

Bioinformatics-I (BIF401)- HELP

Assignment No.01

Total Marks 15

Note: There are two questions. First one is theoretical and second one is practical. You are supposed to do your assignment by yourself. Any kind of plagiarism will be marked straight zero.

Question 1: What you have learnt in basics of Bioinformatics. Also enlist and discuss the tools software and methods you have studied till now. (7)

Question 2: Select a sequence of a candidate gene. You are required to find: (8)

1. Full scientific name of the candidate gene.
2. Top 3 alignment result of the candidate gene with the complete information like accession number, Locus, Coding region etc.
3. Three more relevant ortholog.
4. Name of the tools and databases which you are using during this process.

Hint: (Orthologs are genes in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution)

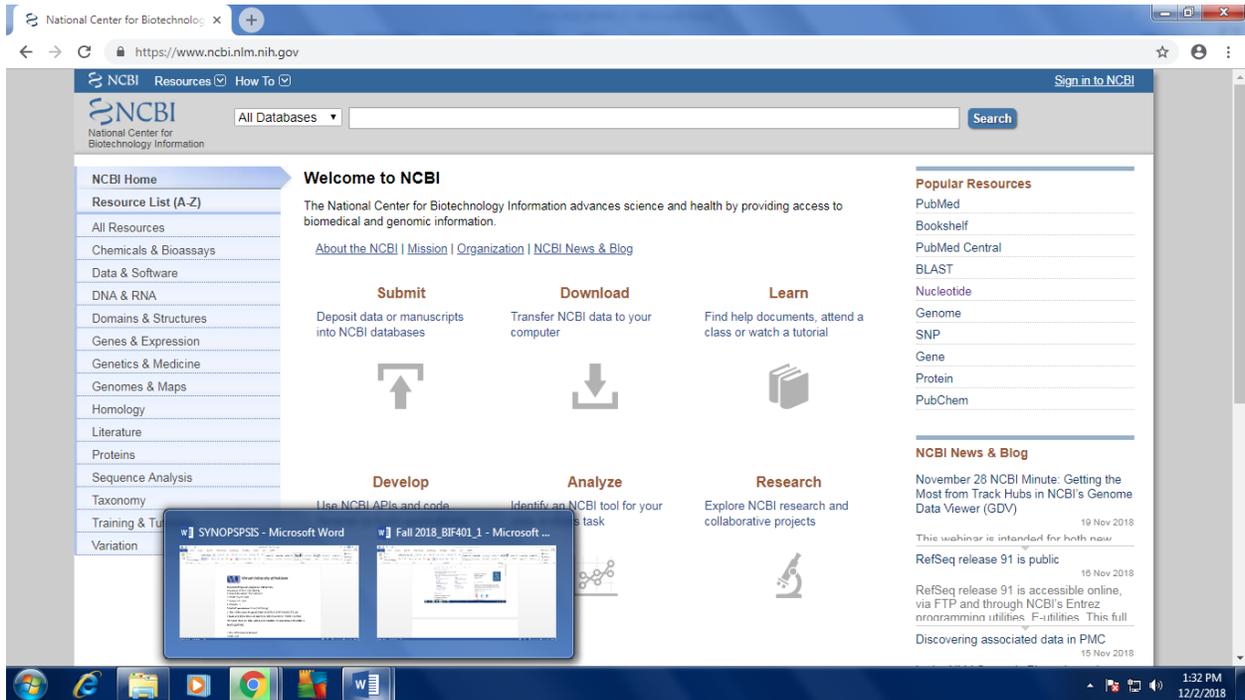
Sample Assignment

1- Go to the NCBI database

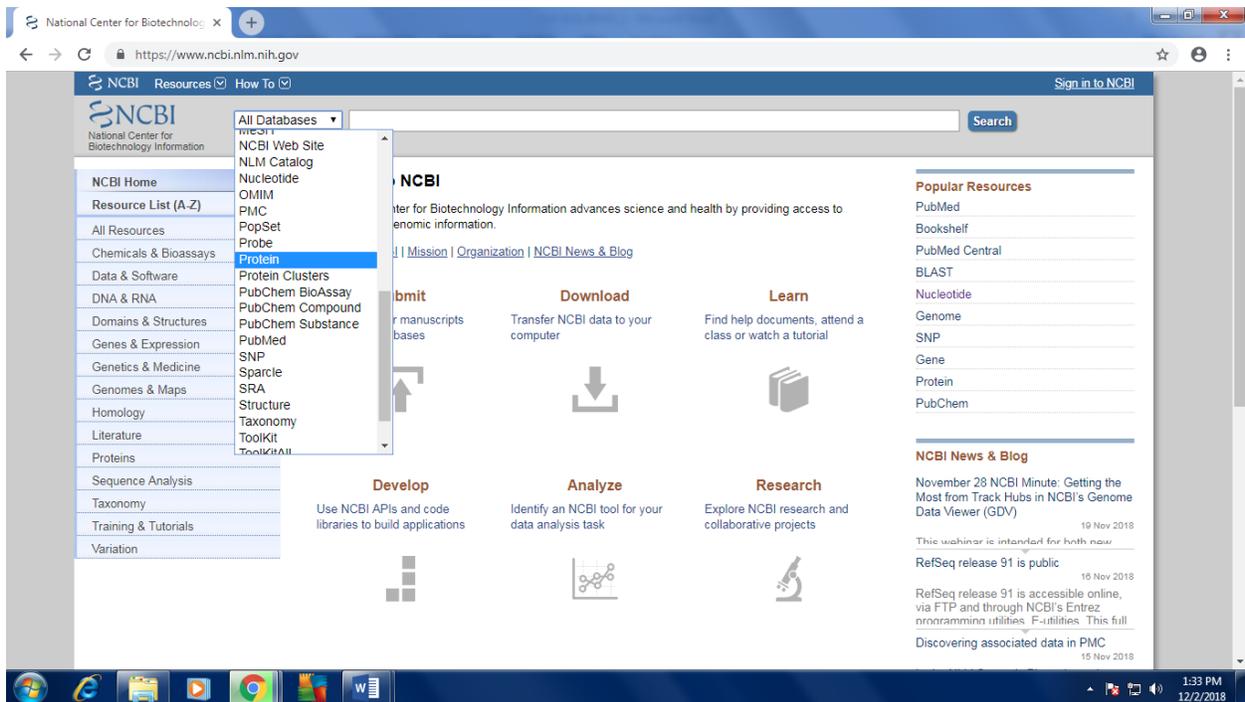
The screenshot shows a Google search for 'ncbi'. The search results page displays the following information:

- Search Results:** About 99,600,000 results (0.35 seconds).
- Top Result:** National Center for Biotechnology Information. URL: <https://www.ncbi.nlm.nih.gov/>. Description: Welcome to NCBI. The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.
- Related Links:** BLAST, PubMed, Nucleotide, Gene, Protein, DNA & RNA.
- Knowledge Panel:** National Center for Biotechnology Information. Company. Website: ncbi.nlm.nih.gov. Description: The National Center for Biotechnology Information is part of the United States National Library of Medicine, a branch of the National Institutes of Health. Headquarters: Bethesda, Maryland, United States. Founded: November 4, 1988.
- Profiles:** YouTube, Twitter, Facebook.
- People also search for:** What does PMC stand for NCBI?, Is NCBI a reliable source?, Is NCBI a journal?

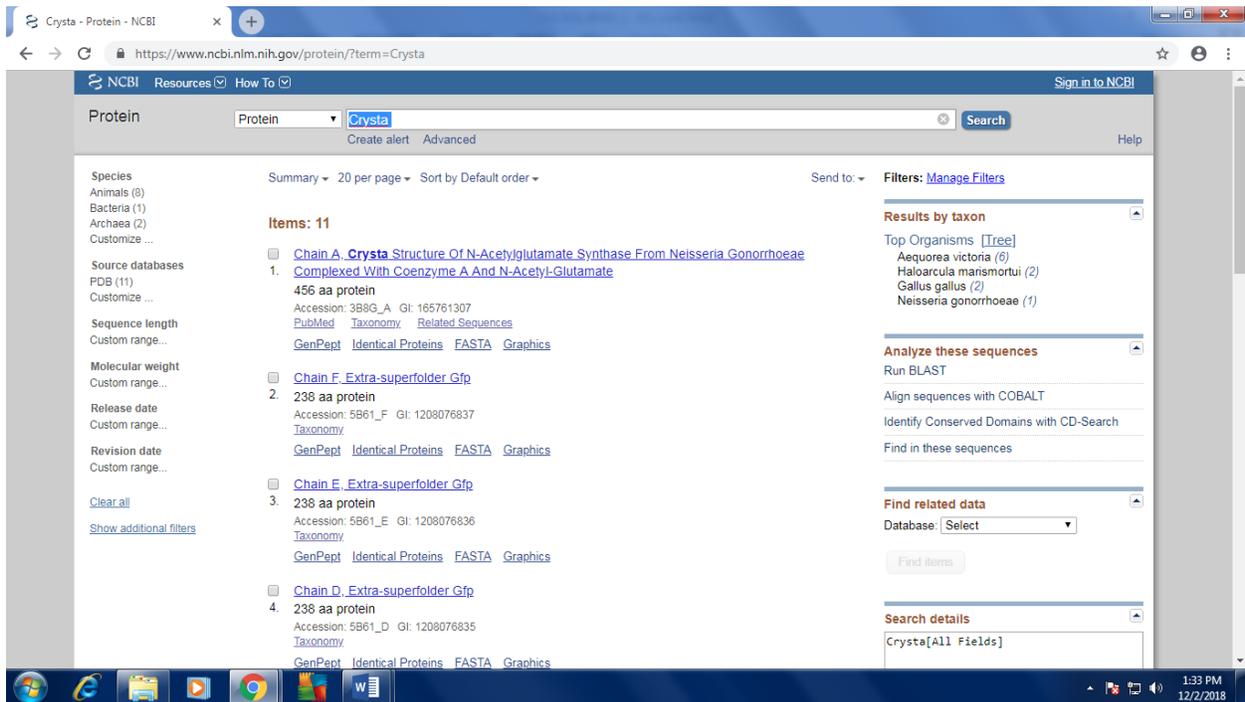
Open NCBI



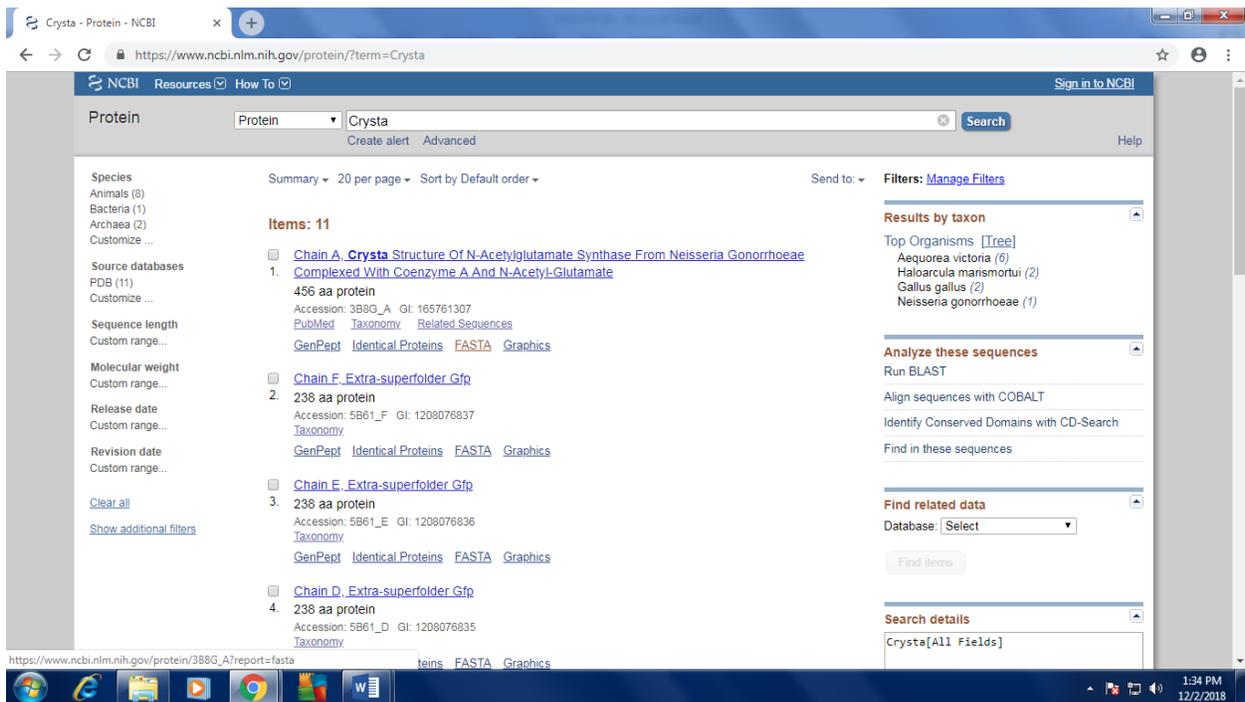
Select Protein database from here



Enter the name of your candidate gene (a gene, which you want to work upon)



Enter it-> it will give you a list of all possible sequence datasets available related to the your gene (protein)



Open it and then you can see the full sequence with all desired information in the first line.

