Biological Database Search and Retrieval of Protein Sequence

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

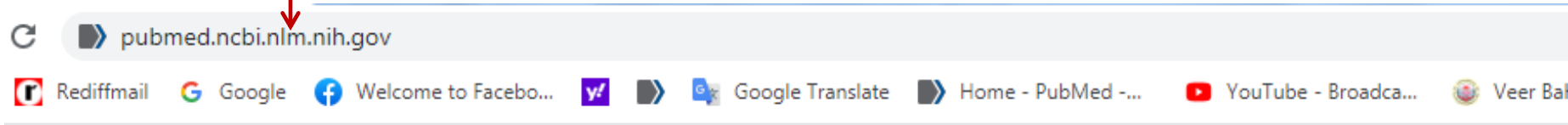
NCBI Protein

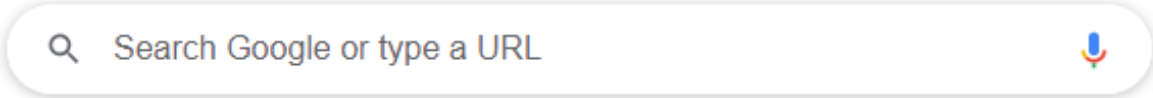
https://www.ncbi.nlm.nih.gov/protein

- 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).

