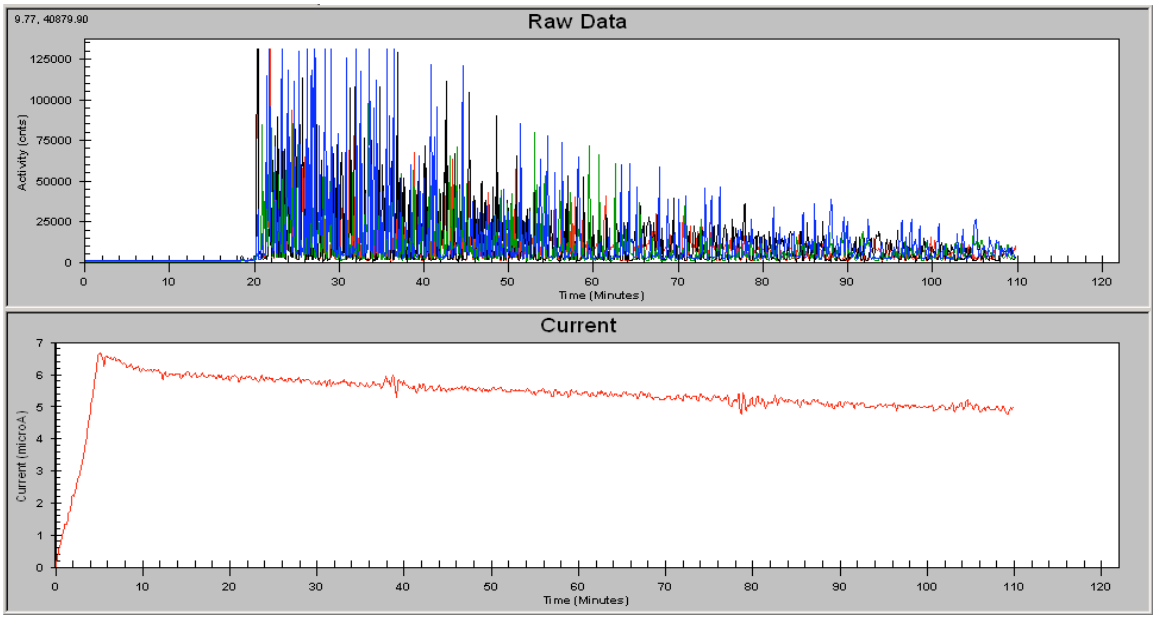
DNA Sequencing Troubleshooting Guide.

There are a number of factors that can lead to less than perfect DNA sequencing results. In this guide, we explain some of the common problems encountered, and outline ways in which these problems can be overcome.

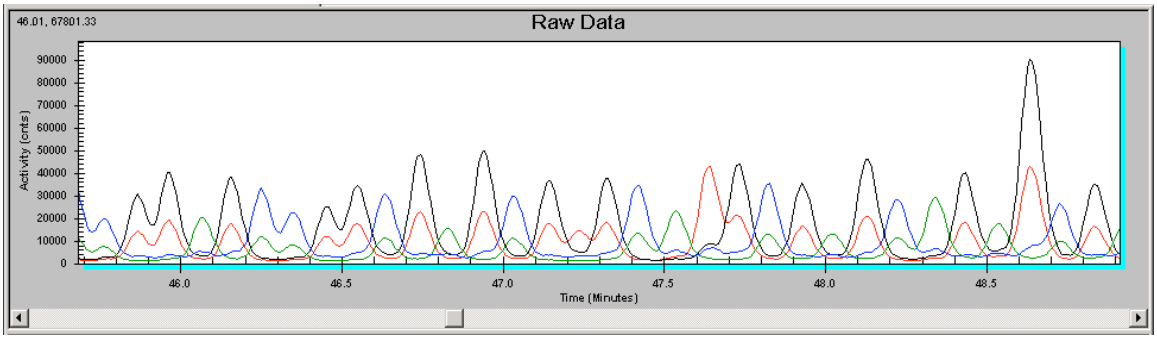
Below is an example of a normal sequencing result. Shown here is the raw data signal chromatogram and the current profile for the run.