Chain A, Crysta Structure Of N-A- x

https://www.ncbi.nlm.nih.gov/protein/3B8G_A?report=fasta

NCBI Resources How To Sign in to NCBI

Protein Protein Search Help

Advanced

FASTA - Send to: - Change region shown

Chain A, Crysta Structure Of N-Acetylglutamate Synthase From Neisseria Gonorrhoeae Complexed With Coenzyme A And N-Acetyl-Glutamate

PDB: 3B8G_A

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

```
>pdb|3B8G|A Chain A, Putative Acetylglutamate Synthase
MGSSHHHHHHSSGLVPRGSHINAPDSFVAHFREAAPYIRQIRGTTLVAGIDGRLLGGT LNKLAADIGLL
SQLGIRLVLIHGAYHFLDRLAAQGRTPHYCRGLRVTDETS L GQAQQFAGTVRSRFEALCGSVGFARA
PSVPLVSGNFLTARP IGVIDGTDMEYAVIRKDTAALRFQLDAGNIVMPP L GHSYGGKTFNLDIVQAA
ASVAVSLQAEKLVVYLLTSDGISRPDGT LAETLSAQEAQSLAEHAASETRRLLSSAVAAL EGGVHRVQILN
GAADGSL LQELFT RINGIGTSIAKEAFVSTRQAHSGDIPHIAALTRPLEEQGILLHRSREYLENHI SEFSI
LEHDGNLYGCAALKTF AEADCGE TACLAVSPQAQGGYGERLLAHTIDKARGIGISRLFALSTNTGEMFA
ERGFQTASEDELPETRKDYRSNGRNSHILVRRLLHR
```

Analyze this sequence

- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence

Protein 3D Structure

Crysta Structure Of N-Acetylglutamate Synthase From Neisseria

PDB: 3B8G

Source: Neisseria gonorrhoeae

Method: X-Ray Diffraction

Resolution: 2.6 Å

See all 4 structures...

