

Biomolecular Core Facility

AI Dupont Hospital for Children, Rockland Center One, Room 214
Core: (302) 651-6712, Office: (302) 651-6707, mbcore@nemours.org

Katia Sol-Church, Ph.D., Director
Jennifer Frenc Holbrook, BS, Assistant Director

Sequencing Guidelines

Adapted from ABI BigDye Terminator v3.1 Cycle Sequencing Kit and Roswell Park Cancer Institute Core Laboratory website

The first and most important factor in automated DNA sequencing is clean, pure template. For a sequencing reaction to be successful all excess primers, dNTPs, salts and residual RNA and proteins must be removed from the sample. The BCL routinely uses QiaQuick PCR cleanup columns from Qiagen. Once your PCR reaction has been cleaned, it is always a good idea to run it on an agarose gel to assure purity. Quantitation of your PCR products after clean up is also helpful since no clean-up protocol results in 100% yield.

Once you have a pure, clean PCR product you are ready to perform a sequencing reaction. A cycle sequencing reaction is very similar to a PCR reaction. Note that only one primer is used for a sequencing reaction. (PCR reactions use two primers, forward and reverse, resulting in a double stranded amplicon.) Usually the forward or reverse primer used for the PCR reaction can be used in the sequencing reaction. However, keep in mind that sometimes they do not perform well under sequencing conditions. We recommend two sequencing reactions for each fragment of interest to insure double strand sequencing of the fragment. Data analysis is not completely accurate for the first and last 25 bases of the sequence. See Applied Biosystems BigDye Terminator v3.1 cycle sequencing manual or the BCL's DNA Sequencing Protocol for sample preparation of the sequencing reaction. ABI does not recommend sequencing fragments less than 100 bp due to the limitations of the BigDye v3.1 Terminator chemistry.

Helpful tips to remember:

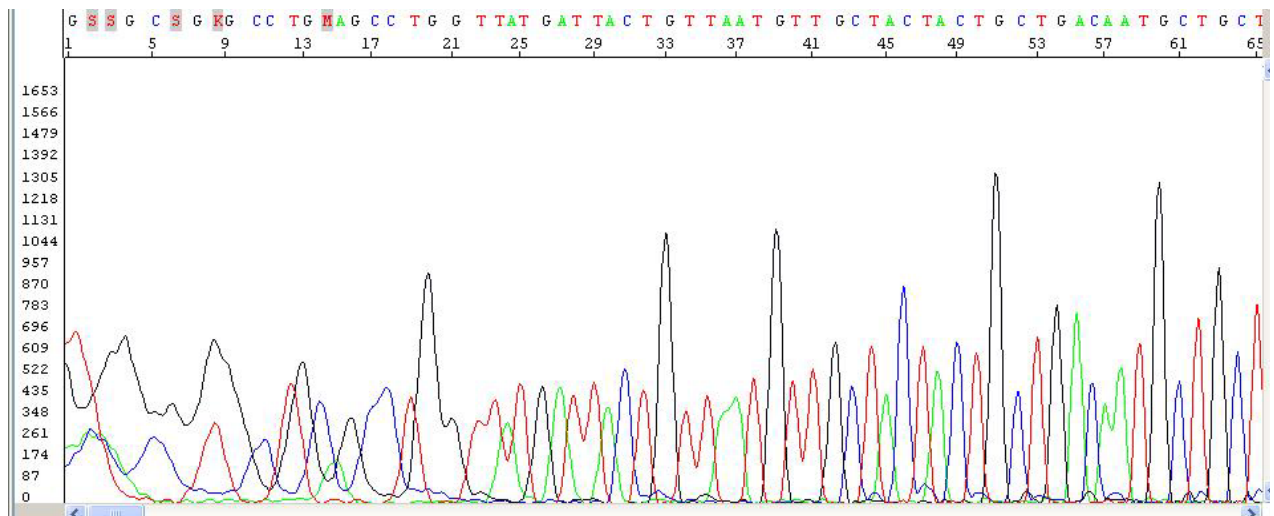
- Crucial for template DNA to be as clean and pure as possible! This cannot be stressed enough!
- Contaminants include residual primers, salts, proteins, RNA, ethanol, detergents, chromosomal DNA, dNTP's, buffer components, etc...
- BigDye Terminator v3.1 can be diluted down to 1/32 reaction with an optimal sequencing template. Note: This dilution is not recommended for PCR product. Plasmids only!
- When sequencing PCR products for the first time, a ¼ dilution of BigDye Terminator mix is a good starting point
- Signal to noise ratio (S/N) should optimally be between 50 and 300 rfu. The goal is to get clean sequence data with low S/N
- If signal is too low...try using more template DNA and/or more BigDye