There are several observations that can be made concerning normal raw data.

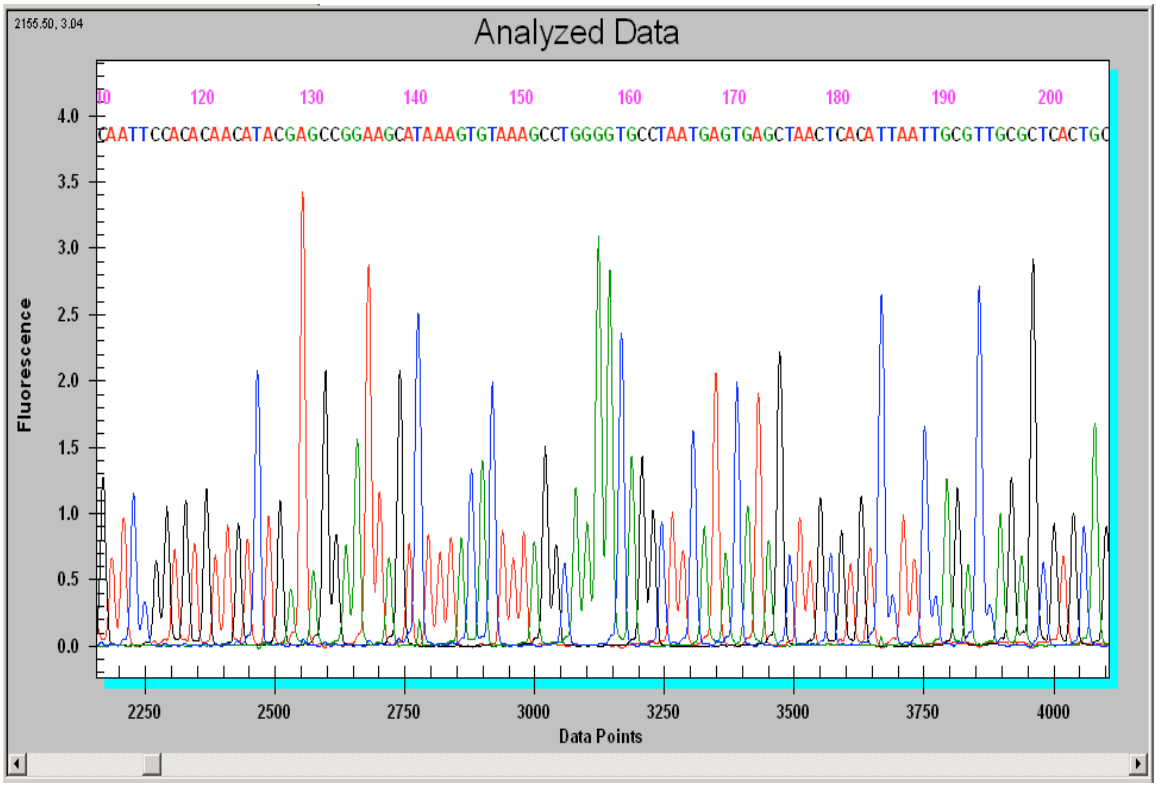
Firstly, the raw data baseline should be uniformly flat across the entire length of the chromatogram. Secondly, the raw data should start at **20 minutes**. The peak height at the start of the data should be around 125,000. The peak height should **gradually** fall as time increases.

This 'close-up' reveals other factors that need to be considered when looking at raw sequencing data.



As you can see, each peak is quite distinct from the next. However, there is an artifact under each peak. This phenomenon is quite normal and follows a specific pattern. Red (A) always 'pairs-up' with black (C), and blue (T) always 'pairs-up' with green (G).

As with raw data, there are several observations that can be made concerning normal processed data.

