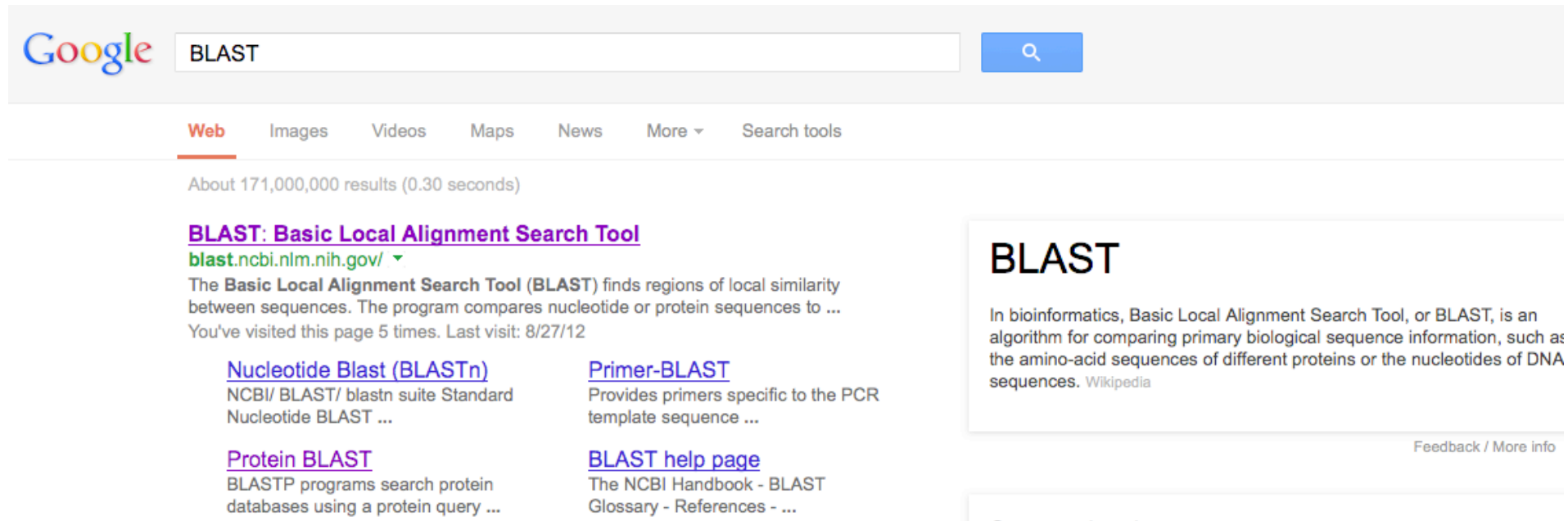


# BLAST assignment instructions

- BLAST 20-mer sequence
- Obtain your protein sequence
- Obtain DNA seq. that codes for protein
- Design primers to clone gene into vector
- Make a multiple sequence alignment
- Make an informative image of protein
- Write 0.5 page summary on protein

# BLAST 20-mer sequence



The image shows a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text "BLAST" and a blue search button with a magnifying glass icon. Below the search bar are navigation tabs for "Web", "Images", "Videos", "Maps", "News", "More", and "Search tools". The "Web" tab is selected and underlined. Below the tabs, it says "About 171,000,000 results (0.30 seconds)".

The main search results area displays a primary result for "BLAST: Basic Local Alignment Search Tool" with a link to [blast.ncbi.nlm.nih.gov/](http://blast.ncbi.nlm.nih.gov/). A snippet of text follows: "The Basic Local Alignment Search Tool (BLAST) finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to ... You've visited this page 5 times. Last visit: 8/27/12".

Below the main result are four related links:

- [Nucleotide Blast \(BLASTn\)](#): NCBI/ BLAST/ blastn suite Standard Nucleotide BLAST ...
- [Primer-BLAST](#): Provides primers specific to the PCR template sequence ...
- [Protein BLAST](#): BLASTP programs search protein databases using a protein query ...
- [BLAST help page](#): The NCBI Handbook - BLAST Glossary - References - ...

On the right side of the search results, there is a knowledge panel titled "BLAST". It contains a brief definition: "In bioinformatics, Basic Local Alignment Search Tool, or BLAST, is an algorithm for comparing primary biological sequence information, such as the amino-acid sequences of different proteins or the nucleotides of DNA sequences. Wikipedia". At the bottom right of the panel, there is a link for "Feedback / More info".

Search for the **B**asic **L**ocal **A**lignment **S**earch **T**ool (BLAST) on the **N**ational **C**enter for **B**io**t**echnology **I**nformation (NCBI) website, and find the Protein BLAST tool

# BLAST 20-mer sequence

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

My NCBI [Sign In] [Register]

NCBI/ BLAST/ blastp suite Standard Protein BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence BLASTP programs search protein databases using a protein query. [more...](#) [Reset page](#) [Bookmark](#)

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#) [Query subrange](#)

DVSEYTKADL FQPGKVTPLA From

To

Or, upload file  No file selected. [?](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database  [?](#)

Organism   Exclude  [?](#)

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude  Models (XM/XP)  Uncultured/environmental sample sequences [?](#)

Input your 20-mer sequence into the “Query Sequence” box ...

The sequence I put here is DVSEYTKADL FQPGKVTPLA

# BLAST 20-mer sequence

Align two or more sequences

### Choose Search Set

**Database**  
Non-redundant protein sequences (nr)

**Organism**  
Optional  
Enter organism name or id—completions will be suggested  Exclude +  
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown.

**Exclude**  
Optional  
 Models (XM/XP)  Uncultured/environmental sample sequences

**Entrez Query**  
Optional  
Enter an Entrez query to limit search

### Program Selection

**Algorithm**

- blastp (protein-protein BLAST)
- PSI-BLAST (Position-Specific Iterated BLAST)
- PHI-BLAST (Pattern Hit Initiated BLAST)
- DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

Choose a BLAST algorithm


**BLAST** Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)  
 Show results in a new window

[+ Algorithm parameters](#)

... and scroll to the bottom and hit the large, blue BLAST (!) button

# BLAST 20-mer sequence

Job Title: Protein Sequence (20 letters)

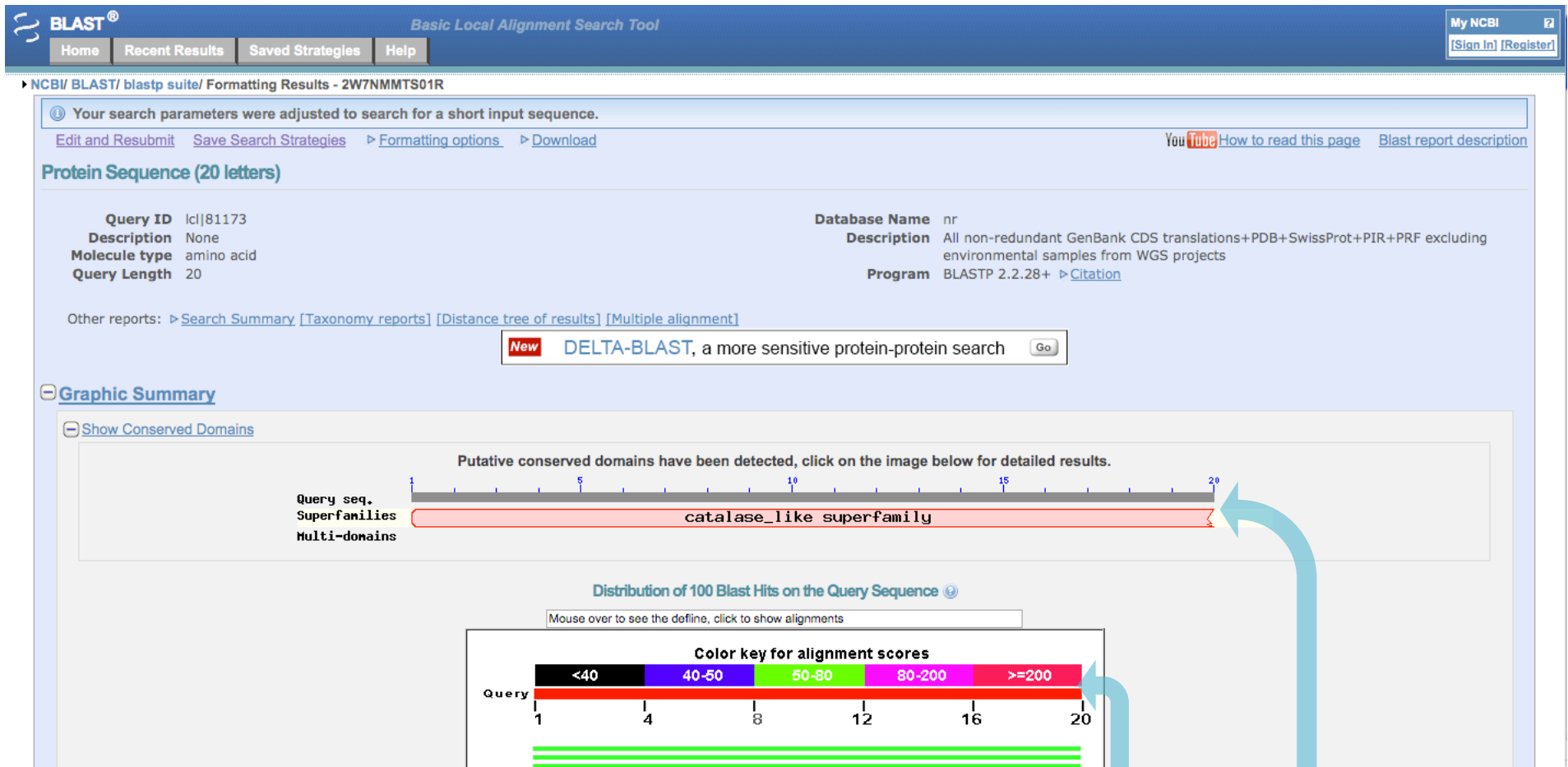
 Your search parameters were adjusted to search for a short input sequence.

Request ID	2W7NMMTS01R
Status	Searching
Submitted at	Mon Sep 9 15:17:08 2013
Current time	Mon Sep 9 15:17:23 2013
Time since submission	00:00:13

This page will be automatically updated in 7 seconds

Be patient while it compares your 20-mer sequence to the entire database and ranks the matches for you. This can take a while during regular work hours.

# BLAST 20-mer sequence



It has finished! It has recognized that it shares similarity to a protein superfamily (catalase in this instance) and has graphically shown the matches and their respective scores.

# BLAST 20-mer sequence

Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium glutamicum ATCC 13032]</a>	65.5	65.5	100%	5e-11	100%	<a href="#">BAB97648.1</a>
<input type="checkbox"/>	<a href="#">Chain A, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F B Chain B, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F C Chain C, Structure Of A Liganded Bacter</a>	65.5	65.5	100%	5e-11	100%	<a href="#">4B7F_A</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_224555.1  catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_007559511.1  catalase [Corynebacteri</a>	65.5	65.5	100%	5e-11	100%	<a href="#">NP_599508.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cgR_0332 [Corynebacterium glutamicum RI] &gt;ref YP_008065043.1  catalase [Corynebacterium glutamicum SCgG1] &gt;ref YP_008068066.1  catalase [Coryr</a>	65.5	65.5	100%	5e-11	100%	<a href="#">YP_001137198.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium callunae DSM 20147] &gt;ref WP_015650135.1  catalase [Corynebacterium callunae] &gt;gb AGG65680.1  catalase [Corynebacterium callunae DSM 20</a>	63.0	63.0	100%	4e-10	95%	<a href="#">YP_007529583.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium efficiens YS-314] &gt;ref WP_006768464.1  catalase [Corynebacterium efficiens] &gt;dbj BAC17034.1  putative catalase [Corynebacterium efficiens YS-3</a>	57.5	57.5	95%	3e-08	89%	<a href="#">NP_736834.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium nuruki]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">WP_010121065.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein jk1994 [Corynebacterium jeikeium K411] &gt;ref WP_005292312.1  catalase [Corynebacterium jeikeium] &gt;emb CAI38176.1  kAtA [Corynebacterium jeikeium ]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">YP_251794.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium bovis]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">WP_010269648.1</a>
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium kroppenstedtii DSM 44385] &gt;ref WP_012730865.1  catalase [Corynebacterium kroppenstedtii] &gt;gb ACR16977.1  Catalase [Corynebacterium kropp</a>	52.8	52.8	100%	1e-06	81%	<a href="#">YP_002905520.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium propinquum]</a>	52.4	52.4	95%	1e-06	84%	<a href="#">WP_018121082.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium halotolerans YIM 70093 = DSM 44683] &gt;ref WP_015399757.1  catalase [Corynebacterium halotolerans] &gt;gb AGF71333.1  catalase [Corynebacteri</a>	51.5	51.5	100%	3e-06	85%	<a href="#">YP_007463695.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium lipophiloflavum] &gt;gb EEI16204.1  catalase [Corynebacterium lipophiloflavum DSM 44291]</a>	50.7	50.7	100%	5e-06	80%	<a href="#">WP_006839270.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium genitalium] &gt;gb EFK55180.1  catalase [Corynebacterium genitalium ATCC 33030]</a>	50.7	50.7	100%	5e-06	80%	<a href="#">WP_005287252.1</a>
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium variabile DSM 44702] &gt;ref WP_014010988.1  catalase [Corynebacterium variabile] &gt;gb AEK37836.1  Catalase [Corynebacterium variabile DSM 44</a>	50.7	50.7	95%	5e-06	79%	<a href="#">YP_004760909.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium urealyticum DSM 7111] &gt;ref WP_015382036.1  catalase [Corynebacterium urealyticum] &gt;gb AGE37398.1  catalase [Corynebacterium urealyticum ]</a>	49.4	49.4	100%	1e-05	81%	<a href="#">YP_007417649.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cur_1884 [Corynebacterium urealyticum DSM 7109] &gt;ref WP_012361117.1  catalase [Corynebacterium urealyticum] &gt;emb CAQ05843.1  unnamed protein</a>	49.4	49.4	100%	1e-05	81%	<a href="#">YP_001801277.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium caspium]</a>	49.0	49.0	100%	2e-05	80%	<a href="#">WP_018340689.1</a>

Scroll down to see a description of the protein matches, an arbitrary score to rank them to each other, the query cover (the % of the input that was found to align with the produced sequence), the expectation value (this should be close to zero), and the % identical residues.