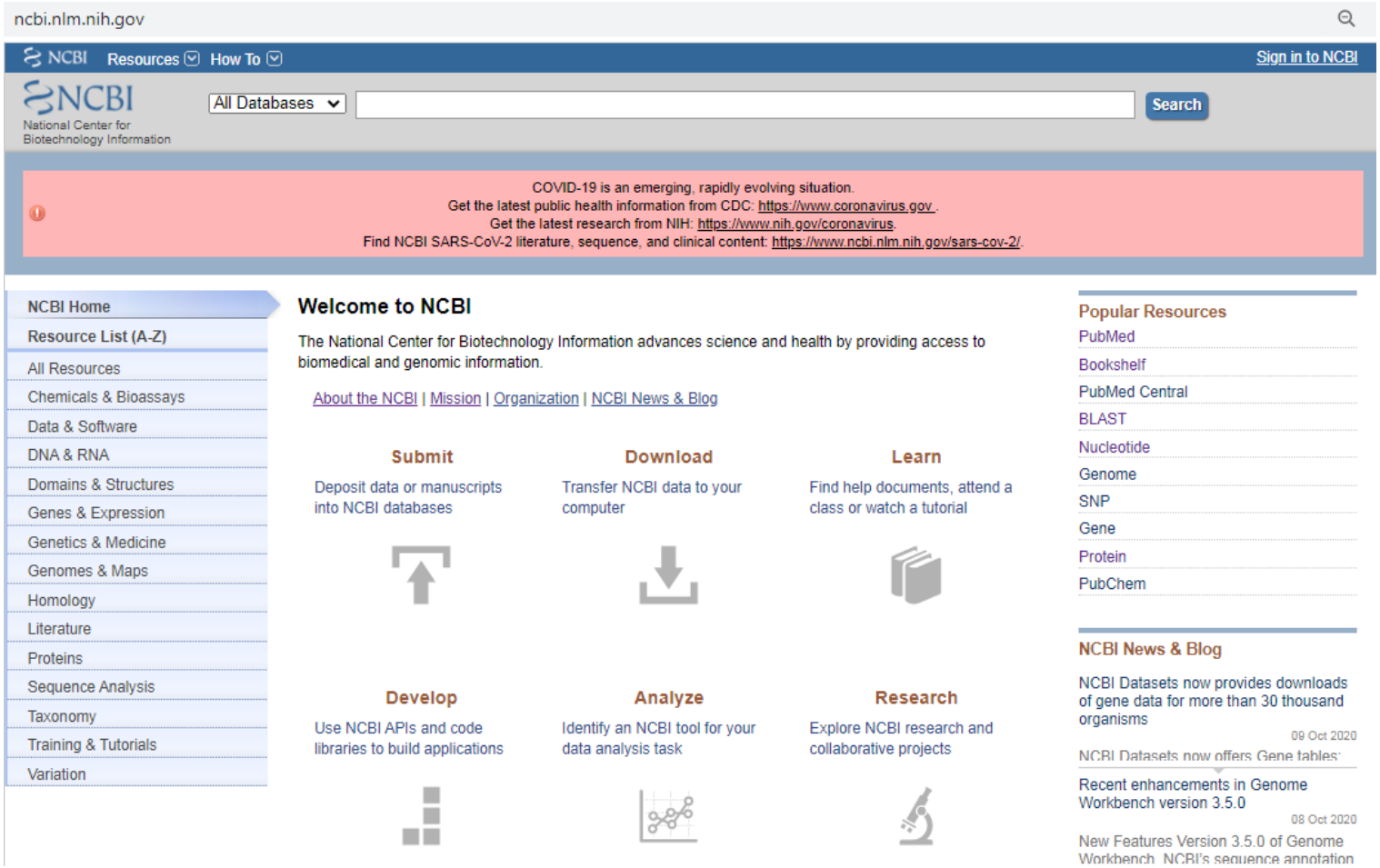
Step 1

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



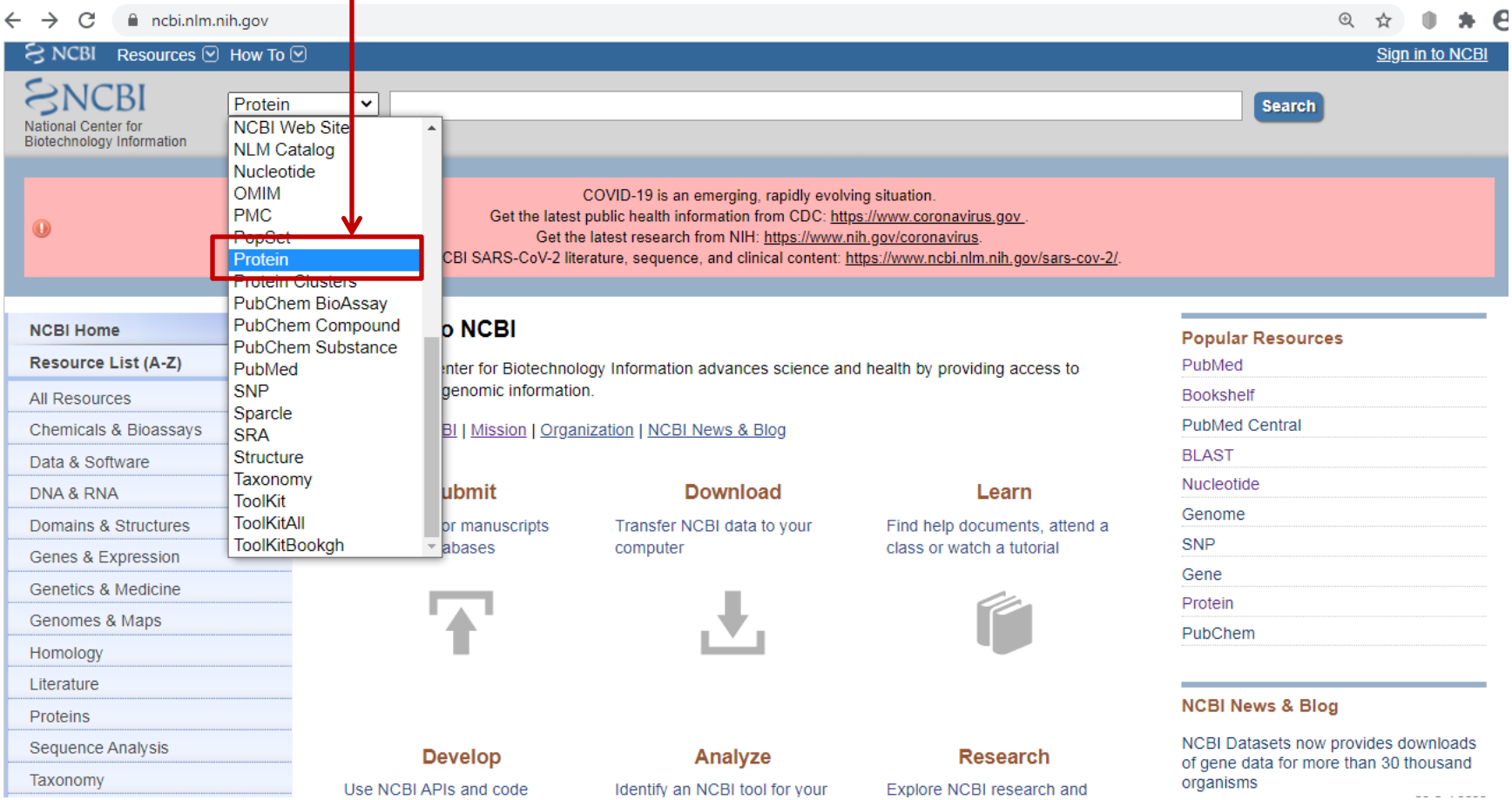


pubmed.ncbi.nlm.nih.gov homepage will open



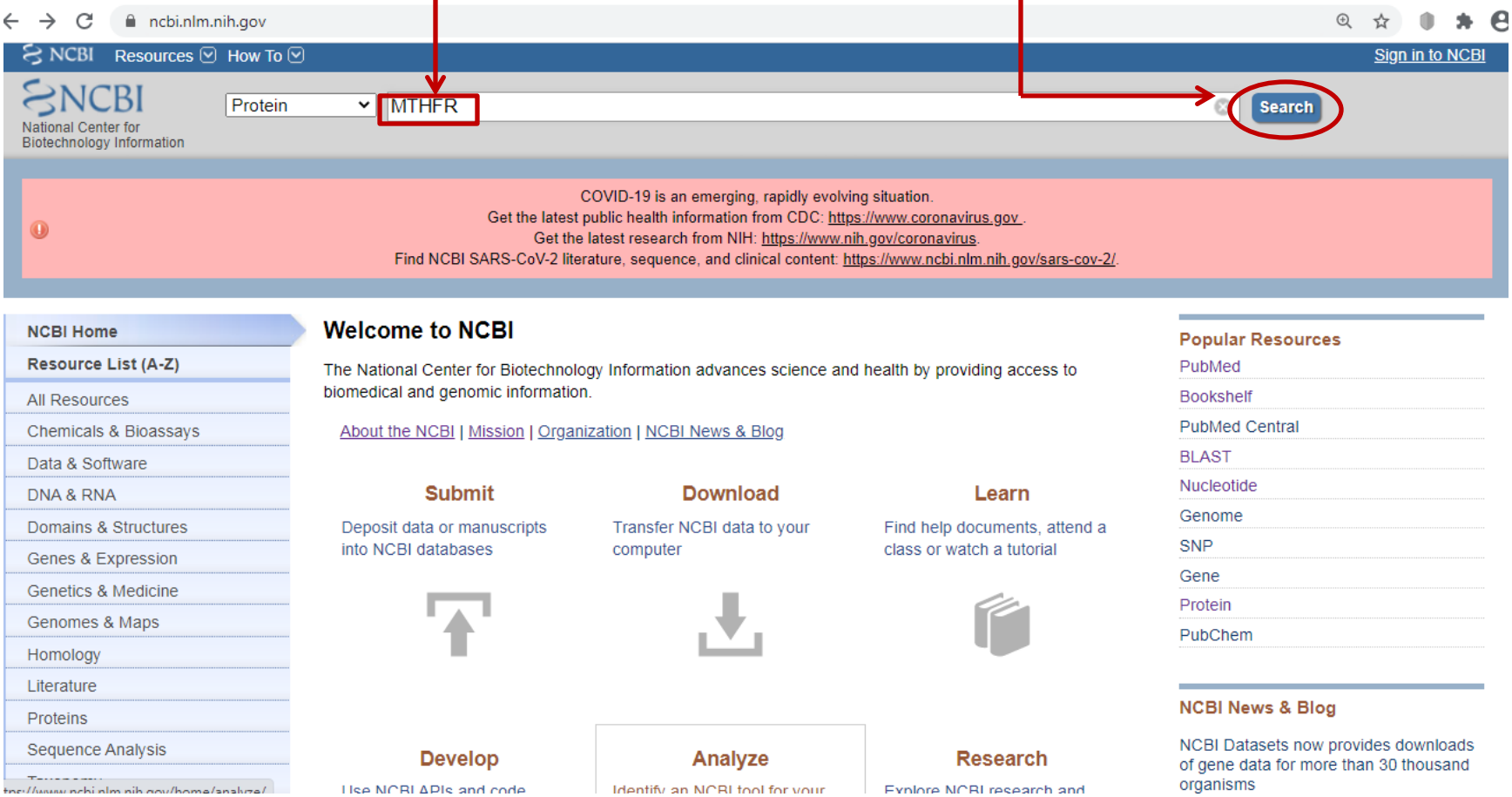
Step 2

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

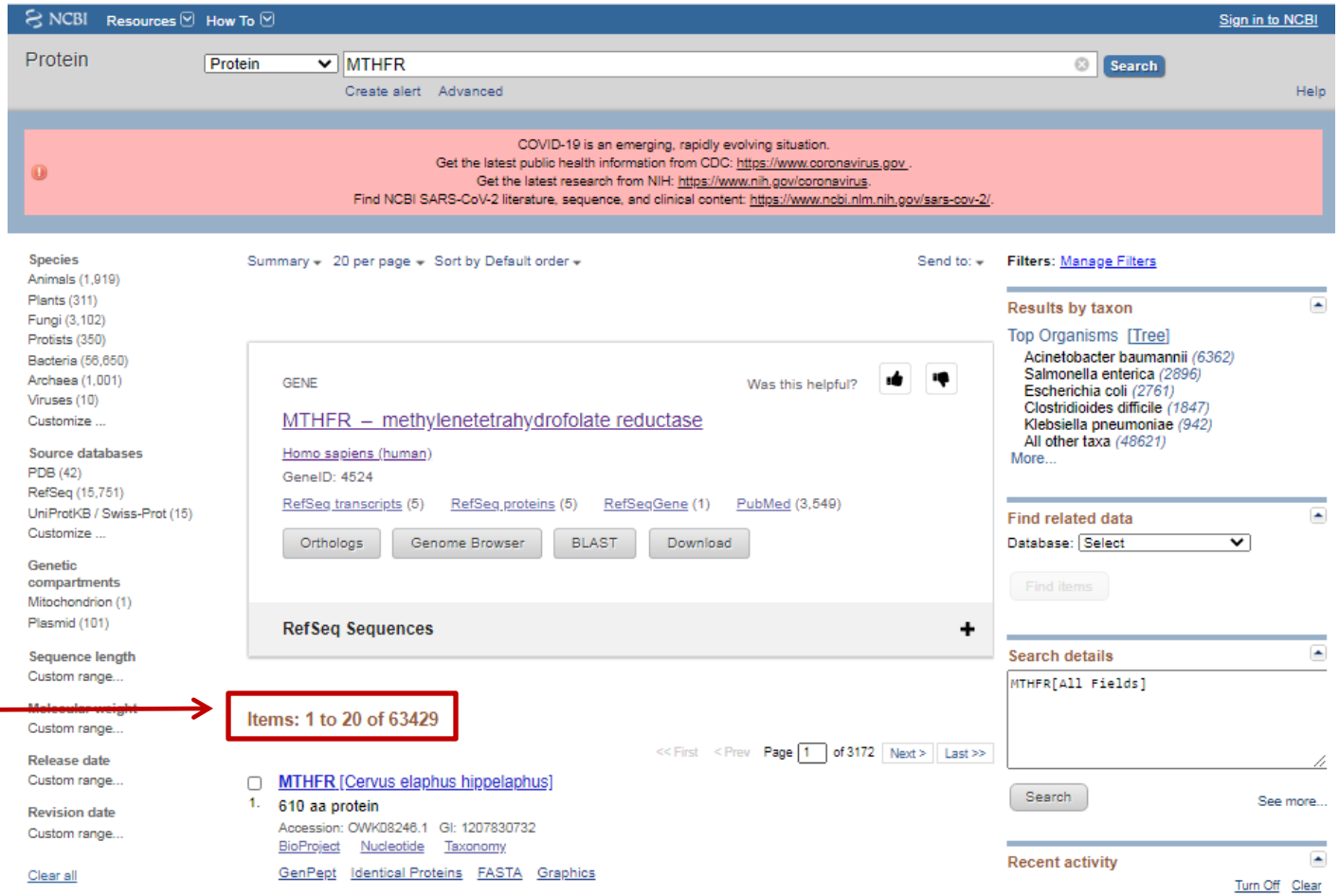


Step 3