- If signal is too high...try using less template DNA and/or diluting BigDye Terminator mix
- ABI recommends no more than 10 freeze/thaw cycles for the BigDye v3.1 Terminator Kit
- Even the best sequencing data is not 100% accurate, always look at your chromatograms
- The capillary electrophoresis platform is very sensitive, and will pick up even the smallest imperfections in your sequencing reaction.
- A hot SDS treatment with column purification or EtOH precipitation will remove dye blobs that can mask some of your sequencing data

Interpreting Your Chromatograms

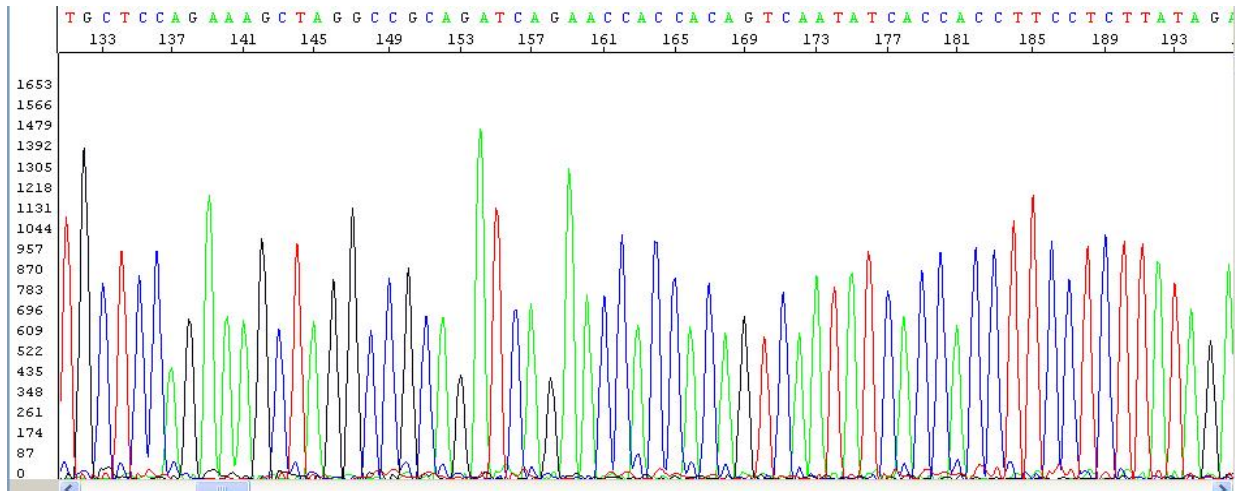
This section gives a brief overview on how to recognize good quality sequence data. The chromatogram images below are from our sequencing standard controls (pGEM) that are run after every array change.

Sequence at the beginning of your chromatogram:



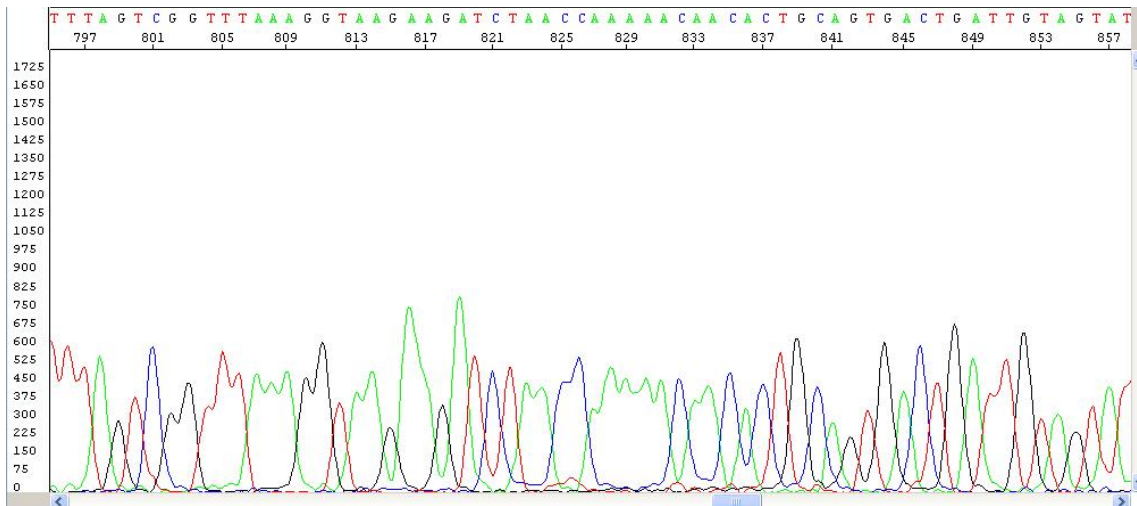
The first 25 bases or so give peaks that are round, crowded, and not resolved. Peak height is also smaller at the beginning (although sometimes very high due to primer dimers). Note that the 2 A's located at bp37 are not clearly resolved from each other. All of this is normal, but reinforces the suggestion of the BCL lab to place your primer at least 40-50 base pairs away from your sequence of interest. As you can see from the example above, even when using a high quality control plasmid, the early data is not very reliable and you will not be able to confidently read your sequence from base #1.