# BLAST 20-mer sequence

Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium glutamicum ATCC 13032]</a>	65.5	65.5	100%	5e-11	100%	<a href="#">BAB97648.1</a>
<input type="checkbox"/>	<a href="#">Chain A, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F B Chain B, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F C Chain C, Structure Of A Liganded Bacter</a>	65.5	65.5	100%	5e-11	100%	<a href="#">4B7F_A</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_224555.1  catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_007559511.1  catalase [Corynebacteri</a>	65.5	65.5	100%	5e-11	100%	<a href="#">NP_599508.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cgR_0332 [Corynebacterium glutamicum RI] &gt;ref YP_008065043.1  catalase [Corynebacterium glutamicum SCgG1] &gt;ref YP_008068066.1  catalase [Coryr</a>	65.5	65.5	100%	5e-11	100%	<a href="#">YP_001137198.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium callunae DSM 20147] &gt;ref WP_015650135.1  catalase [Corynebacterium callunae] &gt;gb AGG65680.1  catalase [Corynebacterium callunae DSM 20</a>	63.0	63.0	100%	4e-10	95%	<a href="#">YP_007529583.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium efficiens YS-314] &gt;ref WP_006768464.1  catalase [Corynebacterium efficiens] &gt;dbj BAC17034.1  putative catalase [Corynebacterium efficiens YS-3</a>	57.5	57.5	95%	3e-08	89%	<a href="#">NP_736834.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium nuruki]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">WP_010121065.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein jk1994 [Corynebacterium jeikeium K411] &gt;ref WP_005292312.1  catalase [Corynebacterium jeikeium] &gt;emb CAI38176.1  kata [Corynebacterium jeikeium ]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">YP_251794.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium bovis]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">WP_010269648.1</a>
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium kroppenstedtii DSM 44385] &gt;ref WP_012730865.1  catalase [Corynebacterium kroppenstedtii] &gt;gb ACR16977.1  Catalase [Corynebacterium kropp</a>	52.8	52.8	100%	1e-06	81%	<a href="#">YP_002905520.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium propinquum]</a>	52.4	52.4	95%	1e-06	84%	<a href="#">WP_018121082.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium halotolerans YIM 70093 = DSM 44683] &gt;ref WP_015399757.1  catalase [Corynebacterium halotolerans] &gt;gb AGF71333.1  catalase [Corynebacteri</a>	51.5	51.5	100%	3e-06	85%	<a href="#">YP_007463695.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium lipophiloflavum] &gt;gb EEI16204.1  catalase [Corynebacterium lipophiloflavum DSM 44291]</a>	50.7	50.7	100%	5e-06	80%	<a href="#">WP_006839270.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium genitalium] &gt;gb EFK55180.1  catalase [Corynebacterium genitalium ATCC 33030]</a>	50.7	50.7	100%	5e-06	80%	<a href="#">WP_005287252.1</a>
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium variabile DSM 44702] &gt;ref WP_014010988.1  catalase [Corynebacterium variabile] &gt;gb AEK37836.1  Catalase [Corynebacterium variabile DSM 44</a>	50.7	50.7	95%	5e-06	79%	<a href="#">YP_004760909.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium urealyticum DSM 7111] &gt;ref WP_015382036.1  catalase [Corynebacterium urealyticum] &gt;gb AGE37398.1  catalase [Corynebacterium urealyticum ]</a>	49.4	49.4	100%	1e-05	81%	<a href="#">YP_007417649.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cur_1884 [Corynebacterium urealyticum DSM 7109] &gt;ref WP_012361117.1  catalase [Corynebacterium urealyticum] &gt;emb CAQ05843.1  unnamed protein</a>	49.4	49.4	100%	1e-05	81%	<a href="#">YP_001801277.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium caspium]</a>	49.0	49.0	100%	2e-05	80%	<a href="#">WP_018340689.1</a>

You should pick a protein that has 100% query cover and 100% identity. The score and the E value don't matter much for this step. There may be many options that fit these categories (either very similar or identical proteins from multiple databases).



# BLAST 20-mer sequence

Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected: 0

Alignments Download GenPept Graphics Distance tree of results Multiple alignment

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium glutamicum ATCC 13032]</a>	65.5	65.5	100%	5e-11	100%	<a href="#">BAB97648.1</a>
<input type="checkbox"/>	<a href="#">Chain A, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F B Chain B, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F C Chain C, Structure Of A Liganded Bacter</a>	65.5	65.5	100%	5e-11	100%	<a href="#">4B7F_A</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_224555.1  catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_007559511.1  catalase [Corynebacteri</a>	65.5	65.5	100%	5e-11	100%	<a href="#">NP_599508.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cgR_0332 [Corynebacterium glutamicum R] &gt;ref YP_008065043.1  catalase [Corynebacterium glutamicum SCgG1] &gt;ref YP_008068066.1  catalase [Coryr</a>	65.5	65.5	100%	5e-11	100%	<a href="#">YP_001137198.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium callunae DSM 20147] &gt;ref WP_015650135.1  catalase [Corynebacterium callunae] &gt;gb AGG65680.1  catalase [Corynebacterium callunae DSM 20</a>	63.0	63.0	100%	4e-10	95%	<a href="#">YP_007529583.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium efficiens YS-314] &gt;ref WP_006768464.1  catalase [Corynebacterium efficiens] &gt;dbj BAC17034.1  putative catalase [Corynebacterium efficiens YS-3</a>	57.5	57.5	95%	3e-08	89%	<a href="#">NP_736834.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium nuruki]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">WP_010121065.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein jk1994 [Corynebacterium jeikeium K411] &gt;ref WP_005292312.1  catalase [Corynebacterium jeikeium] &gt;emb CAI38176.1  kAtA [Corynebacterium jeikeium ]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">YP_251794.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium bovis]</a>	53.2	53.2	100%	8e-07	86%	<a href="#">WP_010269648.1</a>
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium kroppenstedtii DSM 44385] &gt;ref WP_012730865.1  catalase [Corynebacterium kroppenstedtii] &gt;gb ACR16977.1  Catalase [Corynebacterium kropp</a>	52.8	52.8	100%	1e-06	81%	<a href="#">YP_002905520.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium propinquum]</a>	52.4	52.4	95%	1e-06	84%	<a href="#">WP_018121082.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium halotolerans YIM 70093 = DSM 44683] &gt;ref WP_015399757.1  catalase [Corynebacterium halotolerans] &gt;gb AGF71333.1  catalase [Corynebacteri</a>	51.5	51.5	100%	3e-06	85%	<a href="#">YP_007463695.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium lipophiloflavum] &gt;gb EEI16204.1  catalase [Corynebacterium lipophiloflavum DSM 44291]</a>	50.7	50.7	100%	5e-06	80%	<a href="#">WP_006839270.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium genitalium] &gt;gb EFK55180.1  catalase [Corynebacterium genitalium ATCC 33030]</a>	50.7	50.7	100%	5e-06	80%	<a href="#">WP_005287252.1</a>
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium variabile DSM 44702] &gt;ref WP_014010988.1  catalase [Corynebacterium variabile] &gt;gb AEK37836.1  Catalase [Corynebacterium variabile DSM 44</a>	50.7	50.7	95%	5e-06	79%	<a href="#">YP_004760909.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium urealyticum DSM 7111] &gt;ref WP_015382036.1  catalase [Corynebacterium urealyticum] &gt;gb AGE37398.1  catalase [Corynebacterium urealyticum ]</a>	49.4	49.4	100%	1e-05	81%	<a href="#">YP_007417649.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cur_1884 [Corynebacterium urealyticum DSM 7109] &gt;ref WP_012361117.1  catalase [Corynebacterium urealyticum] &gt;emb CAQ05843.1  unnamed protein</a>	49.4	49.4	100%	1e-05	81%	<a href="#">YP_001801277.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium caspium]</a>	49.0	49.0	100%	2e-05	80%	<a href="#">WP_018340689.1</a>

Notice I have four options that have 100% query cover and 100% identity. They are probably identical sequences, as they come from the same organism, *Corynebacterium glutamicum*.

# BLAST 20-mer sequence

Descriptions

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

[Alignments](#) [Download](#) [GenPept](#) [Graphics](#) [Distance tree of results](#) [Multiple alignment](#)

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium glutamicum ATCC 13032]</a>	65.5	65.5	100%	5e-11	100%	<a href="#">BAB97648.1</a>
<input type="checkbox"/>	<a href="#">Chain A, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F B Chain B, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7F C Chain C, Structure Of A Liganded Bacter</a>	65.5	65.5	100%	5e-11	100%	<a href="#">4B7F_A</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_224555.1  catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_007559511.1  catalase [Corynebacteri</a>	65.5	65.5	100%	5e-11	100%	<a href="#">NP_599508.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cgR_0332 [Corynebacterium glutamicum R] &gt;ref YP_008065043.1  catalase [Corynebacterium glutamicum SCgG1] &gt;ref YP_008068066.1  catalase [Coryr</a>	65.5	65.5	100%	5e-11	100%	<a href="#">YP_001137198.1</a>

Also notice that one reference has an associated 3D structure. References from this database have a similar accession number, so watch for alphanumeric codes like 4B7F\_A

BEWARE!! There *should* be a file with a structure in your first BLAST search, but the sequence may not be identical to what is genomically coded. Watch out for partial or mutant proteins or a designation like E72A, which signifies residue #72, Glu, has been mutated to Ala.

# Obtain your protein sequence

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

Display Settings: GenPept Send to:

## Chain A, Structure Of A Liganded Bacterial Catalase

PDB: 4B7F\_A  
[FASTA](#) [Graphics](#)

Go to:

LOCUS 4B7F\_A 515 aa linear BCT 28-AUG-2013  
DEFINITION Chain A, Structure Of A Liganded Bacterial Catalase.  
ACCESSION 4B7F\_A  
VERSION 4B7F\_A GI:534286111  
DBSOURCE pdb: molecule 4B7F, chain 65, release Aug 28, 2013;  
deposition: Aug 20, 2012;  
class: Oxidoreductase;  
source: Mmdb\_id: [112949](#), Pdb\_id 1: 4B7F;  
Exp. method: X-Ray Diffraction.

KEYWORDS .  
SOURCE Corynebacterium glutamicum ATCC 13032  
ORGANISM [Corynebacterium glutamicum ATCC 13032](#)  
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
Corynebacterineae; Corynebacteriaceae; Corynebacterium.

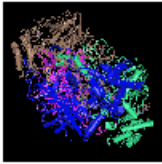
REFERENCE 1 (residues 1 to 515)  
AUTHORS Gumiero,A. and Walsh,M.  
TITLE Structure Of A Liganded Bacterial Catalase

Change region shown  
Customize view

### Analyze this sequence

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

### Protein 3D Structure



Structure Of A Liganded Bacterial Catalase  
PDB: 4B7F  
Source: Corynebacterium glutamicum ATCC 13032  
Method: X-Ray Diffraction  
Resolution: 1.76 Å

Clicking on the accession number will bring you to a webpage dedicated to everything associated with this protein in the database, such as the organism and any journal articles that use this protein.

# Obtain your protein sequence

```
.....
Het          bond(352)
             /heterogen="(HEM,1004 )"
SecStr       454..461
             /sec_str_type="helix"
             /note="helix 8"
SecStr       466..480
             /sec_str_type="helix"
             /note="helix 9"
SecStr       488..498
             /sec_str_type="helix"
             /note="helix 10"
SecStr       499..514
             /sec_str_type="helix"
             /note="helix 11"

ORIGIN
   1 seksaadqiv drgmrpklsq nttrhngapv psenisatag pggpnvlndi hlieklahfn
   61 renvperiph akghgafgel hitedvseyt kadlfqpgkv tplavrfstv ageqgsptw
  121 rdvhgfalrf yteegnydiv gntptfflr dgmkfpdfih sqkrlnkngl rdadmqwdfw
  181 trapesahqv tylmgdrgtp ktsrhqdgfg shtfgwinae gkpvwvkyhf ktrqgwdcft
  241 daeaakvage nadyqredly naiengdfpi wdvkvqimpf edaenyrwnp fdltktsqk
  301 dyplipvgyf ilnrnprnff aqieqialdp gnivpgvqls pdrmlqarif ayadqgryri
  361 ganyrdlpvn rpinevntys regsmqyifd aegepsyspn rydkgagyld ngtdsssnt
  421 syggaddiyv npdphgtdlv raayvkhqdd ddfiqpgily revldegeke rladnisnam
  481 qgiseatepr vydywnnvde nlgarvkely lqkka

//
```

The bottom of this page lists the protein sequence. If you picked a structure accession number, notice that the N-terminal methionine is usually missing (and possibly other residues)

# Obtain DNA seq. that codes for protein

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

Display Settings: GenPept Send to: Change region shown Customize view

## Chain A, Structure Of A Liganded Bacterial Catalase

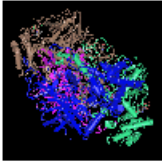
PDB: 4B7F\_A  
[FASTA](#) [Graphics](#)

Go to:

LOCUS	4B7F_A	515 aa	linear
DEFINITION	Chain A, Structure Of A Liganded Bacterial Catalase		
ACCESSION	4B7F_A		
VERSION	4B7F_A GI:534286111		
DBSOURCE	pdb: molecule 4B7F, chain 65, release Aug 28, 2013; deposition: Aug 20, 2012; class: Oxidoreductase; source: Mmdb_id: <a href="#">112949</a> , Pdb_id 1: 4B7F; Exp. method: X-Ray Diffraction.		
KEYWORDS	.		
SOURCE	Corynebacterium glutamicum ATCC 13032		
ORGANISM	<a href="#">Corynebacterium glutamicum ATCC 13032</a> Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.		
REFERENCE	1 (residues 1 to 515)		
AUTHORS	Gumiero,A. and Walsh,M.		
TITLE	Structure Of A Liganded Bacterial Catalase		
JOURNAL	Unpublished		
REFERENCE	2 (residues 1 to 515)		

Find regions of similarity between this sequence and other sequences using BLAST.