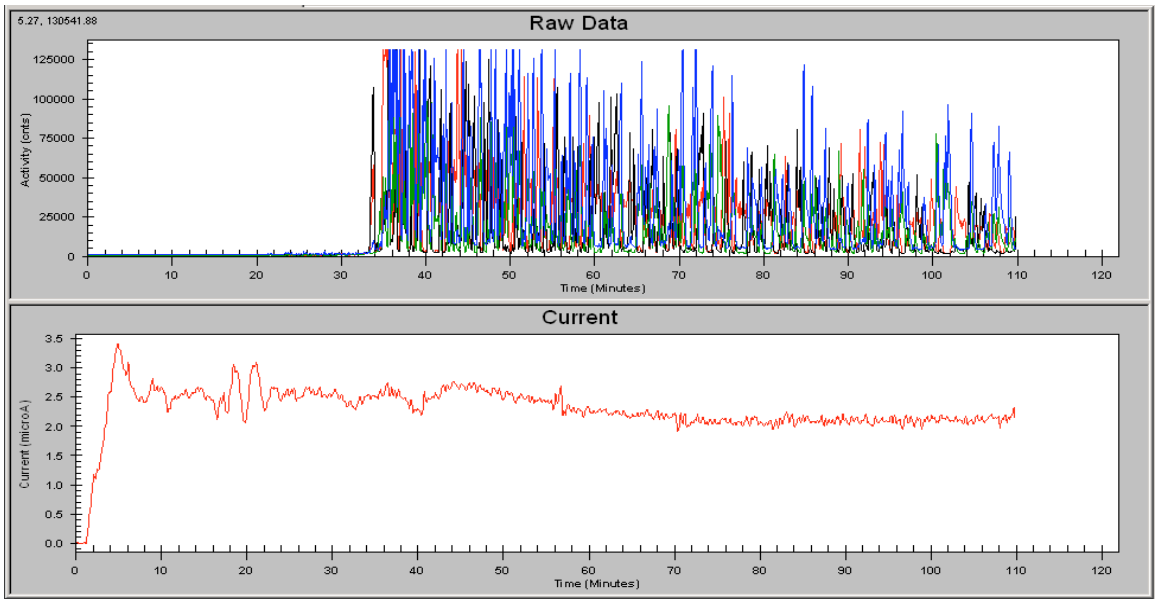


The baseline should be reasonably flat along the length of the chromatogram. Again, the peaks are very distinct and **evenly-spaced**. This time, however, there should be no artifacts under the peaks. The base calls (shown above each peak) should be accurate and should also be evenly-spaced. Peak resolution will naturally diminish towards the end of the chromatogram, and thus, base calling will become less reliable. We aim to provide sequence data that has reliable base calling up to 700 bases into the sequence. We test this on every run by sequencing standard pUC18 template and comparing the acquired sequence to the known sequence. We do not accept any result as reliable unless the control is 98.5% accurate at 700 bases.

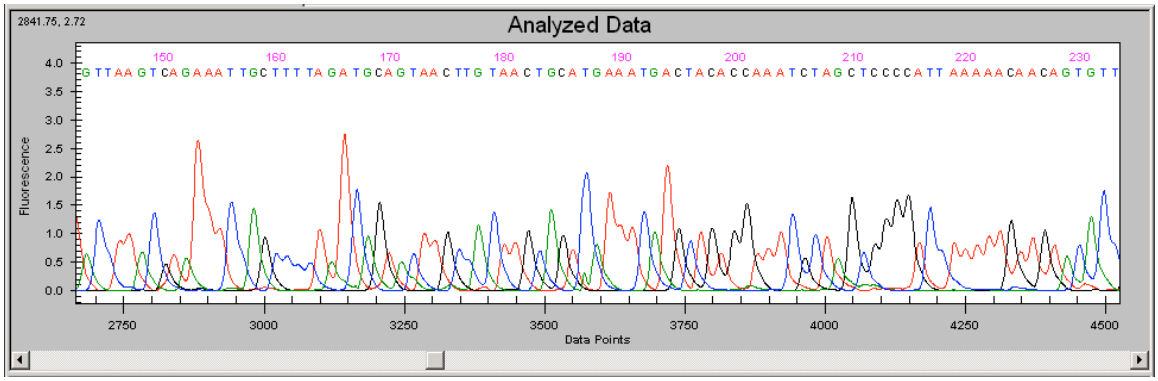
The information above describes a normal result. The following information describes observations we generally make where sequencing data differs from the norm, the factors that cause this, and what steps can be taken to remedy it.

