
WGEELTSEESVFEVFLVYLSGEPNRRNGHKVTCLPWNDEPLAAETSLLKEELLRVNRQGITLINSQPNING
KPSSDPIVGVGSPGGYVFQKAYLEFFTSRETAELQLVKYELRVNYHLVNVKGENITNAPELQPNVAVT
WGIFFGREIIQPTVVDVPSFMFKDEAFALWIERWGKLYEEESPRTIIQYIHDNYFLVNLVDNDFPLDN
CLWQVVEDTLELLNRPQTQARETEAP
```

Change region shown

Analyze this sequence

- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence
- Show in Genome Data Viewer

Protein 3D Structure

Structure of human 5,10-methylenetetrahydrofolate

https://www.ncbi.nlm.nih.gov/protein/NP_005948.3?report=fasta

Step 6

First click on file under **Choose destination** section and then on create file

ncbi.nlm.nih.gov/protein/NP_005948.3?report=fasta

NCBI Resources How To Sign in to NCBI

Protein Protein Search

Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

FASTA

methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]

NCBI Reference Sequence: NP_005948.3

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

```
>NP_005948.3 methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]
MVNEARGNSSLNPCLEGSASSGSESSKDSRRCSTPGLDPERHERLREKMRRRLES GDKWFSLEFFPPRTA
EGAVNLIISRFDRMAAGGPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILHMTCCRQRLEEITG
HLHKAKQLGLKNIMALRGDPIDGQWEEEEE GGFNYAVDLVKHIRSEFGDYFDCVAGYPKGHPAEGSFEAD
LKLHLEKVSAGADFIIITQLFFEADTFFRFVKACTDMGITCPIVPGIFPIQGYHS LRQLVKLSKLEVPQEI
KDVIEPIKDNDAAIRNYGIELAVSLCQELLASGLVPLGHFYTLNREMATTEVLKRLGMWTEPRRPLWA
LSAHPKRREEDVRPIFWASRPKSYIYRTQEWDEFNGRWGNSSSPAFGELKDYVLFYKSKSPKEELLKM
WGEELTSEESVFEVFLVYLSGEPNRRNGHKVTCLPWNDEPLAAETSLLKEELLRVNRQGITLINSQPNI
KPSSDPVINGWPGSGGYVFQKAYLEFFTSRETAEALLQVLKYE LRVNYHLVNVKGENITNAPELQPNVAVT
WGI FPGREIIQPTVVDPVSFMFWKDEAFALWIERWGKLYEEESPRTIIQYIHDNYFLVNLVDNDFPLDN
CLWQVVEDTLELLNRPTQNA RETEAP
```

Send to

Choose Destination

File Clipboard
 Collections Analysis Tool

Download 1 item.

