

Using public databases to gather information required for our sequence analysis of the rifR mutant lines

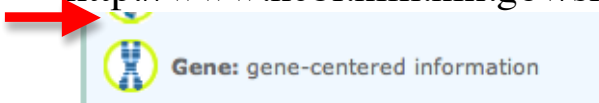
1. Download nucleotide and amino acids sequences of the E. coli rpoB gene
2. Align rpoB primers with the nucleotide sequence
3. Download sequence of PCR product and translate
4. Determine what portion of the rpoB protein will be analyzed in our our experiment by aligning translated PCR product with rpoB amino acid sequence
5. Learn about the structure and biochemistry of the the E. coli RNA polymerase holoenzyme

Working through this exercise

- Open a Word file and use it to take notes, save sequences and generally keep track of what you have done. This file should serve as a substitute for recording this exercise in your lab notebook. You will be required to submit these notes in digital form
- Please include your last name in the name of the file that you submit
- If you are instructed to save a record of your analysis in the form of an image, use Print Screen which sends a copy of whatever is on the screen to the clipboard. If you want to copy only the active window, hold down the Alt key, then hit the Print Screen key. It will copy only that window that is active, not the whole desktop.

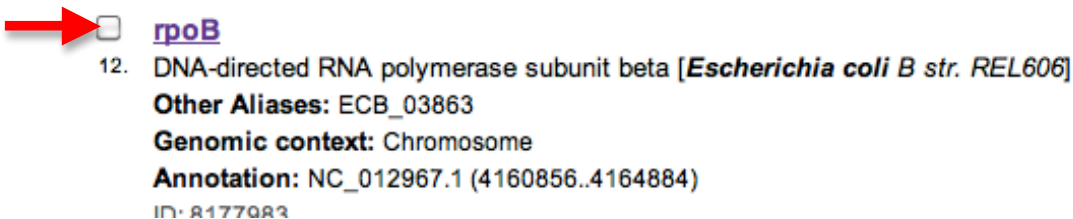
START: Go to Entrez and click on Gene

<http://www.ncbi.nlm.nih.gov/sites/gquery>



Enter Escherichia coli AND rpoB in the search box

Scroll down until you find the E. coli B strain info



Carefully Inspect information available on the Gene page

→Scroll down and read summary info and mRNA/protein info carefully

Summary

Gene symbol [rpoB](#)
Gene description DNA-directed RNA polymerase subunit beta
Locus tag [ECB_03863](#)
Gene type protein coding
RefSeq status Reviewed
Organism [Escherichia coli B str. REL606 \(strain: REL606\)](#)
Lineage Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia

→NOTE rpoB and rpoC (you wil meet the latter later on) are members of the same operon:

Genomic context

Sequence : [NC_012967.1](#) (4160856..4164884)



***→Scroll down to mRNA and Protein(s) under NCBI Reference Sequences (RefSeq)
Click on [YP 0030...](#)***

mRNA and Protein(s)

1. [YP_003047036.1](#) DNA-directed RNA polymerase subunit beta [Escherichia coli B str. REL606]
[UniProtKB/TrEMBL](#) [C6UJL6](#)
[Conserved Domains \(4\) summary](#)

Click on FASTA: this give you the sequence in fasta format:

> Name
SEQUENCE

DNA-directed RNA polymerase subunit beta [Escherichia coli B str.

NCBI Reference Sequence: [YP_003047036.1](#)

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

Go to:

LOCUS [YP_003047036](#) 1342 aa linear BCT 11-APR-2011
DEFINITION DNA-directed RNA polymerase subunit beta [Escherichia coli B str. REL606].
ACCESSION [YP_003047036](#)

Send to file or clipboard and save sequence as a Word file

Display Settings: FASTA

Send to:

DNA-directed RNA polymerase subunit beta [Escherichia coli B str. REL606]

NCBI Reference Sequence: YP_003047036.1

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

```
>gi|254163928|ref|YP_003047036.1| DNA-directed RNA polymerase subunit beta
[Escherichia coli B str. REL606]
MVYSYTEKKRIRKDFGKRPQVLDVPYLLSIQLDSFQKFIEQDPEGQYGLEAAFRSVFPIQSYSGNSELQY
VSYRLGEPVFDVQECQIRGVTYSAPLRVKLRRLVIYEREAPECTVKDIKEQEVYMGEIPLMTDNGTFVING
TERVIVSQLHRSPGVFFDSDKDKGKTHSSGKVLNARIIPYRGSWLDFEFDKDNLFVRIIDRRRKLPAIIL
RALNYTTEQILDLFPEKVIPEIRDNKLQEMELVPERLRGETASFDIEANGKVYVEKGRRITARHIRQLEKD
DVKLIIEVPVEYIAGKVVAKDYIDESTGELICANMELSLDLLAKLSQSGHKRIETLFTNDLDHGPYISET
LRVDPTNDRLSALVEIYRMRPGEPPTREAAESLFENLFFSEDRYDLSAVGRMKFNRSLLREEIEGSGIL
-----
```

Go back to **GenPept (full)** display of protein page, scroll down and click on CDS (coding sequence) and then at the bottom of the page click on FASTA

```
1..1342
/gene="rpoB"
/locus_tag="ECB_03863"
/coded_by="NC_012967.1:4160856..4164884"
/note="DNA-dependent RNA polymerase catalyzes the
transcription of DNA into RNA using the four
ribonucleoside triphosphates as substrates; beta subunit
is part of the catalytic core which binds with a sigma
factor to produce the holoenzyme"
/transl_table= 11
/db_xref="GeneID: 8177983 "
```

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

→ Save nucleotide sequence in Fasta format (Note first and last codon of sequence)

NEXT STOPS:

Find rpoB primers that we will use in our PCR at this link

<http://fire.biol.wvu.edu/trent/trent/primersequence.htm>

*And align on rpoB CDS sequence using **Sequence Extractor**:*

<http://www.bioinformatics.org/seqext/>

- First clear default sequence from boxes
- Follow the instructions on this site for entering all three primer sequences (for both PCR and sequencing)
- *To keep the output easy to read, select these options and then click on **SUBMIT***

Use the following options to alter the output of Sequence Extractor. For more details about individual options, see the [help](#).