Analyze this sequence  
Run BLAST  
Identify Conserved Domains  
Highlight Sequence Features  
Find in this Sequence

Protein 3D Structure  
  
Structure Of A Liganded Bacterial Catalase  
PDB: 4B7F  
Source: Corynebacterium glutamicum ATCC 13032  
Method: X-Ray Diffraction  
Resolution: 1.76 Å

Identical proteins for 4B7F\_A  
Chain D, Structure Of A Highdose Ligar[4B7H\_D]

Now we need to find the DNA sequence for the protein. You can directly BLAST this protein by clicking this link found at the top of the page.

# Obtain DNA seq. that codes for protein

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

Alignments						
Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/> Chain A, Structure Of A Liganded Bacterial Catalase >pdb 4B7FIB Chain B, Structure Of A Liganded Bacterial Catalase >pdb 4B7FIC Chain C, Structure Of A Liganded Bacteri	1072	1072	100%	0.0	100%	<a href="#">4B7F_A</a>
<input type="checkbox"/> catalase [Corynebacterium glutamicum ATCC 13032] >ref YP_224555.1  catalase [Corynebacterium glutamicum ATCC 13032] >ref YP_007559511.1  catalase [Corynebacteriu	1071	1071	100%	0.0	99%	<a href="#">NP_599508.1</a>
<input type="checkbox"/> hypothetical protein cgR_0332 [Corynebacterium glutamicum RI] >ref YP_008065043.1  catalase [Corynebacterium glutamicum SCqG1] >ref YP_008068066.1  catalase [Coryn	1071	1071	100%	0.0	99%	<a href="#">YP_001137198.1</a>
<input type="checkbox"/> Catalase [Corynebacterium glutamicum ATCC 13032]	1046	1046	97%	0.0	99%	<a href="#">BAB97648.1</a>
<input type="checkbox"/> catalase [Corynebacterium callunae DSM 20147] >ref WP_015650135.1  catalase [Corynebacterium callunae] >gb AGG65680.1  catalase [Corynebacterium callunae DSM 201	1021	1021	100%	0.0	94%	<a href="#">YP_007529583.1</a>
<input type="checkbox"/> catalase [Corynebacterium efficiens YS-314] >ref WP_006768464.1  catalase [Corynebacterium efficiens] >dbj BAC17034.1  putative catalase [Corynebacterium efficiens YS-3	964	964	100%	0.0	88%	<a href="#">NP_736834.1</a>
<input type="checkbox"/> catalase [Corynebacterium pilosum]	883	883	100%	0.0	81%	<a href="#">WP_018581253.1</a>
<input type="checkbox"/> catalase [Corynebacterium casei] >emb CCE53994.1  catalase [Corynebacterium casei UCMA 3821]	876	876	99%	0.0	81%	<a href="#">WP_006821561.1</a>
<input type="checkbox"/> catalase [Corynebacterium ammoniagenes] >gb EFG82166.1  catalase [Corynebacterium ammoniagenes DSM 20306]	875	875	99%	0.0	81%	<a href="#">WP_003846068.1</a>
<input type="checkbox"/> catalase [Corynebacterium halotolerans YIM 70093 = DSM 44683] >ref WP_015399757.1  catalase [Corynebacterium halotolerans] >gb AGF71333.1  catalase [Corynebacteriu	872	872	99%	0.0	81%	<a href="#">YP_007463695.1</a>
<input type="checkbox"/> catalase [Corynebacterium lubricantis]	866	866	100%	0.0	80%	<a href="#">WP_018297951.1</a>
<input type="checkbox"/> catalase [Corynebacterium ulcerans 0102] >ref WP_014835811.1  catalase [Corynebacterium ulcerans] >dbj BAM26465.1  catalase [Corynebacterium ulcerans 0102]	851	851	99%	0.0	79%	<a href="#">YP_006493700.1</a>
<input type="checkbox"/> catalase [Corynebacterium ulcerans BR-AD22] >ref YP_005709860.1  catalase [Corynebacterium ulcerans 809] >ref WP_013910628.1  catalase [Corynebacterium ulcerans] >	849	849	99%	0.0	79%	<a href="#">YP_004628860.1</a>
<input type="checkbox"/> catalase [Corynebacterium pseudotuberculosis Cp162] >ref WP_014799916.1  catalase [Corynebacterium pseudotuberculosis] >gb AFM06551.1  Catalase [Corynebacterium p	842	842	99%	0.0	78%	<a href="#">YP_006436344.1</a>
<input type="checkbox"/> catalase [Corynebacterium pseudotuberculosis FRC41] >ref YP_005122300.1  catalase [Corynebacterium pseudotuberculosis 3/99-5] >ref YP_005374218.1  katA gene produc	842	842	99%	0.0	78%	<a href="#">YP_003782575.1</a>
<input type="checkbox"/> catalase [Corynebacterium pseudotuberculosis 31] >ref WP_014654992.1  catalase [Corynebacterium pseudotuberculosis] >gb AFH90022.1  Catalase [Corynebacterium pseu	840	840	99%	0.0	78%	<a href="#">YP_006212712.1</a>
<input type="checkbox"/> catalase [Corynebacterium diphtheriae INCA 402] >ref WP_014302799.1  catalase [Corynebacterium diphtheriae] >gb AEX45515.1  catalase [Corynebacterium diphtheriae INC	839	839	99%	0.0	77%	<a href="#">YP_005126717.1</a>

Here you have all the proteins that are nearly identical to yours! I BLASTed a protein with an associated structure, so notice that the next few matches are the identical protein without the N-terminal methionine, so it is 99% identical over 100% of the query.

# Obtain DNA seq. that codes for protein

Download ▾ GenPept Graphics ▾ Next ▲ Previous ▲ Descriptions

catalase [Corynebacterium glutamicum ATCC 13032]  
Sequence ID: [ref|NP\\_599508.1](#) Length: 516 Number of Matches: 1  
[▶ See 7 more title\(s\)](#)

Range 1: 2 to 516 [GenPept](#) [Graphics](#) ▾ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
1071 bits(2771)	0.0	Compositional matrix adjust.	514/515(99%)	515/515(100%)	0/515(0%)

Query	1	SEKSAADQIVDRGMRPKLSGNTTRHNGAPVPS	ENISATAGPQGPVNLNDIHLIEKLAHFN	60
Sbjct	2	SEKSAADQIVDRGMRPKLSGNTTRHNGAPVPS	ENISATAGPQGPVNLNDIHLIEKLAHFN	61
Query	61	RENVPERIPHAKGHGAFGELHITEDVSEYTKADLFQPGKVTPLAVRFSTVAGEQGSPTW	RENVPERIPHAKGHGAFGELHITEDVSEYTKADLFQPGKVTPLAVRFSTVAGEQGSPTW	120
Sbjct	62	RENVPERIPHAKGHGAFGELHITEDVSEYTKADLFQPGKVTPLAVRFSTVAGEQGSPTW	RENVPERIPHAKGHGAFGELHITEDVSEYTKADLFQPGKVTPLAVRFSTVAGEQGSPTW	121
Query	121	RDVHGFAFRFYTEEGNYDIVGNNTPTFFLRDGMKFPDFIHSQKRLNKNGLRDADMQWDFW	RDVHGFAFRFYTEEGNYDIVGNNTPTFFLRDGMKFPDFIHSQKRLNKNGLRDADMQWDFW	180
Sbjct	122	RDVHGFAFRFYTEEGNYDIVGNNTPTFFLRDGMKFPDFIHSQKRLNKNGLRDADMQWDFW	RDVHGFAFRFYTEEGNYDIVGNNTPTFFLRDGMKFPDFIHSQKRLNKNGLRDADMQWDFW	181
Query	181	TRAPESAHQVTYLMGDRGTPKTSRHQDGFSGHTFQWINAEGKPVVWKYHFKTRQGWDCFT	TRAPESAHQVTYLMGDRGTPKTSRHQDGFSGHTFQWINAEGKPVVWKYHFKTRQGWDCFT	240
Sbjct	182	TRAPESAHQVTYLMGDRGTPKTSRHQDGFSGHTFQWINAEGKPVVWKYHFKTRQGWDCFT	TRAPESAHQVTYLMGDRGTPKTSRHQDGFSGHTFQWINAEGKPVVWKYHFKTRQGWDCFT	241
Query	241	DAEAAK VAGENADYQREDLYNAIENGDFPIWDVKVQIMPFEDAENYRWNPFDLTKTWSQK	DAEAAK VAGENADYQREDLYNAIENGDFPIWDVKVQIMPFEDAENYRWNPFDLTKTWSQK	300
Sbjct	242	DAEAAK VAGENADYQREDLYNAIENGDFPIWDVKVQIMPFEDAENYRWNPFDLTKTWSQK	DAEAAK VAGENADYQREDLYNAIENGDFPIWDVKVQIMPFEDAENYRWNPFDLTKTWSQK	301
Query	301	DYPLIPVGYFILLNRNPRNFFAQIEQIALDPGNIVPGVGLSPDRMLQARIFAYADQQRRI	DYPLIPVGYFILLNRNPRNFFAQIEQIALDPGNIVPGVGLSPDRMLQARIFAYADQQRRI	360
Sbjct	302	DYPLIPVGYFILLNRNPRNFFAQIEQIALDPGNIVPGVGLSPDRMLQARIFAYADQQRRI	DYPLIPVGYFILLNRNPRNFFAQIEQIALDPGNIVPGVGLSPDRMLQARIFAYADQQRRI	361
Query	361	GANYRDLPVNRPINEVNTYSREGSMQYIFDAEGEPSYSPNRYDKGAGYLDNGTDSSSNHT	GANYRDLPVNRPINEVNTYSREGSMQYIFDAEGEPSYSPNRYDKGAGYLDNGTDSSSNHT	420
Sbjct	362	GANYRDLPVNRPINEVNTYSREGSMQYIFDAEGEPSYSPNRYDKGAGYLDNGTDSSSNHT	GANYRDLPVNRPINEVNTYSREGSMQYIFDAEGEPSYSPNRYDKGAGYLDNGTDSSSNHT	421
Query	421	SYGQADDIVNPDPHGTDLVRAAYVKHQDDDDFIQPGILYREVLDGEGERLADNISNAM	SYGQADDIVNPDPHGTDLVRAAYVKHQDDDDFIQPGILYREVLDGEGERLADNISNAM	480
Sbjct	422	SYGQADDIVNPDPHGTDLVRAAYVKHQDDDDFIQPGILYREVLDGEGERLADNISNAM	SYGQADDIVNPDPHGTDLVRAAYVKHQDDDDFIQPGILYREVLDGEGERLADNISNAM	481
Query	481	QGI SEATEPRVYDYWNNVDENLGARVKELYLQKKA	515	
Sbjct	482	QGI SEATEPRVYDYWNNVDENLGARVKELYLQKKA	516	

**Related Information**

- [Gene](#) - associated gene details
- [Identical Proteins](#) - Proteins identical to the subject

Obtaining the DNA sequence can be tricky, so it is a good idea to have a few entries to look through. Be sure that this sequence that you pick is identical to your 20-mer peptide sequence (scroll down to see pairwise alignments).

# Obtain DNA seq. that codes for protein

```
CONTIG      join(WP_011013509.1:1..516)
ORIGIN
1  msekсадqi vdrgrmpkls gnttrhngap vpsenisata gpggpnvln d ihlieklahf
61 nrenvperip hakghgafge lhitedvsey tkadlfqpgk vtplavrfst vageggspdt
121 wrdvhgfalr fyteegnydi vgnntptffl rdgmkfpd fi hsqkrlnkng lrdadmqwdf
181 wtrapesahq vtylmgdrgt pktsrhqdgf gshtfgwina egkpvvvykh fktrggwdcf
241 tdaaaakvag enadyqredl ynaiengdfp iwdvkvqimp fedaenyrrn pfdltktsq
301 kdyp lipvgy filnrnprnf faieqlald pgnivpgvgl spdrmlqari fayadqgryr
361 iganyrdlpv nrpinevnty sregsmqyif daegepsysp nrydkgagyl dngtdssnh
421 tsyggaddiy vnpdphgtdl vraayvkhqd dddfiqpgil yrevldegek erladnisna
481 mggiseatep rvydywnnvd enlgarvkel ylqkka
//
```

- Nucleotide
- Protein Clusters
- PubMed
- PubMed (RefSeq)
- PubMed (Weighted)
- Related Structures (List)
- Related Structures (Summary)
- Structure
- Taxonomy

- Recent activity Turn Off Clear
- catalase [Corynebacterium glutamicum ATCC 13032] Protein
  - pdb|4B7F|A (515 letters) BLAST
  - Corynebacterium glutamicum ATCC 13032 taxonomy
  - Related Sequences for Protein (Select 534286111) (12) Protein
  - Identical Proteins for Protein (Select 534286111) (12) Protein
- See more...