Sequence in the middle of your chromatogram:



Peaks are sharp, well defined, with even spacing between them. Peak height is higher than the earlier data with little or no background interference at the baselines. You will be able to confidently read this sequence manually with 100% accuracy.

Sequence towards the end of your chromatogram

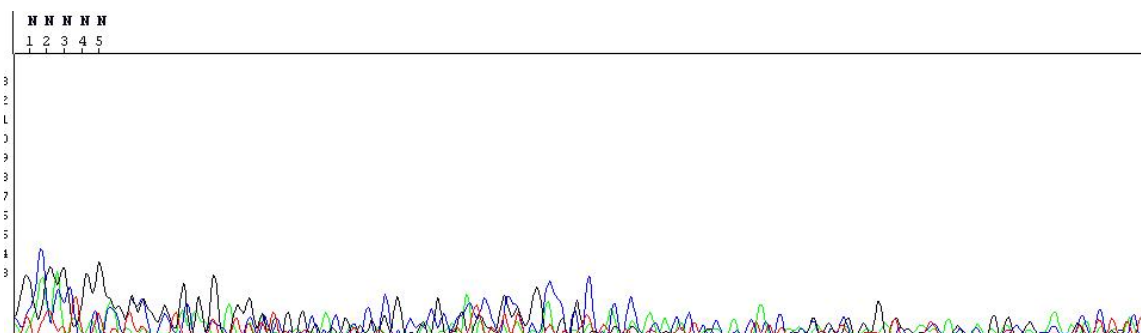


This is where the sequence resolution begins to deteriorate. Peaks are broad and rounded in shape (they are also referred to as "rolling hills"), especially when there are multiples of the same nucleotide in a row. This region will still be called with 100% accuracy by the 3130 base-calling algorithm (KB base caller), although it may be difficult to manually interpret any heterozygote calls, especially in a sequence where high background is present. The sequencing control example above gave a LOR (Length Of Read, which is defined as the read where the average Quality Value (QV) score is >20) of 1068 base pairs.

Troubleshooting Sequencing Data

This section contains examples of some of the most common sequencing issues that are encountered. The examples and possible causes listed below are not all inclusive and there may be other reasons for poor quality sequencing results. Remember, the two most important factors in obtaining good quality sequencing data are DNA purity and template concentration.

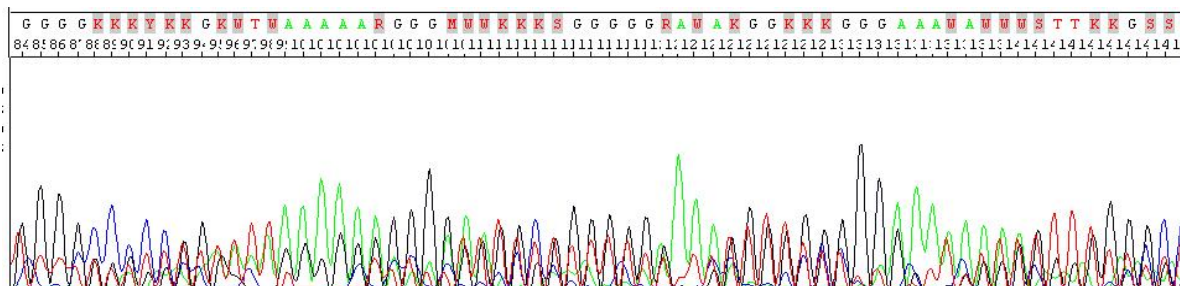
No sequence information:



Possible Causes:

- **Priming site not present.**
- **Not enough or no DNA/primer in sample tube.**
- **Inhibitory contaminant.** See possible contamination list at the end of this section.
- **Expired reagents.**
- **Incorrect thermocycling conditions.**
- **Extension products lost during column clean up.** Make sure sample doesn't run down side of the column and centrifuge at correct speed. Do not overload the cleanup columns.