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



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



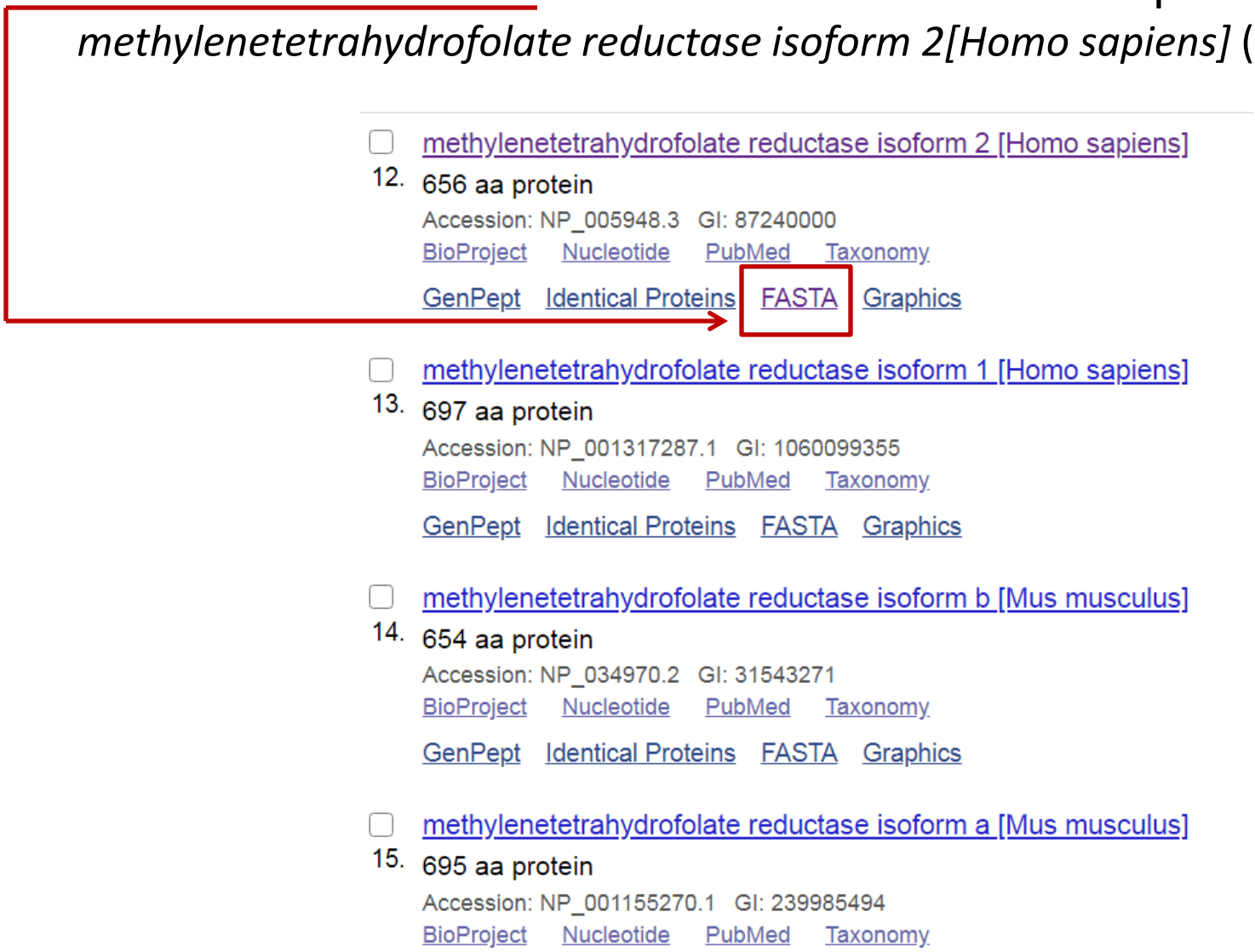
Total number of items available in the protein database related to our search i.e.”MTHFR”

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)



12. methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]