If you click on the accession number, it will bring you to the new protein page and highlight the protein sequence. Clicking on CDS (**CoDing Sequence**) or the GeneID to get a tool bar to display on the bottom of your screen. Click on FASTA.

You are here: NCBI > Proteins > Protein Database

GETTING STARTED	RESOURCES	POPULAR	FEATURED
NCBI Education	Chemicals & Bioassays	PubMed	Genetic Testing Registry
NCBI Help Manual	Data & Software	Bookshelf	PubMed Health
NCBI Handbook	DNA & RNA	PubMed Central	GenBank

NP\_599508 : 1 segment

1..516  
/locus\_tag="NCg10251"  
/gene\_synonym="Cg10255"  
/coded\_by="NC\_003450.3:274324..275874"  
/transl\_table= 11  
/db\_xref="GeneID: 1021318 "

Details ⌵ Display: **FASTA**



# Obtain DNA seq. that codes for protein

## Corynebacterium glutamicum ATCC 13032, complete genome

NCBI Reference Sequence: NC\_003450.3

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

```
>gi|58036263:274324-275874 Corynebacterium glutamicum ATCC 13032, complete genome
ATGTCTGAGAAGTCAGCAGCAGACCAGATCGTAGATCGCGGAATGCGTCCAAAGCTTCTGGAAACACTA
CCCCGCCAACGGAGCACCAGTTCCATCTGAGAACATCTCCGCAACCGCAGGCCACAGGGTCCAAACGT
TCTCAATGACATTCACCTCATTTGAAAAGCTCGCACACTTTAACC CGGAGAACGTTCCAGAGCGTATCCCT
CACGCAAAGGGCCACGGCGCTTTCGGTGAGCTGCACATCACCGAGGACGTATCCGAATACACCAAGGCAG
ACCTGTTCCAGCCTGGTAAGGTCACCCCGCTGGCTGTTTCGCTTCTACTGTTGCTGGTGAGCAGGGCTC
CCCAGATACCTGGCGCGACGTCACGGCTTCGCTCTTCGCTTCTACACCGAAGAGGGCAACTACGACATC
GTGGGTAACAACACCCCAACCTTCTTCCTTCGTGACGGCATGAAGTTC CCGGACTTCATCCACTCACAGA
AGCGTCTCAACAAGAAGCGTCTGCGCGATGCAGACATGCACTGGGATTTCTGGACCCGCGCACCTGAATC
TGCACACCAGGTGACCTACCTGATGGGTGACCGCGGTACCCCTAAGACCTCCCGCCACCAGGACGGCTTC
GGTCCCACACCTTCCAGTGGATTAACGCTGAAGGTAAGCCAGTTTGGGTTAAGTACCAC TTC AAGACCC
GCCAGGGCTGGGATTGCTTCACCGATGCAGAAGCAGCAAAGGTTGCAGGGCAGAACGCTGACTACCAGCG
CGAAGACCTCTACAACGCTATTGAAAACGGCGACTTCCCAATCTGGGACGTC AAGGTT CAGATCATGCCT
TTCGAGGATGCAGAGAACTACCGCTGGAACCCATTGACCTGACCAAGACCTGGTCCCAG AAGGATTACC
CACTGATCCCAGTCCGTTACTTCATCCTGAACCGCAACCCACGCAACTTCTTCGCTCAGATCGAGCAGCT
TGCCTGGATCCAGGCAACATCGTTCTCGCGCTCGGCTGTCCCCAGACCGCATGCTC CAGGCACGTATC
TTCGCATACGCTGACCAGCAGCGTTACCGCATCGGCGCTAACTACCGCGACCTGCCAGTGAACCGTCCAA
TCAACGAGGTCAACACCTACAGCCGCGAAGGTTCCATGCAGTACATCTTCGACGCTGAGGGCGAGCCTTC
CTACAGCCCTAACCGCTACGACAAGGGCGCAGGCTACCTGGATAACGGTACGGATTCCCTCCTCAACCAC
ACCTCCTACGGCCAGGCTGATGACATCTACGTC AACCAGACCCACACGGCACCGACCTGGTTCGTCGCTG
CTTACGCTCAAGCACCAGGATGATGACGACTTCATCCAGCCAGGCATCTATACCGCGAGGTCCTGGATGA
GGGCGAGAAGGAGCGATTGGCAGACAACATCTCCAACGCAATGCAGGGCATCTCTGAGGCAACCGAGCCA
CGCGTCTACGACTACTGGAACAACGTTGATGAGAACCTCGGCGCTCGCGTCAAGGAGCTTTACCTCCAGA
AGAAGGCTTAA
```

Selected region  
from: 274324 to: 275874  
[Update View](#)

Customize view

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

LinkOut to external resources

[REBASE enzyme CglI](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme M.CglI](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme CglORF3009P](#)  
[REBASE - The Restriction Enzy...]

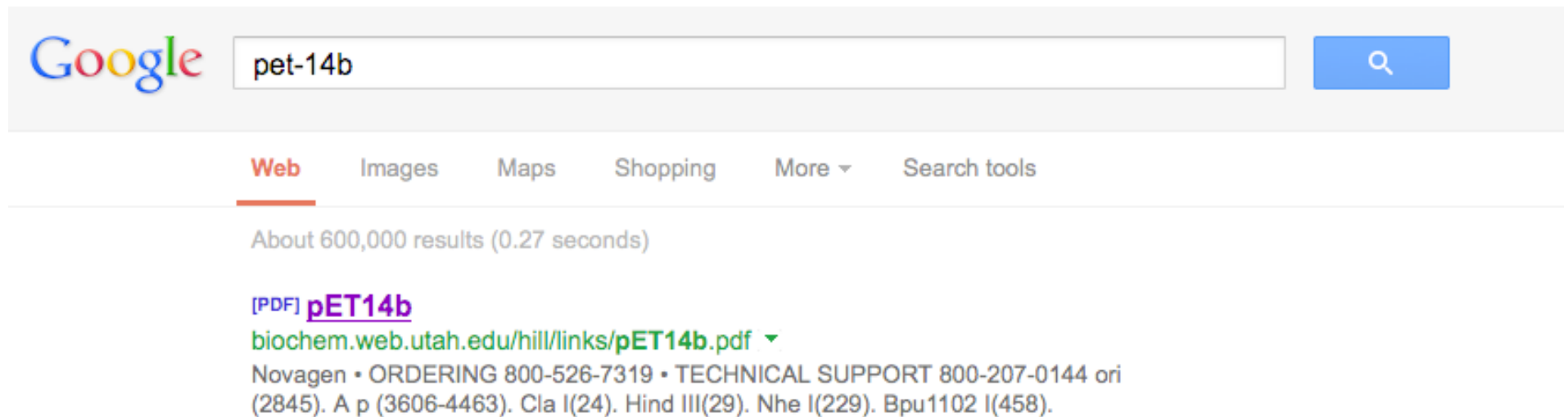
[REBASE enzyme CglORF706P](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme M.CglORF866P](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme CglORF1151P](#)

Here is the DNA sequence! Be sure that it begins with a start codon (ATG) and ends with a stop codon (TAA, TAG, or TGA)

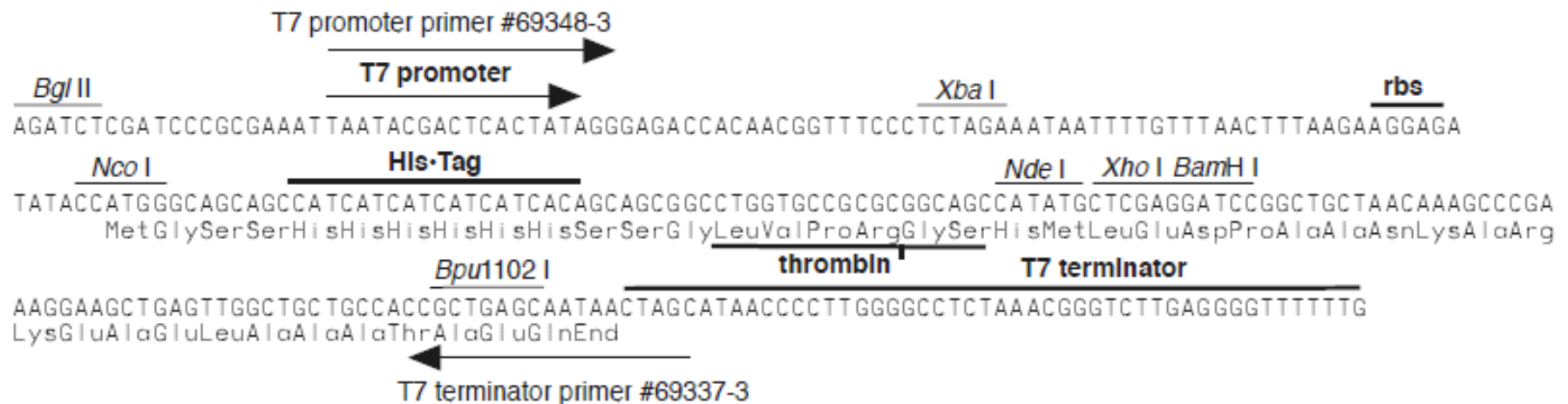
# Design primers to clone gene into vector



The image shows a Google search interface. The search bar contains the text "pet-14b". Below the search bar, the "Web" tab is selected. The search results show "About 600,000 results (0.27 seconds)". The first result is a PDF document titled "[PDF] pET14b" from the URL "biochem.web.utah.edu/hill/links/pET14b.pdf". Below the URL, there is a list of restriction enzymes: "Novagen • ORDERING 800-526-7319 • TECHNICAL SUPPORT 800-207-0144 ori (2845). A p (3606-4463). Cla I(24). Hind III(29). Nhe I(229). Bpu1102 I(458)."

Now that we have the DNA sequence, let's find a vector suitable for protein expression. pET-14b is a very simple vector that allows for a His-tagged protein.

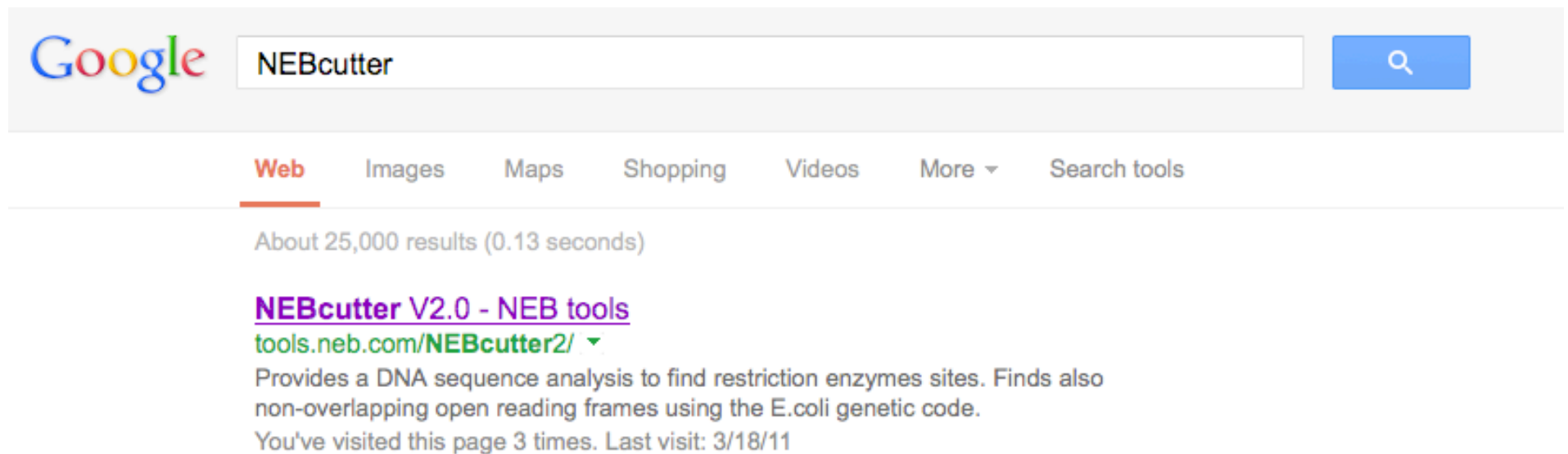
# Design primers to clone gene into vector



## pET-14b cloning/expression region

The link will open up a .pdf file, and above is a part of this file that shows the cloning/expression sequences of this vector. We want our protein to have a His-Tag, so we have to use two restriction sites after the His-Tag. This leaves *Nde*I, *Xho*I, and *Bam*HI.

# Design primers to clone gene into vector



The image shows a Google search interface. The search bar contains the text "NEBcutter". Below the search bar, there are navigation tabs for "Web", "Images", "Maps", "Shopping", "Videos", "More", and "Search tools". The "Web" tab is selected. Below the tabs, it says "About 25,000 results (0.13 seconds)". The first search result is for "NEBcutter V2.0 - NEB tools" with the URL "tools.neb.com/NEBcutter2/". The description of the tool states: "Provides a DNA sequence analysis to find restriction enzymes sites. Finds also non-overlapping open reading frames using the E.coli genetic code." It also notes: "You've visited this page 3 times. Last visit: 3/18/11".