Multiple peaks within sequence:

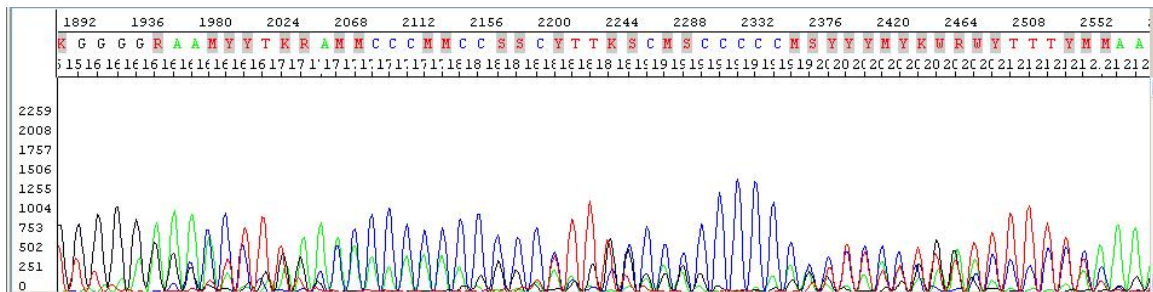


Possible Causes include:

- **Multiple Priming sites in sequence reaction.** Are there multiple bands on a gel? Optimize your PCR product. Are your primers annealing to multiple sites on a single gene or a single site on multiple genes? You may need to redesign your primers. Does the vector site exist in more than one location? Pick a different vector primer.
- **Primer acting as forward and reverse.** Redesign your PCR primers.

- **Residual PCR primers or primer dimers.** Incomplete removal of your PCR primers or primer dimers may allow for amplification of these primers during the sequencing reaction. Try cleaning with Guanidine HCL (protocol in Qiagen QiaQuick PCR Purification protocol).
- **Primer has high T_M .** These primers often do not perform well as sequencing primers. If redesign is not possible, try performing a two-step sequencing program (eliminate the annealing step).
- **Primer has n-1 population.** “Shadow” peaks (true peak plus the peak directly to the right of it) on your chromatogram may be a result of this.
- **Homopolymeric regions in sequence.** Try adjusting cycle sequencing conditions to higher annealing temperature and longer extension time or redesign the primer to be closer to homopolymeric region.
- **Mixed plasmid prep.**
- **Compression caused by secondary structure of DNA.** Alter cycle sequencing conditions or use additive (DMSO, betaine).
- **Frame Shift mutation.**

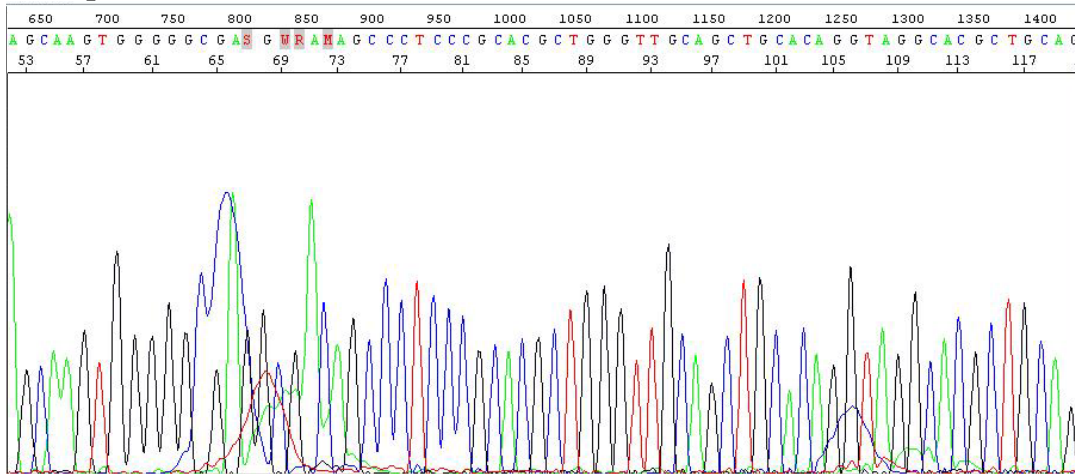
Sawtooth pattern:



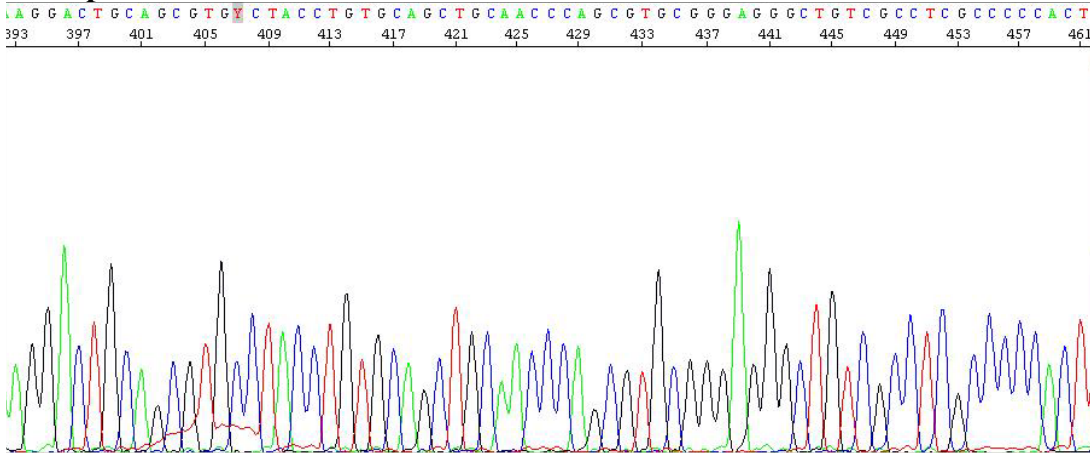
Although this example also falls under multiple peaks within the sequence, it is worth noting because it has a very recognizable pattern of homopolymeric repeats right next to each other throughout the entire sequenced region. The repeats are due to polymerase slippage. Sometimes the only way to resolve this is to sequence in the reverse direction.

Dye blob artifacts:

Example #1

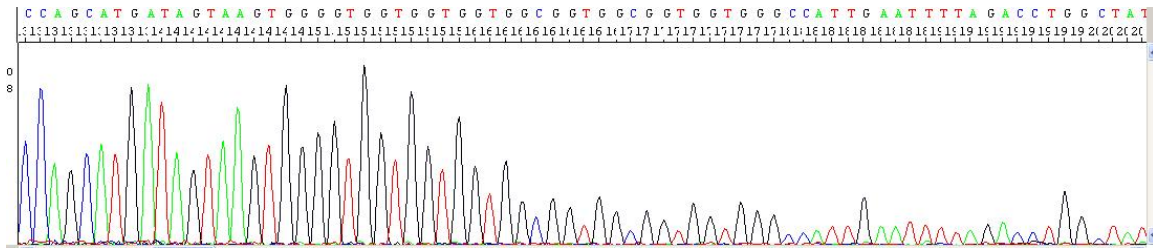


Example #2



- **Non-removal of excess dye terminators will cause this artifact.** Dye blobs run at 60-80 bp, with a small artifact at 100-115bp (sometimes this is very small and may not be noticed or interfere with the sequence). If the dye terminator contamination is severe you may also see peaks or shoulders at 170-190bp and again at 400-410bp (see above example #2). Samples can be re-column purified, then re-submitted. Be sure to load your sample on the center of the resin bead and spin according to the manufactures instructions.

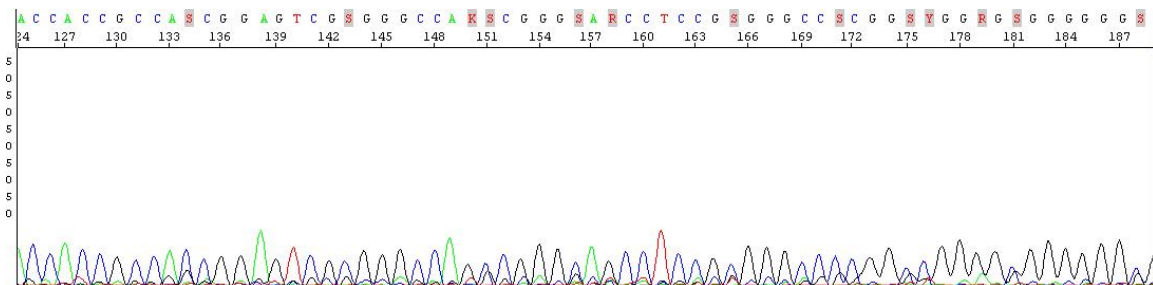
Truncated sequence:



Possible Causes:

- **Secondary structure** (G/C or A/T rich sequence). Try adding DMSO or betaine, increase denaturation temperature, or place primer closer to hairpin loop.
- **Too much template DNA.** See below under inaccurate primer concentration.
- **Inaccurate primer concentration.** Truncated sequences that abruptly end could be caused by the addition of too much primer or template DNA. Either will cause all the dNTP's to be used up in the first few hundred bases and are therefore not available to continue strand synthesis.
- **Salt contamination.** Salt has an inhibitory effect on the Taq polymerase. Perform ethanol precipitation.
- **Repetitive regions.** Same as secondary structure.

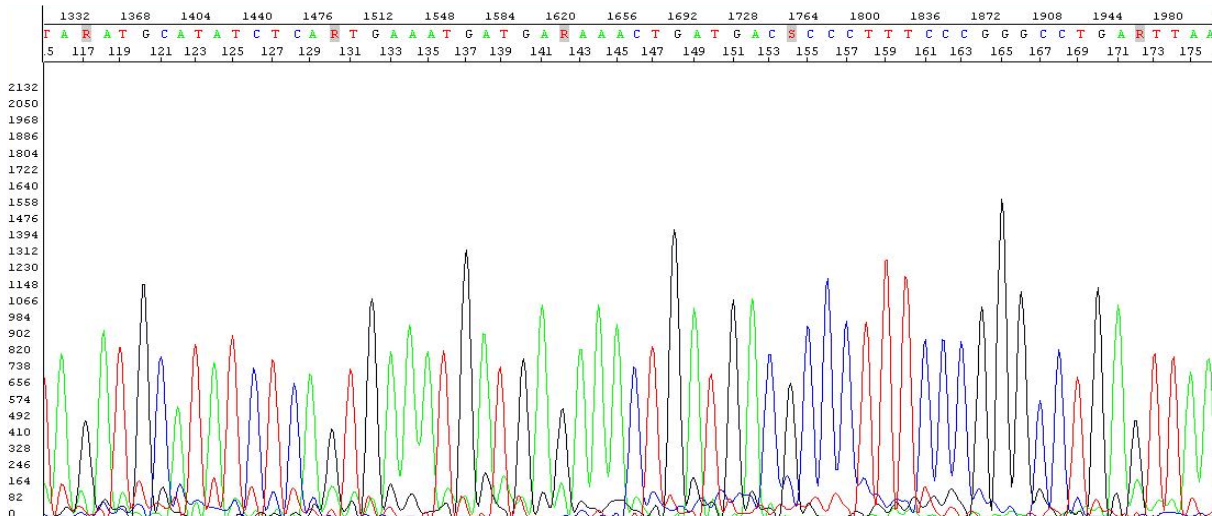
Noisy data with low S/N (>20):



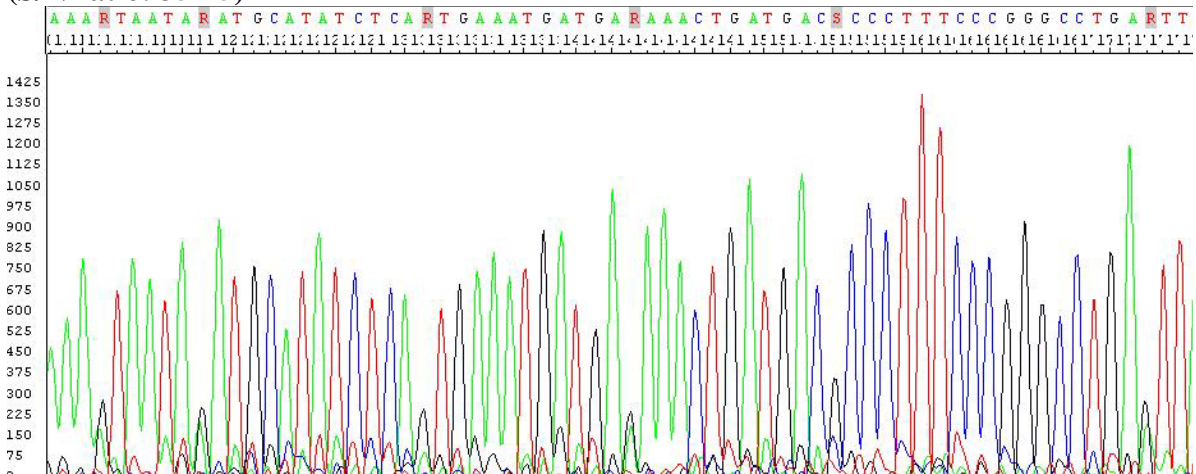
Possible Causes:

- **Not enough or degraded template DNA/primer.**
- **Expired reagents.**
- **Incorrect thermocycling conditions.**
- **Salt contamination.**

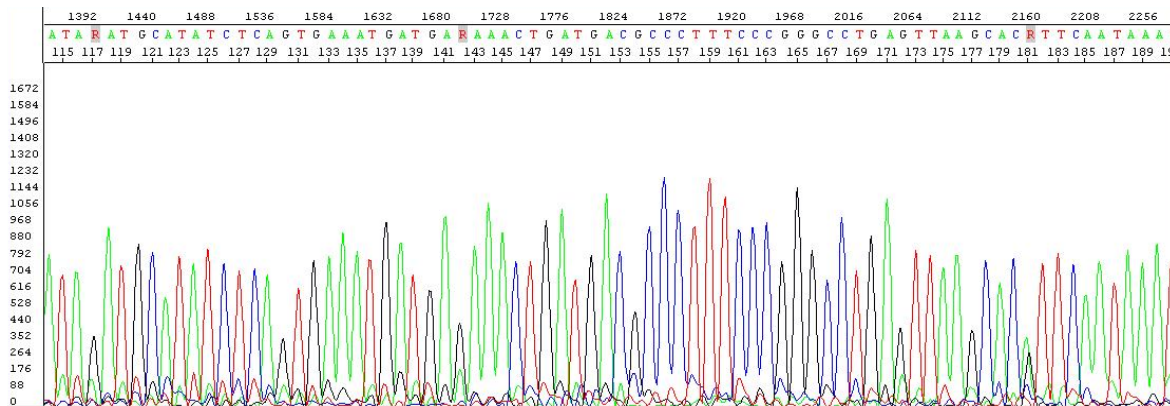
Three examples of noisy data with good S/N ratio (20-500):



(S/N ratio: 35-75)



(S/N ratio: 130-330)



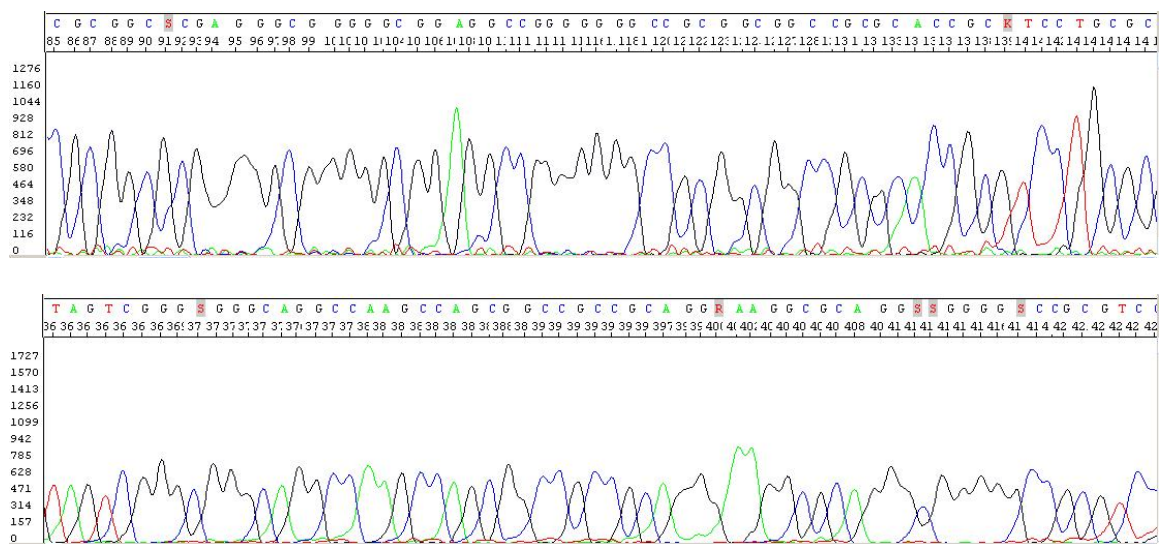
(S/N ratio: 280-550)

There is a wide range of good signal to noise ratios. As you can see from the above examples, over that wide range of good S/N ratio's, background problems can exist that interfere with proper base calling.