Related information

- Similar protein sequences using SmartBlast
- PubMed
- Taxonomy

1:34 PM 12/2/2018

Click on the BLAST icon present at the top of right menu **OR** you can copy the sequence and open BLAST -> Protein to Protein (blastp) and past the sequence in input section.

Chain A, Crysta Structure Of N-A- x

https://www.ncbi.nlm.nih.gov/protein/3B8G_A?report=fasta

NCBI Resources How To Sign in to NCBI

Protein Protein Search Help

Advanced

FASTA - Send to: - Change region shown

Chain A, Crysta Structure Of N-Acetylglutamate Synthase From Neisseria Gonorrhoeae Complexed With Coenzyme A And N-Acetyl-Glutamate

PDB: 3B8G_A

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

```
>pdb|3B8G|A Chain A, Putative Acetylglutamate Synthase
MGSSHHHHHHSSGLVPRGSHINAPDSFVAHFREAAPYIRQIRGTTLVAGIDGRLLGGT LNKLAADIGLL
SQLGIRLVLIHGAYHFLDRLAAQGRTPHYCRGLRVTDETS L GQAQQFAGTVRSRFEALCGSVGFARA
PSVPLVSGNFLTARP IGVIDGTDMEYAVIRKDTAALRFQLDAGNIVMPP L GHSYGGKTFNLDIVQAA
ASVAVSLQAEKLVVYLLTSDGISRPDGT LAETLSAQEAQSLAEHAASETRRLLSSAVAAL EGGVHRVQILN
GAADGSL LQELFT RINGIGTSIAKEAFVSTRQAHSGDIPHIAALTRPLEEQGILLHRSREYLENHI SEFSI
LEHDGNLYGCAALKTF AEADCGE TACLAVSPQAQGGYGERLLAHTIDKARGIGISRLFALSTNTGEMFA
ERGFQTASEDELPETRKDYRSNGRNSHILVRRLLHR
```

Analyze this sequence

- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence

Protein 3D Structure

Crysta Structure Of N-Acetylglutamate Synthase From Neisseria

PDB: 3B8G

Source: Neisseria gonorrhoeae

Method: X-Ray Diffraction

Resolution: 2.6 Å

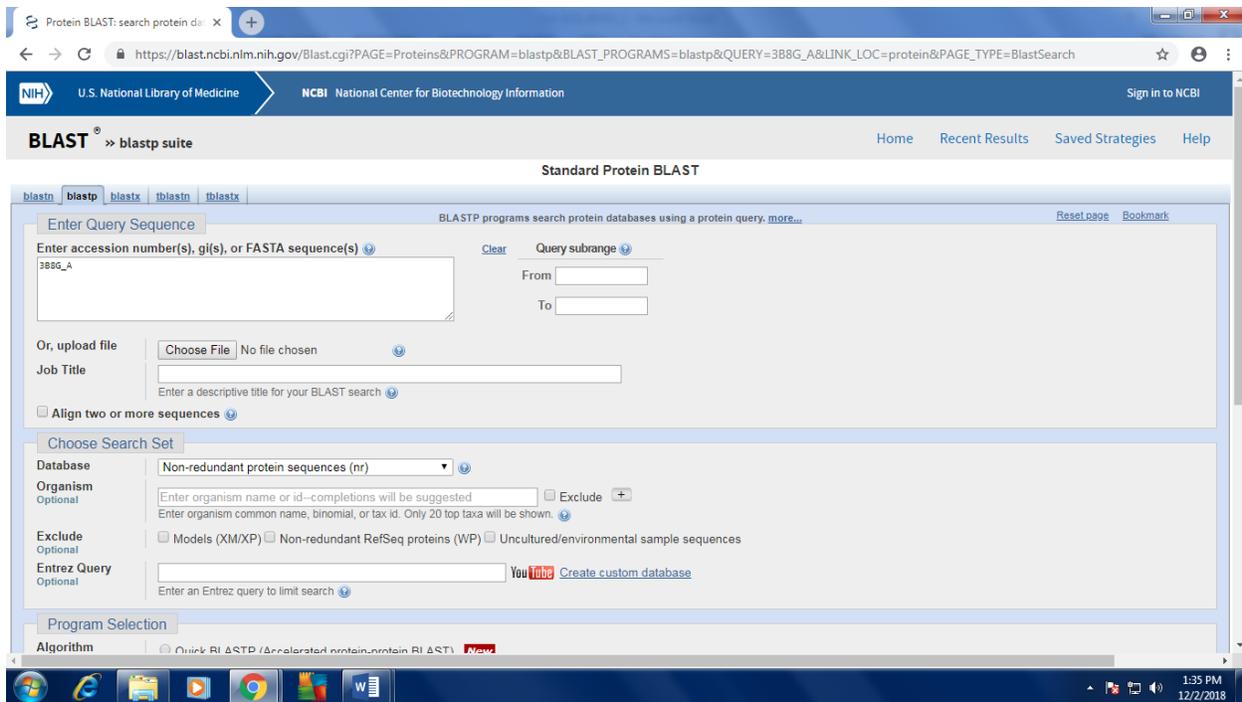
See all 4 structures...