Too much DNA

Highly concentrated plasmid preparations can have a highly detrimental effect on the electrophoresis of sequencing reaction products. They cause the capillary to become blocked, inhibiting the current and causing the reaction products to pass through the capillary slower than normal. Detection of these products begins later, resulting in the following chromatogram.



As you can see, the raw data starts later than normal (>20 minutes). As the length of the run is fixed at 110 minutes, the read length becomes truncated. The current is very uneven. The following diagram shows the effect that the inhibited current has on peak resolution in the processed sequence data.

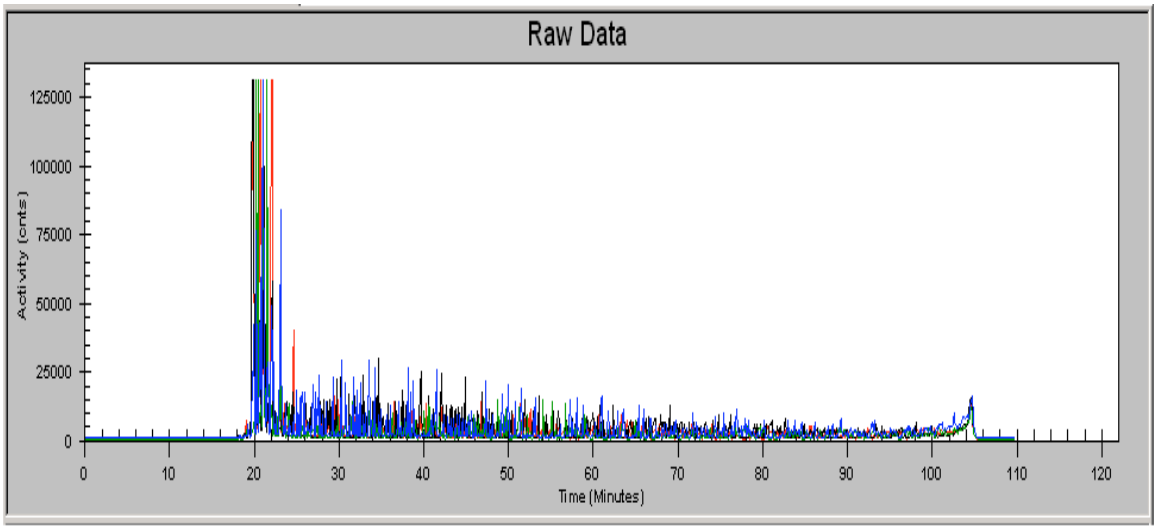


With these things in mind, it should be very apparent that quantification of template DNA prior to submission is **VERY IMPORTANT**. We recommend agarose electrophoresis for accurate quantification.

We also possess a Nanodrop spectrophotometer. It is capable of quantifying DNA and RNA using just 1µl of sample and is very accurate! If you would like to discuss using this nifty piece of kit, please don't hesitate to get in touch.

Low signal

Low raw data signal can be demonstrated by the following chromatogram.