Format
FASTA

Show GI

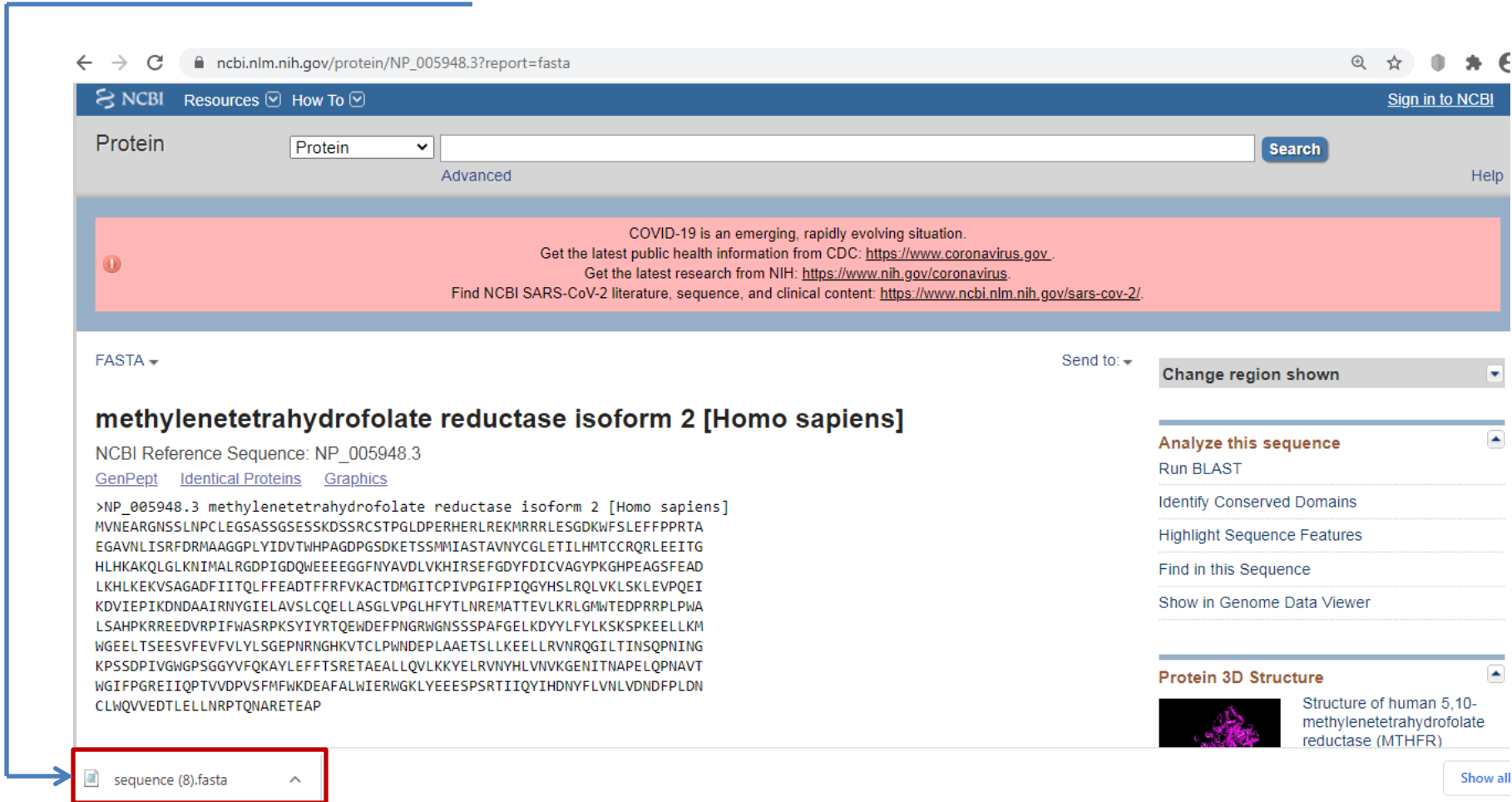
Create File

Protein 3D Structure
Structure of human 5,10-methylenetetrahydrofolate

https://www.ncbi.nlm.nih.gov/protein/NP_005948.3?report=fasta

Step 6

The file will be downloaded in your computer and will open in notepad, click on the downloaded file to open



The screenshot shows the NCBI protein page for NP_005948.3. The page includes a search bar, a navigation menu, and a main content area. A red box highlights the download button for the FASTA sequence file, labeled "sequence (8).fasta".

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>.