Related information

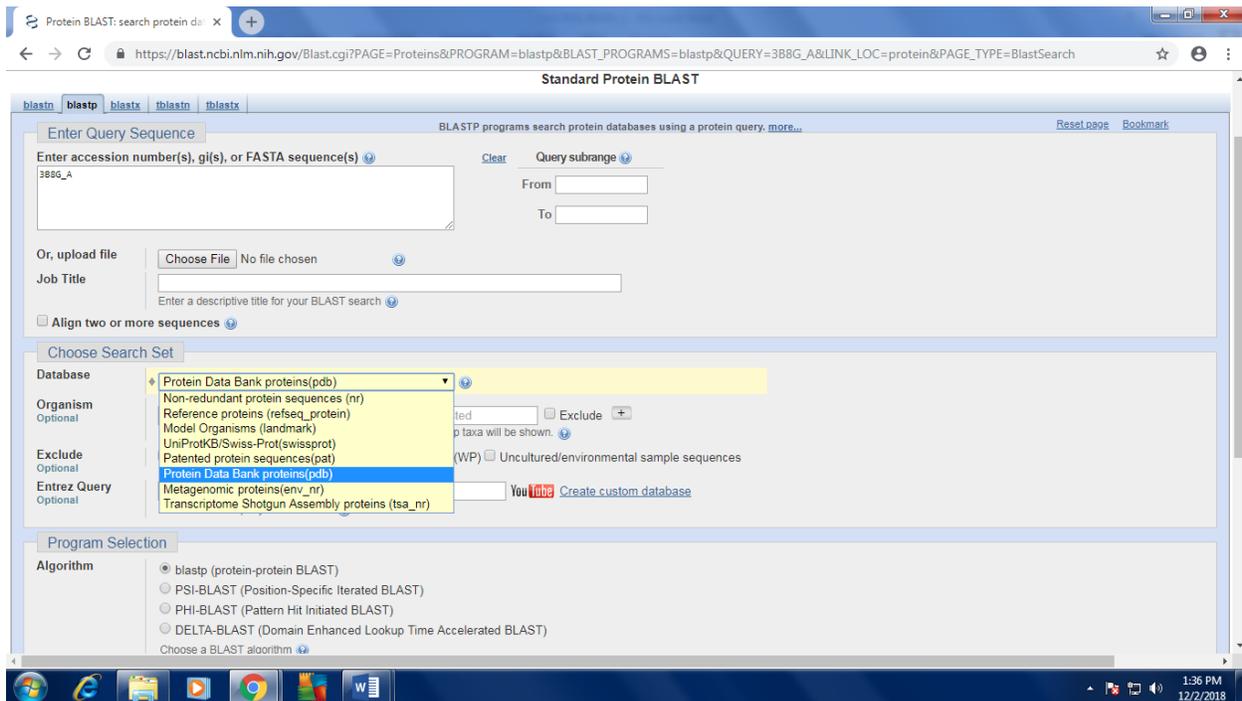
- Similar protein sequences using SmartBlast
- PubMed
- Taxonomy

https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins&PROGRAM=blastp&BLAST_PROGRAMS=blastp&QUERY=3B8G_A&LINK_LOC=protein&PAGE_TYPE=BlastSearch

