390216-AB
July 2003

CEQ™ 8000 Series

Genetic Analysis System

Sequence Analysis Troubleshooting Guide

Beckman Coulter, Inc.
4300 North Harbor Boulevard, Fullerton, CA 92835

Copyright, Licenses and Trademarks

Table of Contents

Section I: General Diagnostic Guidelines

The CEQ System has many useful features that allow the user to identify and resolve problems quickly. The two most useful features in this regard are the **Raw Data** and the **Current Profile**. In certain cases, the analyzed data can be used, but ***only after*** the raw data and current profile have been shown to be good.

Raw Data - Good Raw Data Signal

The raw data generated by the CEQ System is displayed during the actual run and is also accessible to the user in the "Analysis Application." The major diagnostic to use when looking at raw data is the signal strength. In general, signal strength will gradually decline as a run progresses as shown below:

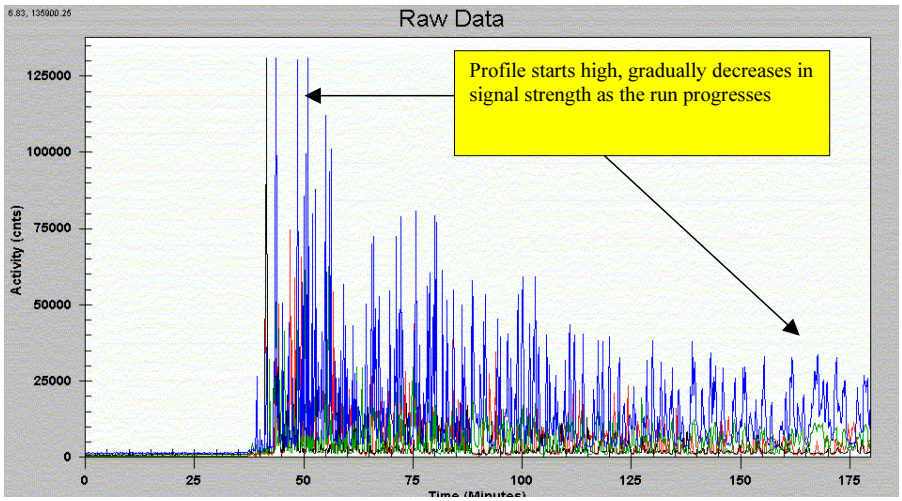


Figure 1: Good Raw Data Signal

In the example shown in Figure 1, with a start time of 20 minutes for the LFR-1 method, the signal strength at the beginning of the run is approximately 70,000 counts, decreasing to about 10,000 counts at the end of the run. Signal strength is important because without sufficient signal, it is very unlikely that accurate base calls can be made. The above example shows excellent signal strength; however, it is not necessary for each sample to have such high signal in order to obtain accurate base calls. The important number to remember is signal to noise ratio. With a signal to noise ratio of about 3:1, the software should give an accurate base call. For example, a profile with a starting signal strength of 25,000 counts and an ending signal strength of 5,000 (assuming a background of about 1,500) should still give an accurate base call.

Current Profiles – Normal Current Profile

The CEQ System continuously monitors current in all 8 capillaries while the system is running. This current trace can be extremely useful in diagnosing certain problems. The current profile should look similar to Figure 2.

The current ramps up to the final level in one stage. The final current level should be approximately 5–9µA. It is determined by the separation voltage, which is set in the method and is maintained throughout the run.

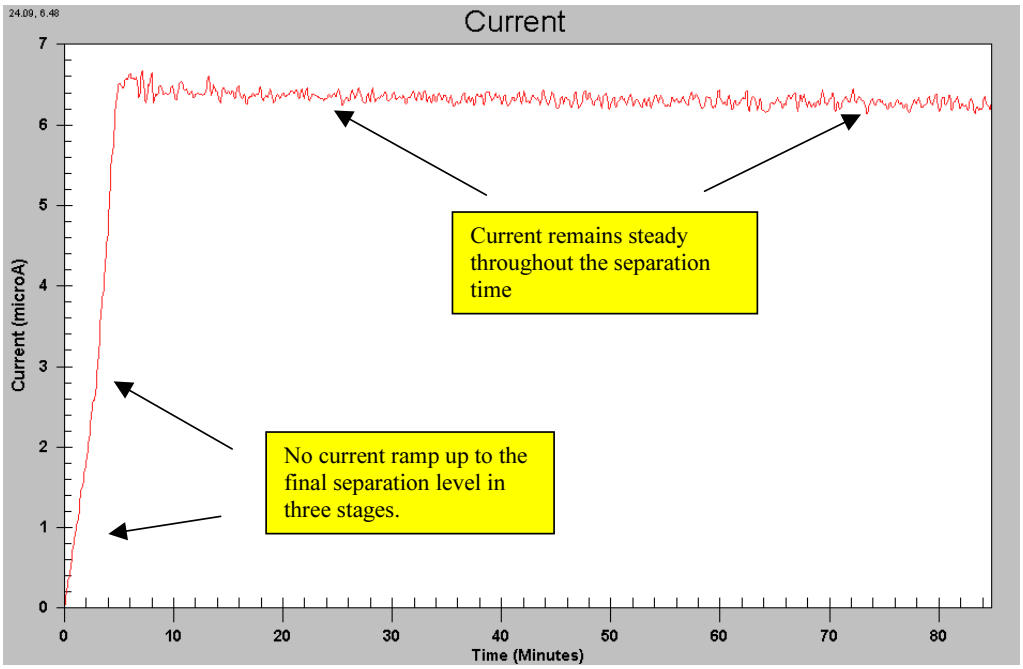


Figure 2: Typical LFR-1 Current Profile

The software allows for the development of custom separation methods. If a significantly higher separation voltage (6–12KV) is desired, then the advanced features of assigning the separation conditions should be employed. The advanced features allow for a gradual, stepped ramp to the final separation voltage. Such a gradual voltage increase allows the linear polyacrylamide (LPA) to slowly heat up and thereby reduces possible thermal expansion out of the capillary caused by the rapid exposure to high voltages. An example of such a three step ramp is given in Figure 3.

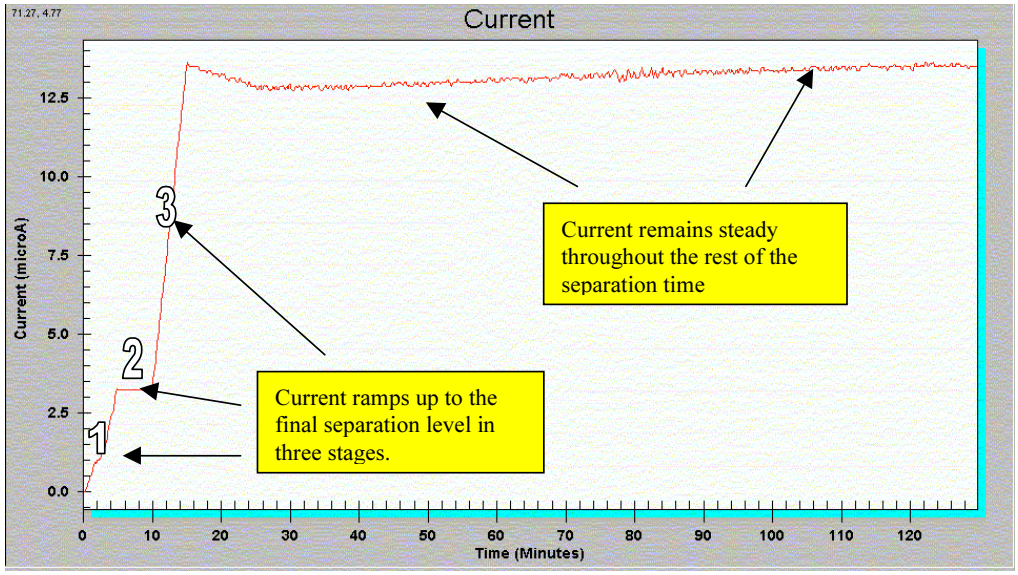


Figure 3: Example of a three step ramp

Section II: Troubleshooting a CEQ System Problem

Is it an instrument problem or a chemistry problem?

A systematic approach to troubleshooting should be adopted to effectively diagnose problems with the CEQ System. The first test that should be done on the CEQ System is to run the CEQ Sequencing Test Sample. Upon successful completion of CEQ Sequencing Test Sample analysis, sequencing of the pUC18 control DNA should be performed. Finally, customer samples should be sequenced. These three tests rely on the analysis of **Raw Data signal, Current profiles and Analyzed Data integrity**. Listed below are more detailed descriptions of the tests in the 3-step diagnostic process.

Diagnostics using CEQ Sequencing Test Sample

The CEQ Sequencing Test sample is used to confirm the integrity of the instrument and various reagents used by the system. This test uses a "pre-made" DNA sequencing reaction prepared using the same pUC18 DNA template as that provided in the CEQ 2000 DTCS kit. The CEQ Sequencing Test Sample should produce data that has a beginning fluorescence of approximately 50,000–120,000 units and then declines gradually until the end of the run (viewed in the Raw Data view of the Sequencing Analysis application). In addition, the current profiles should also be normal (see section "Current Profiles – Normal Current Profile" on page 2). Finally, the actual base calls are checked to see if the instrument passes the system specification of 98% accuracy at 700 bases (if using the LFR-1 method). You should perform the Service Alignment with the control template pUC18dG. If the CEQ Sequencing Test Sample provides acceptable results (good raw data signal, normal current profile and meets the system specifications of 98% accuracy at 700 bases), then the CEQ instrument and reagents are fine and further diagnostics of the sequencing chemistry are needed. If the CEQ Sequencing Test Sample does not yield an acceptable result try testing the other system consumables such as the separation gel, the capillary array and another lot of the CEQ Sequencing Test Sample. The individual components should be tested in separate experiments to identify the faulty component. If an acceptable result is still not achieved after substituting all new components call the Beckman Coulter Field Service Representative for service on the CEQ system.

Validation of the DTCS kit components

After determining that the CEQ System hardware, gel and capillary are functioning properly, prepare sequencing reactions with the pUC18 control template and the -47 sequencing primer provided in the DTCS kit. This test is designed to try to discriminate if the chemistry problems are due to the Beckman Coulter, Inc. provided reagents or customer supplied reagents. Results that pass the system specification of 98% accuracy at 700 bases (using the LFR-1) confirm the quality of the Beckman Coulter, Inc. reagents and the customer supplied materials except for primer and template. If the test does not pass the 98% accuracy at 700 bases (with good raw data signal and normal current profiles) perform the test again using a new DTCS kit (preferably of a different lot number). Failure to pass the 98% accuracy with a second kit would indicate that customer supplied reagents or the post sequencing reaction cleanup method are suspect. If the cause of the problem is customer supplied reagents proceed to the next section.

Suspect Customer Supplied Reagents

The customer uses several reagents that may have an impact on the raw data signal and the current profile. These consist mainly of the chemicals used for the ethanol precipitation of the finished sequencing reactions such as sodium acetate, EDTA and ethanol. In some cases with older lots of DTCS kits customer supplied formamide may also be an issue. The usual symptoms for poor quality sodium acetate, EDTA, ethanol or formamide are low raw data signal and occasionally erratic current profiles. Using sodium acetate and EDTA solutions supplied by Sigma as well as the Sample Loading Solution (SLS) supplied in the DTCS kits will help eliminate problems caused by poor quality reagents. If one of these reagents is suspect, replace each reagent in a systematic approach (one at a time) to identify the faulty reagent. If the control template passes the system specification of 98% accuracy at 700 bases (using the LFR-1 method) and the Raw Data Signal and Current profiles look good, but the customers templates do not yield acceptable results proceed to the next section.

Identification of Customer Supplied Template or Primer Problems

After demonstrating that pUC18 control template and the -47 primer function correctly, customer DNA templates should be sequenced. If possible, reagents should be tested individually. This will allow the most efficient identification of problem reagents. For example, if the customer is using a DNA template that can accommodate the -47 sequencing primer, then side by side comparisons of sequencing reactions performed with customer provided primer and the Beckman Coulter, Inc. provided primer can be conducted.

Using the previously described 3-step process, a CEQ System user should be able to identify and correct many of the most common problems found in CE based DNA sequencing. Listed below are several DNA sequencing chemistry problem areas with their corresponding symptoms and possible corrective actions.

Current problems

Electrokinetic Injection theory

Current problems are associated with the electrokinetic injection of sample and the subsequent separation of the DNA sequencing fragments in the Linear Polyacrylamide (LPA) matrix. The electrokinetic injection system will "load" negatively charged molecules on to the LPA matrix at the tip of the capillary. These negatively charged molecules include the dye-terminated sequencing fragments which are desirable, and molecules like the DNA template, proteins, salts, dNTPs and the unincorporated dye-terminators that are not desirable. Because the CEQ System uses electrokinetic injection to load DNA sequencing fragments on to the LPA in the capillary, it is susceptible to "overloading" problems. The most problematic of these molecules is the DNA template from the DTCS reaction. Listed below are examples of current problems caused by DNA templates, and to a much lesser extent proteins and other large, charged macromolecules.

Diagnosing Current Problems (What does the current profile look like?)

A gradually decreasing current

Supercoiled DNA templates, like plasmid DNA molecules, will load on to the LPA matrix in the "inlet side" of the capillary and impede the flow of current through the capillary. When the supercoiled DNA molecules load on to the LPA in the tip of the capillaries they become "plugged" and do not allow the free flow of electrical current through the capillary. Plugged capillaries exhibit two different current problems. In the less severe case where the current profile shows a gradual decline without any erratic fluctuations in the current, short reads are observed. These short reads are due to the general slowing of the electrophoretic separation of the DNA fragments causing fewer dye-labeled DNA fragments to pass through the detection window so fewer bases are separated. In addition, other macromolecules that co-purify with the DNA template can also cause current problems. These macromolecules include proteins, glycoproteins, bacterial chromosomal DNA, RNA and cell wall debris. In general, linear DNA molecules such as PCR products or restriction endonuclease digestion products do not cause current problems.

