Microsynth

Troubleshooting Guide: Understanding Your Sequencing Results

Dear Customer

Microsynth always strives to make the most of your DNA samples. Optimized processing steps, proven protocols and an overall experience of 25 years in the area of Sanger sequencing usually result in long reading lengths and clean sequences. Nevertheless, there are factors, which can affect the successful generation of a high quality sequence. The reasons for that are of different nature. Some causes can already be identified at the chromatogram level, while for others additional information is required.

Among the most common causes for poor sequencing results are:

1) Poor DNA quality

Even more important than quantity is the quality of your sample DNA. The cleaner the DNA the better; good results can be achieved even with small amounts of DNA. Conventional isolation methods (e.g. alkaline lysis followed by column purification) usually provide DNA samples with sufficient quality to obtain excellent sequencing results with the quantities indicated by Microsynth (see below). "Low copy" vectors however prove to be more problematic, because they inherently have a worse contaminant to DNA ratio. In this case a repeated purification step is often necessary.

2) Low or high DNA concentration

Inadequate DNA concentration can be easily avoided by measuring an aliquot of your DNA sample. Microsynth recommends fluorimetric measurement. But also photometric measurement or length and quantity DNA standards in a gel are appropriate. Please send us 40-100 ng/µl (plasmids) or 1.5 ng/µl per 100 bases (PCR-products).

3) Unfavorably designed primers

A proper primer design is essential for high quality sequence data. The design of a sequencing primer is more demanding than the design for primers used in a polymerase chain reaction. The reason lies within the actual reaction mechanisms. The polymerase chain reaction follows a logarithmic kinetic, allowing an exponential amplification of DNA molecules per cycle step, whereas during a sequence reaction only one set of new DNA molecules is produced in each cycle. If the primer efficiency is low, not enough fluorescent labeled material is generated. Thus, a good primer for PCR is not necessarily a good primer to sequence with.

Important information for a good primer design:
- avoid wobble positions
- balanced GC content (50%)
- avoid hairpins, palindromes, or primer dimer formation
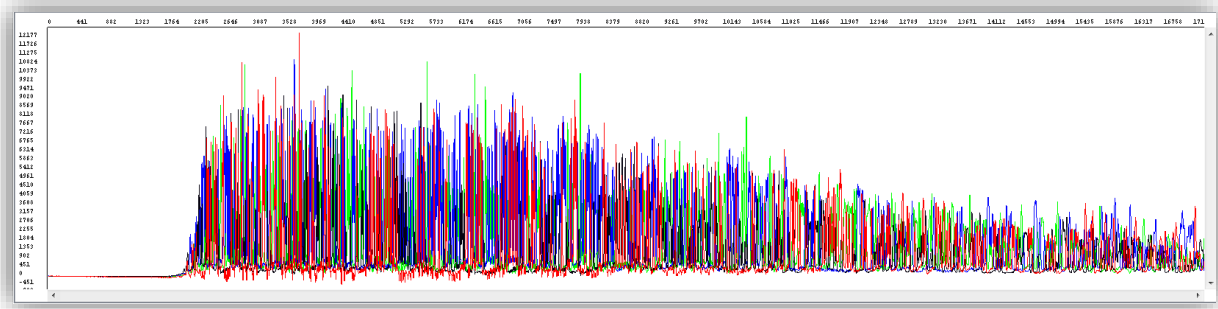- optimum length is 20 bases (+/-2), max. 25

- no second binding site (7 bases from 3' end)
- 50-60°C melting temperature (formula: Tm = 2°C (A + T) + 4°C (G + C))
- G or C at 3' end
- well-chosen primer binding position (sequence starts 20-60 bases from 3' end of primer)

See pages 4 to 16 for more comprehensive discussion of major sequencing failures and the related troubleshooting actions!
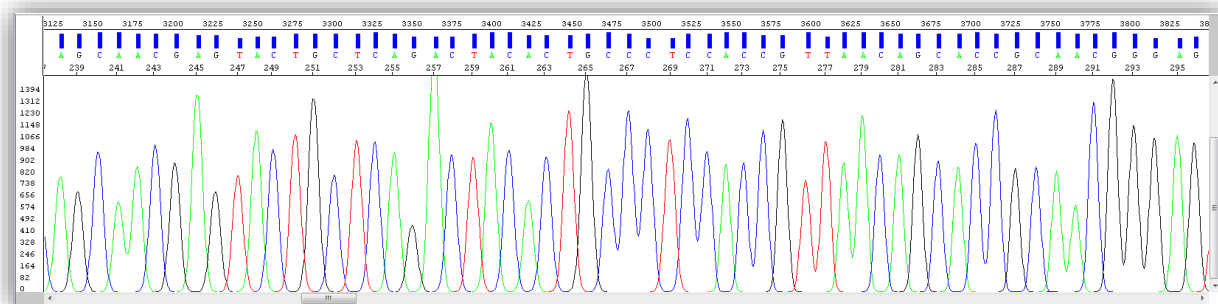
Index

Case 1: High-Quality Sequence Data

Figure 1-1: Raw data of a high-quality DNA sequence


Figure 1-2: Electropherogram of a high-quality DNA sequence reaction. It shows distinct peaks and no or very low background noise.