FASTA Send to: Change region shown

methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]

NCBI Reference Sequence: NP_005948.3
[GenPept](#) [Identical Proteins](#) [Graphics](#)

```
>NP_005948.3 methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]
MVNEARGNSSLNPCLEGSASSGSESSKDSRCSTPGLDPERHERLREKMRRLLESGDKWFSLEFFPPRTA
EGAVNLSRFDRMAAGGPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILHMTCCRQRLLEEITG
HLHKAKQLGLKNIMALRGDPIGDQWEEEEEGLFNAYVDLVKHIRSEFGDYFDICVAGYKPGHPEAGSFEAD
LKHLKEKVSAGADFIITQLFFEADTFFRFVKACTDMGITCPIVPGIFPIQGYHSRQLVKLSKLEVPQEI
KDVIEPIKNDAAIRNYGIELAVSLCQELLASGLVPLGHFYTLNREMATTEVLKRLGMWTEDEPRRPLPWA
LSAHPKRREEDVRPIFWASRPKSYIYRTQEWDEFNPNRGNWGNSSSPAFGELKDYYLFYLSKSPKEELLKM
WGEELTSEESVFEVFLVLYSGEPNRRNGHKVTCLPWNDEPLAAETSLLKEELLRVNRQGILTINSQPNING
KPSSDPIVGNWGPSGGYVFKAYLEFFTSRETAEALLQVLKKYELRVNYHLVNVKGENITNAPELQPNAVT