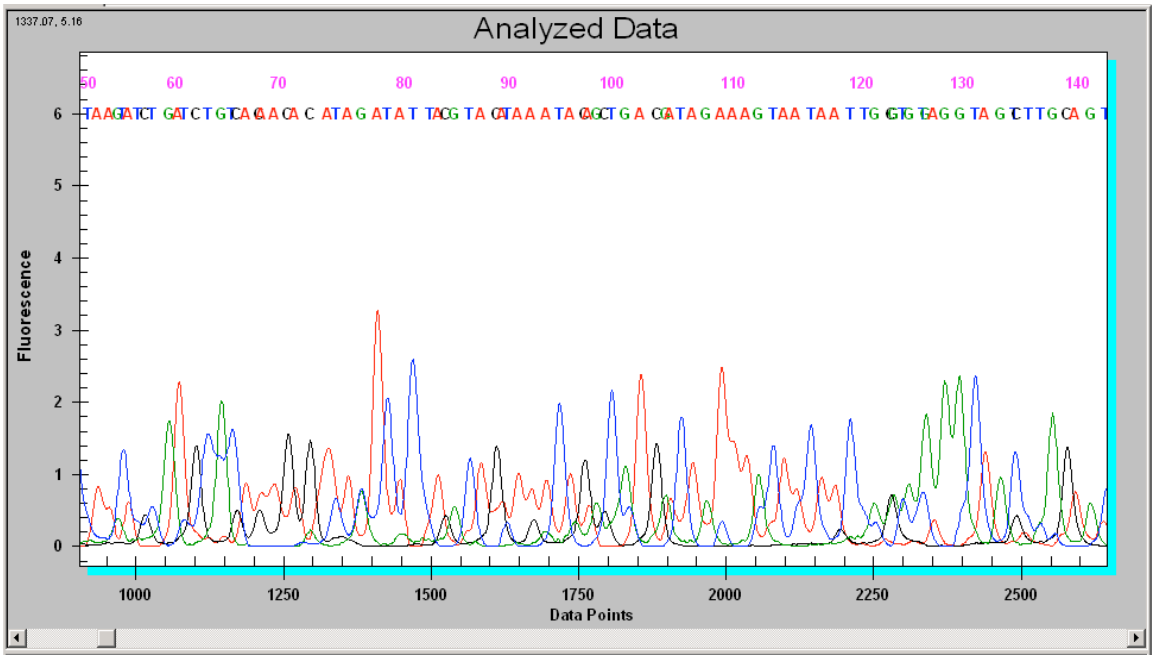
The initial peaks represent the unincorporated dye peaks. The raw data signal then drops sharply and then smoothes out to a gradual decline as seen in normal chromatograms.

When the raw data signal is as low as this, the software encounters difficulty interpreting this signal into a processed sequence. Sequence reliability is reduced, especially at the beginning and end of the read.

Low raw data signal can be caused by insufficient template or primer in the reaction or poor primer design, leading to an inefficient sequencing reaction. **Again, we must stress just how important accurate quantification of template and primer is if sequencing is to be successful.** With regards to primer design, there are several factors that determine whether or not a primer will be effective in the sequencing reaction. Our page on primer design outlines these factors.

Multiple signals

The presence of more than one set of reaction products is indicated by the following type of trace.



As you can see, the processed data is very different to the normal processed data discussed earlier. The peaks are not evenly spaced, they overlap and there are artifacts beneath them.

There can be several explanations for the presence of more than one set of reaction products. These are:

- Contaminated template or primer
- Poor quality primers e.g. preparations with a high degree of n-1 species
- Multiple priming

In order to avoid these issues, it is important to provide well-prepared template DNA. We recommend the use of the following:

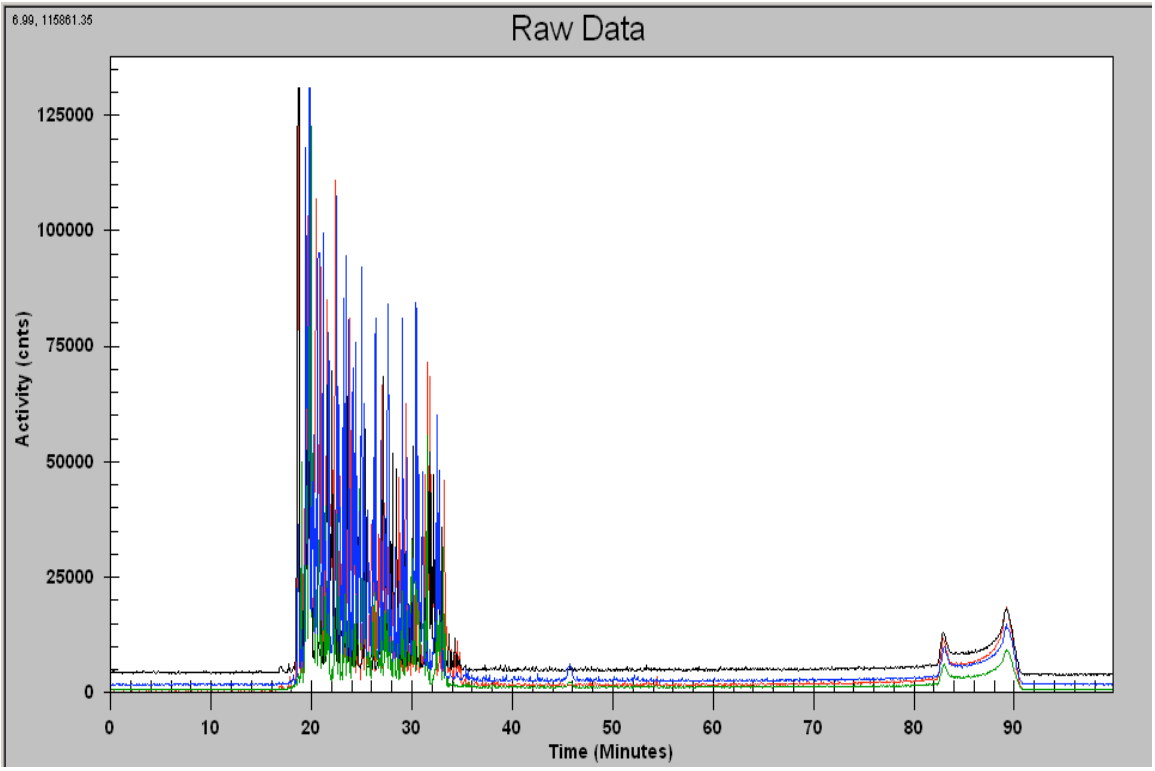
- QIAGEN Qiawell and QIAprep (Mini-Prep) DNA isolation protocols (dsDNA and ssDNA).
- QIAGEN Qiaquick PCR purification protocol (PCR products).