Case 2: No Sequence Reaction, No Sequence Data
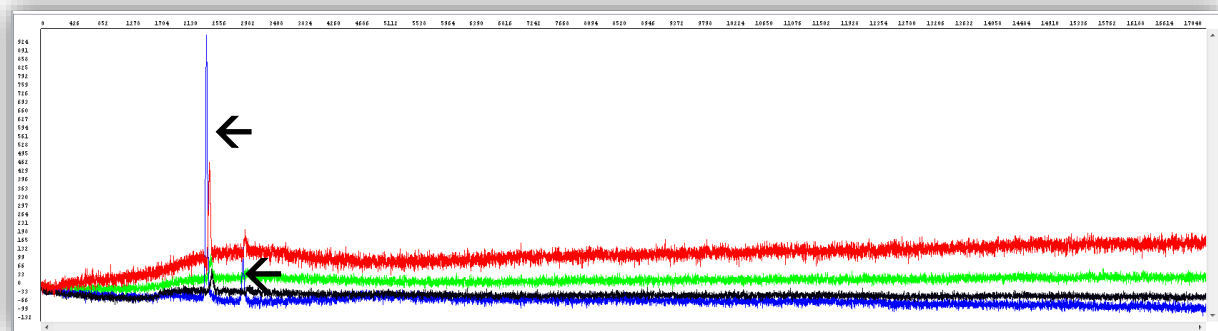
Figure 2-1: Raw data of a failed sequence reaction. The otherwise almost invisible spikes (dye blobs), represent unincorporated fluorescent nucleotides and appear very prominent in this case (arrows).
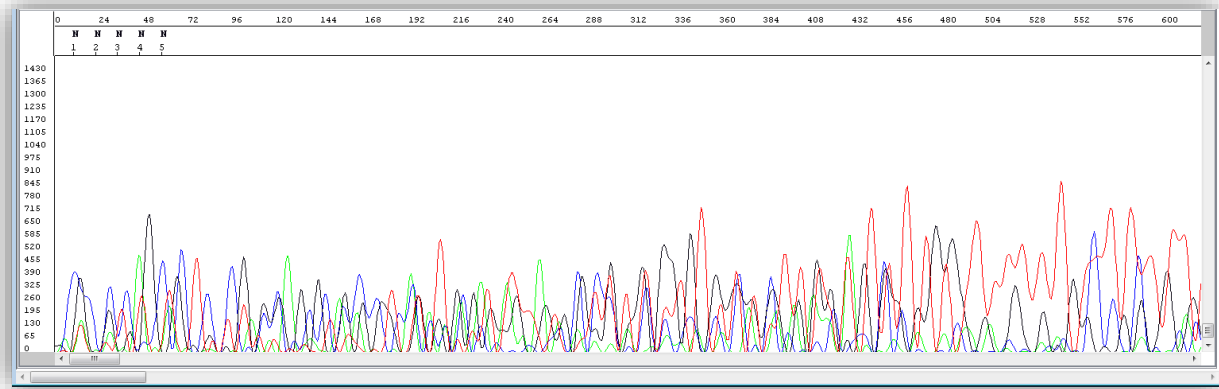
*Figure 2-2: Electropherogram of the starting point of the reaction depicted in figure 2-1.
No sequence information could be deduced, which means that the basecalling failed, only NNNNN
is annotated.*
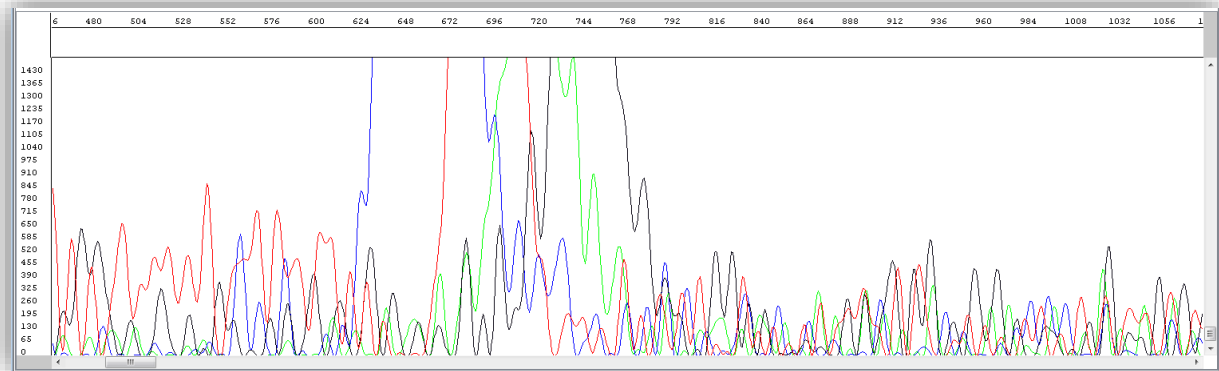
*Figure 2-3: Electropherogram of the failed reaction in figure 2-1. The large signals represent the
unincorporated fluorescent nucleotides, also visible in figure 2-1. They are often referred to as "Dye
Blobs". Visible Peaks represent background noise, amplified by the analysis software.*