We are going to use restriction enzymes to clone our gene into pET-14b, so we have to be sure that these restriction enzymes won't cut our gene.

# Design primers to clone gene into vector



## NEBcutter V2.0

[Program Guide](#)

[Help](#)

[Comments](#)

This tool will take a DNA sequence and find the large, non-overlapping open reading frames using the E.coli genetic code and the sites for all Type II and commercially available Type III restriction enzymes that cut the sequence just once. By default, only enzymes available from NEB are used, but other sets may be chosen. Just enter your sequence and "submit". Further options will appear with the output. **The maximum size of the input file is 1 MByte, and the maximum sequence length is 300 KBases.**

[What's new in V2.0](#)   [Citing NEBcutter](#)

The screenshot shows the NEBcutter V2.0 web interface. At the top, there are input fields for "Local sequence file" (with a "Browse..." button and "No file selected." text) and "GenBank number" (with a "[Browse GenBank]" link). Below these is a text area for "or paste in your DNA sequence: (plain or FASTA format)". The text area contains a DNA sequence with red dashed lines indicating restriction sites. To the right of the text area is a "Submit" button, which is circled in orange. Below the text area are radio buttons for "The sequence is:" (Linear and Circular) and a section for "Enzymes to use:" with radio buttons for "NEB enzymes", "All commercially available specificities", "All specificities", "All + defined oligonucleotide sequences", and "Only defined oligonucleotide sequences". There is also a "[define oligos]" link. At the bottom, there is a "Minimum ORF length to display:" field with the value "100" and "a.a." text. On the right side of the interface, there are "Standard sequences:" dropdown menus for "# Plasmid vectors" and "# Viral + phage", and buttons for "More options" and "Set colors".

NEBcutter will analyze your gene sequence to find restrictions sites in your gene. Paste your sequence in the submission box and hit 'Submit.'

# Design primers to clone gene into vector



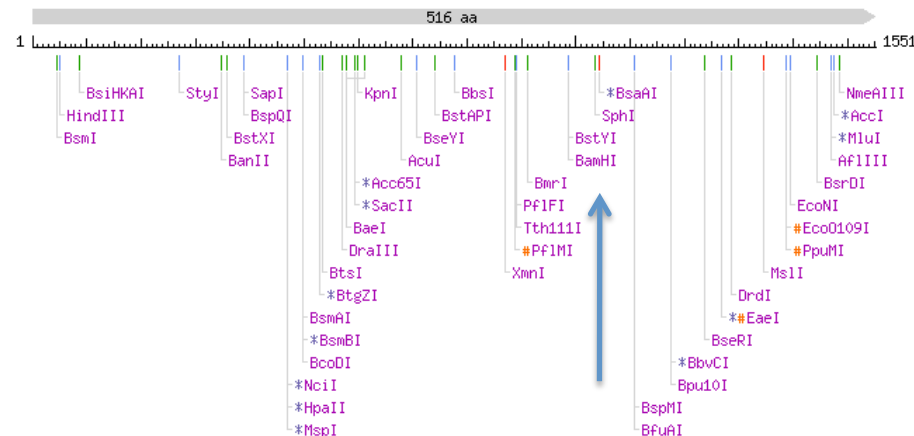
Linear Sequence: *unnamed sequence*

[Help](#) [Comments](#)

**Display:** - NEB single cutter restriction enzymes  
- Main non-overlapping, min. 100 aa ORFs

GC=57%, AT=43%

Cleavage code	Enzyme name code
⌂   blunt end cut	Available from NEB
⌂   5' extension	Has other supplier
⌂   3' extension	Not commercially available
⌂   cuts 1 strand	*: cleavage affected by CpG meth.
	#: cleavage affected by other meth.
	(enz.name): ambiguous site



- Main options
- New DNA
  - Custom digest
  - View sequence
  - ORF summary
  - Save project
  - Print

- Availability
- All commercial
  - All

- Display
- 2 cutters
  - 3 cutters

- Zoom
- Zoom in
  - More...

- List
- 0 cutters
  - 1 cutters
  - All sites
  - Save all sites
  - Flanking enzymes

NdeI, XhoI, or BamHI anywhere? Don't use it. If you look carefully you see BamHI will cut this gene. NEBcutter will also list the enzymes that will and won't cut the gene. Do you have two sites in your gene? Time for a new vector. Try any from the pGEX or pMAL series, or pET21a.

# Design primers to clone gene into vector

## Corynebacterium glutamicum ATCC 13032, complete genome

NCBI Reference Sequence: NC\_003450.3

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

```
>gi|58036263:274324-275874 Corynebacterium glutamicum ATCC 13032, complete genome
ATGTCGTGAGAAGTCAGCAGCAGACCAGATCGTAGATCGCGGAATGCGTCCAAAGCTTCTGGAAAACACTA
CCCCGCCAACCGGAGCACCAGTTCCATCTGAGAACATCTCCGCAACCGCAGGCCACAGGGTCCAAACGT
TCTCAATGACATTCACCTCATTTGAAAAGCTCGCACACTTTAACCGCGAGAACGTTCCAGAGCGTATCCCT
CACGCAAAGGGCCACGGCGCTTTCGGTGAGCTGCACATCACCGAGGACGTATCCGAATACACCAAGGCAG
ACCTGTTCCAGCCTGGTAAGGTCACCCCGCTGGCTGTTTCGCTTCTCTACTGTTGCTGGTGAGCAGGGCTC
CCCAGATACCTGGCGCGACGTCACGGCTTCGCTCTTCGCTTCTACACCGAAGAGGGCAACTACGACATC
GTGGGTAACAACACCCCAACCTTCTTCCTTCGTGACGGCATGAAGTTCCTGGACTTCATCCACTCACAGA
AGCGTCTCAACAAGAACGGTCTGCGCGATGCGAGACATGCACTGGGATTTCTGGACCCGCGCACCTGAATC
TGCACACCAGGTGACCTACCTGATGGGTGACCGCGGTACCCCTAAGACCTCCCGCCACCAGGACGGCTTC
GGCTCCACACCTTCCAGTGGATTAACGCTGAAGGTAAGCCAGTTTGGGTTAAGTACCAC TTC AAGACCC
GCCAGGGCTGGGATTGCTTCACCGATGCAGAAGCAGCAAAGGTTGCAGGGCAGAACGCTGACTACCAGCG
CGAAGACCTCTACAACGCTATTGAAAACGGCGACTTCCCAATCTGGGACGTC AAGGTT CAGATCATGCCT
TTCGAGGATGCAGAGAACTACCGCTGGAACCCATTGACCTGACCAAGACCTGGTCCCAGAAGGATTACC
CACTGATCCCAGTCGGTTACTTCATCCTGAACCGCAACCCACGCAACTTCTTCGCTCAGATCGAGCAGCT
TGCCTGATCCAGGCAACATCGTTCCCTGGCGTCGGCTGTCCCCAGACCGCATGCTCAGGCACGATC
TTCGCATACGCTGACCAGCAGCGTTACCGCATCGGGCGTAACCTACCGCGACCTGCCAGTGAACCGTCCAA
TCAACGAGGTCAACACCTACAGCCGCGAAGGTTCCATGCAGTACATCTTCGACGCTGAGGGCGAGCCTTC
CTACAGCCCTAACCGCTACGACAAGGGCGCAGGCTACCTGGATAACGGTACGGATTCCCTCCTCAACCAC
ACCTCCTACGGCCAGGCTGATGACATCTACGTC AACCAGACCCACACGGCACCGACCTGGTTCGTCCTG
CTTACGCTCAAGCACCAGGATGATGACGACTTCATCCAGCCAGGCATCCTATAACCGGAGGTCTCGGATGA
GGGCGAGAAGGAGCGATTGGCAGACAACATCTCCAACGCAATGCAGGGCATCTCTGAGGCAACCGAGCCA
CGCGTCTACGACTACTGGAACAACGTTGATGAGAACCTCGGGCTCGCGTCAAGGAGCTTTACCTCCAGA
AGAAGGCTTAA
```

Selected region  
from: 274324 to: 275874  
[Update View](#)

Customize view

Analyze this sequence

[Run BLAST](#)

[Pick Primers](#)

[Highlight Sequence Features](#)

LinkOut to external resources

[REBASE enzyme CglI](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme M.CglI](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme CglORF3009P](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme CglORF706P](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme M.CglORF866P](#)  
[REBASE - The Restriction Enzy...]

[REBASE enzyme CglORF1151P](#)

My goal now is to design primers that are specific for this gene sequence, and to contain the NdeI and XhoI restriction sites so the gene can be cloned into the pET-14b vector.

# Design primers to clone gene into vector

```
ATGTCTGAGAAGTCAGCAGCAGACCAGATCGTAGATCGCGGAATGCGTCCAAAGCTTTCTGGAAACACTA
CCCGCCACAACGGAGCACCAGTTCCATCTGAGAACATCTCCGCAACCGCAGGCCACAGGGTCCAAACGT
TCTCAATGACATTCACCTCATTGAAAAGCTCGCACACTTTAACCGGAGAACGTTCCAGAGCGTATCCCT
CACGCAAAGGGCCACGGCGCTTTCCGGTGAGCTGCACATCACCGAGGACGTATCCGAATACACCAAGGCAG
ACCTGTTCCAGCCTGGTAAGGTCACCCCGCTGGCTGTTCCGCTTCTCTACTGTTGCTGGTGAGCAGGGCTC
CCCAGATACCTGGCGCGACGTCCACGGCTTCGCTCTTCGCTTCTACACCGAAGAGGGCAACTACGACATC
GTGGGTAACAACACCCCAACCTTCTTCCTTCGTGACGGCATGAAGTCCCGGACTTCATCCACTCACAGA
AGCGTCTCAACAAGAACGGTCTGCGCGATGCAGACATGCAGTGGGATTTCTGGACCCGCGCACCTGAATC
TGCACACCAGGTGACCTACCTGATGGGTGACCGCGGTACCCCTAAGACCTCCCGCCACCAGGACGGCTTC
GGCTCCACACCTTCCAGTGGATTAACGCTGAAGGTAAGCCAGTTTGGGTTAAGTACCACTTCAAGACCC
GCCAGGGCTGGGATTGCTTCACCGATGCAGAAGCAGCAAAGGTTGCAGGCGAGAACGCTGACTACCAGCG
CGAAGACCTCTACAACGCTATTGAAAACGGCGACTTCCCAATCTGGGACGTCAAGGTTTCAGATCATGCCT
TTCGAGGATGCAGAGAACTACCGCTGGAACCCATTGACCTGACCAAGACCTGGTCCCAGAAGGATTACC
CACTGATCCCAGTCGGTTACTTCATCCTGAACCGCAACCCACGCAACTTCTTCGCTCAGATCGAGCAGCT
TGCCTGGATCCAGGCAACATCGTTCCTGGCGTCCGGCCTGTCCCAGACCCGCATGCTCCAGGCACGTATC
TTCGCATACGCTGACCAGCAGCGTTACCGCATCGGCGCTAACTACCGCGACCTGCCAGTGAACCGTCCAA
TCAACGAGGTCAACACCTACAGCCGCGAAGGTTCCATGCAGTACATCTTCGACGCTGAGGGCGAGCCTTC
CTACAGCCCTAACCGCTACGACAAGGGCGCAGGCTACCTGGATAACGGTACGGATTCTCCTCCAACCAC
ACCTCCTACGGCCAGGCTGATGACATCTACGTCAACCCAGACCCACACGGCACCGACCTGGTTTCGTGCTG
CTTACGTCAAGCACCAGGATGATGACGACTTCATCCAGCCAGGCATCCTATAACCGGAGGTCTTGATGA
GGGCGAGAAGGAGCGATTGGCAGACAACATCTCCAACGCAATGCAGGGCATCTCTGAGGCAACCGAGCCA
CGCGTCTACGACTACTGGAACAACGTTGATGAGAACCCTCGGCGCTCGCGTCAAGGAGCTTTACCTCCAGA
AGAAGGCTTAA
```

Primer design is an art (and a bit voodoo). For an initial attempt, I design primers to have 4-6 codons (15 nt) of specificity on either end of the gene. Then, in order of importance, the primers should have nearly identical melting temperatures, end on G/C, and have 50% G/C content. You may break these “rules.”



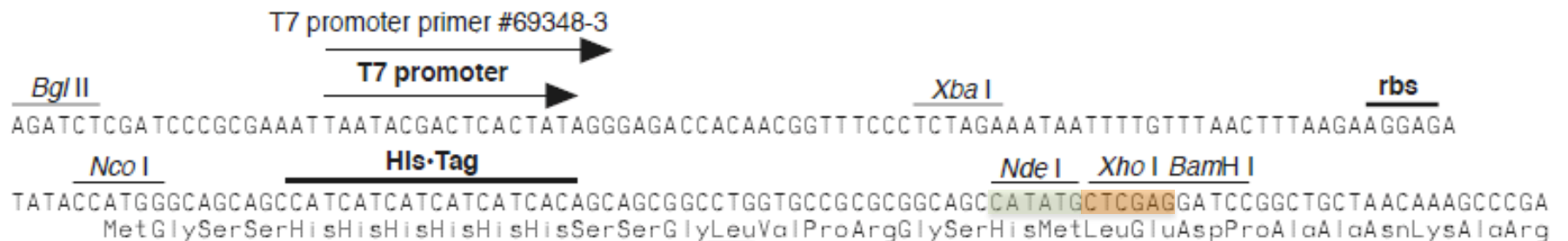
# Design primers to clone gene into vector

5' ATGTCTGAGAAGTCAGC.....GENE.....CCAGAAGAAGGCTTAA 3'  
 3' TACAGACTCTTCAGTCG.....GENE.....GGTCTTCTTCCGAATT 5'

The gene sequence is above, and it is only showing the ~15 nt on the ends. For correct directionality, we want to clone the gene so the NdeI site is at the 5' end and the XhoI site is at the 3' end.