WGIFPGREIIQPTVVDVPSFMFWKDEAFALWIERWGLYEEESPRTIIQYIHDNYFLVNLVDNDFPLDN
CLWQVVEDTLELLNRPTQNA RETEAP
```

sequence (8).fasta

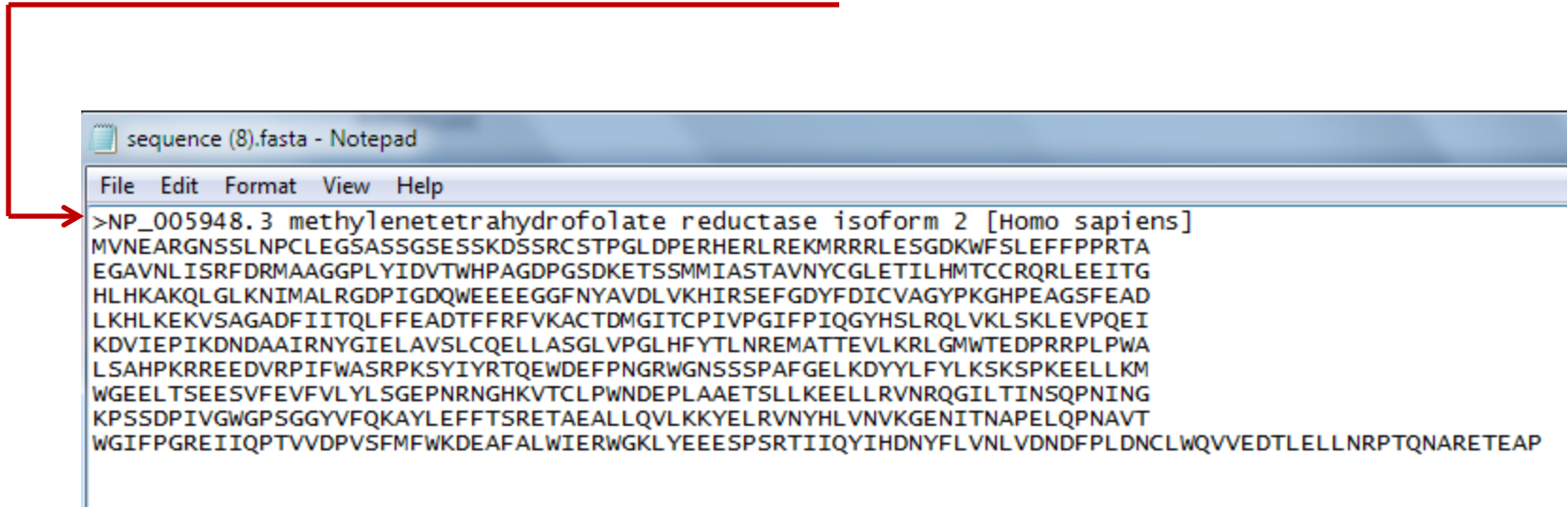
Analyze this sequence
Run BLAST
Identify Conserved Domains
Highlight Sequence Features
Find in this Sequence
Show in Genome Data Viewer

Protein 3D Structure
Structure of human 5,10-methylenetetrahydrofolate reductase (MTHFR)

Show all

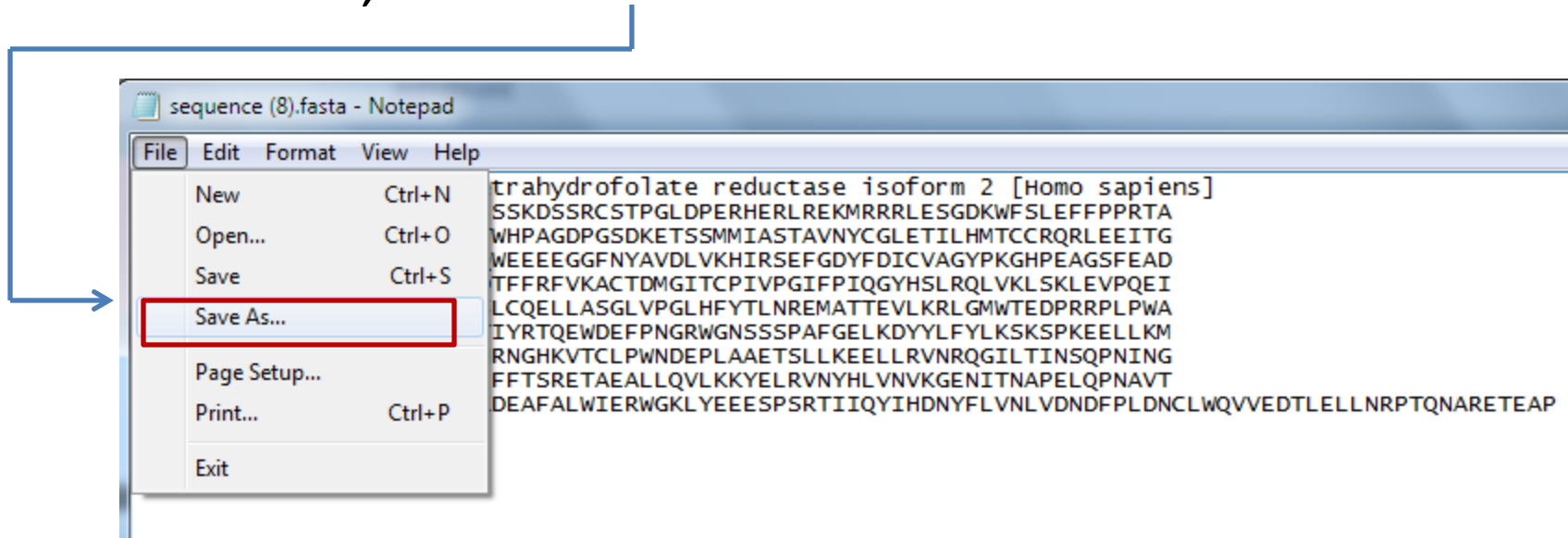
https://www.ncbi.nlm.nih.gov/protein/NP_005948.3?report=fasta

The Protein sequence in FASTA will open in notepad

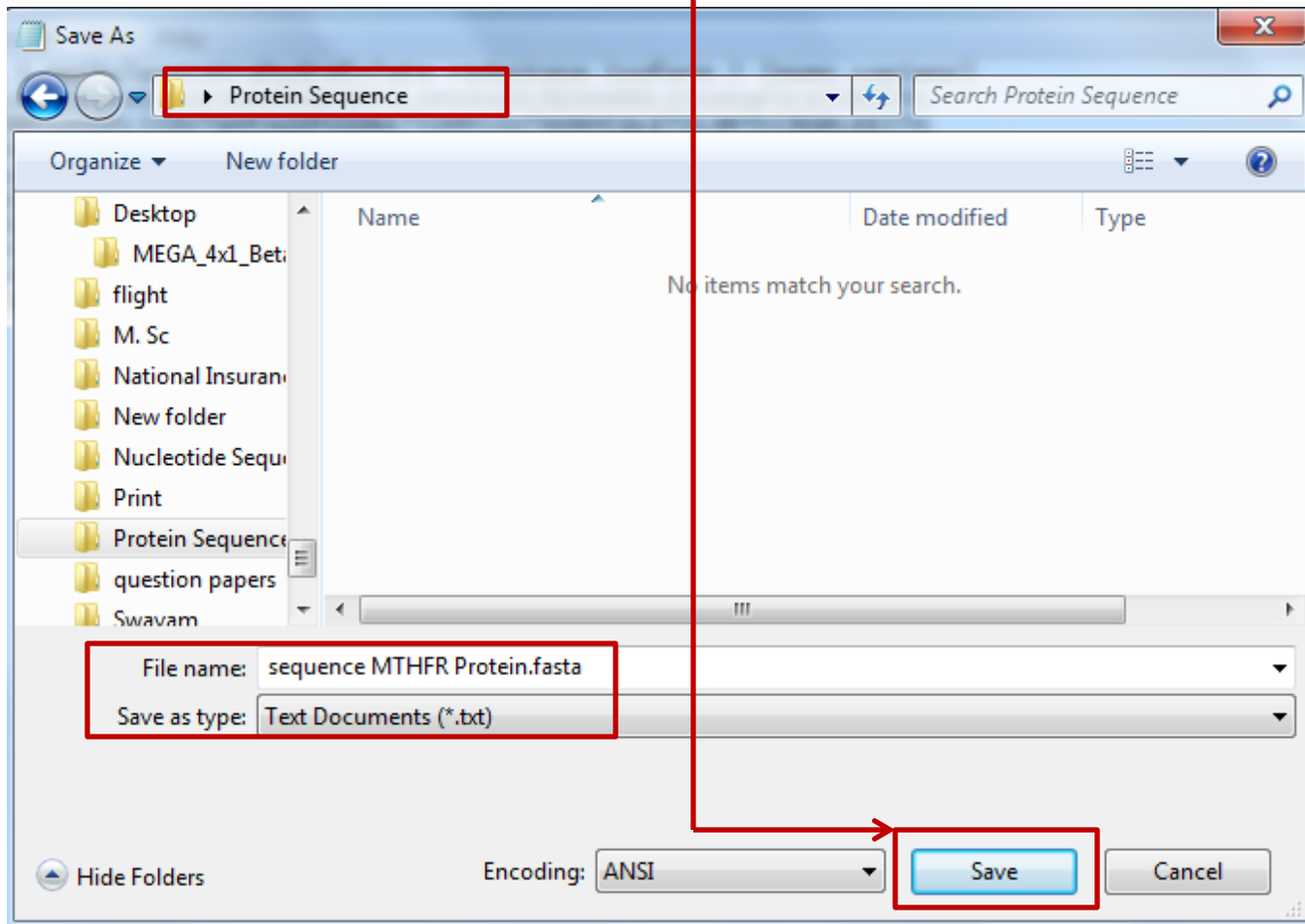


```
sequence (8).fasta - Notepad
File Edit Format View Help
>NP_005948.3 methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]
MVNEARGNSSLNPCLEGSASSGSESSKSSRCSTPGLDPERHERLREKMRRRLES GDKWFSLEFFFPRTA