Possible cause 1: No primer binding site.
Solution: A frequent cause when sequencing plasmids. If you have chosen one of our
standard primers, please check whether it fits your vector. Recheck your plasmid maps and -
sequences.
If you have designed your own, sequence-specific primer from existing sequencing data,
please make sure that the primer sequence was derived from a reliable source. Please see
also the instructions for primer design above.

Possible cause 2: Inappropriate amounts of DNA and / or primer
Solution: Check your concentration measurements. Even if a sequencer has a high degree of
sensitivity and can detect a wide range of DNA concentrations, there is still a "minimum
amount" of template and primer, which needs to be present in the reaction in order to
ensure a successful sequencing. The same is true for excess amounts of DNA. The sequence
reaction can then fail completely (figure 2-4) or distorted peaks show in the
electropherogram (see also case 12).
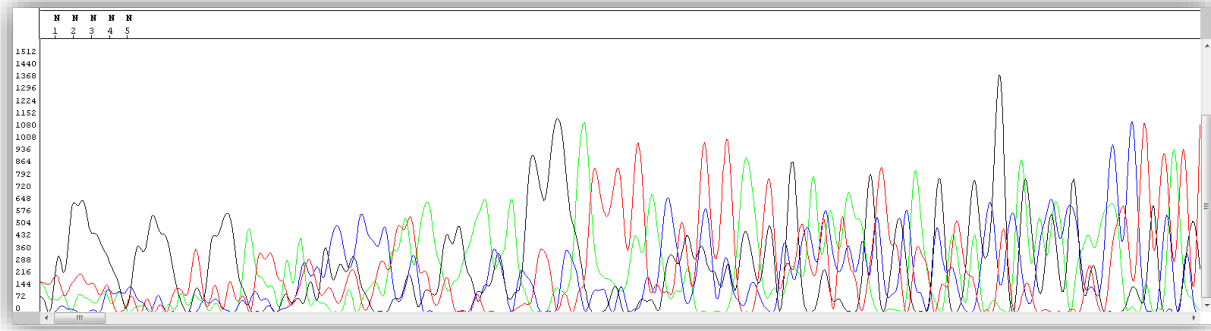
Figure 2-4: Electropherogram of a failed sequence reaction due to excess amounts of DNA.
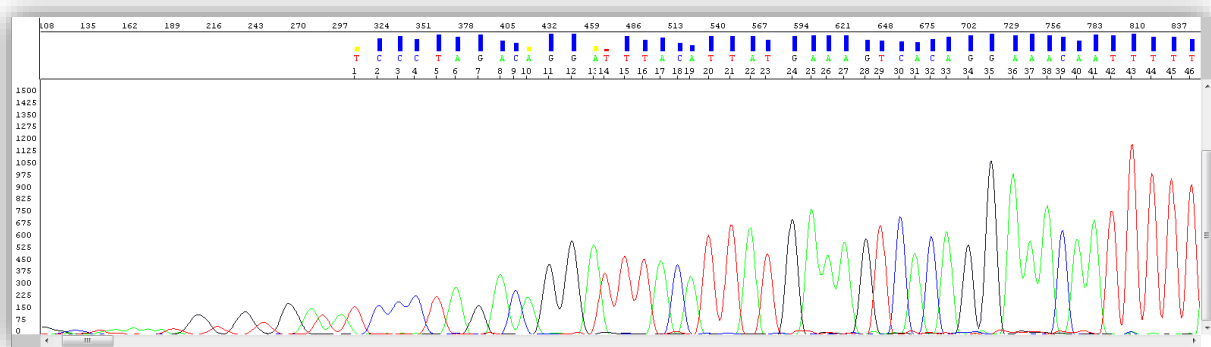
Figure 2-5: Electropherogram of the same sample as in figure 2-4, but after diluting it 2-fold.