5' CATATGTCTGAGAAGTCAGC.....GENE.....CCAGAAGAAGGCTTAACTCGAG 3'  
 3' GTATACAGACTCTTCAGTCG.....GENE.....GGTCTTCTTCCGAATTGAGCTC 5'

The sequence above is what we want our product from PCR to look like. Notice that the start codon is included in the NdeI recognition sequence. There is, however, one last thing we want to consider.



# Design primers to clone gene into vector

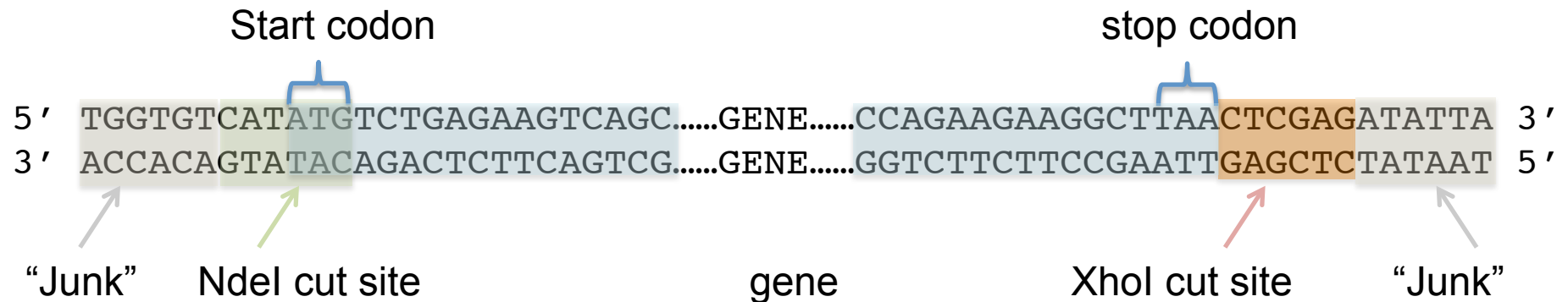
The screenshot shows the New England BioLabs website header with the logo and navigation menu. The main content area is titled 'Cleavage Close to the End of DNA Fragments' and includes a breadcrumb trail: 'Home > Tools & Resources > Cleavage Close to the End of DNA Fragments'. Below the title are icons for 'EMAIL', 'ADD TO MY NEB', 'PRINT', and 'PDF'.

**Note:** As a general rule and for enzymes not listed below, 6 base pairs should be added on either side of the recognition site to cleave efficiently. The extra bases should be chosen so that palindromes and primer dimers are not formed. In most cases there is no requirement for specific bases.

```
5' CATATGTCTGAGAAGTCAGC.....GENE.....CCAGAAGAAGGCTTAACTCGAG 3'
3' GTATACAGACTCTTCAGTCG.....GENE.....GGTCTTCTTCCGAATTGAGCTC 5'
```

DNA digestion works more efficiently if the restriction enzyme has more DNA bases on either side of the restriction site. These are meaningless, “junk,” sequences only to provide space for the enzyme to bind.

# Design primers to clone gene into vector



So this is what we'd actually like the PCR product to look like.

PCR amplifies DNA 5' -> 3', and we need two primers to synthesize the coding strand (top) and the complementary strand (bottom), and we commonly provide primers 5' -> 3'

Our two primers, with a melting temperature of ~ 62 °C (<http://www6.appliedbiosystems.com/support/techtools/calc/>), are:

FWD: 5' TGGTGTTCATATGTCTGAGAAGTCAGC 3'

REV: 5' TAATATCTCGAGTTAAGCCTTCTTCTTGG 3'

# Make a multiple sequence alignment

Sequences producing significant alignments:

Select: [All](#) [None](#) Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
<input type="checkbox"/>	<a href="#">Chain A, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7FIB Chain B, Structure Of A Liganded Bacterial Catalase &gt;pdb 4B7FIC Chain C, Structure Of A Liganded Bacteri</a>	1072	1072	100%	0.0	100%	<a href="#">4B7F_A</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_224555.1  catalase [Corynebacterium glutamicum ATCC 13032] &gt;ref YP_007559511.1  catalase [Corynebacteriu</a>	1071	1071	100%	0.0	99%	<a href="#">NP_599508.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein cgR_0332 [Corynebacterium glutamicum RI] &gt;ref YP_008065043.1  catalase [Corynebacterium glutamicum SCgG1] &gt;ref YP_008068066.1  catalase [Coryn</a>	1071	1071	100%	0.0	99%	<a href="#">YP_001137198.1</a>
<input type="checkbox"/>	<a href="#">Catalase [Corynebacterium glutamicum ATCC 13032]</a>	1046	1046	97%	0.0	99%	<a href="#">BAB97648.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium callunae DSM 20147] &gt;ref WP_015650135.1  catalase [Corynebacterium callunae] &gt;gb AGG65680.1  catalase [Corynebacterium callunae DSM 201</a>	1021	1021	100%	0.0	94%	<a href="#">YP_007529583.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium efficiens YS-314] &gt;ref WP_006768464.1  catalase [Corynebacterium efficiens] &gt;dbj BAC17034.1  putative catalase [Corynebacterium efficiens YS-3</a>	964	964	100%	0.0	88%	<a href="#">NP_736834.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium pilosum]</a>	883	883	100%	0.0	81%	<a href="#">WP_018581253.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium casei] &gt;emb CCE53994.1  catalase [Corynebacterium casei UCMA 3821]</a>	876	876	99%	0.0	81%	<a href="#">WP_006821561.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium ammoniagenes] &gt;gb EFG82166.1  catalase [Corynebacterium ammoniagenes DSM 20306]</a>	875	875	99%	0.0	81%	<a href="#">WP_003846068.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium halotolerans YIM 70093 = DSM 44683] &gt;ref WP_015399757.1  catalase [Corynebacterium halotolerans] &gt;gb AGF71333.1  catalase [Corynebacteriu</a>	872	872	99%	0.0	81%	<a href="#">YP_007463695.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium lubricantis]</a>	866	866	100%	0.0	80%	<a href="#">WP_018297951.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium ulcerans 0102] &gt;ref WP_014835811.1  catalase [Corynebacterium ulcerans] &gt;dbj BAM26465.1  catalase [Corynebacterium ulcerans 0102]</a>	851	851	99%	0.0	79%	<a href="#">YP_006493700.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium ulcerans BR-AD22] &gt;ref YP_005709860.1  catalase [Corynebacterium ulcerans 809] &gt;ref WP_013910628.1  catalase [Corynebacterium ulcerans] &gt;</a>	849	849	99%	0.0	79%	<a href="#">YP_004628860.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium pseudotuberculosis Cp162] &gt;ref WP_014799916.1  catalase [Corynebacterium pseudotuberculosis] &gt;gb AFM06551.1  Catalase [Corynebacterium p</a>	842	842	99%	0.0	78%	<a href="#">YP_006436344.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium pseudotuberculosis FRC41] &gt;ref YP_005122300.1  catalase [Corynebacterium pseudotuberculosis 3/99-5] &gt;ref YP_005374218.1  katA gene produc</a>	842	842	99%	0.0	78%	<a href="#">YP_003782575.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium pseudotuberculosis 31] &gt;ref WP_014654992.1  catalase [Corynebacterium pseudotuberculosis] &gt;gb AFH90022.1  Catalase [Corynebacterium pseu</a>	840	840	99%	0.0	78%	<a href="#">YP_006212712.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium diphtheriae INCA 402] &gt;ref WP_014302799.1  catalase [Corynebacterium diphtheriae] &gt;gb AEX45515.1  catalase [Corynebacterium diphtheriae INC</a>	839	839	99%	0.0	77%	<a href="#">YP_005126717.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium diphtheriae 31A] &gt;ref WP_014316262.1  catalase [Corynebacterium diphtheriae] &gt;gb AEX40965.1  catalase [Corynebacterium diphtheriae 31A]</a>	838	838	99%	0.0	77%	<a href="#">YP_005156992.1</a>
<input type="checkbox"/>	<a href="#">catalase [Corynebacterium diphtheriae PW8] &gt;ref WP_014309894.1  catalase [Corynebacterium diphtheriae] &gt;gb AFX68957.1  catalase [Corynebacterium diphtheriae PW8]</a>	838	838	99%	0.0	77%	<a href="#">YP_005141862.1</a>

A multiple sequence alignment will allow us to determine conserved residues that are important for this enzyme to function. Go back to your BLAST results and pick proteins that are related to your own. Let's add this protein to our multiple sequence alignment.

# Make a multiple sequence alignment

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

Display Settings: GenPept Send to: Change region shown Customize view

## catalase [Corynebacterium lubricantis]

NCBI Reference Sequence: WP\_018297951.1

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

Go to:

LOCUS	WP_018297951	517 aa	linear	BCT 28-JUN-2013
DEFINITION	catalase [Corynebacterium lubricantis].			
ACCESSION	WP_018297951			
VERSION	WP_018297951.1 GI:517109133			
KEYWORDS	RefSeq.			
SOURCE	Corynebacterium lubricantis			
ORGANISM	<a href="#">Corynebacterium lubricantis</a> Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacterineae; Corynebacteriaceae; Corynebacterium.			
COMMENT	REFSEQ: This record represents a single, non-redundant, protein sequence which may be annotated on many different RefSeq genomes from the same, or different, species. COMPLETENESS: full length.			
FEATURES	Location/Qualifiers			
source	1..517 /organism="Corynebacterium lubricantis" /db_xref="taxon:541095"			
Protein	1..517			

Analyze this sequence

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

Related information

- BioProject
- CDD Search Results
- Conserved Domains (Concise)
- Conserved Domains (Full)
- Domain Relatives
- Genomic records
- Nucleotide

Click on the accession number and get the FASTA format for the sequence of this and the other proteins.

# Make a multiple sequence alignment

NCBI Resources How To Sign in to NCBI

Protein Protein Search Advanced Help

Display Settings: FASTA Send to: Change region shown

### catalase [Corynebacterium lubricantis]

NCBI Reference Sequence: WP\_018297951.1

[GenPept](#) [Graphics](#)

```
>gi|517109133|ref|WP_018297951.1| catalase [Corynebacterium lubricantis]
MADKSVEDIVARGERPAGNGQTTRHNGTPVPSSENI SVTAGPQGPVVLNDIHLIEKLAHFNRENVPERIPH
AKGHGAFGELHITEDVVSQYTKAKLFQKDAVTPLAIRFSTVAGEAGSPDTWRDVHGFALRFYTEDGNYDIV
GNNTPTFFFLRDGIKFADF IHSQKRNPATGLRSAEMQWDFWTRTPESAHQVTYLMGDRGTPKTSRHQDGF
SHTFQWINEEGKPVVVKYHFKTRQGWETFTDEEA AVVAGQNADYQREDLYNSIANGDYPIWDVKVQIMPV
EEAENYRFNPFDLTKTWSQKDYPLIDVG YFVLRNPNKFNHQAIEQLALDP SNLVPVGLSPDRMLQARVF
AYADQQR YRIGPNYRDIPVNRPI NEVNTY SERGSMAYFFN ESEPNYTPNSYSK GAGFLDNGEDSSSNHT
EYGQ GADLYVNPEPHGSDIGRYAYVKHEEDDDFGQACTLYR DVFDDGCEKERLVHNI TNAMNGITNKDIEE
RVYQYWTNV DENLGQKVRESLAKKRG A
```

**Analyze this sequence**

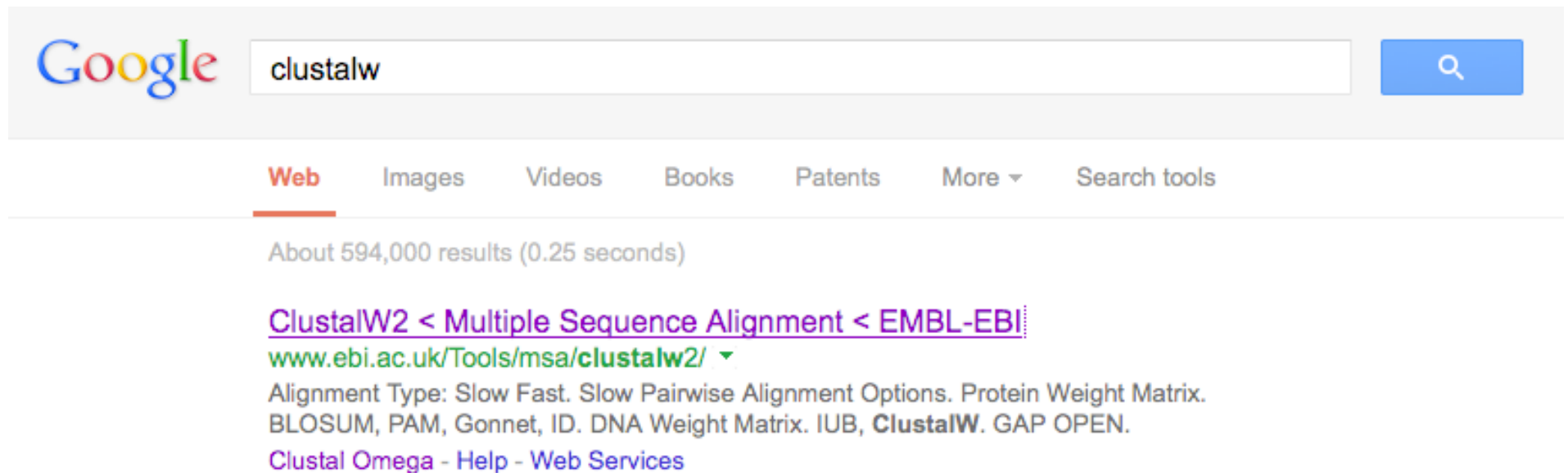
- Run BLAST
- Identify Conserved Domains
- Highlight Sequence Features
- Find in this Sequence

**Related information**

- BioProject
- CDD Search Results
- Conserved Domains (Concise)
- Conserved Domains (Full)
- Domain Relatives

Copy the whole sequence, including the title line. Remember that everything that follows the carrot > IS A TITLE  
After a return is the sequence.