656 aa protein
Accession: NP_005948.3 GI: 87240000
BioProject Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

13. methylenetetrahydrofolate reductase isoform 1 [Homo sapiens]
697 aa protein
Accession: NP_001317287.1 GI: 1060099355
BioProject Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

14. methylenetetrahydrofolate reductase isoform b [Mus musculus]
654 aa protein
Accession: NP_034970.2 GI: 31543271
BioProject Nucleotide PubMed Taxonomy
GenPept Identical Proteins FASTA Graphics

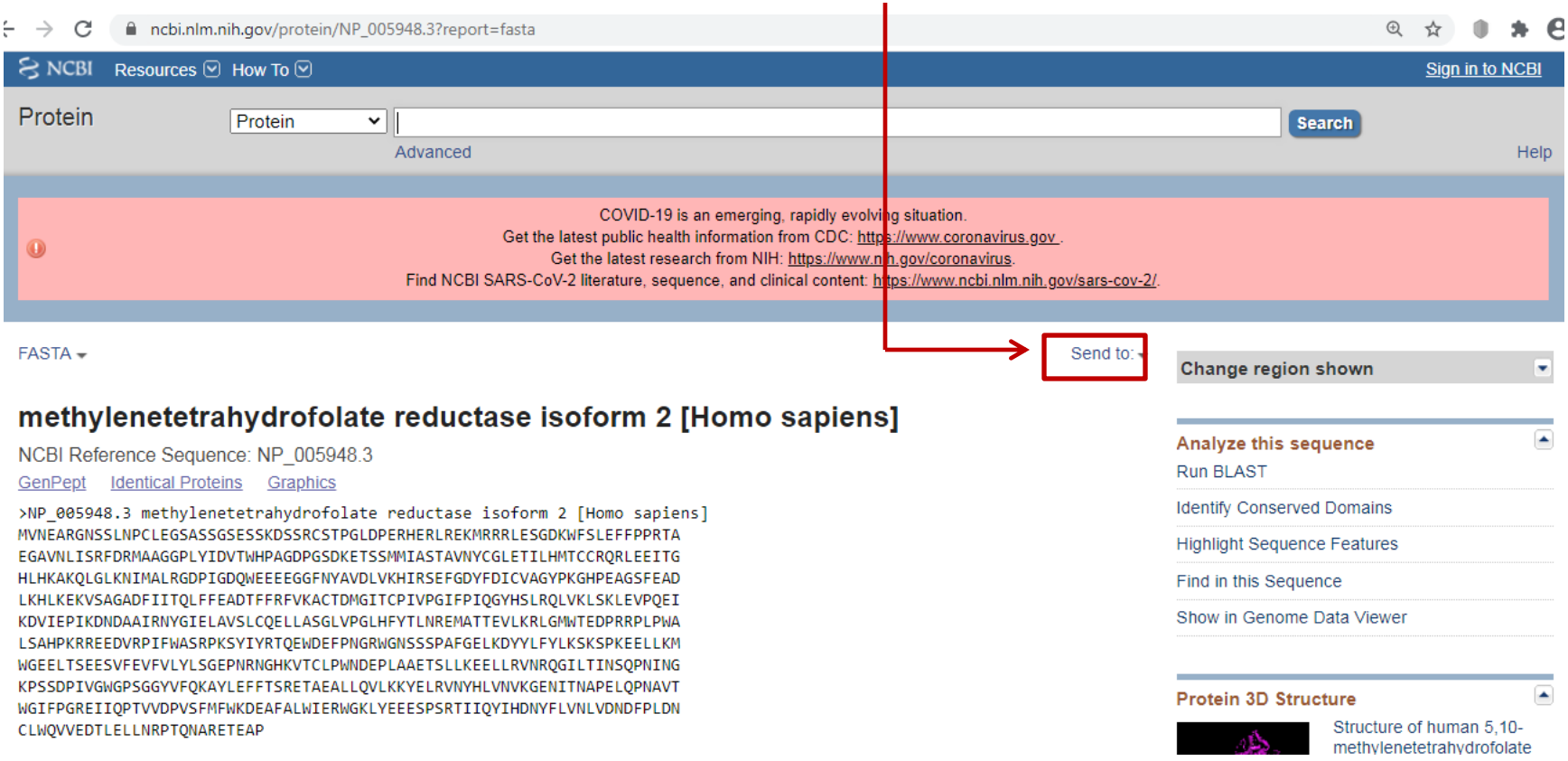
15. methylenetetrahydrofolate reductase isoform a [Mus musculus]
695 aa protein
Accession: NP_001155270.1 GI: 239985494
BioProject Nucleotide PubMed Taxonomy