Possible cause 3: Inhibitory contaminants
Solution: The cycle sequencing reaction is very sensitive to the presence of certain contaminants, some of which may inhibit polymerase activity completely. Among the most common contaminants that have a negative effect on the sequencing reaction are salts, EDTA, alcohol, proteins, RNA, detergents, cesium and phenol. Under certain circumstances, the template needs to be re-prepared or purified through a column to remove one or more contaminants sufficiently and thus enable successful sequencing.

Case 3: Superimposed Signals

The occurrence of superimposed signals within a sequence can have different reasons. In order to determine the responsible factor, please pay attention to:
1) the position from where the peaks overlap and
2) the average signal strength of the sample. Samples with low signal strength can generate an artificially increased background noise that looks like superimposed signals. However, if the average signal strength is high, this possibility can be ruled out.

A) Superimposed signals from the beginning of the sequence:
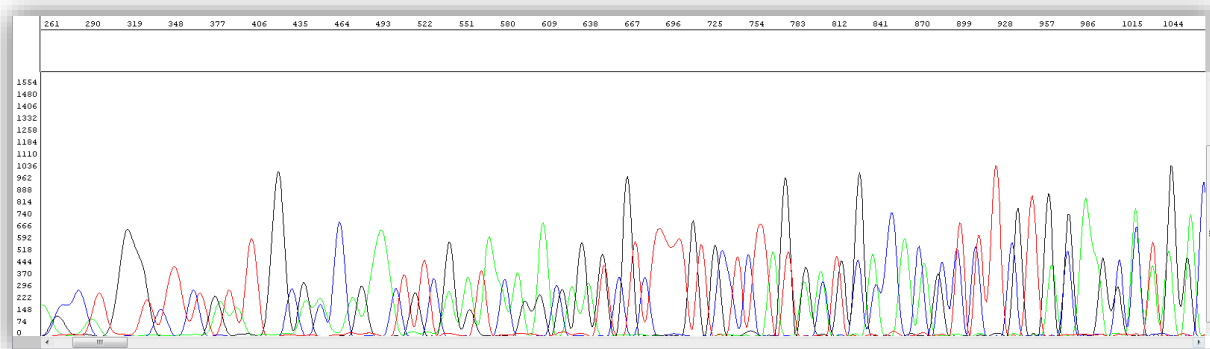
Figure 3-1: Superimposed signals from the beginning of the sequence.

Possible cause 1: Multiple primer binding sites
The sequencing primers could have a second binding site on the plasmid, which is either identical or very similar to the target sequence. The nucleotide sequences that originate from these two binding sites are different, which leads to superimposed signals in the chromatogram. If the primer binding sites are identical, the double peak signals are of approximately equal intensity throughout the entire sequence. The fragments can also show a different running behavior, leading to double peaks that are not congruently superposed, but slightly shifted from each other. Sometimes the second primer binding site of the target sequence differs in a few internal positions. In this case, the primer does not bind with the same efficiency as it would do with the target sequence. Nevertheless, a binding and subsequent extension is possible. The outcome is a second sequence of lower intensity, which runs along below the main sequence.
Solution: In both cases it is important to look at the entire template construct for sequences that are identical to the primer sequence or at least very similar. If so, another sequencing primer should be selected or constructed. If the construction of an alternative primer is difficult (e.g. "primer walking" through a repetitive sequence region), try to find a primer with one (or better several) bases at the 3 'end that are specific to the target sequence and thus act as a sort of "anchor".

Possible cause 2: Multiple primer binding sites while generating the PCR product.
Solution: During PCR one or both PCR primers bind to more than one position on the template DNA leading to multiple, mixed PCR products. In general, such undesirable by-products can be identified when separating an aliquot of the PCR reaction in an agarose gel (or even better a polyacrylamide gel). If there are clear differences between the PCR products in size, the target template should be purified from the gel. PCR products of similar or even identical length do not separate properly in the gel. In this case, the optimization of PCR conditions, a redesign of the primer or cloning of the product should be considered.

Possible cause 3: One of the PCR primers works as forward <u>and</u> reverse primer while generating the PCR product.
Solution: The consequence is a double sequence with one primer, while the other primer leads to a negative sequencing result. In that case, the PCR primers need to be redesigned.

Possible cause 4: Residual primers and / or dNTPs from the PCR reaction
**Solution:** It is vitally important to remove PCR primers and dNTPs from the PCR reaction, because they can lead to signal interference. If a direct sequencing of a PCR product is to be made without prior purification, it is essential to optimize the PCR reaction in so far as primers and dNTPs are not added in excess so that they are consumed almost completely at the end of the PCR reaction. Attempting to reduce the concentration of residual primers and dNTPs by diluting an aliquot of the PCR reaction is a method that we expressly do not recommend.