Corrective actions to solve minor current problems

A gradual decrease in the current throughout the run indicates that there may only be a slight excess of the template DNA in the sequencing reaction or the template is very highly supercoiled. There are several solutions to this problem (listed from simplest to most elaborate):

1. If sufficient raw data signal is present, often just reducing the injection time is adequate to improve the sequence call and read length. Simply rerunning the sample and injecting for half the time may solve the problem (i.e. if the first injection was 2kV for 15 seconds, use 2kV for 7.5 seconds).

2. If sufficient raw data signal is present, reduce the amount of template present in the cycle sequencing reaction. In this case, just adding half the amount of template may eliminate the current problem (i.e. if the original template amount was 100 femtomoles (fmoles), use 50 fmoles).
3. Use the preheat treatment of the template DNA prior to the cycle sequencing reaction. The preheat treatment (see CEQ DTCS Chemistry Guide, Beckman Coulter, Inc. P/N 718119), nicks the template DNA which alleviates the supercoiling of the plasmid DNA. As noted above, supercoiled DNA templates are the primary cause of current problems.
4. Use combinations of 1, 2 and 3. Using various combinations of the first three recommendations may also eliminate current problems. For example, use the preheat treatment on 50 fmoles of template.
5. Use a DNA template purification process that eliminates macromolecules such as proteins, glycoproteins, bacterial chromosomal DNA, RNA and cell wall debris. Most commercial DNA preparation methods yield template DNA that is essentially free of contaminating macromolecules.
6. Remove the template from the sequencing reaction. This is a very elaborate process requiring special primers and reagents, but also yields sequencing reactions that produce no current problems. The most common technique is to use a biotinylated sequencing primer and the DynaPureII™ sequencing reaction cleanup kit (Dynal, Inc. Oslo, Norway, product number 603.05) to purify sequencing reaction products from contaminants such as the template DNA, unincorporated dye-terminators, dNTPs and other salts.

An Erratic current or “Crashed Current”

In the more severe case where the current profile is erratic, the passage of dye-labeled DNA fragments through the detection window does not pass at a constant rate and sequencing errors are introduced. The sequencing software expects the fluorescent molecules to be detected at a fairly constant rate, this makes predictions of the base spacing in the base calling process. If the peak spacing is “expanded” (too few bases per minute) as found in regions were the current has dropped quickly, the software may insert extra bases into the sequence. Shown below are examples of erratic current and crashed current, respectively.

If the current is unstable or erratic, then the migration times of the bases will be adversely affected and this can lead to inaccuracies in the base calling. During the time that the current is not being maintained, as shown in Figure 4, the base migration will be slowed and the accuracy of base calling will be low. In addition to inaccuracies in the base call the total read length will also be diminished simply due to the fact that the migration of the bases has been slowed from the normal run conditions.

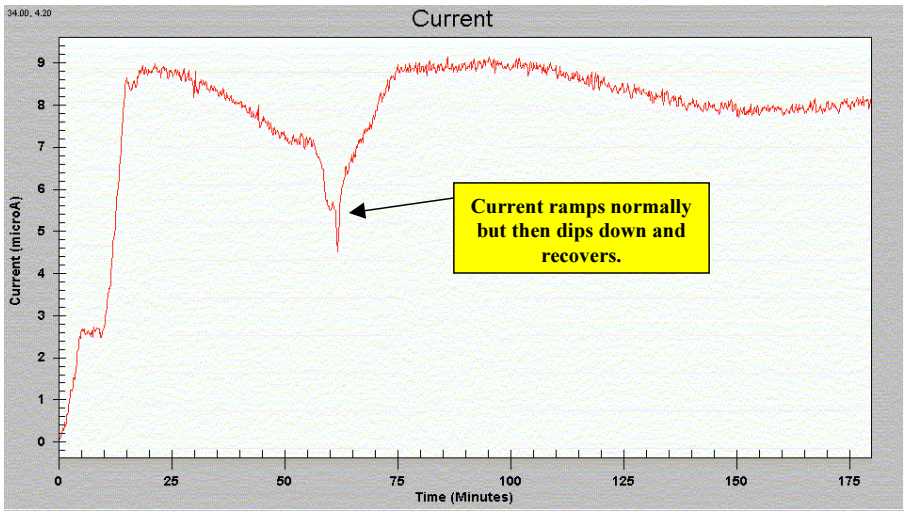


Figure 4: Erratic Current

In other cases, current failures can be more significant. Figure 5 shows a current profile that fails to reach the final running current after the normal ramping. Note that the current never reaches the normal running current of 5–9 µAmps and also declines rapidly to less than 2 µAmps.

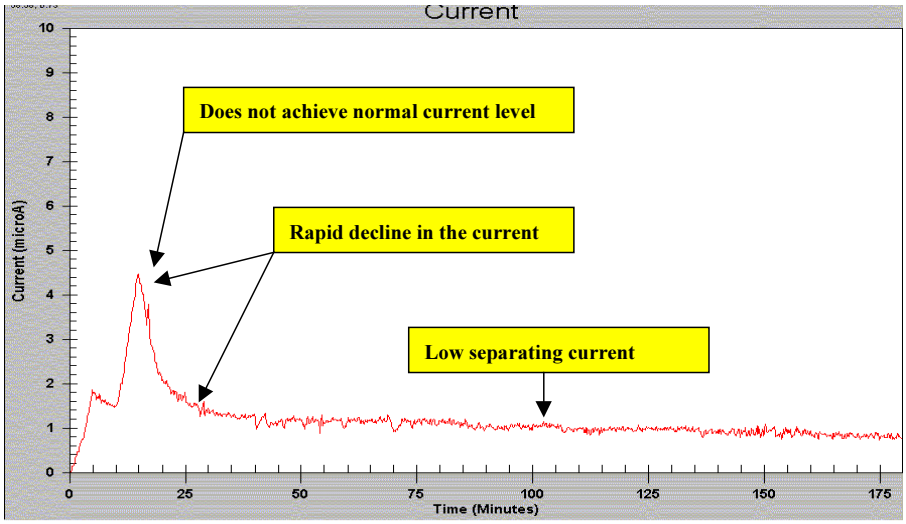


Figure 5: Crashed Current

Corrective actions to solve erratic current or current crashes

The occurrence of erratic current profiles indicates that there is a more severe DNA template-overloading problem. In this case more drastic measures may need to be taken. In the situation were erratic current is observed, simply rerunning the sample and injecting for half the time will not solve the problem. Listed below are some of the steps the should be made to eliminate the erratic current problem:

1. If sufficient raw data signal is present, reducing the amount of template present in the cycle sequencing reaction is an obvious first step. In this case, adding half the amount of template may eliminate the current problem (i.e. if the original template amount was 100 fmoles, use 50 fmoles).
2. Use of the preheat treatment of the template DNA prior to the cycle sequencing reaction is highly recommended for supercoiled plasmid samples and is required to solve erratic current problems or current crashes. Once again, the preheat treatment (see CEQ DTCS Chemistry, Beckman Coulter, Inc. P/N 718119) nicks the template DNA to alleviate the supercoiling of the plasmid DNA and nicked DNA does not “clog” the capillaries during injection. In many cases the preheat treatment of plasmid templates at 96°C for 1 minute prior to adding the template to the sequencing reaction can eliminate current irregularities (see Figure 6).

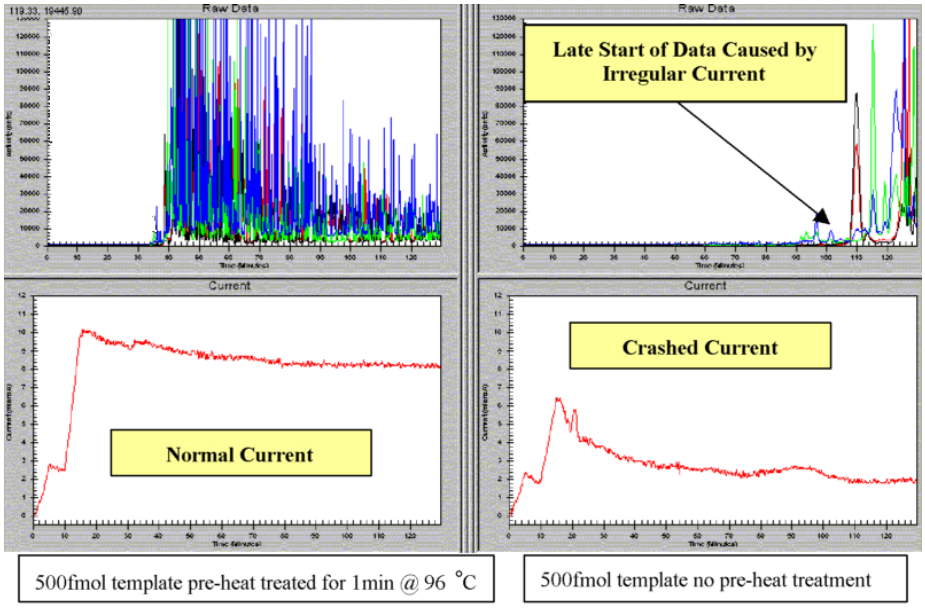


Figure 6: Preheat treatment of plasmid templates

3. An alternative to preheat treatment is to cleave the plasmid with a restriction endonuclease before adding the template to the sequencing reaction. In effect, this method converts a supercoiled plasmid into a linear molecule, which causes the DNA template to perform like a PCR product (no current problems observed). This method, although very effective, requires some knowledge of the DNA sequence to know where to "cut" the plasmid DNA. The plasmid needs to be cut at a position that is 5' to the sequence that is desired. For example if the primer binds to the plasmid at position 100 and sequence from position 100 to 600 is desired then the restriction endonuclease recognition site should be at position 601 or greater.
4. Removing the template from the sequencing reaction is the only guaranteed method for curing the current instability problem. This is a very elaborate process requiring special primers and reagents, but also yields sequencing reactions that produce no current problems. The most common technique is to use a biotinylated sequencing primer and the DynaPureII™ sequencing reaction cleanup kit (Dynal, Inc., Oslo, Norway, product number 603.05). This purifies the sequencing reaction products from contaminants such as the template DNA, unincorporated dye-terminators, dNTPs and other salts.

An Identical Erratic current in all eight Capillaries

If all eight capillaries have identical erratic current problems, then the cause of the problem is a block in the electrical path on the "manifold side" of the CEQ gel system. A bubble forming in the electrode chamber which blocks the flow of current through the system most often causes this problem. The result of the blocked current flow is that all eight capillaries will have reduced read lengths and probable base calling errors. This is also an indication that there is a system problem and not a problem with the individual samples. The bubbles may be caused by a variety of reasons: excessive removal and installation of gel cartridges, improper sealing of the manifold plug of the capillary array in the gel manifold gasket, a cracked electrode housing, etc.

Shown in Figure 7 are the current profiles of a set of capillaries with an air bubble on the manifold side of the CEQ gel system.

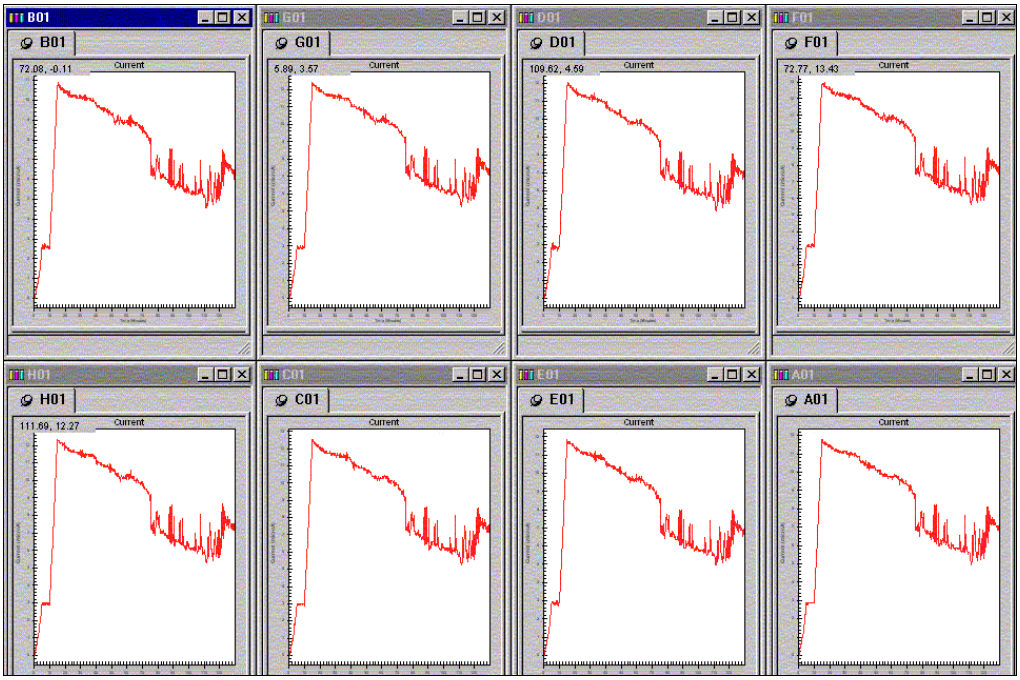


Figure 7: Manifold bubble current problem

Corrective actions to solve manifold bubble current problems

1. Do 1–2 manifold purges from the Direct Control menu on the CEQ Run application.
2. Check the sample plate for air bubbles before loading it onto the CEQ.
3. Call Beckman Coulter, Inc. service if the first two actions do not fix the problem.

Raw Signal Problems

Types of Raw data signal problems

There are several types of raw data signal problems. The most common problem is signal levels that are too low. In this case there is insufficient fluorescence signal (shown as Activity (cnts) on the electropherogram) to generate an accurate base call. Another possible problem is when there is too much raw data signal, which causes the photomultiplier detection system to be saturated. This is caused by too many fluorescently labeled DNA sequencing fragments. Another category of raw data signal problem is too many fluorescent peaks in the electropherogram. This can be caused by two independent sequencing events present in the same sample. This phenomenon is caused by either having two or more templates in the reaction or multiple (2 or more) primer binding sites on a DNA template. A variation on the previous problem is referred to as a "pre-peak problem". Prepeaks are caused by poor quality primer synthesis where a high proportion of n-1 primer (primers which are one base short of the desired length) are present. The previously mentioned problems, as well as several other possible problems, will be discussed in more detail in the following sections.