EGAVNLI SRFDRMAAGGPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILHMTCCRQRLEEITG
HLHKAKQLGLKNIMALRGDPIGDQWEEEEGGFNAYVDLVKHIRSEFGDYFDICVAGYPKGHPEAGSFEAD
LKHLKEKVSAGADF IITQLFFEADTFFRFVKACTDMGITCPIVPGIFPIQGYHSLRQLVKLSKLEVPQEI
KDVIEPIKDNDAAIRNYGIELAVSLCQELLASGLV PGLHFYTLNREMATTEVLKRLGMWTE DPRRPLPWA
LSAHPKRREEDVRPIFWASRPKSYIYRTQEWDEF PNGRWGNSSSPAFGELKDYYLFYLSKSPKEELLKM
WGEELTSEESVFEVFLYL SGEPNRNGHKVTCLPW NDEPLAAETSLLKEELLRVNRQGILTINSQP NING
KPSSDP IVGWGPGSGGYVFQKAYLEFFTSRETA EALLQVLKKYELRVNYHLVNVKGENITNAPELQ PNAVT
WGIFPGREIIQPTVVDPV SFMFWKDEAFALWIER W GKLYEEESPRTIIQYIHDNYFLVNLVDNDFPLDNCLWQ VVEDTLELLNRPTQNARETEAP
```

To rename the file and save it in your computer in any folder ,click on file and then on save as



The file will be saved in a specific folder in your computer.