Possible Causes:

- **Inhibitory contaminate in template.** See contamination list at end of section.
- **Multiple templates in reaction.**
- **Primer carryover from PCR.** Try a cleanup with GnHCL
- **Multiple priming sites.**
- **N-1 primer contamination.** Request MALDI-TOF documentation from MWG
- **High signal saturating the detector.** Use less DNA in the sequence reaction.

Loss of Resolution:

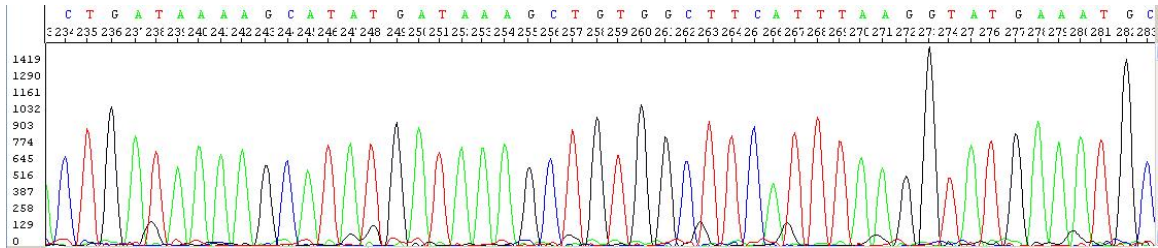


As you can see from the two examples above, loss of resolution is primary defined as peaks that become broad and not well resolved and progressively deteriorate through out the run. They may be wavy or have shoulders. Loss of resolution may start from bp 1 or somewhere in the middle of your sequence.

Possible cause:

- **High salt concentration.**
- **DNA concentration to high.**
- **Poor template quality.**
- **Unknown contaminant.** Other Core labs have speculated that a currently unknown contaminant may show up when columns that are used in some commercially available miniprep kits are heavily overloaded.
- **Retarded capillary.** Delaying electrophoresis may cause diffusion of the sequencing products resulting in loss of resolution. When loss of resolution is accompanied by a retarded capillary, the sample is re-run.

Breakdown of G (black) peaks:



- Expired Big Dye Terminator v3.1 mix.** The “G” peak is the first dye that is affected when the Big Dye chemistry starts to break down. This will show up as little “G” peaks under your sequencing reaction. ABI recommends that the BigDye Terminator v3.1 Cycle sequencing reaction mix undergo no more than 5-10 freeze thaw cycles. Aliquot the reagent in smaller amounts and store the kit at -15 to -25 °C.

IUPAC Table

IUPAC nucleotide code	Base
A	Adenine
C	Cytosine
G	Guanine
T (or U)	Thymine (or Uracil)
R	A or G
Y	C or T
S	G or C
W	A or T
K	G or T
M	A or C
B	C or G or T
D	A or G or T
H	A or C or T
V	A or C or G
N	any base

Contamination list:

CONTAMINANT	AMOUNT TOLERATED IN SEQUENCING REACTION
RNA	1 ug
PEG	0.3%
NaOAc	5-10mM
Ethanol	1.25%
Phenol	0%
CsCl	5mM
EDTA	0.25mM

Helpful References

- ❑ Automated DNA Sequencing Chemistry Guide
ABI PN 4305080
PDF available from BCL by request

- ❑ ABI Prism 3100 Genetic Analyzer Sequencing Chemistry Guide
ABI PN 4315831
PDF available from BCL by request

- ❑ ABI BigDye Terminator v3.1 Cycle Sequencing Kit Protocol Manual
ABI PN 4337035
BCL has a copy available for reference

- ❑ Roswell Park Cancer Institute Core Facility website
www.roswellpark.org/document.asp?lid=3632

- ❑ DSRG (DNA Sequencing Research Group) Troubleshooting Resource
Web page.
www.abrf.org
Click on References (tab located on left side of page)
Click on DNA Sequencing Troubleshooting Resource
You can search for your specific problem, or browse all.