

Data Interpretation

The DNA Sequencing Facility recommends that three aspects of each sequence be checked prior to using the data. Checking each sequence is easy to do, takes very little time, and will result in more accurate data while preventing wasted time spent analyzing poor quality, unreliable sequence. The basecaller computer program is not 100% accurate and **an experienced DNA sequencer can often do a much better job than the program can!**

- 1 Check the total signal strength to ensure it is above 400.
- 2 Locate any "dye blobs" at the beginning of the sequence and ignore any sequence obscured by them.
- 3 Scan the printed sequence above the peaks and note any areas of compressed bases, gaps or N's. Examine these areas in more detail to see if the basecaller program has made any errors.

Explanations:

1 Signal Strength

Adequate signal strength is essential for optimal DNA sequencing and should be checked on each sample. The signal strength is listed in the seventh line in the chromatogram details box of each chromatogram's results page or in the Annotation window of the electropherogram file. For optimal results, the signal for each base should be at least 1000. Samples with values below 200 per base, although sometimes of good quality, often contain many errors due to low signal to noise ratios. Blank samples have a signal strength of about 100 for each base.

2 Compressions, Gaps, or N's

The basecaller program prints the letter of each base directly over the top of each peak. High quality sequence should have very evenly spaced peaks and therefore evenly printed bases. Any alteration in the pattern of printed bases, either a compression (two or more bases in the same location) or a gap (missed base) suggests a basecalling error and should be carefully checked. N's result when the basecaller program is unable to determine the correct peak, either because the peak(s) are too low or there are multiple peaks in the same location. The most common cause of N's in good quality sequence is due to "low G after A" phenomenon.

3 Low G peaks after A:

Due to steric hindrance, the incorporation of a fluorescently-labeled ddGTP terminator after an A is inhibited up to 95%. This gives rise to the very

predictable “low G after A” phenomenon and the commonly seen miscall of N after an A by the basecaller program. Manual editing of your sequence can easily correct these miscalls and increase the accuracy of your sequence significantly.

4 Blank Samples

Blank samples are samples that have an extremely low signal intensity along with chromatograms which look "Noisy". Blank samples are caused by a number of factors. Some of which include: a) the primer did not anneal properly to the template. b) your sample was gel purified, meaning that there are impurities in most agarose gels which inhibit the primer from annealing. c) Your primer has a specific annealing temperature, if PCR is not set to this temperature your primer will not anneal.

How do you correct Blank samples? Check your template and primers, make sure they are compatible. Check the annealing temperature. When gel purifying try to use a higher grade of agarose.

5 "Line Through My Sequence Text"

We are often asked: "What does the line through my sequence text mean?" The line is a quality value score that the FinchTV software uses to tell you what it thinks is low quality data. You can still use this data if you would like to "Blast" the sequence text.