You can rename the file (e.g “Sequence MTHFR Protein.fasta” in this case)
Click on save to save the file



NCBI Resources How To Sign in to NCBI

Protein Search

sequence (8).fasta - Notepad

```
File Edit Format View Help
>NP_005948.3 methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]
MVNEARGNSSLNPCLEGSASSGSESSKDSRSTPGLDPERHERLREKMRRLLESGDKWFSLEFFPPRTA
EGAVNLISRFDRMAAGGPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILHMTCCRQLEEITG
HLHKAKQLGLKNIMALRGDPIGDQWEEEEEKGFNYAVDLVKHIRSEFGDYFDICVAGYKGGHPEAGSFEAD
LKHLKEKVSAGADFIITQLFFADTFFRFVKAC TDMGITCP IVPGFPIQGYHSLRQLVKLSKLEVPQEI
KDVIEPIKNDAAIRNYGIELAVSLCQELLASGLVPGLHFYTLNREMATTEVLKRLGMWTEDPRRPLPWA
LSAHPKRREEDVRPIFWASRPKSYIYRTQEWDEFNPNRGNSSSPAFGELKDYLYFYLSKSPKEELLKM
WGEELTSEESVFEVFLYLSGEPNRRNGHKVTCLPWNDEPLAAETSLKEELLRVNRQGILTINSOPNING
KPSSDPVIGWGPSGGYVFQKAYLEFFT SRETAEALQVLKYYELRVNYHLVNVKGENITNAPELQPNNAV
WGIFFGREIIQPTVVDPVFSFMFKDEAFALWIERWGKLYEEESPRTIIQYIHDNYFLVNLVDNDFPLDNCLWQVVEDTLELLNRPTQNARETEAP
```

FASTA

methylenetetra

NCBI Reference Sequences

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

```
>NP_005948.3 methylen
MVNEARGNSSLNPCLEGSASS
EGAVNLISRFDRMAAGGPLYI
HLHKAKQLGLKNIMALRGDP
LKHLKEKVSAGADFIITQLFF
KDVIEPIKNDAAIRNYGIEL
LSAHPKRREEDVRPIFWASRP
WGEELTSEESVFEVFLYLSG
KPSSDPVIGWGPSGGYVFQK
WGIFFGREIIQPTVVDPVFS
CLWQVVEDTLELLNRPTQNA
```

References/Acknowledgments

- <https://www.ncbi.nlm.nih.gov>
- <https://www.ncbi.nlm.nih.gov/protein>
- <https://www.ncbi.nlm.nih.gov/genbank/>
- <https://www.ncbi.nlm.nih.gov/refseq/>
- <https://www.ncbi.nlm.nih.gov/genbank/tpa/>
- <https://www.uniprot.org/uniprot/?query=swiss+prot&sort=score> ,
- <https://proteininformationresource.org/> ,
- <https://www.prf.or.jp/aboutdb-e.html>
- www.rcsb.org/pdb.
- <https://www.ncbi.nlm.nih.gov/protein/?term=MTHFR>