

## Biology 322 Fall 2012 *Sequence Analysis of a portion of the rpoB gene in rifamycin-resistant mutants of E. coli: Part 4*

### Working through this exercise

- Open this document online and use imbedded links.
- Open a Word file and use it to take notes and 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** (in addition to a summary of what you have learned about the rifR mutants—see other handout). **Please include your last name in the name of the file that you submit via email**
- Browse through this handout before you start the exercise. You will want to set up a table to organize your observations about each mutation. See the data workup handout for suggestions as to how to organize your table.
- If you are instructed to save a figure or 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 (Fn + F11). It will copy only that window that is active, not the whole desktop.
- 🍏 If you are working on an Apple computer use these commands:
  - Snapshot of Screen: saved to clipboard: *command (apple) \* control \* shift \* 3*
  - Snapshot of selected portion of screen: saved to clipboard: *command (apple) \* control \* shift \* 4*

---

**First: Think through the steps of this experiment**, which has spanned a number of weeks. How did we get to this point? List each step:

1. First we did this
2. Then we did that
3. Then we or somebody else did this other thing
4. Etc.

---

**Second: Review the results of the previous online exercise** that introduced you to the *E. coli rpoB* gene and its protein product:

- Reexamine the alignment of the translated PCR product with the AA sequence of the rpoB gene. What portion of the rpoB gene did we target with our PCR/sequence analysis?
- Orient your self with respect to the nucleotide sequence: in other words, find the nucleotide coordinates of clusters I, II and III in the ***E. coli B rpoB PCR product*** – start with AA coordinates for I, II & III in your translated PCR product and then calculate the NT coordinates. No introns here to complicate the NT and AA alignments!

```
>EcoB PCR product      SEQ primer  I  II  III
RRIRSVGEMAENQFRVGLVVRVERAVKERLSLGDLDLTLMPQDMINAKPISA AVKEFFGSSQ
LSQFMDQNNPLSEITHKRRISALGPGGLTRERAGFEVRD VHPH YGRVCPIETPEGPNIG
LINSLSVYAQTNEYGFLETPYRKVTDGVVTDDEIHYLSAIEEGNYVIAQANSNLDEEGHFV
EDLVTCRSKGESSLFSRDQVDYMDVSTQQVVS VGASLIPFLEHDDANRALMGANMOROAV
PTLRADKPLVGTGMERAVAVDSGVTAVAKRGGVVQYVDASRIVIKVNEDEMYPGE
```

I II III more info at link below  
<http://fire.biol.wvu.edu/trent/trent/rpoBrifRregion.htm>

**Third:** Before you start this analysis, take note of the position of the rpoB-S primer relative to the 5' end of the PCR product. –Note in the alignment (below) of sample sequence data: the sequence each rifR PCR product will start 3' of this rpoB-S primer – in variable positions. NOTE the nt register refers to the position in the full-length PCR product.

```

      130      140      150      160      170
EcoliB/1-885  TCAACGCCAAGCCGATTTCCGCAGCAGTGAAAGAGTTCCTTCGGTTC
5-rpoB-S/1-670  -----GCCAAGCCGATTTCCGCAGCAGTGAAAGAGTTCCTTCGGTTC
10-rpoB-S/1-729  ---C-----GCCAAGCCGATTTCCGCAGCAGTGAAAGAGTTCCTTCGGTTC
18-rpoB-S/1-664  -----A-----GTTCTTCGGTTC

```

**Here is a link to the nucleotide sequences**

<http://fire.biol.wvu.edu/trent/trent/F12rpoBseqdata.docx>

The first sequence represents the *E. coli B rpoB*<sup>+</sup> PCR product (you should have the same file from your Part 3 exercise)

**Note the sequence data are in the so-called Fasta format:**

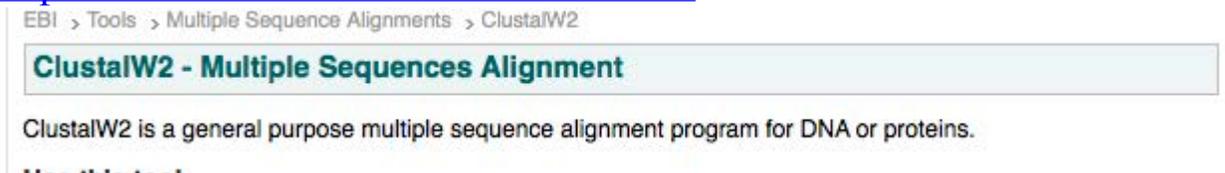
- name of file is preceded by a >
- sequence starts on the next line

Also note that rpoB-S is the name of the primer used for the sequencing reaction. The third number indicates the number of bases in the sequence file.

See link at end of handout for instructions for accessing the sequence files online to view chromatograms and sequence quality

**You will do your alignments at this site**

<http://www.ebi.ac.uk/Tools/msa/clustalo/>



**Open up the site and inspect the page.....**

1. First do an alignment of *E. coli B rpoB+* PCR product with the first wild-type sequence run (# >1) at the beginning of the sequence document.