We also recommend using primers from a reliable supplier and advise carefully considering the design of the primer. Again, see our page on primer design.

This is extremely important!!!

Secondary structure

The following diagram shows the effect of secondary structure on a sequencing reaction.



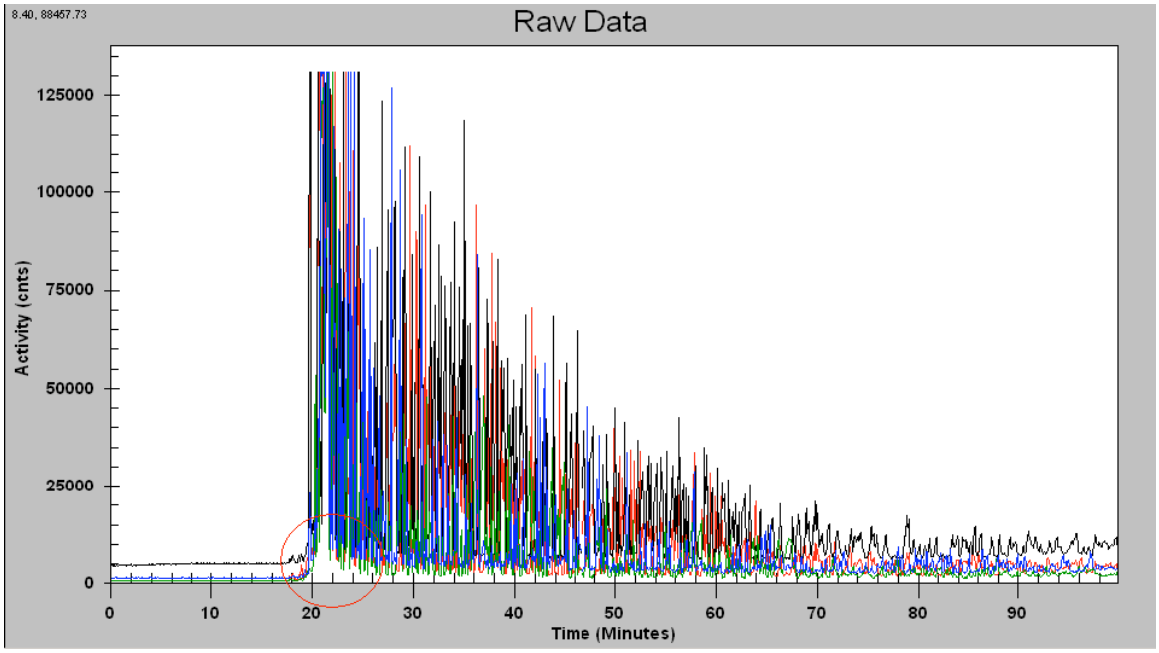
As you can see, the reaction has stopped suddenly, leading to the sudden drop in raw data. This is caused by the inability of the dNTP and ddNTP molecules to be incorporated into the secondary structure during the sequencing reaction.