Diagnosing signal level problems

The first step in diagnosing a Raw Data signal problem is to look at the fluorescent signal (Raw Data signal) in the Sequencing Analysis application. The data, which is plotted as Activity (cnts) versus Time (Minutes), should show fluorescent signal (Activity cnts) in the range of 5,000–125,000 cnts. The signal will be high in the beginning and gradually decline throughout the run.

Causes of Low Raw Data Signal

Low Raw Data Signal due to not enough template

Low raw data signal can be caused by a variety of issues. One of the most common causes is lack of sufficient DNA template in the cycle sequencing reaction. It is vital to the success of CE sequencing to have the correct amount of template in the reaction. It is recommended that 50–100 fmoles of plasmid DNA be used in the cycle sequencing reaction. This provides enough template to generate an adequate amount of fluorescently labeled DNA sequencing fragments yet not so much as to cause current problems (see "Current problems" on page 6). Half this amount of DNA template (25–50 fmoles) should be used for single stranded DNA templates such as M13 phage DNA and even less DNA is needed for small PCR products (10–50 fmoles for PCR products less than 3KB in length). In many cases the amount of template added to the reaction is not determined and therefore, insufficient template is present. In other cases, an incorrect approximation of the DNA concentration is made. Spectrophotometric estimation of DNA samples is only valid if the DNA is pure (as in the case of commercial DNA template purification methods). Crude preparation of DNA templates which have substantial amounts of protein and/or RNA will over estimate the concentration of the template and cause the user to add too little DNA to the sequencing reaction (as in the case of crude alkaline lysis minipreps).

Shown, in Figure 8, is an example of a DNA sequencing sample where not enough template DNA was added. In this example less than 25 fmoles of template was added to the sequencing reaction. Notice the very low Raw Data signal throughout the run and particularly the low Raw Data signal in the red, green and black traces. An easily identifiable indicator of low template concentration and, in general, a poor sequencing reaction is a high amount of unincorporated dye-terminators (note the large peaks at around 40 minutes).

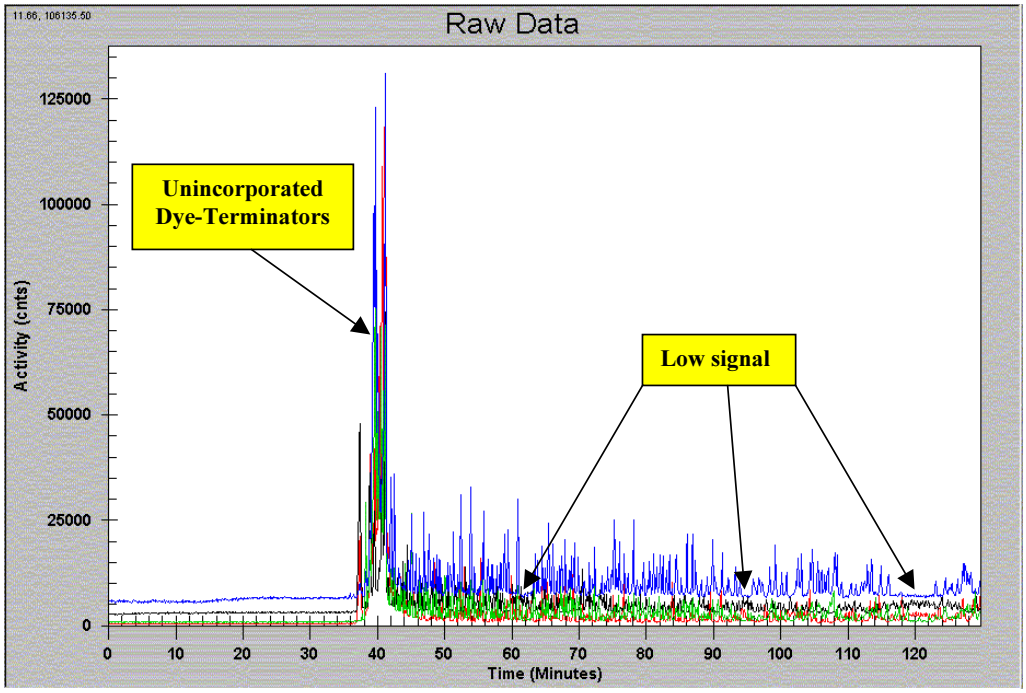


Figure 8: Low Template amount

Corrective actions

1. Add the correct amount of the template DNA to the reaction. This will require quantitation of the template DNA by spectrophotometry (in the case of commercial DNA preparations) or by estimation using agarose gel electrophoresis and comparison to a know quantity of DNA.
2. Alternatively, the user could try a dilution series with the same template starting with an amount that is obviously too high and ending with an amount which is much too low. This method assumes that the user knows the approximate amount of template added to the reaction (this may be from previous work using similar DNA preparation methods).
3. Use the preheat treatment for highly supercoiled plasmids. Most commercial DNA preparations yield highly supercoiled plasmids. The preheat treatment will knick the supercoiled plasmid which yields much more efficient DNA sequencing (linear molecules sequence better than supercoiled molecules).

4. If the correct amount of template was added and the preheat treatment does not yield a substantial increase in Raw Data signal increase the number of cycles in the thermal cycling program from 30 to 40 or 50.
5. If the correct amount of template was added and the Preheat Treatment and / or increasing cycle number does not yield a substantial increase in Raw Data signal, a DNA Polymerase inhibitor may be present (do not resuspend DNA in DEPC treated water). In this situation further purification of the DNA template may be required. In some cases a simple ethanol precipitation of the plasmid will remove the inhibitor, whereas other situations may require the use of commercial DNA preparation methods such as the Qiagen QiaQuick kit.

Low Raw Data Signal Due to Primer Quantity and Quality Problems

The amount and purity of sequencing primers can also cause low raw data signal. For example, having insufficient quantities of the primer or having a poor quality primer (e.g. a primer that has not been totally deprotected after synthesis) will yield low Raw Data signal. Many suppliers of oligonucleotides provide sufficiently pure preparations of primers. Most of these vendors will provide “desalted primers” which function well for DNA sequencing. However, further purification of the primers by TSP (Trityl Specific Purification) or HPLC (High Performance Liquid Chromatography) is highly recommended when the highest quality DNA sequencing results are desired. Also, primers purified by denaturing gel electrophoresis have been shown to contain inhibitors of the DNA Polymerase and therefore, are not recommend for use in DNA sequencing reactions.

Primer Design Problems that Cause Low Raw Data Signal

Another cause of primer related low raw data signal is poor primer design. Primers should be designed so that they have reasonably high T_m’s (T_m greater than 55°C), no internal secondary structures (such as hairpin loops) and a low probability of forming “primer dimers” (the 3' ends of the primers binding to each other and are extended). The T_m of the primer must be at least 5°C higher than the annealing temperature of the cycle sequencing program (50°C) to adequately bind to the template DNA and be extended by the polymerase. Internal secondary structures and primer dimers are thermodynamically favored over primer- template binding so they reduce the effective concentration of the primer in the sequencing reaction.

In the following examples (Figure 9 and Figure 10), the template is the same, but two different primers have been used. Primer 1 has a T_m of 43.5°C, a 31.6% GC content and is 19 bases long. Primer 2 has a T_m of 79.8°C, a 62.5% GC content and is 24 bases long. The effect on the data obtained from the CEQ is clear. With a less than efficient reaction, the signal level is low and the base calling accuracy is not optimal.

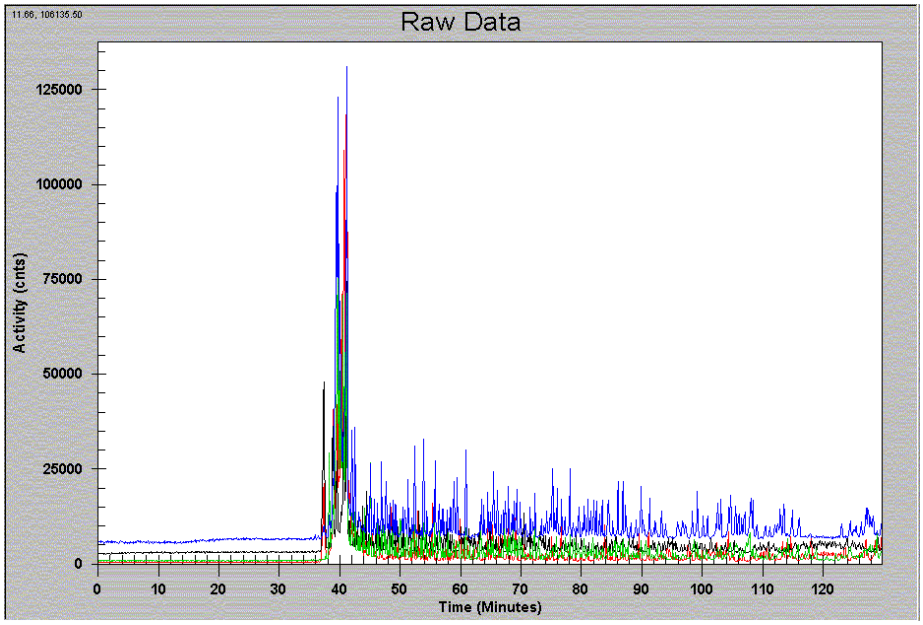


Figure 9: Primer 1

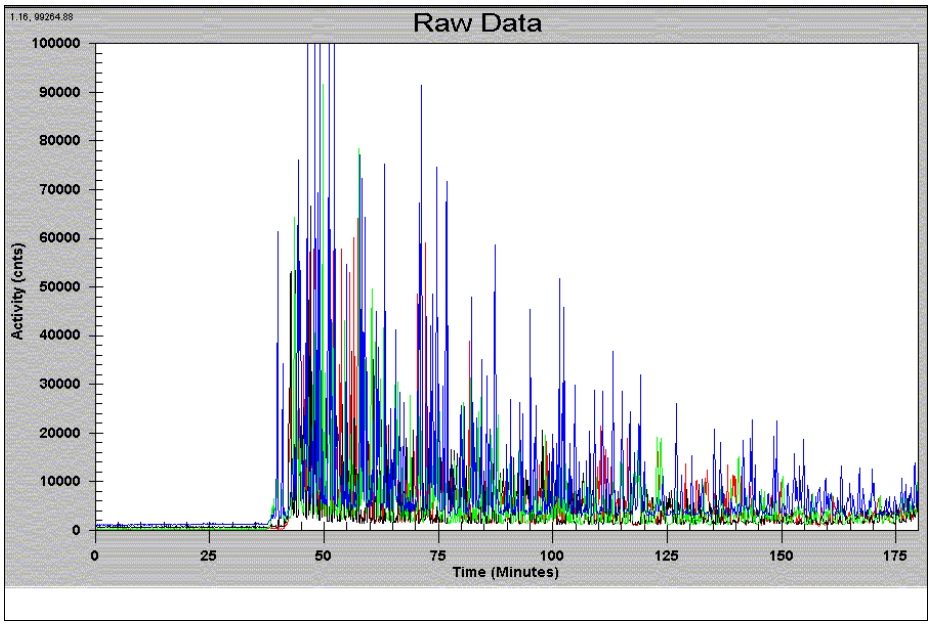


Figure 10: Primer 2

Corrective Actions

1. Add the correct amount of the primer to the reaction. This will require quantitation of the primer by spectrophotometry. The recommended amount of primer is at least 40-fold molar excess in relation to the template. For sequencing of most template DNAs, 3–10 pmoles of primer is sufficient for a successful sequencing reaction.
2. Make sure that the primer preparation is pure. Using reputable oligonucleotide vendors is your best way to get high quality primers. If possible have the primers purified by TSP or HPLC.
3. Always design the sequencing primers for use in Cycle sequencing reactions. This means designing primers with high T_m's and no possibility of forming secondary structures or primer dimers. There are many DNA analysis programs available for checking primer design.
4. Make sure that the primer is resuspended in a solution that has no EDTA or DEPC present. Molecular biology grade water or 10mM Tris pH 8.0 are good resuspension solutions.
5. Increase the number of cycles in the thermal cycling program from 30 to 40 or 50.

Low Raw Data Signal Due to "Bad Formamide"

Formamide is used to resuspend the DNA sequencing fragment prior to loading on to the CEQ. The formamide solution must be prepared and stored properly to achieve high quality sequencing data. If the formamide is not deionized and stored properly it will decompose into ammonia and formic acid. The formic acid then destroys the fluorescent dyes and produces low Raw Data signal. Shown in Figure 11 is a sample that was resuspended in formamide that had not been deionized. Note that the resulting base calling accuracy for this sample was less than 90% (>40 errors) at 500 bases.

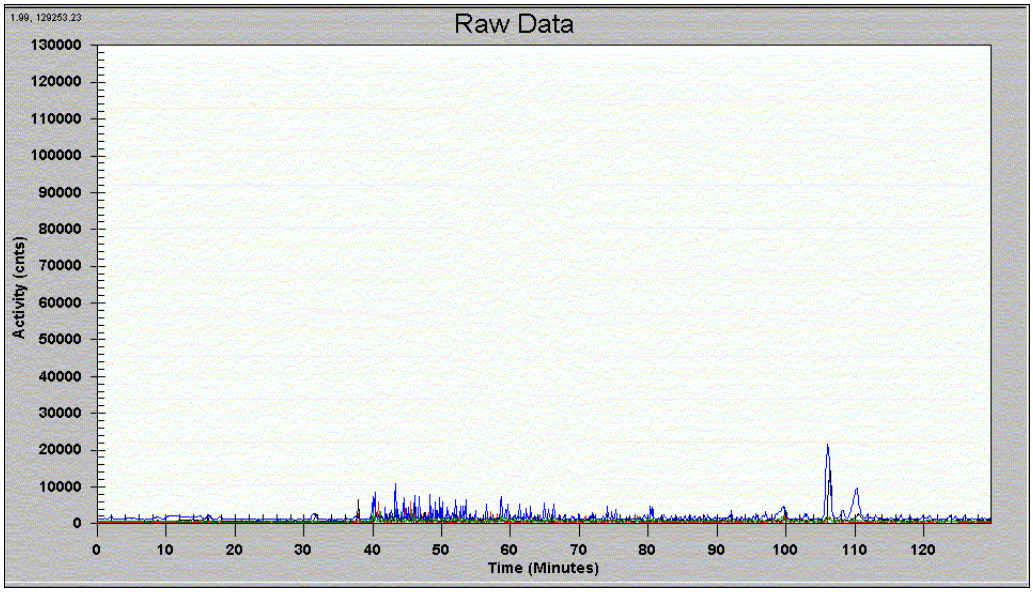


Figure 11: Sample resuspended in formamide that has not been deionized

Corrective actions

1. Use the Beckman Coulter, Inc. supplied Sample Loading Solution (SLS).
2. Do not freeze-thaw the SLS or formamide solutions. Store aliquots at –20°C in a non-frost free freezer and use the aliquots only once.