# Make a multiple sequence alignment



The image shows a Google search interface. The search bar contains the text "clustalw". Below the search bar, the "Web" tab is selected. The search results show "About 594,000 results (0.25 seconds)". The first result is a link to "ClustalW2 < Multiple Sequence Alignment < EMBL-EBI" with the URL "www.ebi.ac.uk/Tools/msa/clustalw2/". Below the link, there is a description: "Alignment Type: Slow Fast. Slow Pairwise Alignment Options. Protein Weight Matrix. BLOSUM, PAM, Gonnet, ID. DNA Weight Matrix. IUB, ClustalW. GAP OPEN." and a link to "Clustal Omega - Help - Web Services".

Google clustalw

Web Images Videos Books Patents More Search tools

About 594,000 results (0.25 seconds)

[ClustalW2 < Multiple Sequence Alignment < EMBL-EBI](http://www.ebi.ac.uk/Tools/msa/clustalw2/)

[www.ebi.ac.uk/Tools/msa/clustalw2/](http://www.ebi.ac.uk/Tools/msa/clustalw2/)

Alignment Type: Slow Fast. Slow Pairwise Alignment Options. Protein Weight Matrix. BLOSUM, PAM, Gonnet, ID. DNA Weight Matrix. IUB, ClustalW. GAP OPEN.

[Clustal Omega - Help - Web Services](#)

Use any program you like to make the multiple sequence alignment.  
The 'Clustal' series are popular.

# Make a multiple sequence alignment

Note: **ClustalW2 is no longer being maintained.** Please consider using the new version instead: [Clustal Omega](#)

## STEP 1 - Enter your input sequences

Enter or paste a set of Protein sequences in any supported format:

```
>gil534286111|pdbj4B7FIA Chain A, Structure Of A Liganded Bacterial Catalase
SEKSAADQIVDRGMRPKLSGNTTRHNGAPVPSENISATAGPOGPNVLNDIHLIEKLAHFNRENVPERIPH
AKGHGAFGELHITEDVSEYTKADLFQPGKVTPLAVRFSTVAGEQGSPTWRDVGAFALRFYTEEGNYDIV
GNNTPTFFLRDGMKFPDFIHSQKRLNKNGLRDADMQWDFWTRAPESAHQVTYLMGDRGTPKTSRHQDGFG
SHTFQWINAEGKPVVVKYHFKTRQGWDCFTDAEAAKVAGENADYQREDLYNAIENGDFPIWDVKVQIMPF
EDAENYRWNPFDLTKTWSQKDYPLIPVGYFILNRNPRNFFAQIEQIALDPGNIVPGVGLSPDRMLQARIF
AYADQQRVIRIGANYRDLPVNRPINEVNTYSREGSMQYIFDAEGEPSYSPNRVYDKGAGYLDNGTDSSSNHT
SVCGADDDVAIPDDUCTSLVBAVWZUQDDDDDEIQDCILYBEVLDECEVEELADNLSNAMOCISEATEFB
```

Or, upload a file:  No file selected.

## STEP 2 - Set your Pairwise Alignment Options

Alignment Type:  Slow  Fast

*The default settings will fulfill the needs of most users and, for that reason, are not visible.*

*(Click here, if you want to view or change the default settings.)*

## STEP 3 - Set your Multiple Sequence Alignment Options

*The default settings will fulfill the needs of most users and, for that reason, are not visible.*

*(Click here, if you want to view or change the default settings.)*

## STEP 4 - Submit your job

Be notified by email *(Tick this box if you want to be notified by email when the results are available)*

Paste your sequences, one after another, and be sure to keep the >title lines  
This way the program knows when a sequence begins and ends.

Hit 'Submit' and wait for the program to finish.



# Make a multiple sequence alignment

```
gi|517109133|ref|WP_018297951.  -----MADKS-VEDIVARGERPAGNGQTRHNGTPVPSENISVTAGPQ 42
gi|517408857|ref|WP_018581253.  -----MADKSAENVVSRGERAAGDGTTRLNGAPVPSENISVTAGPQ 43
gi|534286111|pdb|4B7F|A         -----SEKSAADQIVDRGMRPKLSGNTRHNGAPVPSENISATAGPQ 42
gi|516831493|ref|WP_018121082.  -----MTKN-IDDKLDQGGREDAPGTTTRQGGQPIASENISITAGPQ 41
gi|497955492|ref|WP_010269648.  MTDNTPAGNPAGSDDVAMRGVCP-VTGHSSNINGAPV RTEEHSVTVGAQ 49
                                ::  :*      *  ::. .* *; :*: * *.*

gi|517109133|ref|WP_018297951.  GPVVLNDIHLIEKLAHFNRENVPERIPHAKGHGAFGELHITEDVVSQYTKA 92
gi|517408857|ref|WP_018581253.  GPNVLDDIHMIEKLAHFNRENVPERIPHAKGHGAFGELHITEDVVSQYTKA 93
gi|534286111|pdb|4B7F|A         GPNVLNDIHLIEKLAHFNRENVPERIPHAKGHGAFGELHITEDVSEYTKA 92
gi|516831493|ref|WP_018121082.  GPNVLNDLQLIEKLSFNREVRPERNPHAKGHGAFGEFHVTEDEVSAYTKA 91
gi|497955492|ref|WP_010269648.  GPIALNDVHLIEKHAHFNRERIPERNVHAKGSGAFGELTVTEDEVSKYTKA 99
** .*::::***  *****:***  *****  *****: ***** **

gi|517109133|ref|WP_018297951.  KLFQKDAVTPLAIRFSTVAGEAGSPDTRDVRDVGHFALRFYTEDGNYDIVGN 142
gi|517408857|ref|WP_018581253.  KLFQKGTVTPMAGRFSTVAGEAGSPDTRDVRDVGHFALRFYTEDGNYDIVGN 143
gi|534286111|pdb|4B7F|A         DLFQPGKVTPLAVRFSTVAGEQGSPTDTRDVRDVGHFALRFYTEEGNYDIVGN 142
gi|516831493|ref|WP_018121082.  DLFQPNKVTPMGIRFSTVAGEQGSPTDTRDVRDVGHFALRFWTQEGNFDIVGN 141
gi|497955492|ref|WP_010269648.  DLFQPCRVTPLARFSTVAGEQGYPTDTRDVRGFSLKFYEQEGNYDLVGN 149
.*** . ***:  ***** * *** ***:**:*:*:*:*:*:*:*

gi|517109133|ref|WP_018297951.  NTPTFFLRDGIKFADF IHSQKRNPATGLRSAEMQWDFWTRTPESAHQVTY 192
gi|517408857|ref|WP_018581253.  NTPTFFLRDAIKFPDF IHSQKRNPASGLRDEMOWDFWTRTPESAHQVTY 193
gi|534286111|pdb|4B7F|A         NTPTFFLRDGKMFPDF IHSQKRLNKNGLRDADMOWDFWTRAPESAHQVTY 192
gi|516831493|ref|WP_018121082.  NTPTFFLRDGIKFPDF IHSQKRTPGASGLRDADMOWDFWTRTPESAHQVTY 191
gi|497955492|ref|WP_010269648.  NTPVFLRDGIKFPDF IRSQKRLHGPGLQSDADMOWDFWTRSPESAHQVTY 199
***.*****:***.***:***  **:. :*****:*****

gi|517109133|ref|WP_018297951.  LMGDRGTPKTSRHQDGFSGSHTFQWINEEGKPVVVKYHFKTRQGWCFTDE 242
gi|517408857|ref|WP_018581253.  LMGDRGTPKTRRNDQDGFSGSHTFQWINEEGTPVWVKYHFKTRQGWDCFTDE 243
gi|534286111|pdb|4B7F|A         LMGDRGTPKTSRHQDGFSGSHTFQWINAEGKPVVVKYHFKTRQGWDCFTDA 242
gi|516831493|ref|WP_018121082.  LMGDRGTPKTSRHQDGFSGSHTFQWVNDKGEAFVVKYHFKTQQGWECFTDE 241
gi|497955492|ref|WP_010269648.  LMGDRGIPDTRHMDGFSSTHYQWINADNERFVVKYHFKTRQGWKYFTDE 249
***** * . *: ***.***:*** .. .*****:***. ***

gi|517109133|ref|WP_018297951.  EAAVVAG-QNADYQREDLYNSIANGDYPIDVVKVQIMPVEEAENYRNPFP 291
gi|517408857|ref|WP_018581253.  EAEEMAG-KNADYHRQDLYEAIERGDYPIWVVKVQIMPFEAENYRWNPFP 292
gi|534286111|pdb|4B7F|A         EAAKVG-ENADYQREDLYNAIENGDFPIWVVKVQIMPFEAENYRWNPFP 291
gi|516831493|ref|WP_018121082.  EAAEMAG-QNADYHREDLFKAIENGDPYRWDVYVQIMPFEAENYKFNFP 290
gi|497955492|ref|WP_010269648.  EASQVLASQDQDHSRKDLWEAIEAGDYPTWVVKVQIMPLDEAEGYRWNPFP 299
** : . : *: *:*:*:* *:* * * * * * * * * * * * * * * *

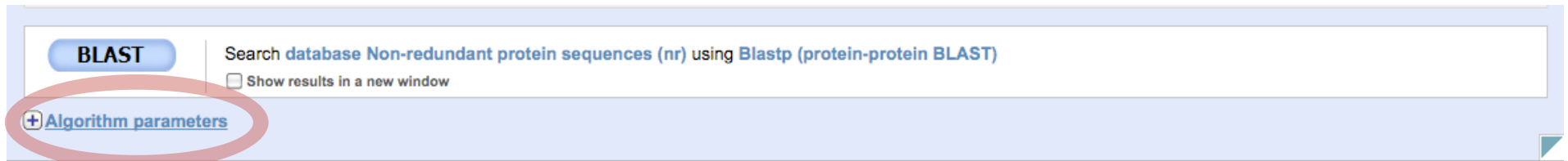
gi|517109133|ref|WP_018297951.  DLTKTWSQKDYPLIDVGYFVLRNRPKNFHAQIEQLALDPSNLVPGVGLSP 341
gi|517408857|ref|WP_018581253.  DLTKTWSQKDYPLHDVGYFVLRNRPKNYHAQIEQMALDPSNLVPGVGLSP 342
gi|534286111|pdb|4B7F|A         DLTKTWSQKDYPLIPVGYFILNRPNPNFFAQIEQIALDPGNIIPVPGVGLSP 341
gi|516831493|ref|WP_018121082.  DLTKVWYTEDYPLQKVGYPFLNRPNPNFFAQIEQIALDPSNIVPGTGLSP 340
gi|497955492|ref|WP_010269648.  DLTKTWSQKDYPLIPVGHFTLRNRPENFFAQIEQAAPSPNIVPGIGFSP 349
***** * :***** **:* ***** * .***** *: *.*:*** *:*
```

This is an example of a bad sequence alignment.

What's wrong?

- 1) These sequences are far too similar. You can see that about 80% of residues here are identical. We need more diversity to discover the truly important residues.
- 2) The titles are meaningless. When you copy/paste, the title begins with the accession number, so you'll need to alter the title.

# Make a multiple sequence alignment

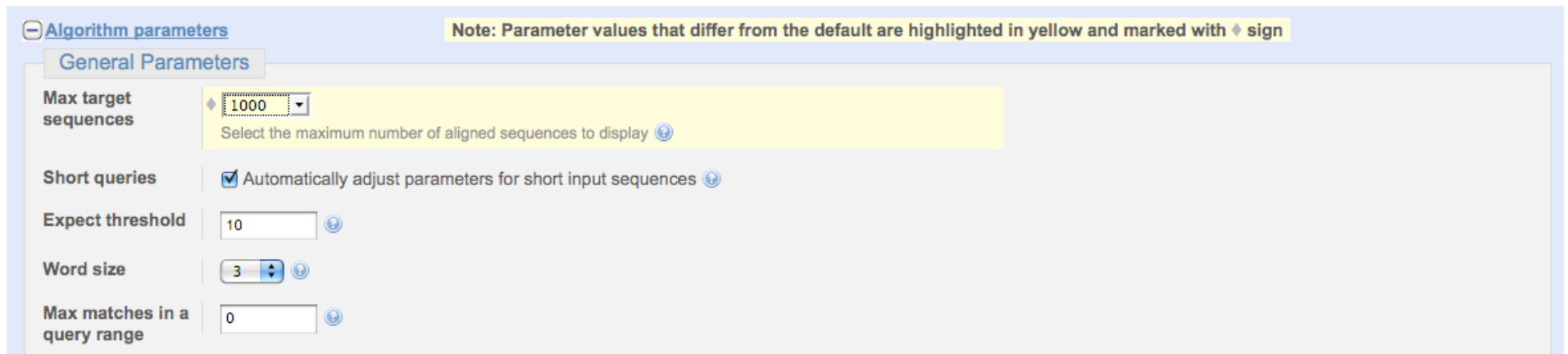


BLAST Search database **Non-redundant protein sequences (nr)** using **Blastp (protein-protein BLAST)**

Show results in a new window

[+ Algorithm parameters](#)

There wasn't enough diversity in the results given the default settings. We'll have to change them. Go into 'Algorithm parameters.'