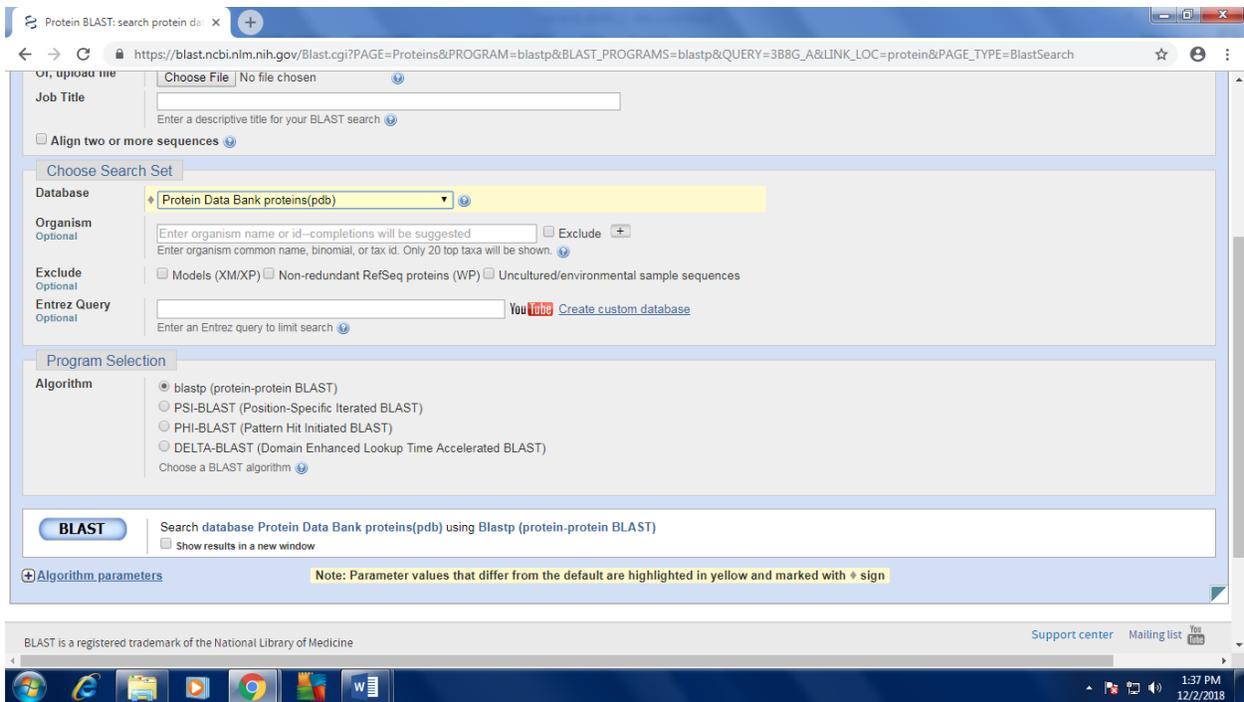
1:34 PM 12/2/2018



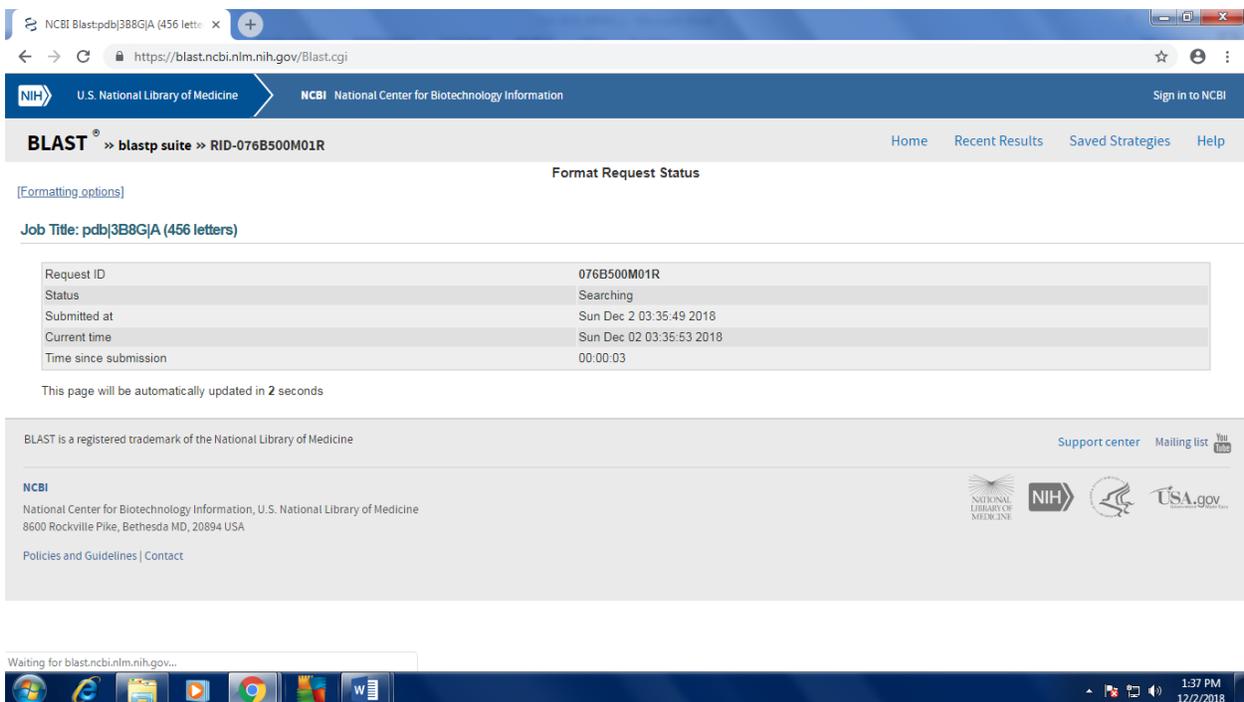
Then select Protein data bank database as reference