Low Raw Data Signal Due to the Use of Water as the Sample Loading Solution

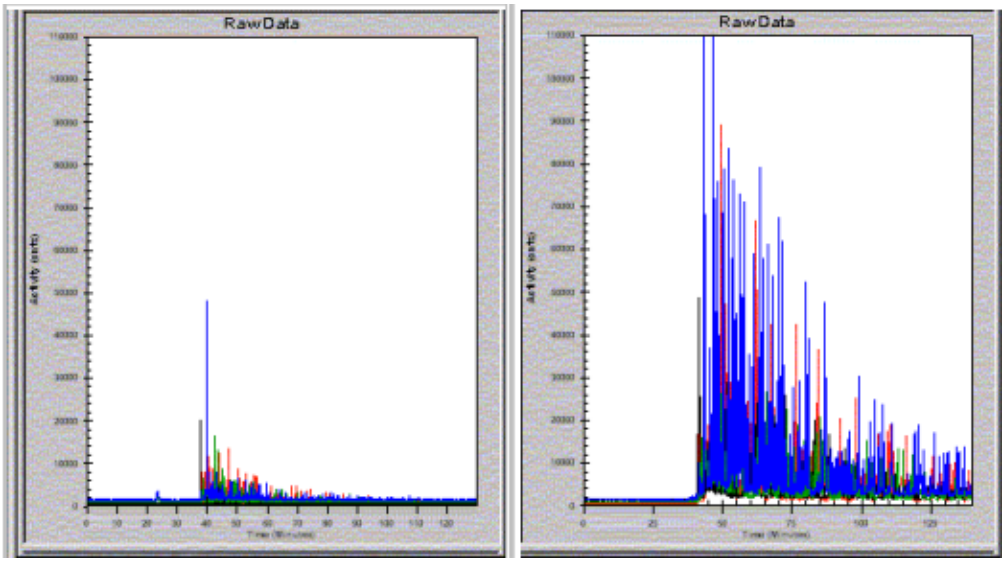
We do ***not*** recommend using water to resuspend the DNA sequencing fragments prior to loading on the CEQ. The Beckman Coulter, Inc. WellRED™ dyes are not stable in pure water solutions and will yield Raw Data signals similar to that of "Bad Formamide" (see prior section).

Corrective actions

1. Use the Beckman Coulter, Inc. supplied Sample Loading Solution (SLS).
2. Do not freeze-thaw the SLS or formamide solutions. Store aliquots at –20°C and use the aliquots only once.

Low Raw Data Signal Due to Insufficient Sample Injection

Poor injection of DNA sequencing fragments onto the CEQ capillaries will lead to low Raw Data signal. Since the CEQ uses electrokinetic injection it is highly sensitive to excess salts in the loading solution. The excess salts compete with the DNA sequencing fragments during injection and result in lower loading of the fluorescently labeled DNA molecules. The sources of the excess salts are improperly purified sequencing reactions and decomposed formamide.



Sequencing Reaction with excess salts

Sequencing Reaction with no impurities

Figure 12: Sequencing reaction

Corrective actions

1. Follow the ethanol precipitation procedure in the CEQ Dye Terminator Cycle Sequencing protocol (Beckman Coulter, Inc. P/N 608019).
2. If using spin column purification methods make sure that the column materials do not contain salts (check with the spin column manufacturer for details for using their products with capillary sequencers).
3. Use Beckman Coulter, Inc. supplied SLS for sample resuspension.

Low Raw Data Signal Caused by Poor Quality Mineral Oil

The mineral oil supplied in the DTCS kit is high quality oil containing no detectable nuclease activity. The use of other lower quality mineral oils can lead to sample degradation and hence low signal as shown below. The red and black dyes are particularly susceptible to this problem.

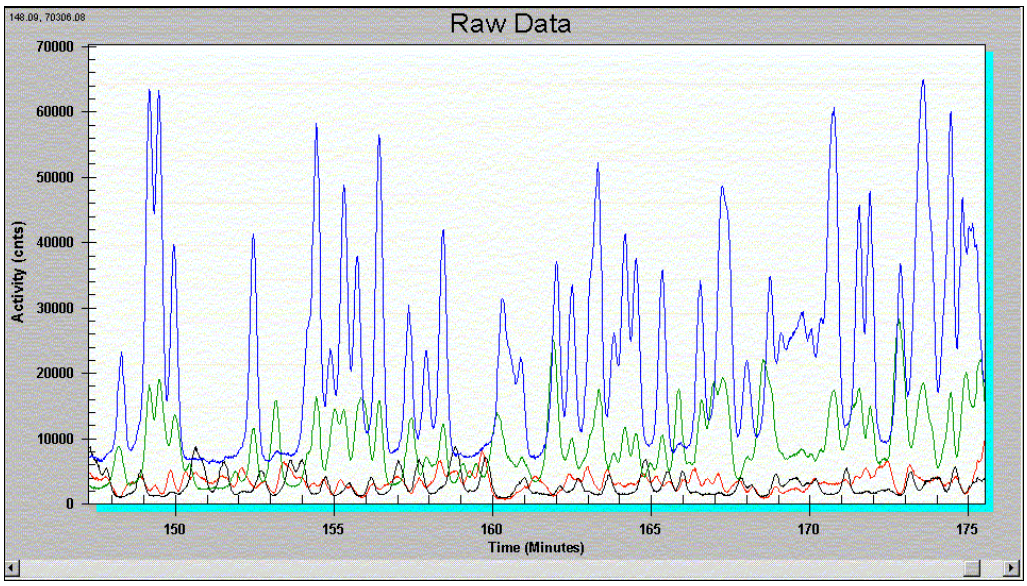


Figure 13: Sample degradation resulting from low quality mineral oils

Corrective actions

1. Repeat the sequencing reactions and use the mineral oil supplied in the Beckman Coulter, Inc. CEQ DTCS kits.
2. Repeat the sequencing reactions and use molecular biology grade mineral oil from Sigma Chemical Company (P/N M-5904)

Low Raw Data Signal Caused by Thermal Cycling Problems

The CEQ DTCS chemistry uses thermal cycling to produce a linear amplification process (called cycle sequencing) to generate sufficient fluorescent signal to generate an accurate sequencing base call. The Thermal Cycler program and performance are critical to achieve optimal results. There are several possible problems that are associated with the thermal cycling process. These include:

- Using the wrong thermal cycler or a poorly performing Thermal Cycler
- Using the wrong ramping setting on a thermal cycler
- Using an incorrect annealing temperature for the primer being used for sequencing
- Using a tube or plate that does not properly fit the thermal cycler
- Using a thermal cycler without a heated lid
- Loose fitting strip caps, crimped sealing film or sealing mats

Any one of the previously mentioned problems will yield low Raw Data signal similar to that shown in Figure 14.

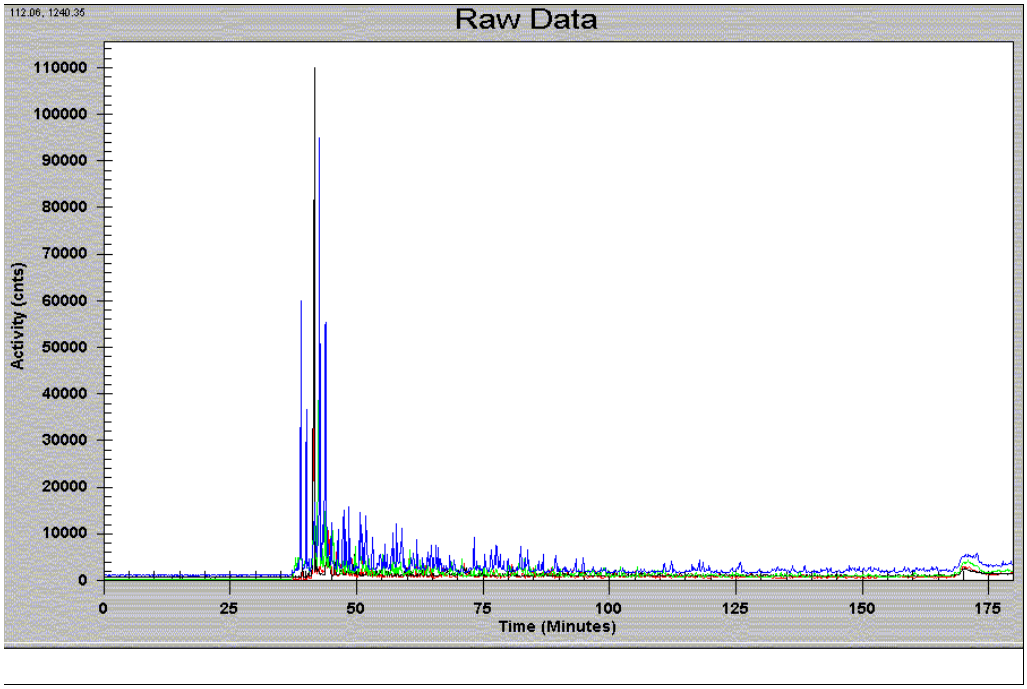


Figure 14: Low Raw Data signal

Corrective Actions

1. Check thermal cycler for correct performance. If you are using a robocycler optimization will be required.
2. Make sure the temperature ramp is set to maximum.
3. Check primer melting temperature, if the annealing temperature is too high for the primer you will need to lower it.
4. Make sure you are using PCR tubes or plates. The CEQ sample plate has been proven to work on the MJ PTC-200 and P/E 9600 thermal cyclers.
5. Make sure strip caps, A sealing film, or sealing mat are put on the PCR plate correctly. The lack of a tight seal will lead to evaporation of water from the tubes / plate which will change the buffer concentration in the reaction mixture.

Raw Data Signal that is Too High

In a few cases, the signal strength can be so high that it saturates the detector. This can lead to an erroneous base call where the software will artificially estimate peak height and position. In this case, the software inserts extra bases into the base sequence. By setting the raw data to full scale (137,000 counts) and looking at the peak shapes the user can determine if peaks are "over-ranged". If the peaks are "squared-off" at the top (see the blue peaks in Figure 15), then the detector is saturated and the peaks are "over-ranged".

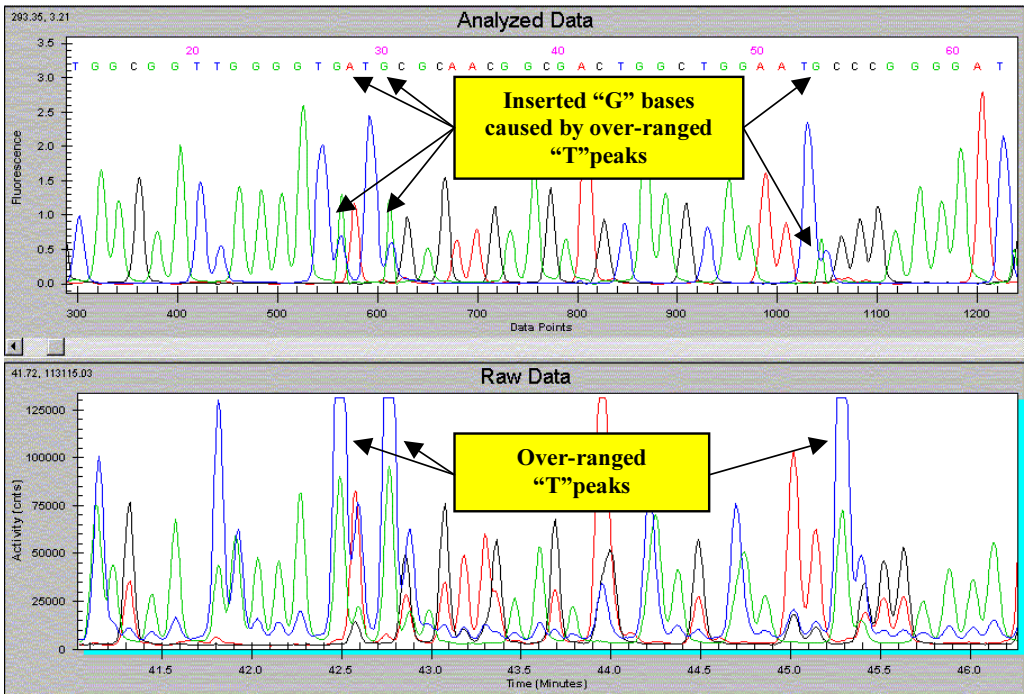


Figure 15: Over-ranged Peaks Causing Inserted Bases

Corrective Action:

1. If the peaks are too high, the simplest solution is to rerun the same sample using a shorter injection time (for example: 7.5 seconds instead of 15 seconds).
2. Use less template DNA or less thermal cycles to decrease the amount of fluorescence signal generated by the sequencing reaction.

Insertions Caused by n-1 primers

The presence of a significant percentage of n-1 primer in a primer preparation can lead to small peaks preceding the true sequence peaks. In certain cases these small peaks will be called as bases in the sequence. This will most often occur in a sequence that has a “G” followed by an “A”. However multiplets (i.e. TT or CC) may also yield insertions. In Figure 16, the presence of a large amount of n-1 primer (~20%) in the sequencing primer causes each true sequencing peak to be preceded by a smaller prepeak. As can be seen in Figure 16, in many cases, these prepeaks are called as bases.

