Ocimum Biosolutions
... enabling R&D™

DNA Sequencing Services

Troubleshooting Sequencing Reactions

This document is a mix of procedural information and troubleshooting hints, and is based on examining each of the components of the sequencing procedure. This is an attempt to describe certain diagnostic features of the trace file of your failed chromatograms .

There are several components in a sequencing reaction—template, primer, sequencing reagent, and "other additives." The mix of these reagents then undergoes 3 steps—cycle sequencing, post sequencing cleanup, and analysis. By examining each of these processes, this will help you to figure out why you're not getting the data you desire.

Template

Template is a very common problem area hence, the primary suspect in failed sequencing. If your chromatogram is blank, has very low signal, or starts well but gradually dies out, the template should be examined.

Plasmids

For plasmids in the 3-10 kb range, 0.2 ug is a good amount of DNA to use. More is almost never better. One problem is that accurate quantification of DNA is not easy. Spectrophotometer readings will invariably overestimate the amount of template DNA, unless you have CsCl banded material. That's because use of almost all miniprep kits on the market will result in some RNA, chromosomal DNA, and other fluorescent cellular material coming through the purification procedure that will absorb UV light. These other chemicals won't necessarily inhibit the sequencing reaction, but they will contribute to the A260 reading. As a result, relying on this spec reading alone will cause you to add less than DNA than you think (midi and maxi prep kits do a better job, giving a higher plasmid:contaminant ratio). It's informative to run the template on an agarose gel, using some standard of known concentration and estimate the relative fluorescent intensity following gel staining. The problem of sequence starting out strong then rapidly dying out is common and is usually attributed to "impurities" in the prep. These impurities can't always be removed by EtOH precipitation and one may need to re-prep the plasmid.

PCR Products

PCR fragments seem to sequence either incredibly well or not at all. PCR fragments generally must be cleaned up following the initial amplification and prior to sequencing to remove the two amplifying primers and the unused nucleotides Common ways to clean up the PCR fragment are listed below, with advantages and disadvantages for each:

1. ***Gel extraction:*** *C*utting the band out of gel using any available kit is recommended if multiple fragments are seen following amplification. It is need to cut out the fragment that's the "right" size. Disadvantages of this procedure include low yields, exposure of the DNA to ethidium and UV light, which damages it, and contamination from other bands. It might be better adjusting the PCR conditions to give one, or at least a dominant, band. One might also consider cloning the products and sequencing individual plasmid clones if several attempts at sequencing a gel-extracted band fails.

2. ***Column base cleanup protocol:*** This typically involves applying the entire PCR to a column, then carrying out washing and elution steps to generate fragment free of contaminating primers and nucleotides. This procedure depends on having a single or predominant band following amplification, and can generate DNA that sequences effectively. It is highly recommended an agarose gel be run following this or the previous method in order to see how much material is present following purification.
3. ***Enzymatic cleanup of the PCR fragment:*** Commercially available shrimp alkaline phosphatase and exonuclease (SAP-EXO) are used to inactivate excess primers and triphosphates. As with the column cleanup, a single or predominant band must be present. This procedure has a number of advantages—it is fast, inexpensive, quantitative, and amenable to scaling up in 96 well format. Most importantly, it yields DNA that sequences very well.

Template composition

Template composition is a factor that can cause sequencing problems. GC rich templates will generally sequence fine, unless there are particular regions of very strong secondary structure. A diagnostic trace pattern for this sort of template would be a chromatogram that looks great to a certain point, then suddenly dies out. Methods for dealing with this are discussed below, but basically these involve the addition of denaturants to the sequence reaction or specially formulated versions of BigDye. Poly A/ Poly T regions will often cause difficulty—the chromatogram will look fine up to the polyT (A) stretch, then either be very noisy peaks under peaks, or just long "rolling hills" of the four chromatogram colors. This results from "polymerase slippage" on the poly T(A) region and is a difficult issue to resolve. The addition of "reaction enhancers" (see below) may help if the problem is not severe. Another strategy is to design "anchored primers," a poly A or poly T sequence with the final 3' base being either G, C, or T (or G, C, and A for a poly T primer). With these primers the poly A/T stretch needs to be at least 17 bases long, and special conditions for annealing (42°) and cycling (52°) are recommended. Sometimes this works well, but it may require the user experimenting with different conditions. Finally, di- and tri-nucleotide repeats can cause poor data. It's usually pretty obvious when such sequences are at fault. Often the use of sequencing reagent enhancers (discussed below) helps a lot.