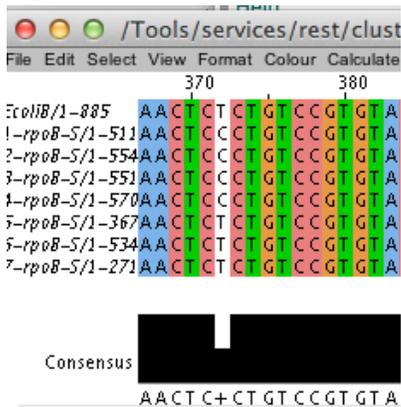
- Copy and past these two sequences into the alignment box.
- Select DNA and use default options.
- Inspect the output. Double-check to confirm that the *E. coli B rpoB+* reference sequence is listed first.
- Note that the sequence alignment starts at about 135 bases 3' of the start of the reference *E. coli B rpoB+* sequence and ends wells before the 3' end of the reference sequence.

2. Next do a multiple NT (nucleotide) alignment of the *E. coli B rpoB+* PCR product with the wildtype sequence runs (#>1 through >7).

SWT = rifS colony taken from non-mutagenized (spontaneous) culture. The *E. coli* strain used for this culture is of unknown ancestry.

EWT = rifS colony taken from EMS-mutagenized culture (known to be a descendant of the wildtype *E. coli B* strain)

- Inspect the output. Click on **Result Summary** and then **start JalView**
- Inspect JalView output for differences between *the reference E. coli B rpoB+ sequence with the* sequences from rifS strains. Keep a close eye on the black consensus bar at the bottom.
- For now just focus on the interval between *nt 185* and *nt 610* using the *E. coli B rpoB+* nucleotide count



• **FOR any sequence differences observed, take a picture of the window** (see example) when you have the sliding bar positioned to indicate the base pair difference. Play around with the JalView window shape and size to best document your observations.

• Save this info for later. The base pair variation shown to the left would be symbolized T372C using *E. coli B* as the reference sequence.

2. Your next task is to examine the sequence from the rifR strains. You can paste all of the sequences in at once if you want including the SWT and EWT (analyzed above) since you will need to keep track of mutations that are in the background of the strain versus resistance causing mutations. Best bet though is to use the *E. coli rpoB+* sequence and one of the SWT (#4 is preferable since it has the longest read) sequences as reference and to omit the other WT sequences as they will tend to clutter up the alignment. You might also want to do the ER and the SR sequences separately but that is your call. It is useful to copy and paste the sequences that you want to align into a separate word file before you past them into the box on the web site. Be sure that *E. coli rpoB+* sequence is the first sequence as this will set the NT register.

- Inspect the output. Click on Result Summary and then start JalView
- Inspect JalView output for differences between the reference wildtype and rifR sequences. Focus on the interval between *nt 185* (or wherever the sequence happens to

start) and *nt 610* (or wherever the sequence happens to end. using the *E. coli B rpoB+* PCR product nucleotide count. **Keep a close eye on the black consensus bar at the bottom.**

- **FOR any base pair difference, take a picture of the window.** Note position and nature of mutation in standard form

### 3. Figure out how many of the *rifR* mutants show the same mutations

### 4. Next determine whether the base pair changes you have observed affect the amino acid sequence of the protein.

- First organize your data. Obviously you only need to do one analysis of the translated sequence data for each mutational site
- READ about sequence conventions and translating your nucleotide sequence at this site: <http://fire.biol.wvu.edu/trent/trent/rpoBseqanalysisORFs.pdf>
- Translate a representative sequence from each category of mutation (mutational site). Be sure to set the translation options as shown below. Direct means that you have pasted in a sequence that read 5'-3' in the same direction as the mRNA.

[http://www.bioinformatics.org/sms2/orf\\_find.html](http://www.bioinformatics.org/sms2/orf_find.html)

Paste the text into the text area below. Input limit is 100000 characters.

```
AGTGACCGACGGTGTGTAAGTACGAAATTCAGTATCTGTCTGCTATCGAAGAAGGC
AACTACGTTATCGCCAGGCGAACTCCAACCTGGATGAAGAAGGCCACTTCGTAGAAG
ACCTGGTAACTTGGCGTAGCAAAGGCGAATCCAGCTTGTTCAGCCGCGACCAGGTTGA
CTACATGGACGTATCCACCCAGCAGGTGGTATCCGTGGTGGTCCCTGATCCCGTTC
CTGGAACACGATGACGCCAACCGTGATTGATGGGTGCGAACATGCA
```

Please check the [browser compatibility page](#) before using this program.

- ORFs can begin with:  .
- Search for ORFs in reading frame  on the  strand.
- Only return ORFs that are at least  codons long.
- Use the  genetic code.

---

### 5. Return to the alignment site

<http://www.ebi.ac.uk/Tools/msa/clustalo/>

- Align translations of the *E. coli B rpoB+* sequence with translation of each selected nucleotide sequence. **Focus on AA 50-240**
- Record your observations as described above noting the nature and location of the amino acid change and documenting your analysis with JalView screen shots. Again, use standard notation: L230Q indicates that glutamine replaced leucine at amino acid position 230.
- What is the significance of the conservation bar in JalView? Review amino acid biochemistry at this link:

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

If you want to check out your own personal sequence data follow this link.....<http://fire.biol.wvu.edu/trent/trent/F12SequencePeek.pdf>