Possible cause 5: Primers with high melting point
Solution: Primers with a very high melting point (Tm> 65 ° C) are not recommended for sequencing. Those primers are often GC rich or very long, two factors that increase the ability of the primer to form secondary structures and/or dimers. We recommend to design primers with a melting point of 50 – 60°C. The following equation can be used to calculate a good approximation of the melting temperature:

$$Tm = 2°C (A + T) + 4°C (G + C)$$

B) Superimposed signals occur within the sequence
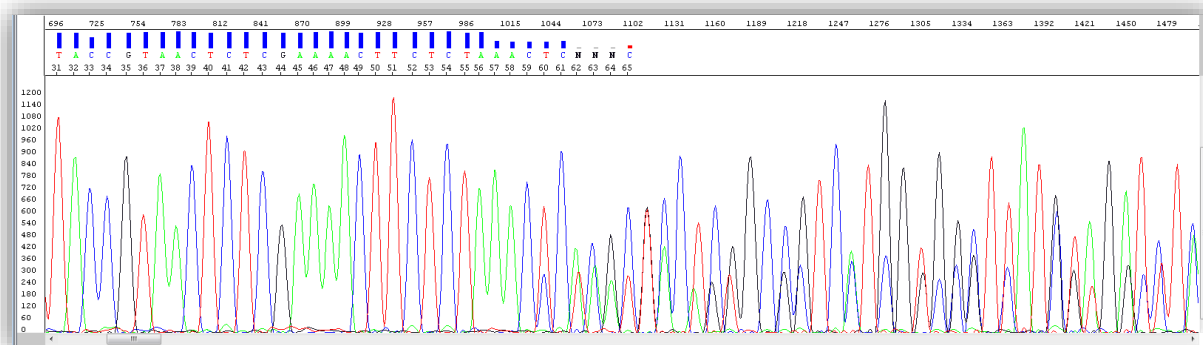
Figure 3-2: Superimposed signals follow a region of well resolved peaks

Possible cause 1: Mixture of plasmids in the sample
Solution: A plasmid preparation that contains more than one plasmid (for example two vectors with different inserts, or one vector with insert and one without) shows a clearly readable sequence with well resolved peaks at the beginning (sequencing section that is the same for both vectors), followed by a region with superimposed signals. (In rare cases, this is the result of unstable inserts or sections of it getting deleted spontaneously during the growth of the host strain.)
It is very important to pick a single (!) colony for plasmid preparation. In case of doubt, another streaking on a plate for single colonies should be performed to make sure that a single colony is picked. After preparation it should be checked again by DNA restriction analysis whether the expected bands show up.

**Possible cause 2: Mixture of PCR-products in a preparation**
For instance if two PCR products differ in one or several deletions or insertions, then superimposed signals occur after this indel.

Possible cause 3: Insert consisting of direct repeats
If the insert consists of direct repeats the primer will bind to every single repeat. The primer binding to the first copy will read into the second copy, whereas the primer binding to the last copy reads into the vector. This leads to superimposed signals from this point on.

Case 4: N-1, N-2, N-3… Primer Signals
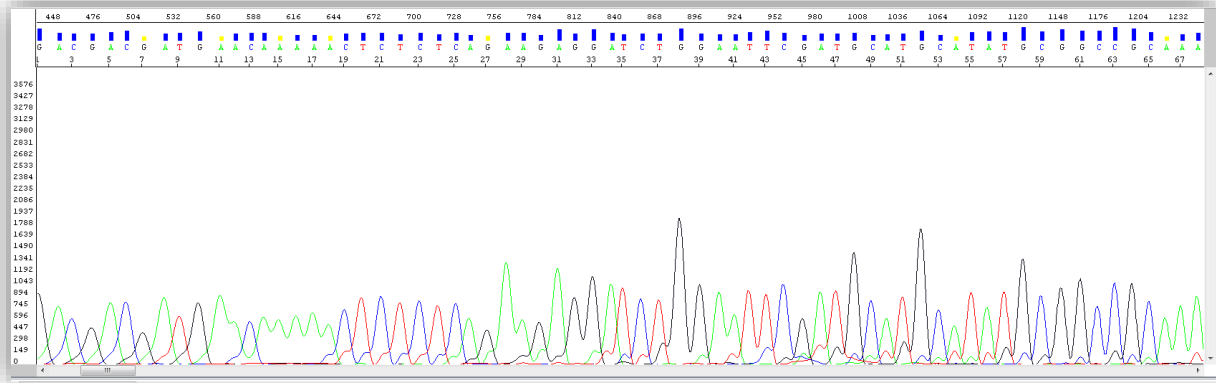
Figure 4-1: Minor N-1 signals
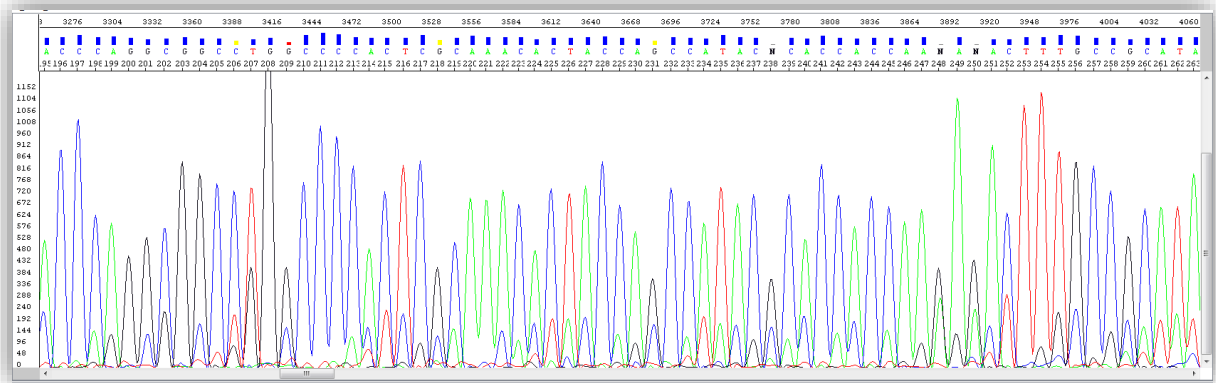
Figure 4-2: Clear N-1 signals