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

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

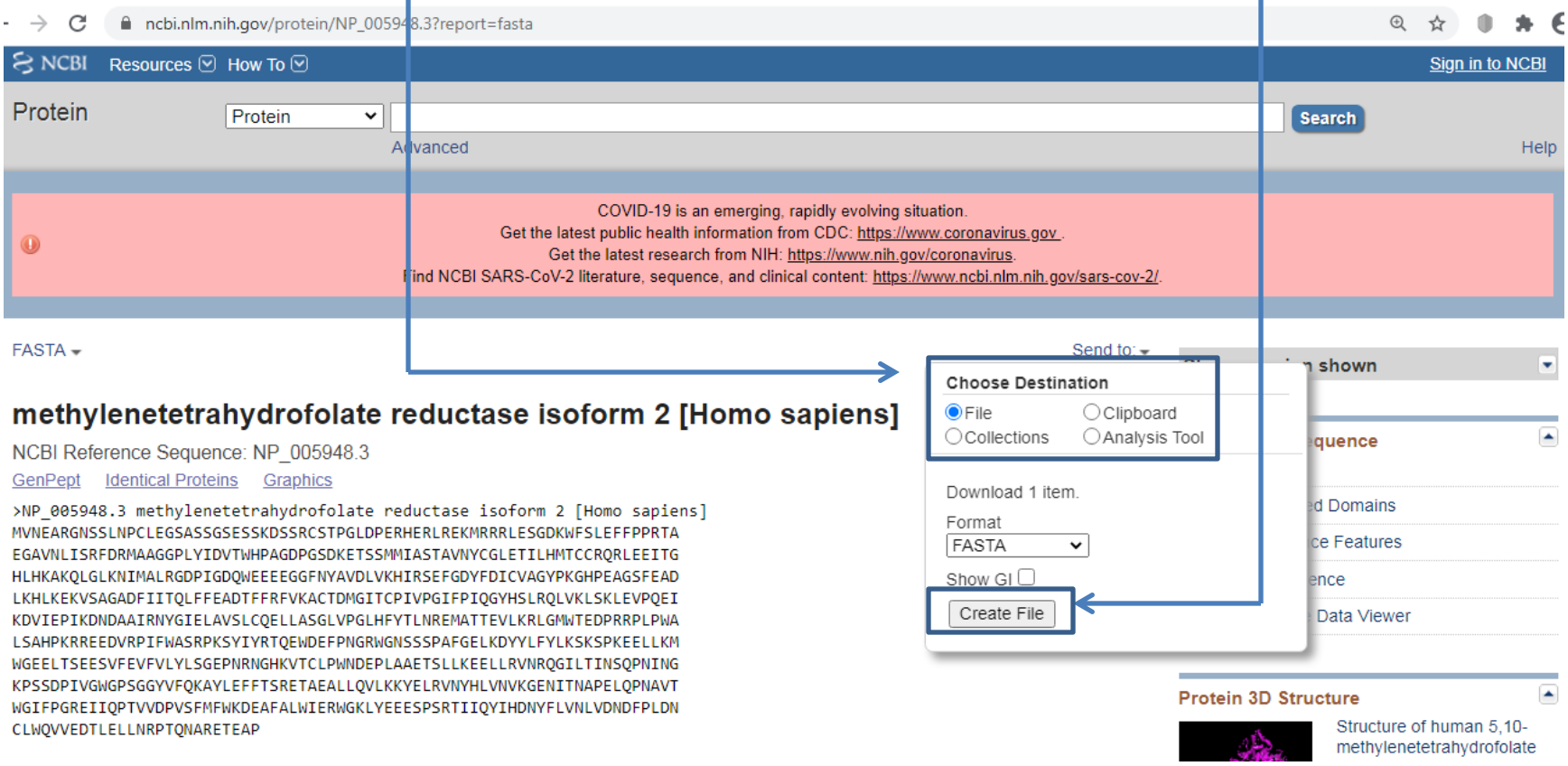
Click on send to



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

Step 6

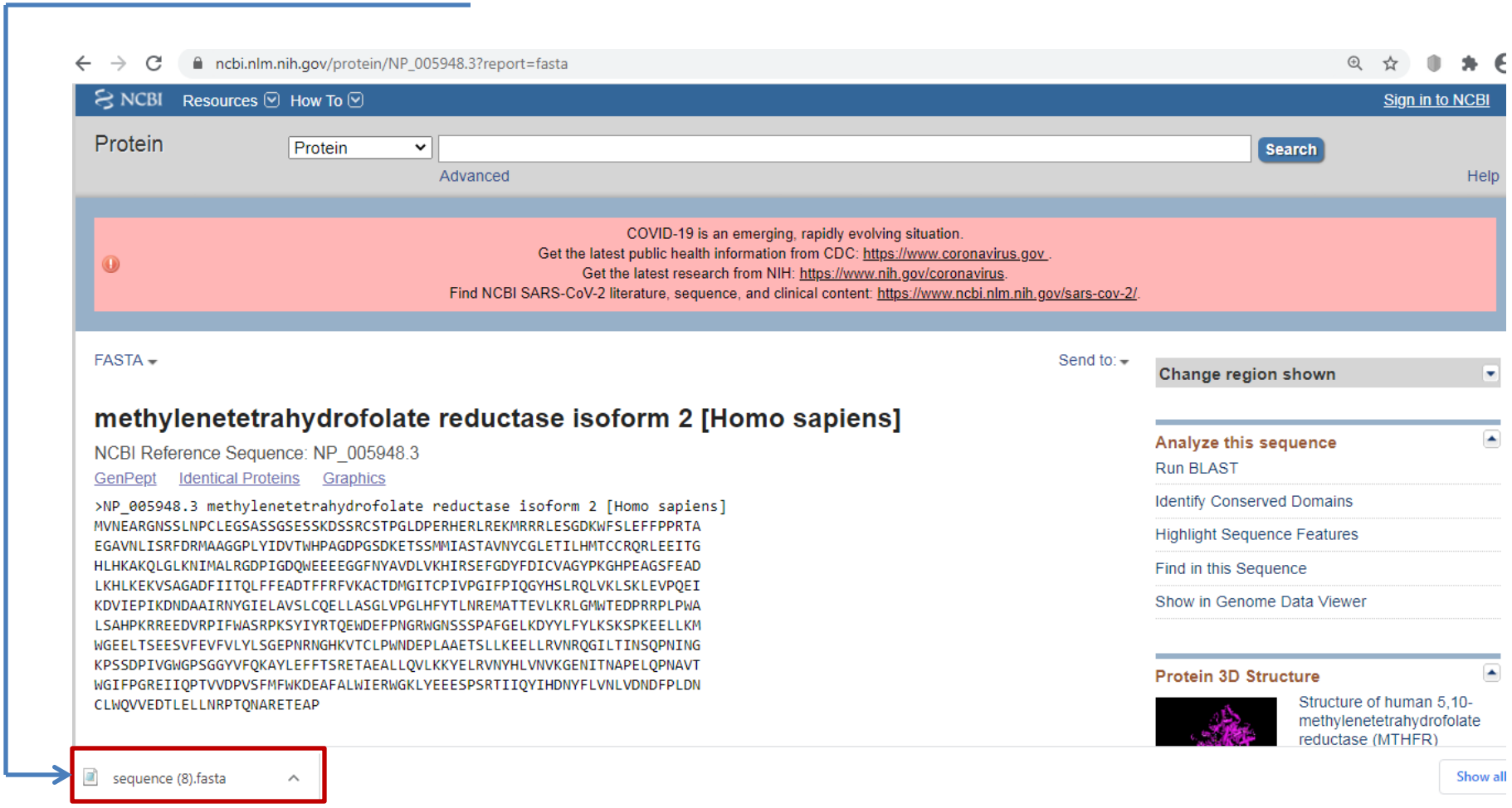
First click on file under **Choose destination** section and then on create file