[Algorithm parameters](#) Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

**General Parameters**

**Max target sequences** ♦ 1000  
Select the maximum number of aligned sequences to display

**Short queries**  Automatically adjust parameters for short input sequences

**Expect threshold** 10

**Word size** 3

**Max matches in a query range** 0

We'll have to get more results, so first try increase the number of result sequences given. BLAST it again.

# Make a multiple sequence alignment

<input type="checkbox"/>	<a href="#">catalase [Yersinia pseudotuberculosis 9110] &gt;ref YP_012749991.1 </a>	<a href="#">catalase [Yersinia pseudotuberculosis] &gt;gb AF017492.1 </a>	<a href="#">Catalase [Yer...</a>	700	700	97%	1e-102	99%	<a href="#"> L_00272997.1</a>
<input type="checkbox"/>	<a href="#">catalase [Pseudomonas syringae]</a>			482	482	93%	1e-162	51%	<a href="#">WP_010438211.1</a>
<input type="checkbox"/>	<a href="#">catalase [Pseudomonas fluorescens] &gt;gb F0158295.1 </a>	<a href="#">catalase KatA [Pseudomonas fluorescens O8r1-96]</a>		482	482	93%	1e-162	51%	<a href="#">WP_003206086.1</a>
<input type="checkbox"/>	<a href="#">catalase [Koseovarius sp. IM1033] &gt;gi EU032765.1 </a>	<a href="#">catalase hydroperoxidase npII(III) protein [Koseovarius sp. IM1033]</a>		453	453	94%	3e-131	40%	<a href="#">WP_006279931.1</a>
<input type="checkbox"/>	<a href="#">catalase [Daphnia magna]</a>			454	454	97%	3e-151	44%	<a href="#">ACU81116.1</a>
<input type="checkbox"/>	<a href="#">catalase [Bordetella pertussis Tohama II] &gt;ref NP_886532.1 </a>	<a href="#">catalase [Bordetella parapertussis 12822] &gt;ref YP_005591739.1 </a>	<a href="#">c...</a>	453	453	96%	3e-151	46%	<a href="#">NP_882347.1</a>

Given 1000 results, there still isn't enough diversity! Remember that my protein was from *Corynebacterium glutamicum*, a Gram positive bacterium, so from the 1000 results, I've included *Pseudomonas syringae*, a Gram negative bacterium, and *Daphnia magna*, a water flea.

I have to go back and change more settings.

# Make a multiple sequence alignment

The screenshot shows the NCBI BLAST search interface. The 'Choose Search Set' section is highlighted with a red oval. It includes a dropdown menu for 'Database' set to 'Non-redundant protein sequences (nr)', a text input for 'Organism' set to 'bacteria (taxid:2)', and a checked 'Exclude' checkbox. Below this are checkboxes for 'Models (XM/XP)' and 'Uncultured/environmental sample sequences', and an 'Entrez Query' field. The 'Program Selection' section shows 'blastp (protein-protein BLAST)' selected. A 'BLAST' button is present, along with a 'Show results in a new window' checkbox. The 'Algorithm parameters' section is expanded to 'General Parameters', showing 'Max target sequences' set to 1000 and 'Short queries' checked. A yellow highlight is under the 'Max target sequences' field.

**Choose Search Set**

Database: Non-redundant protein sequences (nr)

Organism Optional: bacteria (taxid:2)  Exclude +

Exclude Optional:  Models (XM/XP)  Uncultured/environmental sample sequences

Entrez Query Optional:

**Program Selection**

Algorithm:  blastp (protein-protein BLAST)  
 PSI-BLAST (Position-Specific Iterated BLAST)  
 PHI-BLAST (Pattern Hit Initiated BLAST)  
 DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)

**BLAST** Search database Non-redundant protein sequences (nr) using Blastp (protein-protein BLAST)  
 Show results in a new window

**Algorithm parameters** Note: Parameter values that differ from the default are highlighted in yellow and marked with ♦ sign

**General Parameters**

Max target sequences: ♦ 1000  
Select the maximum number of aligned sequences to display

Short queries:  Automatically adjust parameters for short input sequences

I'm still asking for 1000 sequences, but I want to exclude all bacteria from my results in order to get greater diversity. You can also force the program to only give results from whatever taxon you desire.

# Make a multiple sequence alignment

<input type="checkbox"/>	<a href="#">catalase [Danio rerio] &gt;emb CAB64949.1  catalase [Danio rerio]</a>	468	468	93%	8e-157	47%	<a href="#">Q9PT92.1</a>
<input type="checkbox"/>	<a href="#">catalase [Danio rerio] &gt;gb AAH51626.1  Catalase [Danio rerio]</a>	467	467	93%	1e-156	47%	<a href="#">NP_570987.1</a>
<input type="checkbox"/>	<a href="#">catalase [Homo sapiens]</a>	429	429	93%	1e-141	44%	<a href="#">AAK29181.1</a>
<input type="checkbox"/>	<a href="#">catalase isoform 1 [Thamnoophis elegans]</a>	429	429	93%	1e-141	45%	<a href="#">AFS63885.1</a>
<input type="checkbox"/>	<a href="#">catalase (EC 1.11.1.6) catA - rice &gt;dbj BAA06232.1  catalase [Oryza sativa]</a>	360	360	94%	3e-115	41%	<a href="#">CSRZ</a>
<input type="checkbox"/>	<a href="#">catalase 1b [Lotus japonicus] &gt;qbl AAR84579.1  catalase 1b [Lotus japonicus]</a>	358	358	75%	3e-115	47%	<a href="#">AAR84577.1</a>
<input type="checkbox"/>	<a href="#">CG9314-PA [Drosophila melanogaster]</a>	377	377	98%	9e-122	38%	<a href="#">CAR93080.1</a>
<input type="checkbox"/>	<a href="#">Catalase, partial [Chelonia mydas]</a>	380	380	93%	9e-122	41%	<a href="#">EMP28164.1</a>
<input type="checkbox"/>	<a href="#">catalase_Cat [Talaromyces marneffeii ATCC 18224] &gt;gb EEA20965.1  catalase_Cat [Talaromyces marneffeii ATCC 18224]</a>	395	395	94%	4e-129	42%	<a href="#">XP_002151965.1</a>
<input type="checkbox"/>	<a href="#">catalase [Heliconius erato hvdara]</a>	390	390	67%	4e-129	54%	<a href="#">AET96342.1</a>
<input type="checkbox"/>	<a href="#">catalase 2 [Caenorhabditis elegans]</a>	419	419	95%	2e-138	44%	<a href="#">CAA74394.1</a>
<input type="checkbox"/>	<a href="#">hypothetical protein [Monosiga brevicollis MX1] &gt;gb EDQ90517.1  predicted protein [Monosiga brevicollis MX1]</a>	423	423	94%	5e-140	46%	<a href="#">XP_001744568.1</a>
<input type="checkbox"/>	<a href="#">LOC495840 protein, partial [Xenopus laevis]</a>	424	424	93%	7e-140	43%	<a href="#">AAH86479.1</a>

This is better. I've got sequences from Zebra fish, human, rice, fruit fly, a fungus, a worm, and a frog.

# Make a multiple sequence alignment

```

Danio.rerio      MADDREKSTDMKMLWKEGRGSQRPDVLTGAGVPIGDKLNAMTAGPRGPL 50
Homo.sapiens    MADSRDPASDQMQHWEQRAAQKADVLTTGAGNPVGDKLNVIITVGRGPL 50
Daphnia.magna  --MAERDAATNQLKNFGESEKNEAKASVTTAGAPIGFKTASLTAGPRGPM 49
Caenorhabditis.elegans --MPNDPSDNQLKTYKETYP--KQVITTSNGAPIYSKTAVLTAGRRGPM 46
Drosophila     --MCSRDTASNQLIDYKNNDS--EVQREITTSSGTPVGVKDAIQTVGPRGPA 48
Talaromyces.marneffei --MGKD---DEPKTYRYNET----PTYTTSNGCPVFDPESSQRIGKNGPL 41
Pseudomonas.syringae -----MSQKT---LTTASGAPVADNQNSRSAGPRGPL 29
BLASTEDproteinCorynebacterium. ---SEKSAADQIVDR--GMRPKLSGNTTRHNGAPVPSENISATAGPQGN 45
Oryza.sativa   ----MDPCKFRPSSSFDTKT-----TTTNAGAPVWNDNEALTVGPRGPI 40
                *  * * :          * . **

```

```

Danio.rerio      LVQDVVFTDEMAHFDRERIPERVVHAKGAGAFGYFEVTHDITRYSKAKVF 100
Homo.sapiens    LVQNVVFTDEMAHFDRERIPERVVHAKGAGAFGYFEVTHDITKYSKAKVF 100
Daphnia.magna  LLQDHVYIIDEAHFDRERISERVVHAKGSGAFGYFEVTHDISKYCKAAIF 99
Caenorhabditis.elegans LMQDVVYMDEMAHFDRERIPERVVHAKGAGAHGYFEVTHDITKYCKADM 96
Drosophila     LLQDFQFLDEVVMHFDSEIRIPERVAYAKGAGAFGYFECTHDISKFCASIF 98
Talaromyces.marneffei LLQDFHLIDLHLAHFDRERIPERVVHAKGAGAYGEFEVTDIDISDITDML 91
Pseudomonas.syringae LLDDFHLIEKLAHFNRENIPERRVHAKGSAHGTFVTRDISQYTSAKLF 79
BLASTEDproteinCorynebacterium. VLNDIHLIEKLAHFNRENVPERIPHAKGHGAFGELHITEDVSEYTKADLF 95
Oryza.sativa   LLEDYHLIEKVAHFARERIPERVVHARGASAKGFFECTHDVTDITCADFL 90
                :::      :   ** *.:** :*: . * * : * *::      .:

```

```

Danio.rerio      EHVGGKTPIVVRFSTVAGEAGSPDTRDPRGFAVKFYTDEGNWDLTGNN 150
Homo.sapiens    EHIGKKTPIAVRFSTVAGESGSADTVRDPGRGFAVKFYTDEGNWDLVGN 150
Daphnia.magna  SQVGGKTPVAVRFSTVGGESGSADTARDPRGFAVKFYTEEGNWDLVGN 149
Caenorhabditis.elegans NKVGGKTPLLVRFSTVAGESGSADTVRDPGRGFSKLFYTEEGNWDLVGN 146
Drosophila     DKVRKRTAVAMRFSVACGEQGSADTVREQRGFAVKFYTDDGIWDIVGCM 148
Talaromyces.marneffei KGVGKTKLVTRFSTVGGEEKSADSARDPRGFSVKFYTEQGNWDWVFN 141
Pseudomonas.syringae DTVGGKQTPIFLRFSTVGGGERGSADTERDPRGFAIKFYTEEGNWDIVGN 129
BLASTEDproteinCorynebacterium. Q-PGKVTPLAVRFSTVAGEQGSPTWRDVHGFALRFYTEEGNYDIVGN 144
Oryza.sativa   RSPGAQTPVIVRFSTVIHERGSPETIRDPRGFAVKFYTREGNWDLLGN 140
                * :   ***. * **:: * : ****:*** : * : *

```

```

Danio.rerio      PTFPIRDTLLSPSFIHSQKRNPQTHLKDPMVWDFWLSLRPESLHQVSLF 200
Homo.sapiens    PIFPIRDPILFSPSFIHSQKRNPQTHLKDPMVWDFWLSLRPESLHQVSLF 200
Daphnia.magna  PIFPIRDPILFSPSFIHTQKRNPVTHLKDPMFWDPISLRPETTHQVCF 199
Caenorhabditis.elegans PIFPIRDAIHFPNFIHALKRNQPQTHMRDPNALDFWMMNRPSIHQVMFLY 196
Drosophila     PVHYVRDPMFLFSLVHAQKRNPQTHLKDPMFWDFTLRPETLHALLMYF 198
Talaromyces.marneffei PVFFLRDPSKFPVFIHTQKRNPQTNLKDANMFWDYLSHQSALHQVMHLF 191
Pseudomonas.syringae PVFFIRDPMKFPDFIHTQKRLPQTNLKSQMMWDFWWSHSPESALHQVTILF 179
BLASTEDproteinCorynebacterium. PTFFLRDGMKFPDFIHSQKRLNKNGLRDADMQWDFWTRAPESAHQVTYLM 194
Oryza.sativa   PVFFIRDGIKFPDVIHAFKPNPRSHVQYWRVDFLSHHPESLHTFFFLF 190
                * .::** * .::: * . :. :*: * : * .

```

Much better. Fewer conserved residues and some moderately mutated positions.

Catalase is a difficult protein to make a multiple sequence alignment because it is a 'perfect' protein, ubiquitous in nearly all living organisms.

For some proteins, a good set of relatives may only require sequences from different bacteria.

# Make an informative image of protein

See the PyMol Demo on how to make an image.

The most important thing with this is to have the image tell us something useful. Don't make an image just because we told you to do so.

# Write 0.5 page summary on protein

Tell us something interesting about the protein. Say something about the function, and look through the primary literature to discuss something unique. The Pubmed reference for the protein that has a crystal structure is a good place to start looking for information. Wikipedia is a good place to start, but don't stick to that alone.