The sequence shows superimposed signals over the entire length. The second (usually smaller) peak is the same base as the next following real peak / base.

Possible cause: Improperly purified primer
Insufficient purification of primers during the synthesis process, leading to a certain amount of remaining n-1 primers truncated at the 5' end. When the template is sequenced, this remaining portion of n-1 primers bind to the template DNA and results in a sequencing, which is one base shorter than it should be.
Solution: n-1 primers should be resynthesized, to obtain a primer which is suitable for sequencing. Through proper storage the hydrolysis of a primer can be slowed down. We recommend storing the primer as stocks in 10mM Tris pH 8.0 at -20°C and to prepare fresh dilutions of the stocks before each sequencing.

Case 5: High Background / Noisy Signals
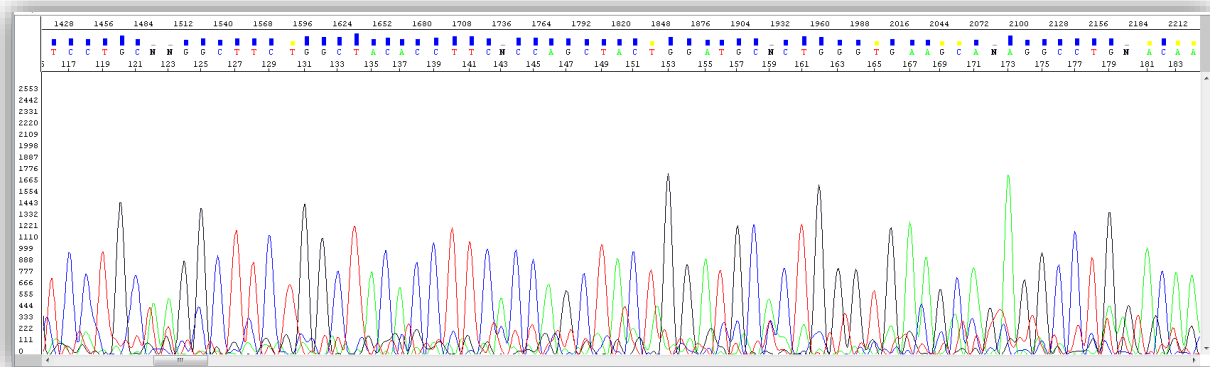
Figure 5-1: Sequence electropherogram depicting peaks within noisy background

"Noisy" sequences are easily recognizable by the multiple peaks and a large number of "N"s (if your sequencing preferences are set to "N for unclear bases").

An "N" can indicate the actual presence of two nucleotides in the case of a heterozygous sample, but is also displayed when multiple products or high background is present. If the signal strength of a sample is weak, the software tries to compensate this by raising all signals to detectable levels. As a result the background signals are raised as well, resulting in a poor signal to noise ratio. Background noise consists of many small unspecific peaks that run below the peaks of the target sequence. This background is always present, but will not be detected in samples with higher signal strength.

Possible cause 1: Not enough DNA
Solution: Check your concentration measurements, the concentrations of the stock solutions, your calculations and dilutions. Please note that 480 – 1200 ng plasmid DNA in a volume of 12µl (depending on the size of the template) are required for a sequence reaction (DNA concentration should be approximately 80ng/µl for plasmids and 1.5ng/µl per 100 bases for PCR products).