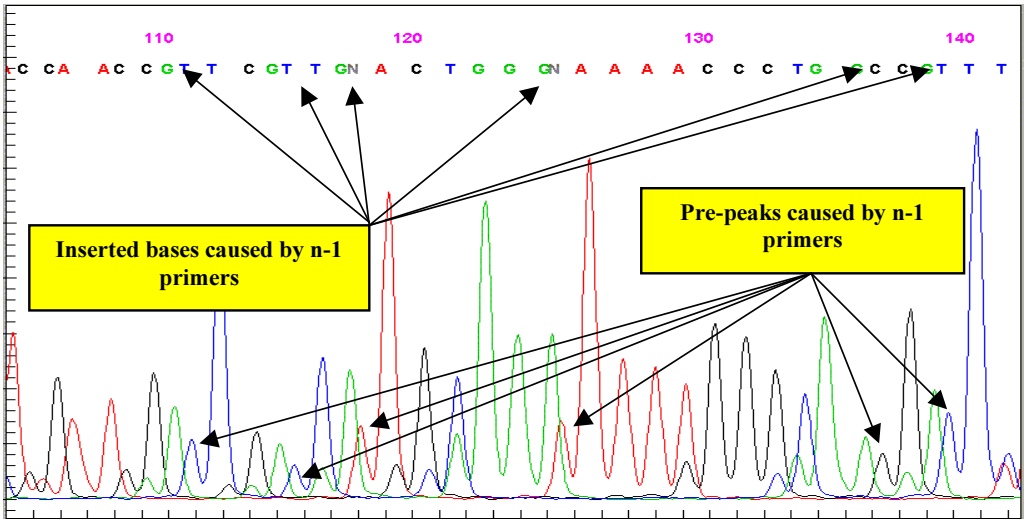


Figure 16: High amount of n-1 primer

Note that the characteristic incorporation effects of thermally stable DNA Polymerase and the WellRED™ dyes can be helpful in diagnosing prepeaks from true peaks. For example; a T followed by a T always has a much larger signal for the first T than the second T (Figure 17). A small T followed by a large T therefore, indicates that the small T is a prepeak and not the correct sequence.

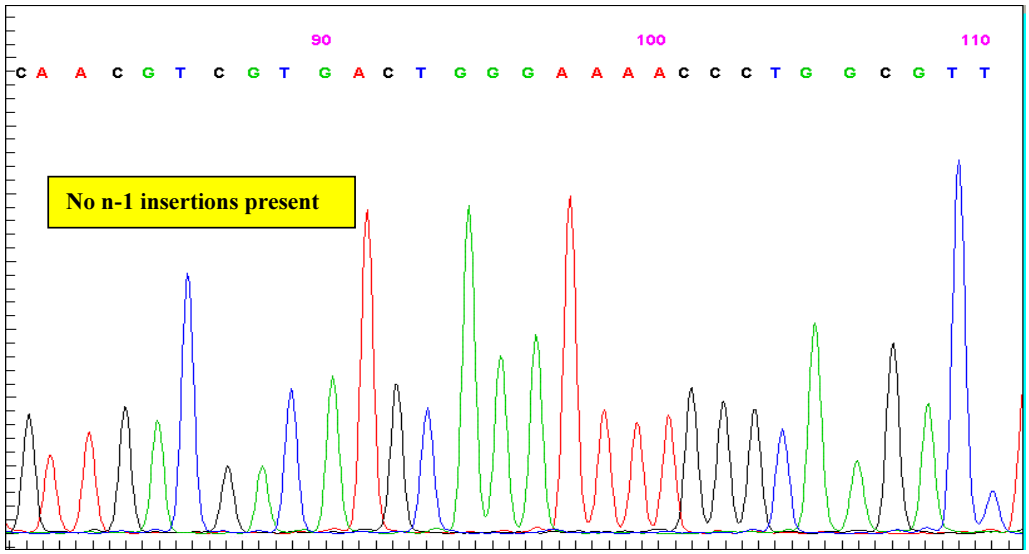


Figure 17: No n-1 primer insertions

Corrective Actions:

The only way to rectify this problem is to obtain high quality oligonucleotides for use as sequencing primers.

1. If you obtain sequencing primers from an external source (oligo vendor, core facility, etc.), request that your oligonucleotide supplier use the highest quality DNA synthesis conditions. Also note that oligonucleotide purification methods that rely on chromatography can not completely separate the n-1 from full-length oligonucleotides and therefore can not “rescue” a contaminated primer preparation.
2. If you synthesize sequencing primers using an automated DNA synthesizer, use the “High quality” settings on your instrument. This usually involves using increased amount of “capping” reagents and chemical coupling times. As is noted above, oligonucleotide purification methods that rely on chromatography can not completely separate the n-1 from full-length oligonucleotides and therefore can not “rescue” a contaminated primer preparation.
3. If you have a software version of 3.0 or higher, check the “pre-peak removal” check box in “Analysis Parameters” upon reanalysis of data. This will help to alleviate base call insertions due to n-1 primers. Although, it may not correct for all possible insertions if the sequencing primer is too greatly contaminated with n-1 products.

Insertions Caused by Mixed Templates, PCR Primer Carryover or Primer Mis-priming

In many instances poor sequencing results are the cause of multiple sequencing events occurring in one sequencing reaction. These multiple sequencing events are usually caused by either mixed templates (2 or more unique templates in a reaction), PCR primer carryover (sequencing reactions primed by the PCR primer in addition to the sequencing primer) or mis-priming of the sequencing primer (the sequencing primer binding to 2 or more sites on the template DNA). All three types of multiple sequencing events have similar symptoms: inserted bases, more than 8–10 bases per minute in the raw data and "peaks under peaks" in the raw and analyzed data (more than one dye labeled sequencing fragment co-migrating).

1. When two or more different populations of DNA molecules with a common priming site are present in the template DNA preparation multiple sequencing events will occur. It is very common to obtain a mixed template from PCR samples cut from gels (e.g. differential display) or from PCR amplified template that contain more than one amplified fragment. It is also seen with plasmid samples when, for example more than one recombinant colony is inoculated into a single culture. In this case, the initial analyzed data is good (common vector sequence) but the inserted region is different and will generate inserted bases.
2. PCR primer carryover and sequencing primer mis-priming are similar in that they will generate inserted bases from multiple primer extension events occurring in the same sequencing reaction. In the former case not all of the primers used to generate a PCR product were removed prior to doing the sequencing reaction and so these leftover primers now initiate a sequencing reaction from places on the template that are different from that of the desired sequencing reaction.
3. In the case of sequencing primer mis-priming, there are 2 or more binding sites on the template DNA for the primer to initiate the sequencing reaction. Primer mis-priming results in too many bases being present in the analyzed data, which then generates inserted bases in the sequence. In some instances the data has so many insertions that the sequence analysis will fail.

Figure 18 is an example of inserted bases caused by contaminated template.

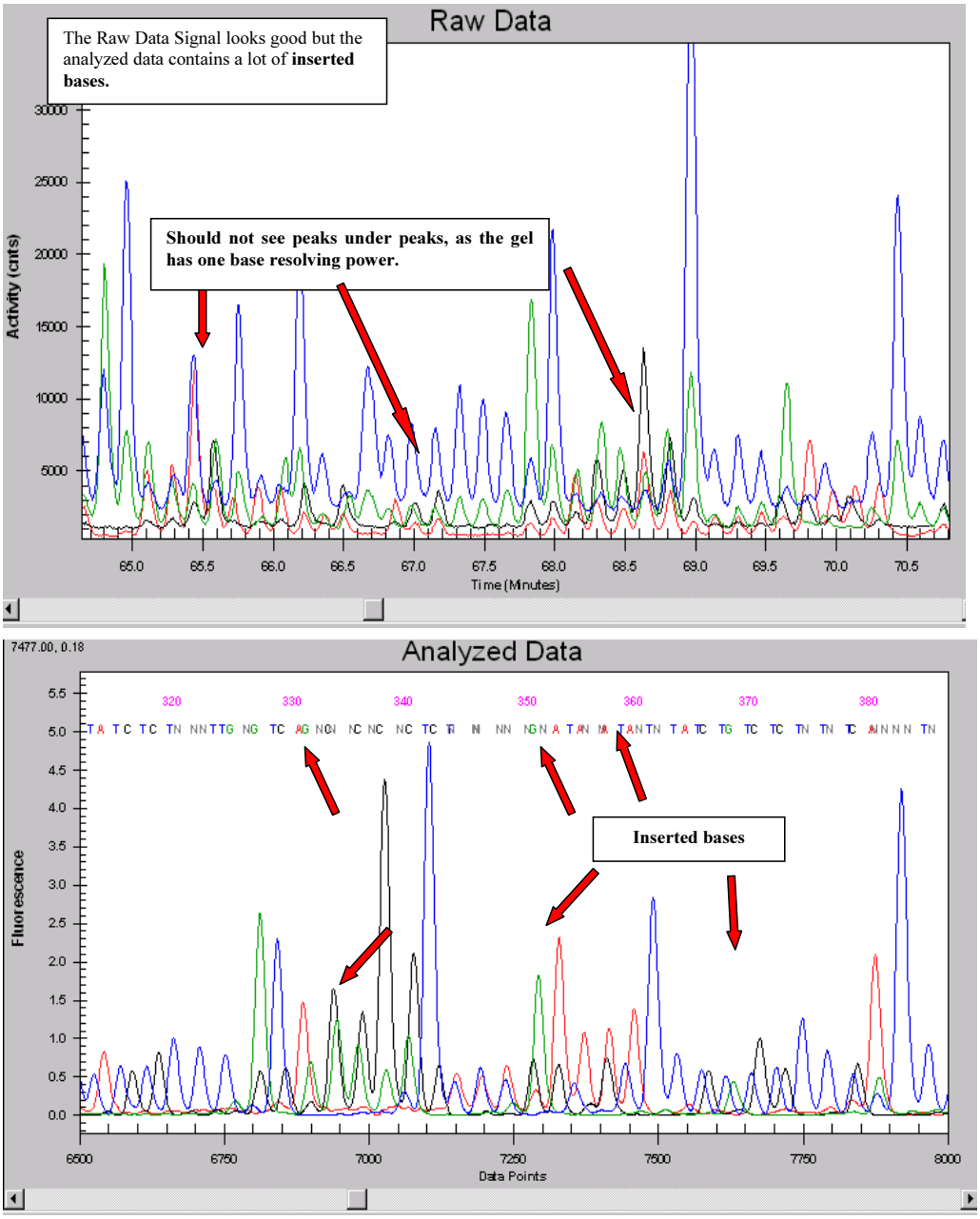


Figure 18: Insertions Caused by Multiple Sequencing Events

Corrective Actions:

1. Prepare the DNA template in such a manner as to avoid the incorporation of mixed templates.
2. In some instances when the contaminating templates are present in very small quantities simple dilution of the sample will decrease the contaminating sequence to a level where an acceptable sequence call can be made.

3. Purify the PCR products prior to using them in DTCS reactions.
4. Use a primer that binds to a site internal to the primers used for PCR (nested primer).
5. Select annealing temperatures that limit the annealing of mismatched primers.
6. Unambiguous sequence regions should be chosen for selecting the priming sites.

High Baseline Levels

High baseline levels can lead to erroneous base calling and short read lengths. Optimal performance of the CEQ System is achieved when the baselines for the 4 channels (blue, green, black and red) are below 6,000 cnts. Shown in Figure 19 are two examples of “dirty capillaries”. The example on the left may yield acceptable quality data. The example on the right shows an extreme case of a dirty capillary.

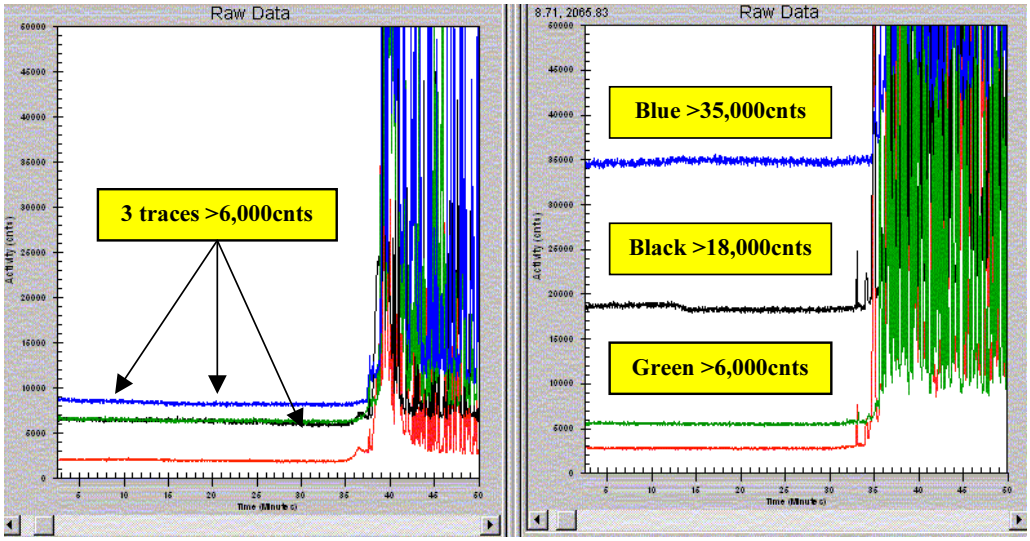


Figure 19: High Baselines

Corrective Action

1. Remove the capillary array at the manifold end and clean the capillary window. Use sterile water. **DO NOT USE METHANOL**. Clean in one direction only.
2. Use “Direct Control” from the “Run” application to purge the manifold and to fill the capillaries with new gel and then clean the capillaries (see step 1 above).