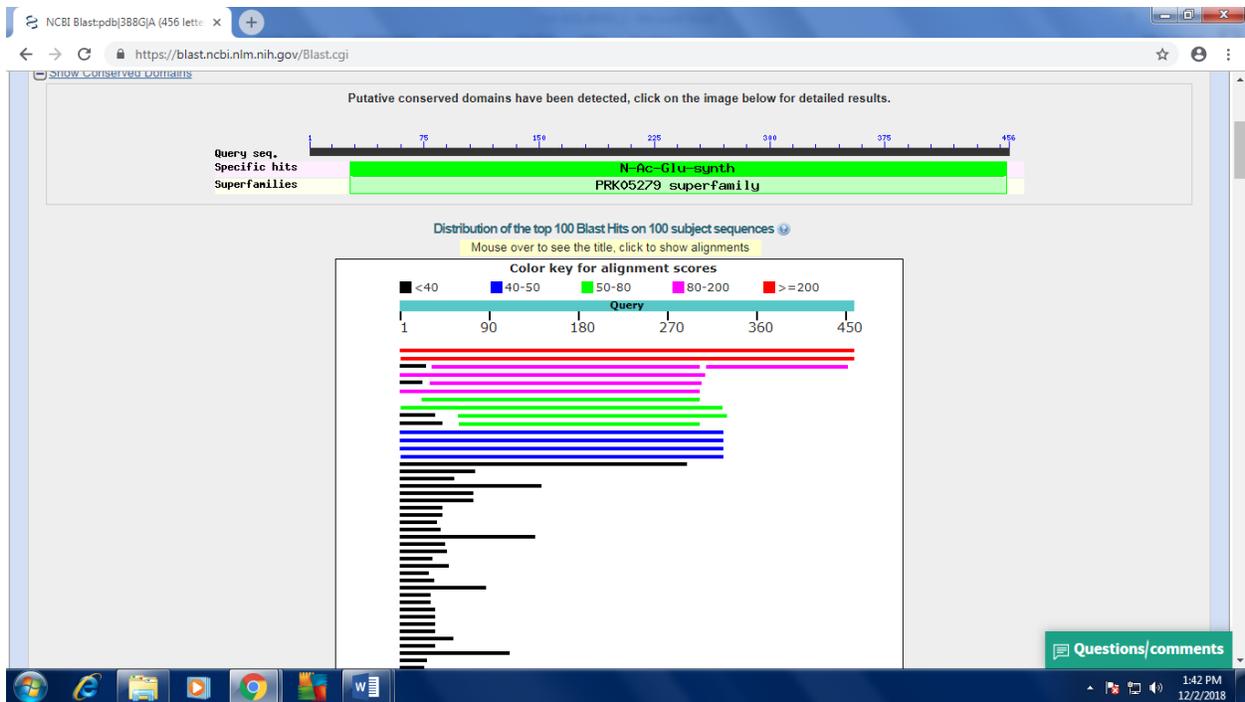
Then Click on the BLAST Icon



It might take some seconds



It will show a page like this



At the lower part of the page you will see a list of reference sequence of protein structure which our sequence has been mapped upon. Where at the right side different columns are giving the information about the identity, Query coverage and Accession ID etc

Description	Max score	Total score	Query cover	E value	Ident	Accession
Chain A, Native Structure Of N-Acetylglutamate Synthase From Neisseria Gonorrhoeae	925	925	100%	0.0	100%	2R8V_A
Chain A, Crystal Structure Of N-Acetylglutamate Synthase (Selenomet Substituted) From Neisseria Gonorrhoeae	909	909	99%	0.0	99%	2R98_A
Chain A, Crystal Structure Of C-Terminal Domain Of N-Acetylglutamate Synthase From Vibrio Parahaemolyticus	110	110	31%	1e-28	43%	3E0K_A
Chain A, Acetylglutamate Kinase From Thermotoga Maritima Complexed With Its Inhibitor Arginine	108	108	58%	2e-26	30%	2BTY_A
Chain A, The Complex Of Pii And Acetylglutamate Kinase From Synechococcus Elongatus Pcc7942	92.4	92.4	67%	2e-20	30%	2JJ4_A
Chain A, Arginine Feed-Back Inhibitable Acetylglutamate Kinase	86.7	86.7	59%	2e-18	26%	2RUF_A
Chain A, N-acetylglutamate Kinase From Arabidopsis Thaliana In Complex With Pii From Chlamydomonas Reinhardtii	85.9	85.9	66%	3e-18	31%	4USJ_A
Chain A, Structural Basis For The Regulation Of N-acetylglutamate Kinase By Pii In Arabidopsis Thaliana	73.9	73.9	61%	3e-14	27%	2RDS_A
Chain A, Crystal Structure Of Xcngas	62.4	62.4	70%	5e-10	27%	3S8K_A
Chain A, Crystal Structure Of A Tetrameric Acetylglutamate Kinase From Saccharomyces Cerevisiae	55.8	55.8	59%	5e-08	24%	3Z7I_A
Chain A, Crystal Structure Of The Amino Acid Kinase Domain From Saccharomyces Cerevisiae Acetylglutamate Kinase Complexed With Its Substrate N-Acetylglutamate	51.2	51.2	53%	1e-06	25%	3Z7E_A
Chain A, Structure mmNAGS bound with L-arginine	47.8	47.8	71%	2e-05	27%	4KZT_A
Chain A, Crystal Structure Of Mmngas In Space Group P2121 At 4.3 A Resolution	47.0	47.0	71%	3e-05	27%	3S7Y_A
Chain A, Crystal Structure Of Native MmngasK	47.0	47.0	71%	4e-05	27%	3S6H_A
Chain A, Crystal structures of Seleno-substituted mutant mmNAGS in space group P212121	41.2	41.2	71%	0.002	26%	3S6G_A
Chain A, The Crystal Structure Of Smu 665 From Streptococcus Mutans Ua159	38.1	38.1	63%	0.014	26%	3L86_A
Chain A, Crystal structure of the beta-glucosidase from Trichoderma harzianum	37.0	37.0	16%	0.053	44%	5BWF_A
Chain A, E. coli LpoA N-terminal domain	35.8	35.8	12%	0.072	50%	2MHK_A
Chain A, Crystal Structure Of Conjugated Polyketone Reductase C2 From Candida Parapsilosis	35.8	35.8	31%	0.096	31%	3VXG_A
Chain A, Structure of the glycoside hydrolase domain of PelA variant E218A from Pseudomonas aeruginosa	35.4	35.4	16%	0.11	46%	5TXY_A
Chain A, Structure of the glycoside hydrolase domain of PelA from Pseudomonas aeruginosa	35.4	35.4	16%	0.12	46%	5TCB_A
Chain C, Crystal Structure Of The Mot1 N-Terminal Domain In Complex With Tbe	34.7	34.7	7%			
Chain A, Small Molecule Inhibitors Of Iao	33.1	33.1	5%			