Possible cause 2: Inhibitory contaminants, such as salts or phenol
Solution: The cycle sequencing reaction is very sensitive to the presence of certain contaminants, some of which are capable to inhibit polymerase activity partially or even completely. Try to prepare the template again to obtain a better sequencing result.

Possible cause 3: Degraded DNA due to the effects of nucleases, repeated freeze/thawing, excessive exposure to UV light or bisulfite treatment.
Solution: Nuclease contamination in a template preparation or repeated freezing and thawing can degrade DNA over time. Even small amounts of nucleases can have a strong effect on DNA degradation, depending on temperature, storage conditions and time. In this case, only a re-isolation and purification of the DNA template will provide a better sequencing result. If PCR products are to be purified from an agarose gel, keep in mind that long exposure to UV light can lead to single strand breaks or degradation. Minimize the exposure time and UV intensity, in order to counteract degradation of DNA. If the DNA is treated with bisulfite for methylation experiments, make sure to avoid long incubation periods at high temperatures, because substantial amounts of DNA can be destroyed in the process.

Possible cause 4: Inefficient primer binding (low T$_m$, degenerate primers, mismatches)
Solution: The T$_m$ of a primer is defined as the temperature at which 50% of the oligonucleotide and its complementary partner form a duplex. The T$_m$ of an oligo can be roughly determined by the following equation: T$_m$ = 2°C (A+T) + 4°C (G+C)
In our standard cycle sequencing protocol annealing takes place at 50°C. If the T$_m$ of the primer used is much lower than 50°C, the hybridization to the complementary sequence of the template is decreased. As a consequence, a smaller amount of elongated fragments are generated. In such a case please increase the T$_m$ of your primer to approximately 52°C-58°C, for example by attaching additional bases to the 3 'or 5' end. Additionally, degenerate primers and primers with mismatches hybridize with lower efficiency due to the reduced stability of the primer to the template. If mismatches, or degenerate bases occur directly or close the 3 'terminus of your primer, it is very likely that the sequencing fails or produces a bad result.
Microsynth does not recommend primers with degenerate bases as sequencing primers.

Case 6: Truncated Sequence (due to secondary structure)
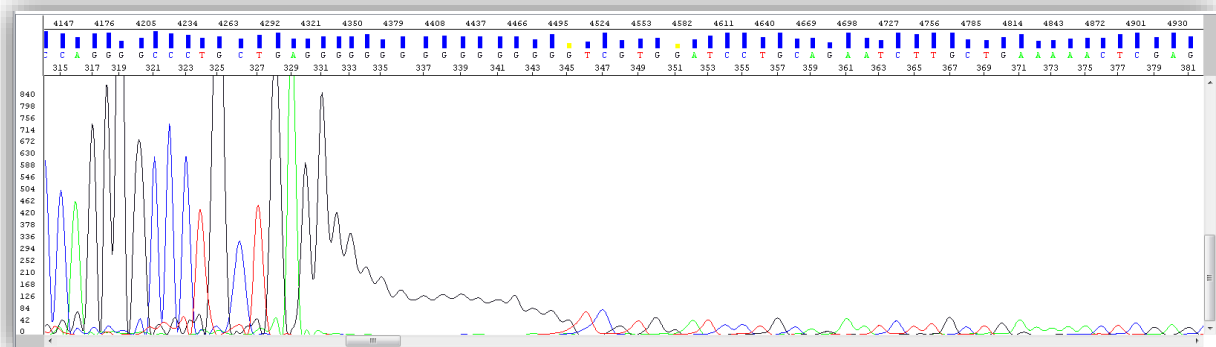
Figure 6-1: Electropherogram of a sequence with secondary structure illustrating a drastic drop in signal intensity. However the sequence stays readable.
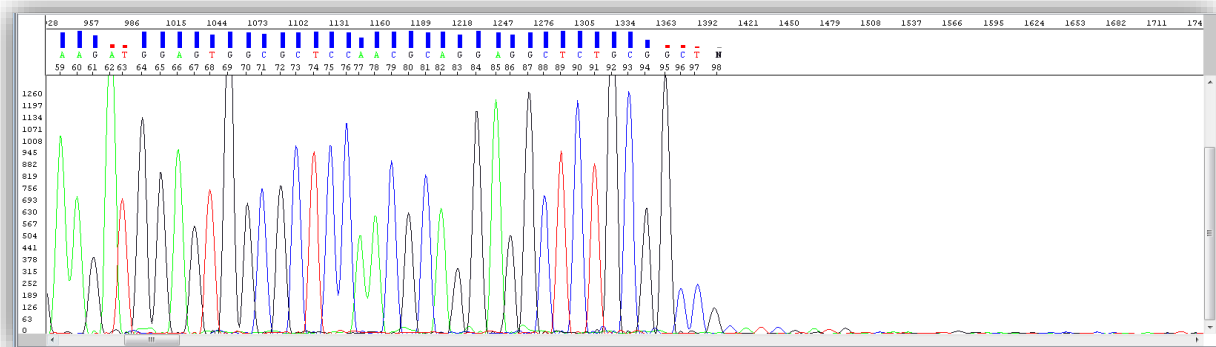
Figure 6-2: Electropherogram of a sequence with pronounced secondary structure. The quality of the sequence decreases immediately at the beginning of the secondary structure.