Section III: Sequencing Troubleshooting Quick Reference Table

PROBLEM	RAW DATA SIGNAL / CURRENT PROFILE	POTENTIAL SOURCE OF PROBLEM	POSSIBLE SOLUTIONS
• **Irreproducibility** • **Short Read lengths** • **Inaccurate base calling** • **Data not analyzed**	**Low Raw Signal and Normal current profile**	Not enough DNA	Increase the [DNA], always quantify the DNA
		Poor Thermal Cycling	Check Thermal Cycler
		Problem with Thermal Cycler	Check that the correct method was chosen on the thermal Cycler
		Bad Formamide	Use supplied SLS for sample resuspension
		Incorrect primer concentration	Check Primer Concentration
		Incomplete resuspension of pellet in formamide	Make sure that the pellet is properly resuspended in formamide ***or*** If the pUC18 Control Template does not meet the specifications
		Problem in post reaction cleanup of samples	Problem with Ethanol precipitation Check that glycogen was added Centrifugation at 4°C pH of Sodium Acetate Concentration of EDTA
		Incorrect mineral oil	Check that supplied mineral oil was added to samples in plate
		No mineral added	Check that mineral oil was added to samples in plate

PROBLEM	RAW DATA SIGNAL / CURRENT PROFILE	POTENTIAL SOURCE OF PROBLEM	POSSIBLE SOLUTIONS
	Low Raw Signal and Erratic or Crashed current profiles	Preheat treatment on plasmid samples was not performed (plasmid/cosmids/YAC/BAC/ PAC samples).	Preheat all plasmids templates
		The preheat treatment was carried out with all the sequencing reagents added.	Preheat ONLY the DNA and WATER
		The cycling program had an extra step in the beginning for the preheat treatment.	***Do not*** add an extra step in the beginning of the cycling program
		Too much DNA loaded onto capillaries.	Reduce the amount of DNA—follow protocol recommendations
		Sample contains contaminating ions that are injected onto capillaries- (ions from contaminating salts, bad formamide etc.).	Check method of DNA purification
		Incorrect method of template purification e.g.- CsCl, Phenol/Chloroform, or insufficient removal of Ethanol during template preparation etc.	**Refer to the CEQ 2000 DTCS chemistry Protocol, 718119-AB, pg. 18 and Application Note A-1872A for detail on the preheat treatment procedure.** **Use supplied SLS for sample resuspension**
	No signal associated with normal Current Profile **(No signal even from the Unincorporated dyes)**	Lost Pellet	Use "gel loading pipette tips" to remove ethanol during precipitation
		Tips of capillary broken	Replace capillary array if tips are broken
		ddNTPs not added	Check that all reagents were added
		Sample not resuspended in formamide	Only use SLS for sample resuspension
		Not enough sample in well (though in most cases will see a current problem)	Use at least 30µl of SLS to resuspend the samples

PROBLEM	RAW DATA SIGNAL / CURRENT PROFILE	POTENTIAL SOURCE OF PROBLEM	POSSIBLE SOLUTIONS
	Only signal of the Unincorporated dyes observed with a normal Current Profile	Failed Enzymatic Reaction	Check T_m of primer and work out the correct annealing temperature
		Incorrect primer	Check sequenced T_m of primer and work out the correct annealing temperature
		Incorrect annealing temperature	Check T_m of primer and work out the correct annealing temperature
		Did not add one of the reagents in the chemistry setup	Check that all reagents were added
		Malfunctioning Thermal Cycler	Check the thermal cycler program for correct program and ramping Check Thermal Cycler
	No signal associated with bad Current profile	See explanation on erratic or crashed current profiles above.	
	Identical Erratic or Crashed Current profile in:		
	All eight capillaries of the array	Air bubbles coming from the Manifold end	Do 2 manual Manifold Purges and 2 manual Gel Fills from the Direct Control section of the Run module and rerun the samples (Note: use a fresh buffer plate)
		Manifold area dirty thus not allowing for the manifold end of the array to be sealed correctly	Clean Manifold area of instrument
		The current can be affected in all eight capillaries if not enough buffer is added to the buffer plate	Fill buffer plate well at least ¾ full
	One capillary	Air Bubbles in the Sample Well	Check sample plate for air bubble before loading onto CEQ
		Too much DNA template (These peaks do not tend to be as sharp as those caused by air bubbles)	Reduce the amount of template in cycle sequencing of template in cycle sequencing reaction Preheat super-coiled template DNA prior to cycle sequencing

PROBLEM	RAW DATA SIGNAL / CURRENT PROFILE	POTENTIAL SOURCE OF PROBLEM	POSSIBLE SOLUTIONS
• **Irreproducibility** • **Short Read lengths** • **Inaccurate base calling;** • **Data not analyzed**	**Jump in the Raw Signal baseline**	Some sort of carry over in the sample from - Salt - Dirty Primer - Too much primer in template PCR reaction	Elute DNA template in sterile water or 10mM Tris pH 8.5 solution Do not use a strong buffer (ionic strength) for elution of DNA in template purification Use HPLC purified primers Use correct concentration of primers Do not add salts of any kind to the sample Reduce amount of primer in template PCR reaction
• **High raw signal but wrong bases called;** • **Sequence contains inserted bases.**	**High Raw signal (off-scale) and normal current profile.**	Too many cycles	Cut down the number of cycles
		Very efficient cycling causing the raw signal to over-range the detector	Reinject the sample for less time—7.5 seconds instead of 15 seconds
		Poor quality primer (n-1 primer)	Improve quality of DNA synthesis
		Degraded primer	
• **Effects of Formamide**	**Loss of dyes** **Overall reduction in signal intensity** **No signal and/or current problems**	Formamide breaks down into formic acid and ammonia, and these ionic products compete significantly with the larger DNA ions for injection and thus cause reduction in signal intensity. In addition they can also cause degradation of the sample	Use SLS for sample resuspension Store and use the SLS solution correctly
• **Primer Artifacts**	**The analyzed data contains a lot of inserted bases.** **Peaks under peaks in the raw data and also in the analyzed data.** **More than 8–10 bases per minute of raw data.**	Mis-priming Mixed templates Poor quality primer	Select annealing temperatures that limit the annealing of mismatched primers Unambiguous sequence regions should be chosen for selecting the priming sites Prepare the DNA template in such a manner to avoid mixed template contamination Use a primer that binds to a site internal to the primers used for PCR In some instances the contaminating template is present in very small quantities and dilution of the sample helps to decrease the concentration of the contaminant even further
• **Salt mobility**	**Compression of bases in the region of 20–70 bases from the primer**	Caused by the migration of salts in the gel	Only seen in software v1.1 Not present in v2.X or greater Check if denaturation has been set to 0 time length

PROBLEM	RAW DATA SIGNAL / CURRENT PROFILE	POTENTIAL SOURCE OF PROBLEM	POSSIBLE SOLUTIONS
• **Poor Resolution**	**Check the accuracy of base calling with the pUC18 Control Template.**		Check pUC18 Control template—if this is okay then the DNA template and/or primer is at fault ***or*** If the pUC18 Control Template does not meet the specifications
		Poor Cycling	Use SLS solution to resuspend samples
		Bad Formamide	Check Thermal Cycler
		Problem in post reaction cleanup of samples	Problem with Ethanol precipitation
		Incorrect mineral oil	Check that glycogen was added
		Capillaries	Centrifugation at 4°C
		Gel	pH of Sodium Acetate
			Concentration of EDTA
			Check if mineral oil was added to samples in plate
			Check the life of the capillaries
			Check life of gel
• **Capillaries**	**Resolution is bad** **Less than 6 peaks per minute** **Spurious peaks** **Late start of data even though current is normal**	The capillary has performed more than 100 runs The capillary coating has been damaged The tips were not kept moist The array was not stored correctly Some type of chemical was used to load the sample that destroyed the coating (DMSO etc.) A sample plate was set up but no corresponding buffer plate was loaded	Purchase a new array
	One or more of the four dye traces are >6000 counts in the beginning of the run	Dirt on the window	Remove the capillary array at the manifold end and clean capillaries. Use sterile water or absolute ethanol. **DO NOT USE METHANOL**. Clean in one direction only
		Manifold area dirty	Clean Manifold area

Section IV: Sequence Analysis FAQ Sheet

General Operation

Can I perform DNA sequencing and fragment analysis on the same plate using the CEQ 8000 Series System?

If you have purchased or have upgraded to the CEQ 8000 Series System it is possible and practical to run both sequencing and fragment analysis on the same plate. The CEQ 8000 Series System upgrade uses the same 33cm capillaries and plenum for both sequencing and fragment analysis. It is still advisable to separate the sample plate into blocks of sequencing and fragment samples if overall run time is critical. Separating the different types of samples into blocks will reduce delays due to possible temperature ramping between methods.

Can I switch to another database to inspect old data while new data is being collected?

No, you may only retrieve data from the active database, and you may not switch the active database while any of the other CEQ software modules (Run, Sample Setup, or Analysis) are open. However, data from different projects within the same database can be examined freely, even while new data is being collected.

What are the differences between the LFR-1, LFR-a, LFR-b and LFR-c methods?.

You can examine all the parameters of each separation method by clicking the Method tab in the CEQ Sample Setup module. The different methods are designed for different read lengths and time constraints. LFR-1 is the standard Long fast read sequencing method for the CEQ System. LFR-a, b and c are alternative methods that can be used, depending on individual application and the desired amount of sequence data.

LFR-a typically yields the longest read lengths but also requires additional separation time. LFR-b can yield approximately 700 bases in a shortened separation time compared to the standard LFR-1. The LFR-b method greatly reduces the separation time compared to LFR-a but at the cost of 50 or so bases. LFR-c is the fastest method, reducing the separation time for applications that require only several hundred bases of sequence data.

The choice of which method to use is dictated by the amount of sequence required and the amount of time the researcher is willing to use. In addition, the researcher has the option to create their own separation methods to better suit their individual needs.

Chemistry

Which primer purification method is best for sequencing primers?

The major factor with a primer that affects DNA sequence quality is the actual synthesis quality. If a synthesized primer contains little n-1, then high quality sequence can be obtained. Using other primer purification methods do not affect the quality of sequence as illustrated below.

Shown below is data generated from a primer that was purified by three different methods: 1) OPC column, 2) Reverse Phase HPLC and 3) gel filtration. In each case, because the synthesis quality was high, the sequencing quality was high and the purification method had little or no effect on the results. The CEQ does, however, require that all sequencing primers (at a minimum) be desalted (i.e. gel filtration) before use.