Primers

Primer problems can also be a cause sequencing failures, and some of these give characteristic chromatogram patterns. A blank chromatogram, which is not that diagnostic of a primer problem specifically, can result from use of the wrong primer, too low a concentration of primer, or simply bad or degraded primer. Primer design is usually not an issue. A wider variety of primer compositions can give excellent sequencing data, compared to PCR.

A "noisy" chromatogram—good signal strength, but peaks under peaks resulting in numerous ambiguities ("N's)—can indicate several primer difficulties such as more than one primer present, more than one primer binding site present, secondary priming at a related sequence, or degraded primer. [This pattern is also evident if a mixed population of template is present, so this needs to be considered as well. It is particularly diagnostic if the peaks under peaks begin at the cloning site used in the construct].

Additional primers in the reaction is usually an accidental result resolved by repeating the reaction, or in the case of sequencing a PCR fragment, re-cleaning the input template to remove amplifying primers. Multiple peaks on the chromatogram can also occur if the primer binding site is within a repeat region on your template. Depending on the structure of your template, this can look pretty strange—you can have good sequence that all of a sudden turns into peaks under peaks, or peaks under peaks that will suddenly resolve into good looking data. If this happens you may need to reexamine the template you're sequencing: investigate what's known about it or analyze it yourself and see if you can detect evidence of a repeat pattern. This can be a real problem—sequencing through long repeat regions is one of the big challenges faced by genomic sequencers.

It is also possible to get secondary primer binding if the primer is very GC rich and is being used to sequence a GC rich template. In these cases, we try raising the annealing temperature and adding a denaturant such as DMSO (to 5% final concentration), formamide (5%), or betaine (1M final concentration) to increase primer:template specificity. Secondary priming giving peaks under peaks, or too strong a signal resulting in peaks under peaks, can also occur if the primer concentration is too high in PCR fragment sequencing. In contrast to sequencing plasmids, where too much primer shows little effect, adding too much primer to a PCR fragment sequencing will frequently result in noisy data. It is important to limit primer amount to 5 pmol when sequencing PCR fragments.

Finally, if the primer is starting to degrade, or if there is a high proportion of n-1 products in the oligonucleotide preparation, you will also see peaks under peaks since you're essentially adding multiple primers to the reaction. Primers can last a long time, but they can also degrade and it's impossible to set an expiration date that covers all primers. It may be worth just having a new one synthesized.

Post Sequencing Cleanup

Following cycle sequencing, reactions must be cleaned up to remove excess dye terminators which don't incorporate into DNA. Ideally, all the excess terminators will be removed and all the extension products will remain. This can be a common source of problems for individuals, either in terms of getting too many unincorporated nucleotides in the sample, or in sample loss during cleanup resulting in lower signal. Low signal is not necessarily a huge problem because the instrumentation is sensitive enough to extract useful information. Another issue is that low signal generally doesn't look good out past 600 bp or so. From the capillary you should ideally get about 800 bp of readable sequence, and this usually depends on starting out with strong signal (because of the way the capillary instruments operate, signal strength will usually fall over the course of the sequence). Another diagnostic sign of cleanup problems is the so-called "dye blobs." In the capillary they can be seen at different places in the chromatogram--sometimes early, at about 70, and also at about 300 bp

They appear as off-scale, or very tall, broad, peaks stretching over what may be other legible peaks below. All colors are sometimes seen, but often they are comprised of just red (the latter resulting from a breakdown product of the T terminators). These peaks confuse the basecalling software, and lead to inaccurate sequence data. Often the true peaks are seen below and the sequence can be manually

edited. However, depending on the signal strength of the other bases, and the severity of the dye blobs, all the T-signal can get "sucked up" by the dye blobs and the chromatogram will appear to have no other T's (this is a function of the way the software works). Ideally, no dye blobs beyond small ones appearing early in the run should be seen.

Analysis
Good signal, but peaks are shifted relative to each other and basecalling results in many N's. This suggests the data were analyzed using the mobility files from the wrong version of BigDye. We can easily reanalyze these files with the correct input mobilities.

Trace file starts out well, but turns into "rolling hills" of color, or starts right out with rolling hills. This is either a capillary or a sample issue, but since it's impossible to tell (though if it starts out nice it's usually the capillary), we generally re-run such samples if we notice this. Charged contaminants in the sample will lead to a pattern of rolling peaks throughout the trace file. If several samples from a particular user show the same pattern, it's not the capillary and you will need to alter your procedures to avoid the problem in the future.

If you get sequence that's from your organism but doesn't seem to be what your primer should have given you, or you weren't expecting exactly that PCR fragment but it's close, please look over all your manipulations before calling. If it's not one of the software or switched tube issues, chances are it's something at your end.