From here click the Accession IDs of top three references one by one and It will lead you to the following page for each time for each reference.

Chain A, Native Structure Of N-Acetylglutamate Synthase From Neisseria Gonorrhoeae
 PDB: 2R8V_A
[Identical Proteins](#) [FASTA](#) [Graphics](#)

Go to:

LOCUS 2R8V_A 456 aa linear BCT 18-OCT-2012
 DEFINITION Chain A, Putative Acetylglutamate Synthase.
 ACCESSION 2R8V_A
 VERSION 2R8V_A
 DBSOURCE pdb: molecule 2R8V, chain 65, release Oct 13, 2012;
 deposition: Sep 11, 2007;
 class: Transferase;
 source: Mmdb_id: [61814](#), Pdb_id 1: 2R8V;
 Exp. method: X-Ray Diffraction.

KEYWORDS
 SOURCE [Neisseria gonorrhoeae](#)
 ORGANISM [Neisseria gonorrhoeae](#)
 Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
 Neisseriaceae; Neisseria.

REFERENCE 1 (residues 1 to 456)
 AUTHORS Shi,D., Sagar,V., Jin,Z., Yu,X., Caldovic,L., Morizono,H.,
 Allwell,N.M. and Tuchman,H.
 TITLE The crystal structure of N-acetyl-L-glutamate synthase from
 Neisseria gonorrhoeae provides insights into mechanisms of
 catalysis and regulation
 JOURNAL J. Biol. Chem. 283 (11), 7176-7184 (2008)
 PUBLISHED [18184560](#)

REFERENCE 2 (residues 1 to 456)
 AUTHORS Shi,D., Sagar,V., Jin,Z., Yu,X., Caldovic,L., Morizono,H. and
 Allwell,N.M.
 TITLE Direct Submission
 JOURNAL Submitted (11-SEP-2007)
 COMMENT Native Structure Of N-Acetylglutamate Synthase From Neisseria
 Gonorrhoeae

Protein 3D Structure
 Native Structure Of N-Acetylglutamate Synthase From Neisseria
 PDB: 2R8V
 Source: Neisseria gonorrhoeae
 Method: X-Ray Diffraction
 Resolution: 2.5 Å
 See all 4 structures...

Related information
 Similar protein sequences using SmartBlast
 PubMed
 Taxonomy
 Related Sequences

From here get the Complete information like accession number, Locus, Coding region etc. of the three references one by one and make a table.

Then go back to the first page where you found your first sequence from.

Crysta - Protein - NCBI
 Chain A, Native Structure Of N-A-
 https://www.ncbi.nlm.nih.gov/protein/?term=Crysta

Protein
 Create alert Advanced Help

Species Summary - 20 per page - Sort by Default order -
 Animals (8)
 Bacteria (1)
 Archaea (2)
 Customize ...

Source databases PDB (11)
 Customize ...

Sequence length Custom range...

Molecular weight Custom range...

Release date Custom range...

Revision date Custom range...

[Clear all](#)
[Show additional filters](#)

Items: 11

[Chain A, Crysta Structure Of N-Acetylglutamate Synthase From Neisseria Gonorrhoeae Complexed With Coenzyme A And N-Acetyl-Glutamate](#)
 456 aa protein
 Accession: 3B8G_A GI: 165761307
[PubMed](#) [Taxonomy](#) [Related Sequences](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[Chain F, Extra-superfolder Gfp](#)
 238 aa protein
 Accession: 5B61_F GI: 1208076837
[Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[Chain E, Extra-superfolder Gfp](#)
 238 aa protein
 Accession: 5B61_E GI: 1208076836
[Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

[Chain D, Extra-superfolder Gfp](#)
 238 aa protein
 Accession: 5B61_D GI: 1208076835
[Taxonomy](#)
[GenPept](#) [Identical Proteins](#) [FASTA](#) [Graphics](#)

Top Organisms [Tree](#)
 Aequorea victoria (6)
 Halobacterium salinarum (2)
 Gallus gallus (2)
 Neisseria gonorrhoeae (1)

Analyze these sequences
 Run BLAST
 Align sequences with COBAL
 Identify Conserved Domains with CD-Search
 Find in these sequences

Find related data
 Database:

Search details
 Crysta[All Fields]

From the above page see the right top encircled area it will give you the information about the orthologs of the Human Gene.

Past the sequences of the same genes extracted from those three orthologs.

At the end enlist all the databases tools software which you will have used.