- Genetic code:
- Restriction set:
- Translate reading frame:
- Topology:
- Allow primers to have mismatched: 5' tails, 3' tails.
Matching bases required when mismatching bases allowed:
- Bases per line:
- Show reverse strand, number line, spacer line.
- Return restriction summary, primer summary, help information, coding sequence links, translation links, options selected.

- *Note at the end of the output is a summary of the primer positions. Record this information*
- *Also notice instructions for obtaining sequence defined by two primers. Generate sequence of PCR product (primers rpoB-L and rpoB-R):*

Mouse over items on the map to view additional information. Click on two restriction sites or two compatible PCR primers to generate a product (requires JavaScript).

*Example of output (**difference gene and primers**)*

PCR Product Window

PCR product from: 18 to 1260.

Length: 1243 bp.

The sequence of the PCR primers replaces the template sequence.

```
cgaggtcgacggtatcgataagccttgatatcgaattcccccgggaccaaaatggcaagta
aatagcctgatgggataggctctaagtccaacgaaccagtgtcaccactgacacaatga
gaaaaccATGTCACATCTCGCAGAACTGGTTGCCAGTGCGAAGGCGGCCATTAGCCAGG
CGTCAGATGTTGCCGCGTTAGATAATGTGCGCGTCGAATATTTGGGTAAAAAAGGGCACT
-----
```

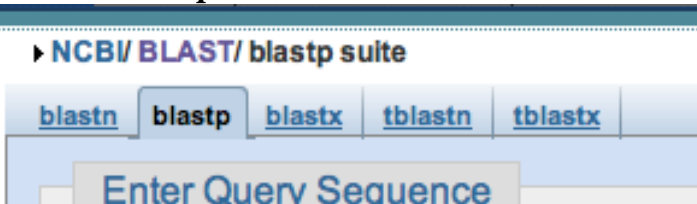
*Copy and paste the sequence of the PCR product into a Word file
Translate it here and save sequence output (longer ORF = open reading frame)
http://www.bioinformatics.org/sms2/orf_find.html*

NEXT: Align translation of PCR Product with amino acid sequence of rpoB gene using BLAST (basic local alignment search tool)
<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Scroll down and under Specialized BLAST, click on Align two or more sequences:

- **Align** two (or more) sequences using BLAST (bl2seq)

Click on blastp tab:



Paste translated PCR sequence into Query box and complete amino acid sequence of the rpoB gene into Subject Sequence

*Click the **BLAST** button and scroll down to alignment*

SAVE a desktop picture of the alignment OR **copy and paste** sequence alignment into a Word file (courier 10 font will align nicely)

NEXT: Examine the *rif* resistance regions of the RNA polymerase beta subunit at this link and *determine which of the four clusters of rifR mutations will be covered by our analysis*

<http://fire.biol.wvu.edu/trent/trent/rpoBrifRregion.htm>

continues on the next page

Learning about RNA polymerase and transcription

Inspect this web page

http://www.ecogene.org/topic.php?topic_id=150

The following table summarizes the subunits of *E. coli* RNA polymerase and their properties:

subunit	size aa	size (Kd)	gene	function
alpha (b)	329	36511	rpoA	required for assembly of the enzyme; interacts with some regulatory proteins; also involved in catalysis
beta (b)	1342	150616	rpoB	involved in catalysis: chain initiation and elongation
beta' (b')	1407	155159	rpoC	binds to the DNA template
sigma (s)	613	70263	rpoD	directs enzyme to the promoter
omega (w)	91	10237	rpoZ	required to restore denatured RNA polymerase in vitro to its fully functional form

Note: Links in the subunit column will take you to the SWISS-PROT database entry on the protein; Links in the gene column will take you to the EMBL database entry on the gene.

The subunits of RNA polymerase assemble into a structure that has the same "hand" like structure as DNA polymerases.

Go back to *rpoB* Gene page (pgs 1-2 of this handout). scroll down and click on RNA polymerase

Additional Links ⬆ ⬇ ?

Gene LinkOut

The following [LinkOut](#) resources are supplied by external providers. These providers are responsible for maintaining the links.

[Molecular Biology Databases](#)

- [PATRIC: PathoSystems Resource Integration Center](#)
[VBIecCol129921_4100](#)
- [Kyoto Encyclopedia of Genes and Genomes](#)
 - [RNA polymerase](#)
 - [Metabolic pathways](#)
 - [Pyrimidine metabolism](#)
 - [Purine metabolism](#)
 - [eco.b3987](#)
- [iHOP - Information Hyperlinked over Proteins](#)
[GENEVESTIGATOR](#)

Save image of RNAPol complex to refer to later