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

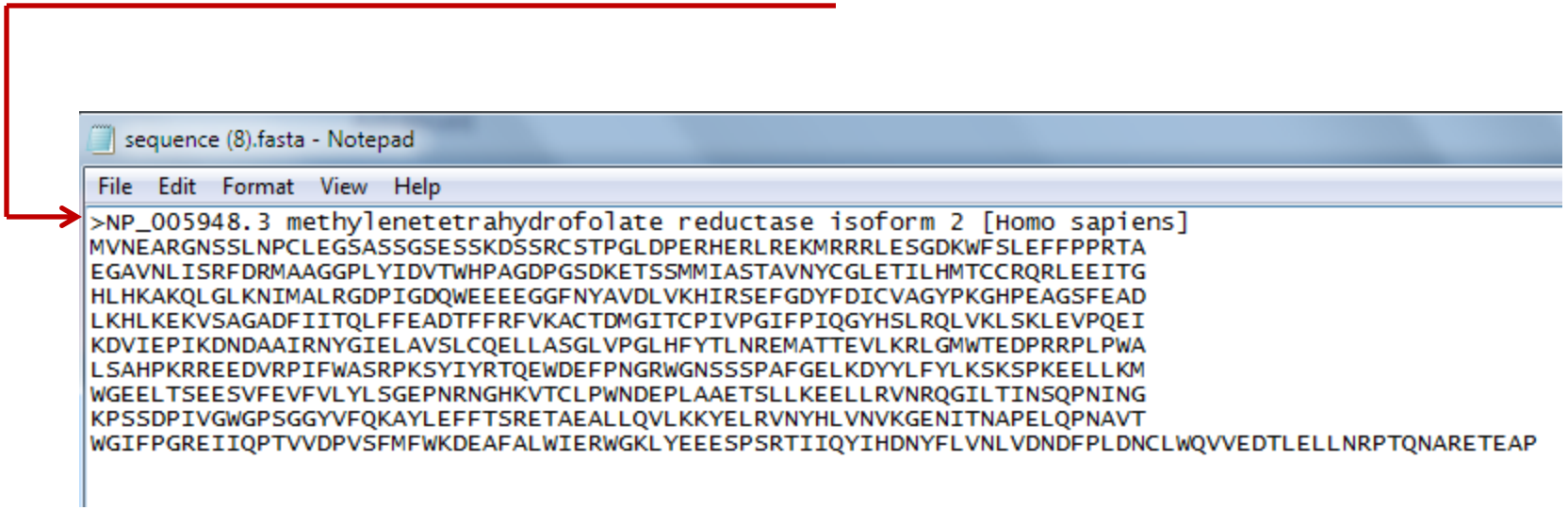
Step 6

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

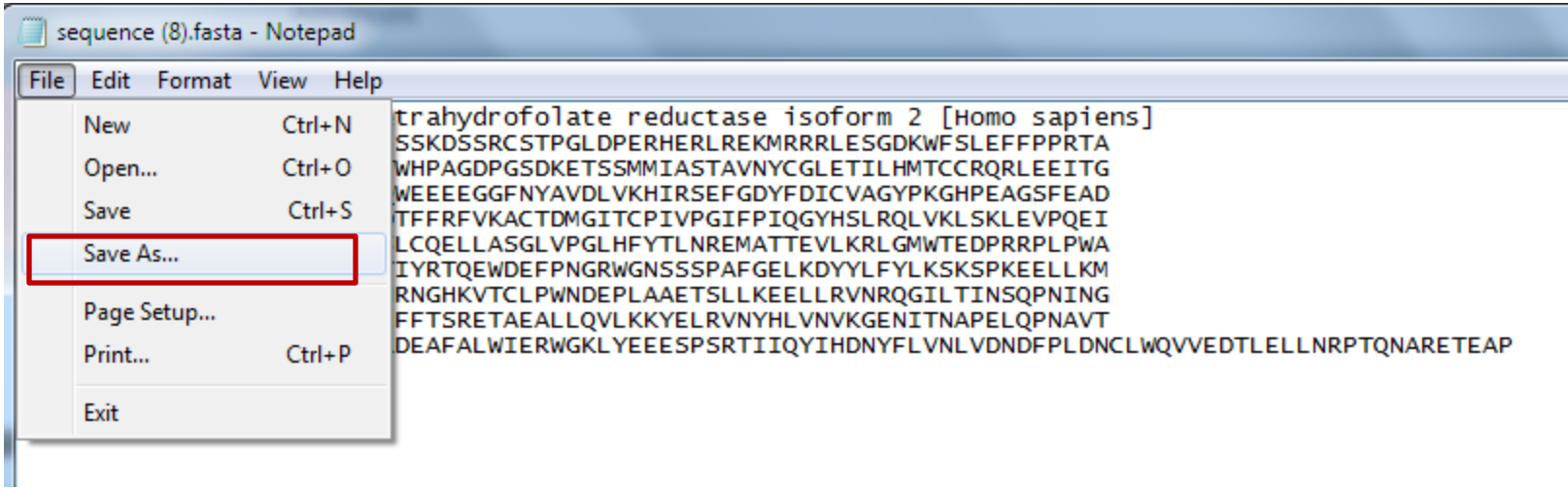


https://www.ncbi.nlm.nih.gov/protein/NP_005948.3?report=fasta

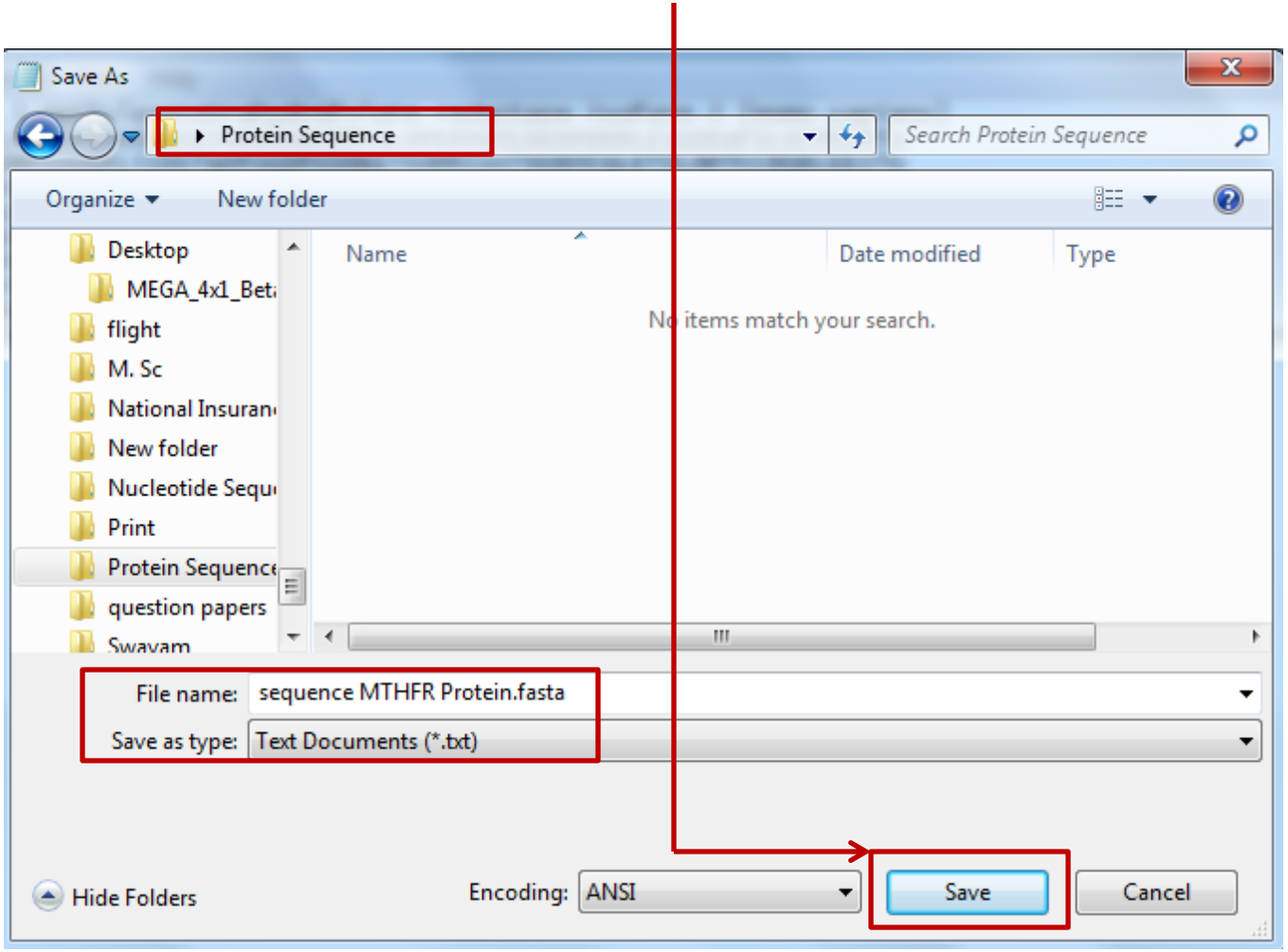
The Protein sequence in FASTA will open in notepad