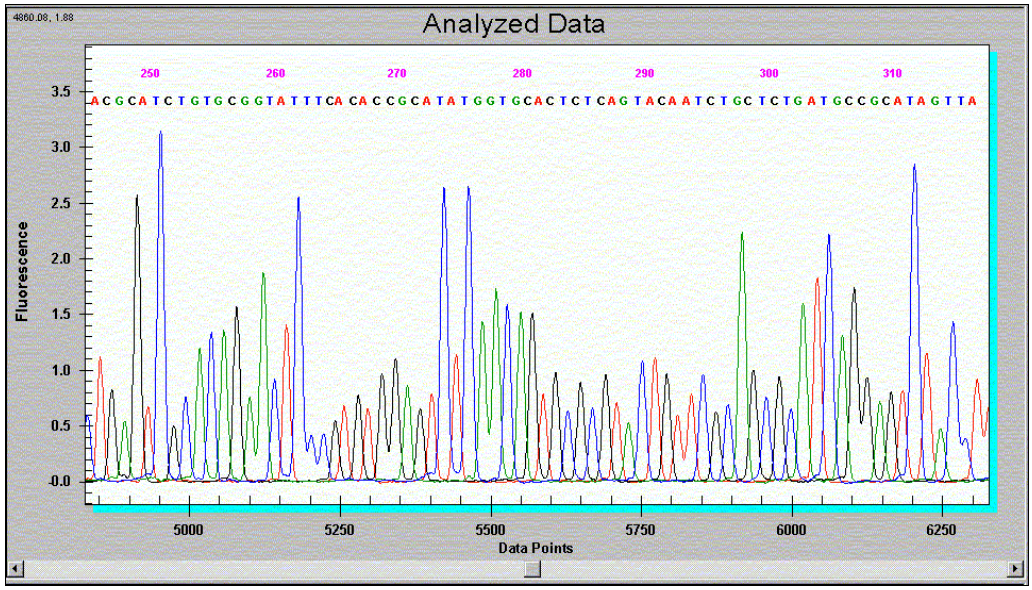


Figure 20: OPC Purified Primer

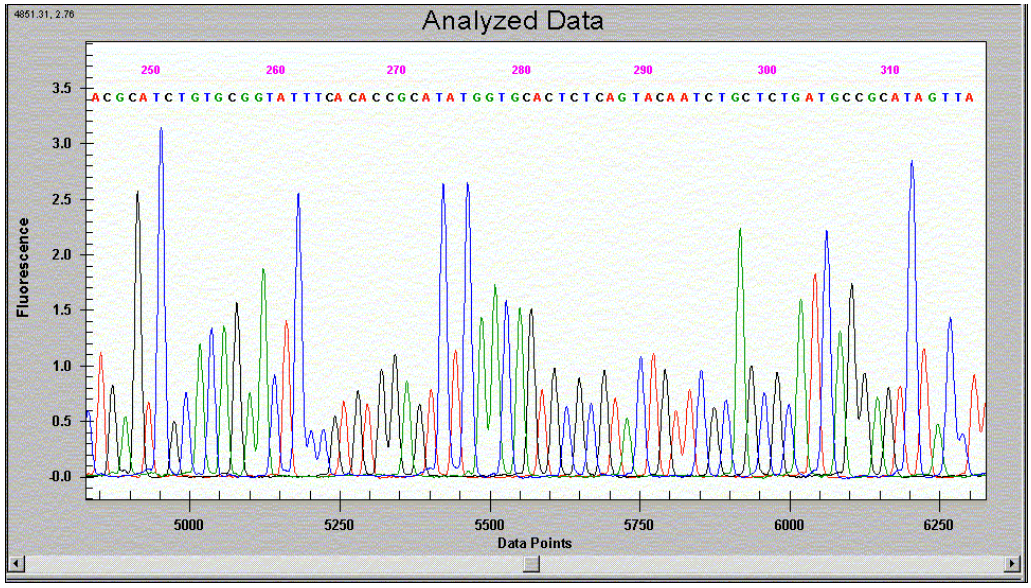


Figure 21: Reverse Phase HPLC Purified Primer

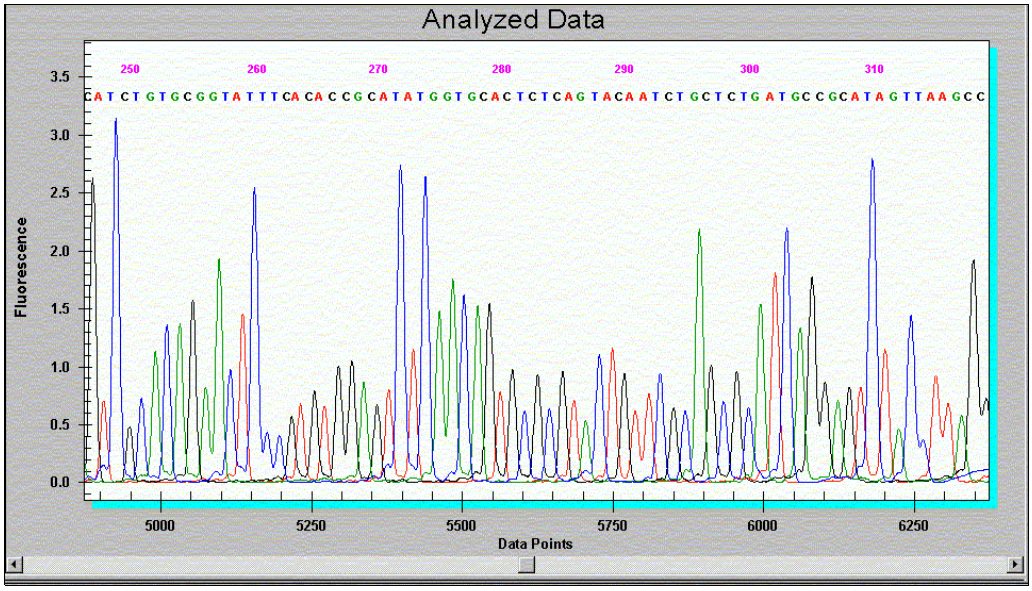
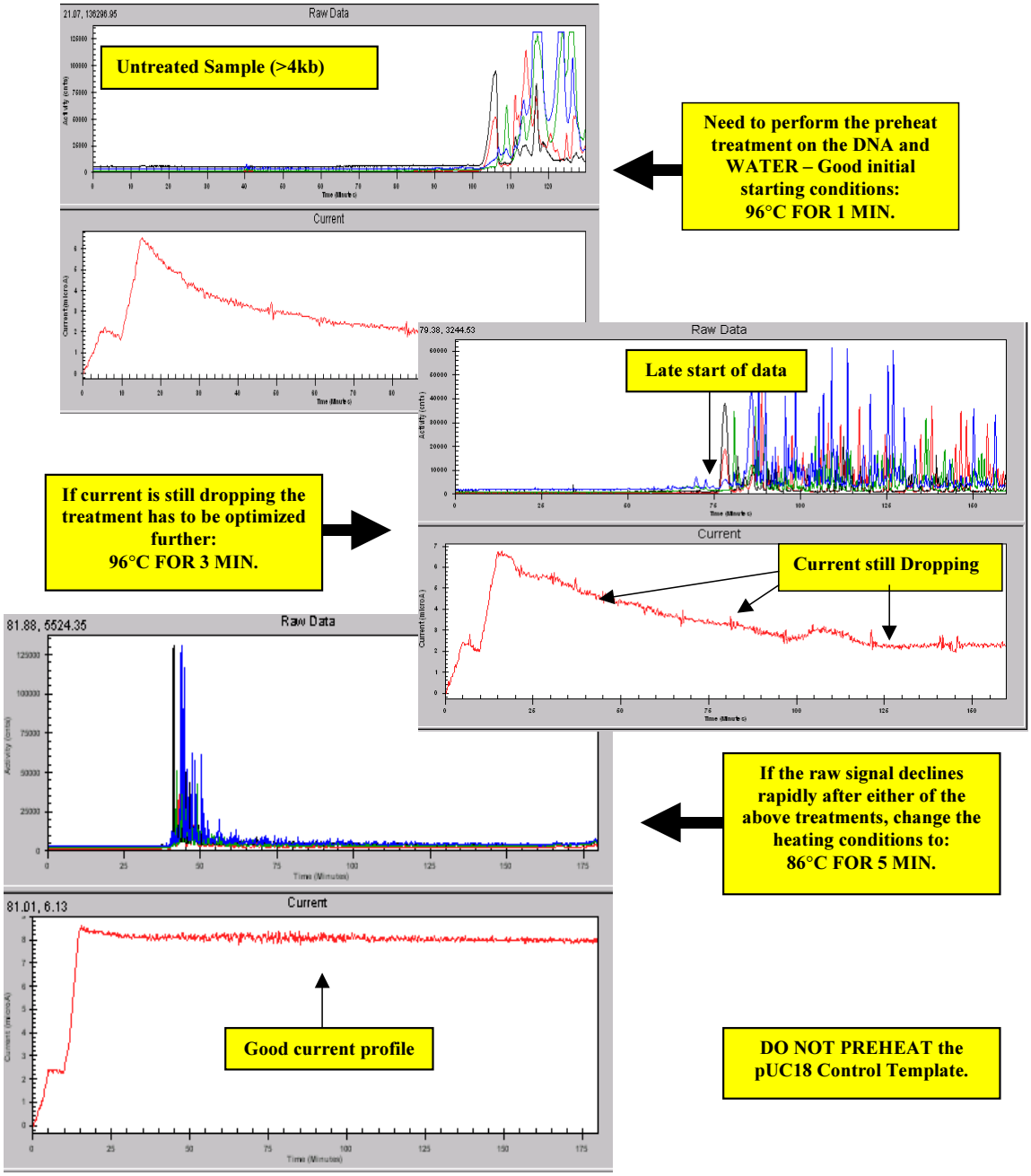


Figure 22: Gel Filtration Primer

How do I optimize the Preheat treatment?

Shown below are some hints for optimizing the preheat treatment of plasmid DNAs.



I have sufficient Raw Data signal but I didn't get a base call. What can I do?

In certain cases, the base calling algorithm cannot find the start of data. If this happens, a default start time is automatically used. The default time is determined by a combination of the separation method parameters (voltage, temperature, and voltage ramping profile) and the length (33cm or 53cm) of the capillary used to collect the data. This may not be the optimal starting point for the data.

Sometimes simply by setting the start of data manually, samples can be analyzed accurately. To manually set the start of data, expand the raw data near where the initial peaks are seen. Examine the data and look for a peak that is DNA (i.e. not unincorporated dye terminators, which will usually be the first eluted peaks and be very large and broader than the DNA peaks) and has a sharp profile. Reanalyze the data using the center of this peak as a start time.

Including peaks from unincorporated dye terminators in the data to be analyzed may result poor basecalling results in as many as the first 50 called bases. If it appears that such peaks have adversely affected the basecalling, try setting the start time later than the unincorporated dye terminator peaks.

What can be done to have the greatest chance of accurately detecting heterozygotes?

While the default settings work well for most good sequencing results, adjustment of the parameters may be required under the following circumstances:

- Heterozygosities that occur before the 50th or after the 500th base
- Raw Signal is too weak or too strong
- Raw signal has contaminating sequence

To optimize the resolution of bases late in the run, it is recommended that the separation method LFR-a be used. Other than this recommendation, there is no way for the user to know how aggressive to be with the heterozygote detection settings on his own data without experimentation. On clean data, false heterozygosities may be absent with % Average peak spacing of 90% and Sensitivity settings as high as 0.75. However, routinely using these settings is likely to cause many false “red” bases (heterozygotes) to appear the sequence results.

Review the examples on the following pages.

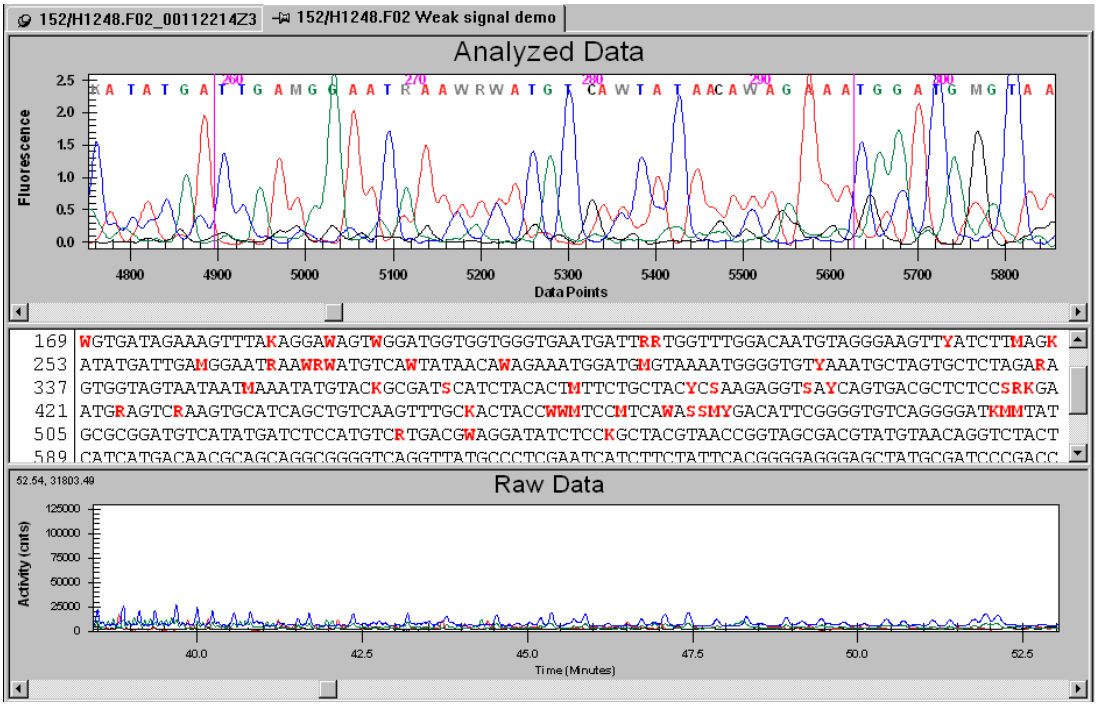


Figure 23: Too little signal. Noise is being recognized as heterozygosity, even at the low Sensitivity setting of 0.1. The Raw Data is scaled from 0 to 130,000 counts.

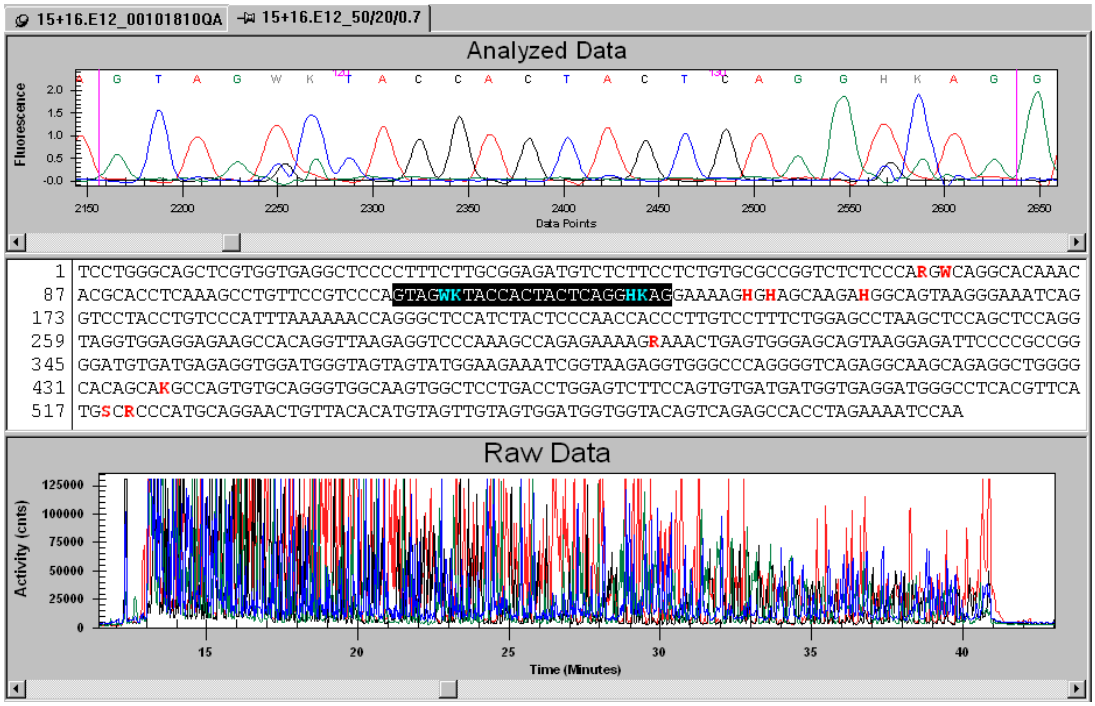


Figure 24: Too much signal can lead to bad color correction, where residual signal in the other channels has not been completely removed, leading to the insertion of peaks due to the poor color correction. Notice that analyzed peaks from Raw Data that is off-scale has slightly more rounded tips than normal peaks. The Raw Data is scaled from 0 to 130,000 counts.