There are steps we can take to overcome this. The DNA sequencing team here at the Wolfson Institute has a lot of experience in troubleshooting these kinds of problems and has developed different protocols that are proven to aid sequencing through regions of secondary structure. We can vary PCR conditions and can use different reaction mixtures in order to overcome this.

GC-rich sequences will have a similar effect on a sequencing reaction. Please specify prior to sequencing if the template you are providing is GC-rich. This will help us tailor the reaction for your needs.

Salt contamination

Salt contamination is a serious problem in DNA sequencing. It has the potential to inhibit everything from the sequencing reaction itself, to the electrophoresis of the reaction products. This is what salt contamination looks like.

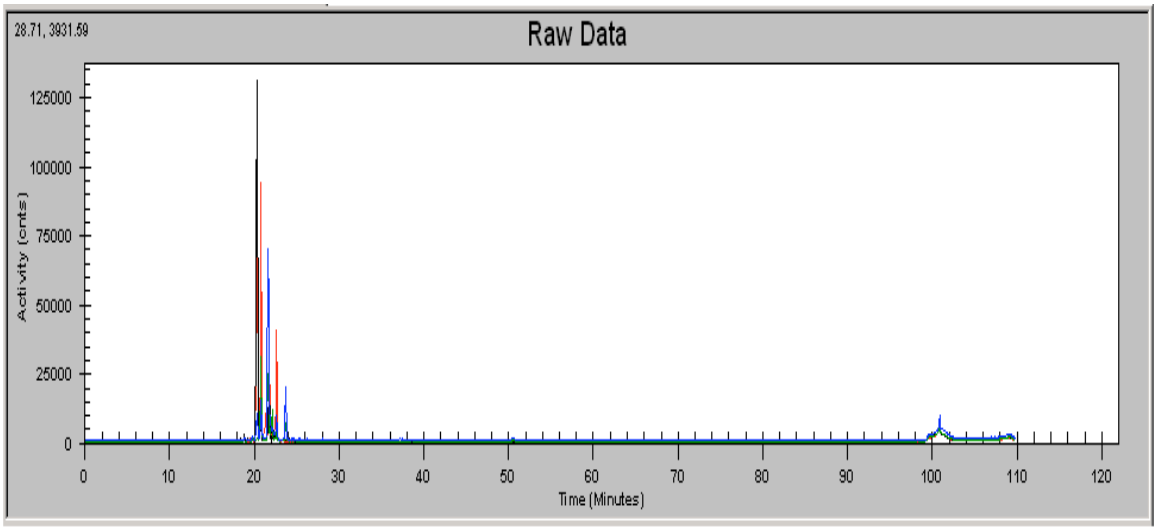


An elevated baseline in the raw data leads to poor sequence reads, in terms of read length and also reliability.

To remedy this problem, please provide well-prepared template and primer as described above in the section on multiple signals. **We cannot stress strongly enough just how important template quality and quantity are, so we'll keep mentioning it!**

Failed reaction

This is an example of a reaction that has failed completely.



Only the unincorporated dye peaks are present. For some reason, the reaction has been inhibited completely. What could this reason be? Well, provided that our controls have worked fine, and can therefore rule out a problem with our chemistry and preparation, we just don't know. Any one of the above factors (if extreme enough) or a combination of those factors could have caused this result.

However, please don't give up! Come and talk to us or give us a call using the details on our home page; we are very happy to work with you to try and overcome any problems you may have with your sequencing.

If the information here hasn't answered your question about our service, or if there are any other issues that you would like to raise with the sequencing team, please do not hesitate to get in touch using the details on our home page.

Alternatively, take a look at the Beckman Coulter CEQ8000 Troubleshooting guide for more detailed information on problems associated with DNA sequencing.