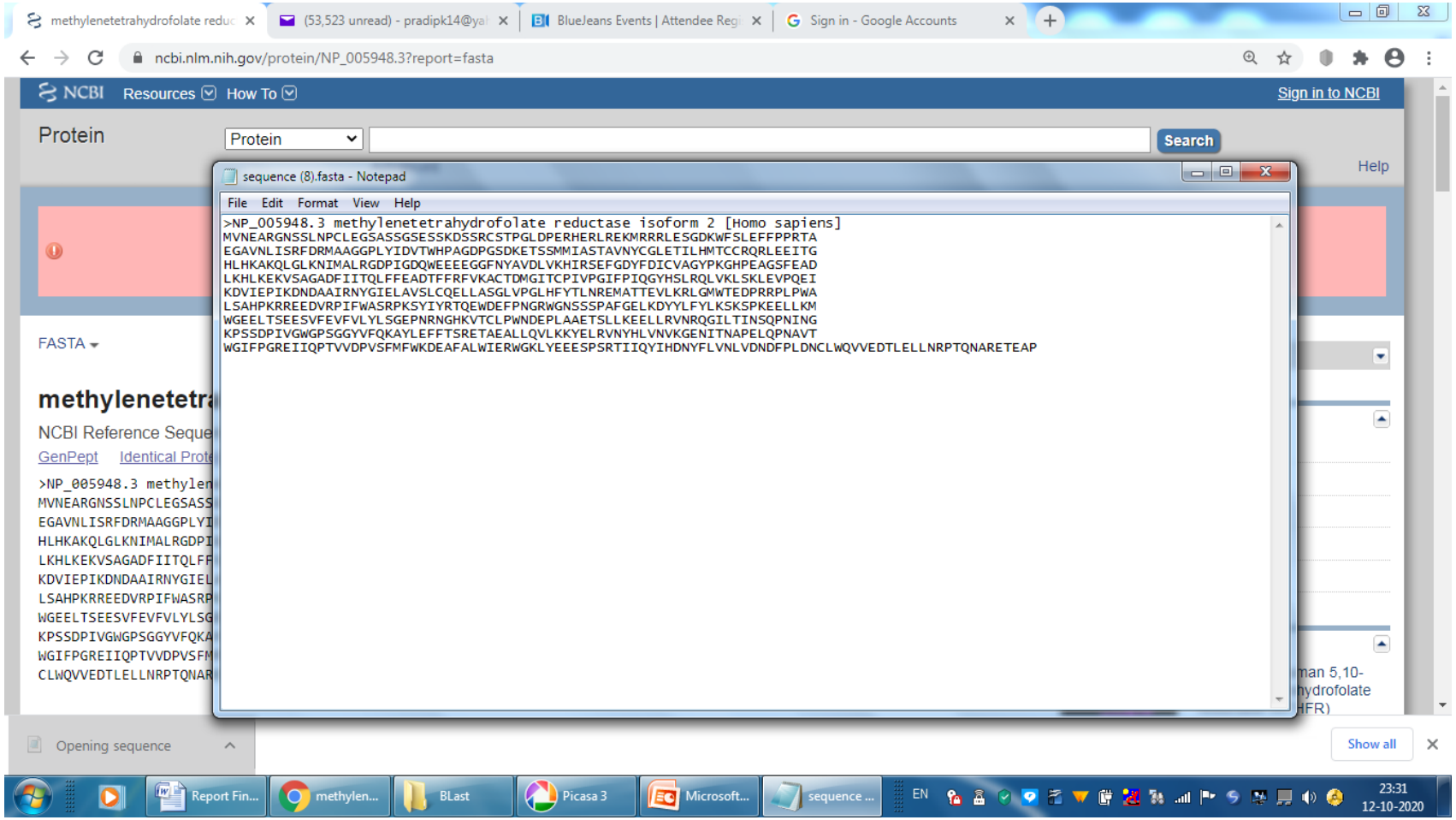


To rename the file and save it in your computer in any folder ,click on file and then on save as



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




methylenetetrahydrofolate reduc
(53,523 unread) - pradipk14@ya
BlueJeans Events | Attendee Reg
Sign in - Google Accounts
ncbi.nlm.nih.gov/protein/NP_005948.3?report=fasta
NCBI
Resources
How To
Sign in to NCBI
Protein
Protein
Search
Help
FASTA
methylenetetra
NCBI Reference Seque
GenPept
Identical Prote
>NP_005948.3 methylen
MVNEARGNSSLNPCLEGSASS
EGAVNLISRFDRMAAGGPLYI
HLHKAKQLGLKNIMALRGDPI
LKHLKEKVSAGADFIITQLFF
KDVIEPIKDNDAAIRNYGIEL
LSAHPKRREEDVRPIFWASRP
WGEELTSEESVFEVFVLYLSG
KPSSDPIVGWGPSGGYVFQKA
WGIFPGREIIQPTVVDPVSFM
CLWQVVEDTLELLNRPTQNAR
sequence (8).fasta - Notepad
File Edit Format View Help
>NP_005948.3 methylenetetrahydrofolate reductase isoform 2 [Homo sapiens]
MVNEARGNSSLNPCLEGSASSGSESSKDSSRCSTPGLDPERHERLREKMRRRLESGDKWFSLEFFPPRTA
EGAVNLISRFDRMAAGGPLYIDVTWHPAGDPGSDKETSSMMIASTAVNYCGLETILHMTCCRQRLEEITG
HLHKAKQLGLKNIMALRGDPIGDQWEEEEGGFNYAVDLVKHIRSEFGDYFDICVAGYPKGHPEAGSFEAD
LKHLKEKVSAGADFIITQLFFEADTFFRFVKACTDMGITCPIVPGIFPIQGYHSLRQLVKLSKLEVPQEI
KDVIEPIKDNDAAIRNYGIELAVSLCQELLASGLVPGLHFYTLNREMATTEVLKRLGMWTEDPRRPLPWA
LSAHPKRREEDVRPIFWASRPKSYIYRTQEWDEFPNGRWGNSSSPAFGELKDYYLFYLKSKSPKEELLKM
WGEELTSEESVFEVFVLYLSGEPNRNGHKVTCLPWNDEPLAAETSLLKEELLRVNRQGILTINSQPNING
KPSSDPIVGWGPSGGYVFQKAYLEFFTSRETAEALLQVLKKYELRVNYHLVNVKGENITNAPELQPNAVT
WGIFPGREIIQPTVVDPVSFMFWKDEAFALWIERWGKLYEEESPSRTIIQYIHDNYFLVNLVDNDFPLDNCLWQVVEDTLELLNRPTQNARETEAP
Opening sequence
Show all
Report Fin...
methylen...
BLast
Picasa 3
Microsoft...
sequence ...
EN
23:31
12-10-2020

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