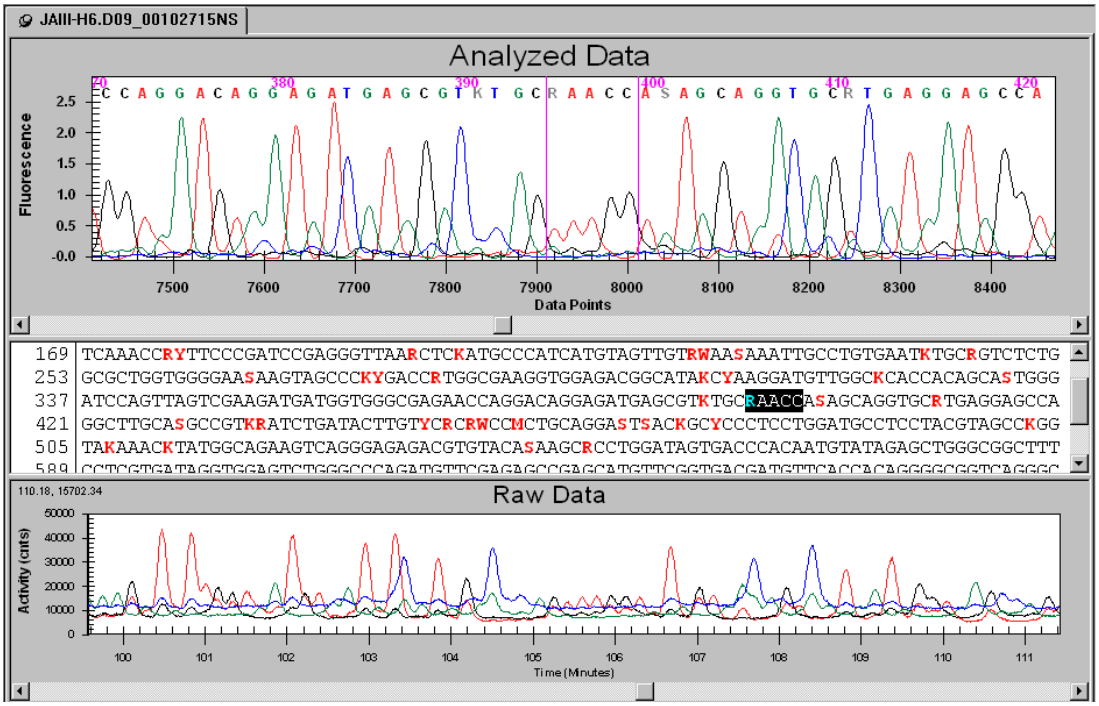


Figure 25: Raw signal strength is fine in this electropherogram (scaled to 50,000 counts). However, the presence of numerous smaller underlying peaks at many positions in the analyzed data generates an abundance of false heterozygotes.

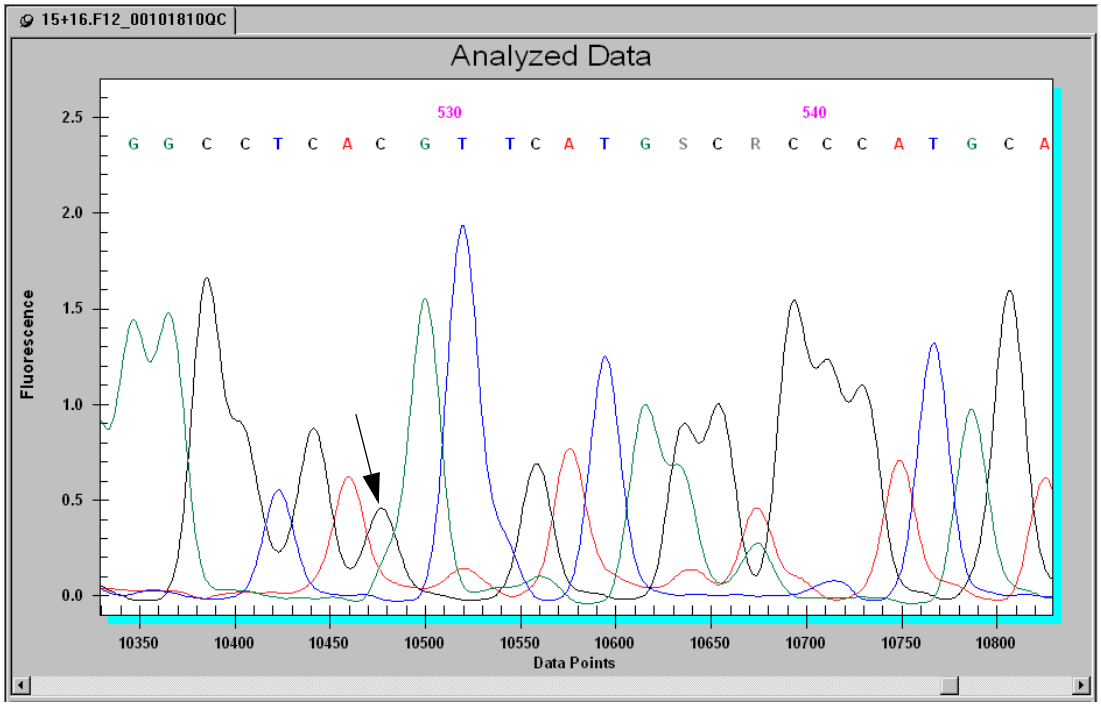


Figure 26: Using the fast method LFR-b, there is not enough resolution to discern the G shoulder, which is heterozygous with the C at position 528. There is no setting in the heterozygote detection software that can identify this heterozygosity.

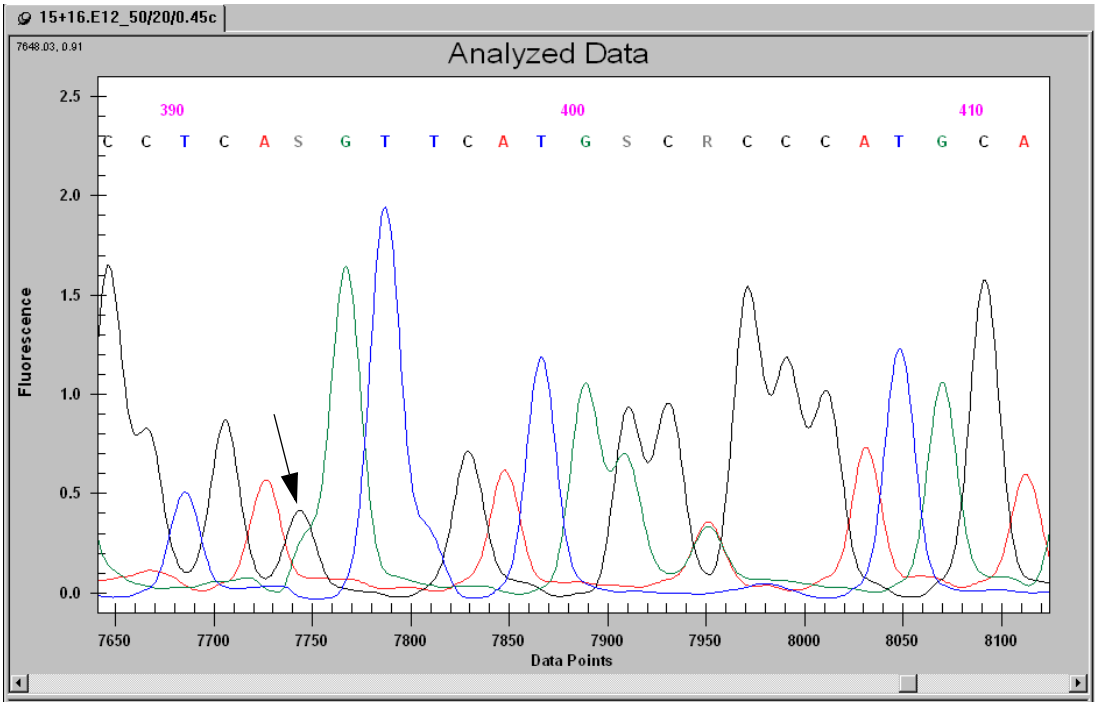


Figure 27: The same heterozygosity run under the same separation method as in Figure 26, LFR-b, using a primer that is closer to the heterozygosity. The resolution is slightly better, and is picked up with settings of 50/20/0.45.

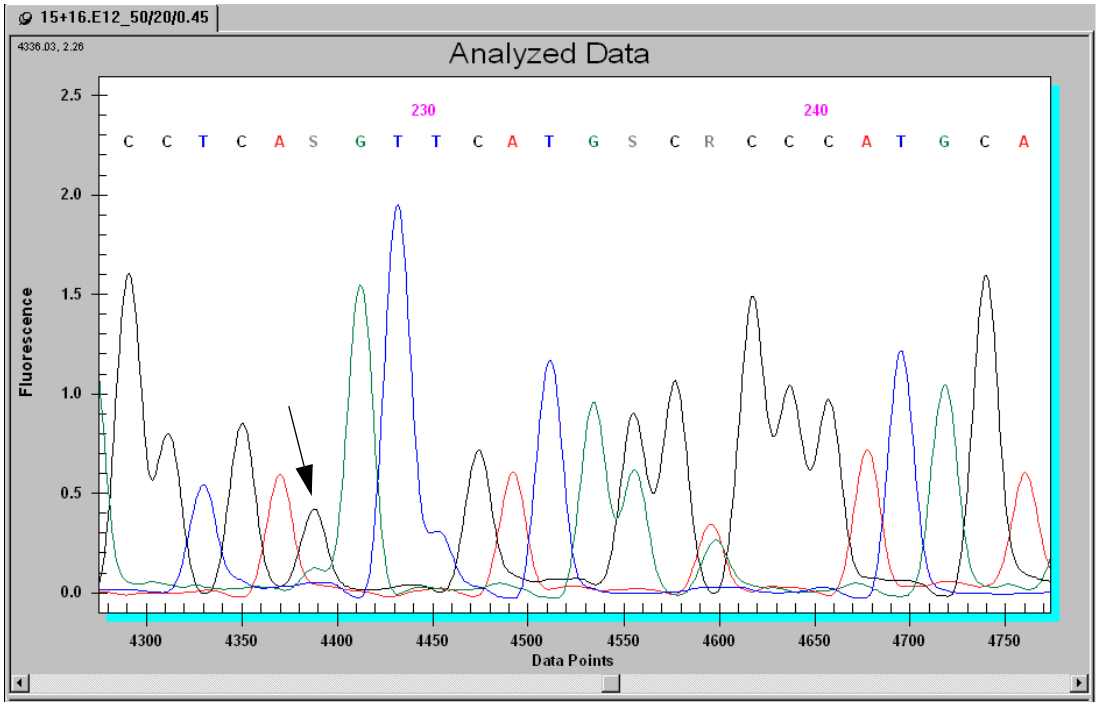


Figure 28: The same heterozygosity as in Figure 26 and Figure 27, using a primer that is still closer to the heterozygosity, placing the position of interest at base 227. The resolution is better than either of the two previous cases, and is picked up with settings of 50/20/0.45.

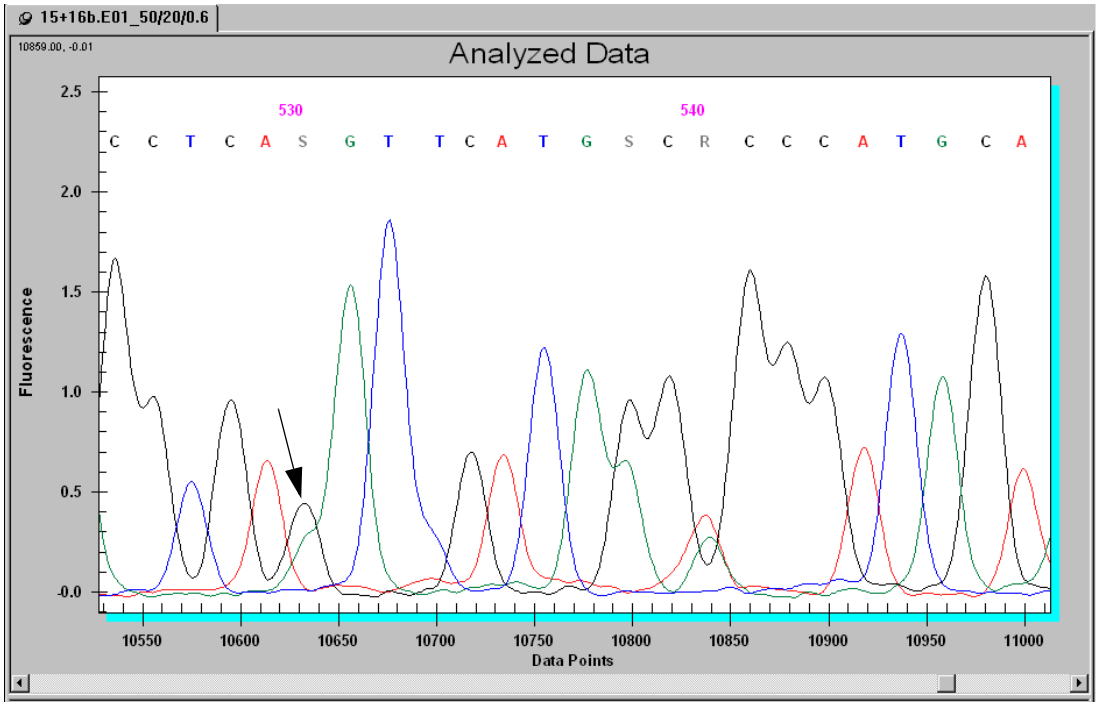


Figure 29: The slower method, LFR-a, was used to increase the resolution at 530 bases. Now settings of 50/20/0.6 identify the heterozygosity correctly.

Appendix A: Technical Bulletins

Listed below are technical bulletins that are currently available for sequencing with the CEQ 8000 Series System:

Publication A-1872A

Improved Sequencing of Plasmids on the CEQ 2000 by a Simple Template Pre-heating Procedure.

Published by Beckman Coulter and available on the Beckman Coulter web site.

Publication A-1726

High Sensitivity Analysis of Nucleotides Using Electrokinetic Injection and Sample Stacking with MECC

Published by Beckman Coulter and available on the Beckman Coulter web site.

Publication A-1766A

High-Throughput DNA Sequencing Reactions Using the Biomek® 2000 BioRobotics System

Published and available at: Published by Beckman Coulter and available on the Beckman Coulter web site.

Publication A-1774A

Quantitative Capillary Electrophoretic Analysis of PCR Products

Published and available at: Published by Beckman Coulter and available on the Beckman Coulter web site.

Publication A-1892B

Detecting Heterozygotes with the CEQ 2000XL

Published by Beckman Coulter and available on the Beckman Coulter web site.

Publication T-1852A

An Integrated System for Automation of DNA Sequencing

Published by Beckman Coulter and available on the Beckman Coulter web site.

Publication T-1853A

Automation of Plasmid Purification with the Biomek® 2000 and QIAGEN's QIAprep 96 Turbo BioRobot Kit for DNA Sequence Analysis on the CEQ™ 2000 DNA Analysis System

Published by Beckman Coulter and available on the Beckman Coulter web site.

Publication T-1874A

Post Sequencing Reaction Purification in 96-well Gel Filtration Column Plates

Published by Beckman Coulter and available on the Beckman Coulter web site.