Abrupt interruptions are characterized by strong, clear signals up to a certain point at which they drop sharply and weaken within a few nucleotides to a no longer detectable signal.

Possible cause 1: Secondary structure within the template DNA.
GC rich and, to a lesser extent, AT rich DNA shows a strong tendency to form secondary structures, because the hydrogen bonds between G and C nucleotides promote the formation of hairpin structures. For example, the DNA can bend or loop, so that complementary segments can anneal. These hairpins can interfere with the function of the polymerase, or inhibit it completely. The enzyme dissolves from the template and the sequencing is terminated. Hairpin structures may not melt during a standard sequencing protocol, so that the following DNA strand is not sequenced. If the secondary structure is not dissolved, abrupt signal loss can be observed and consequently no sequence data is generated. If the secondary structure is at least partly dissolved, a sharp drop in signal intensity after an initial series of clear signals can be seen. With the latest BigDye Terminator sequence chemistry (v3.1), many GC associated problems have been greatly improved, but still not all the problems have been solved yet.

Solution 1: Place the sequencing primer as close as possible to the hairpin area to promote its unwinding, which in some cases can be successful.

Solution 2: Sometimes the sequencing of the complementary strand can give much better results.

Solution 3: We recommend to linearize the DNA with restriction enzymes in order to weaken the hairpin.

Solution 4: Another possibility is to make changes in the cycle sequencing protocol. Microsynth offers the following:

> **"GC-rich Treatment":** Please choose this option only when you are sure that your sample contains GC-rich sequences or hairpins. This service is free of charge.

> **"Premium Run Service":** Please consider re-sending your samples and selecting our http://www.microsynth.ch/de/10183/Premium-Run.html . Special protocols associated with this type of service are successful in more than 98% of all samples containing difficult regions.

Possible cause 2: linearized DNA
If the template DNA was digested with one or more restriction enzymes   or a PCR product is sequenced, the sequence information ends abruptly at the restriction site, or at the end of the PCR product, usually with a very high A signal.

Case 7: Gradual, Early Termination of the Sequence Signal, Repetitive Regions
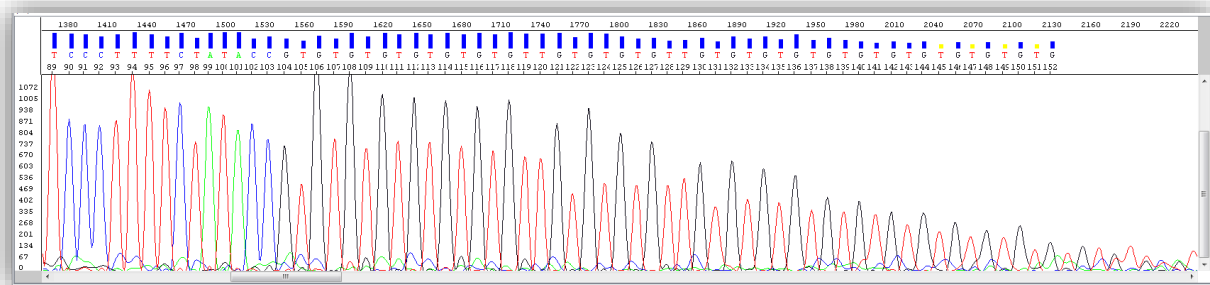
Figure 7-1: Repetitive GT-signals leading to gradually decrease of signal intensity

Repetitive areas (usually mono- or dinucleotide repeats) exhibit strong intensities at the beginning, but will decline during the process until the signals gets too weak.

Possible cause 1: "slides" (slippage) of DNA polymerase on the template during elongation.

Possible cause 2: formation of secondary structures due to the repeats

Possible cause 3: loss of dNTPs
The nucleotide composition, and the size of a repetitive region have a major influence on whether a sequencing through such an area is successful. In general, GC and GT cause the greatest difficulties, although the latest version of the Applied Biosystems BigDye Terminator v3.1 sequencing chemistry includes some modifications that considerable improve the ability to sequence templates previously classified as "difficult".

Case 8: Dye Blobs
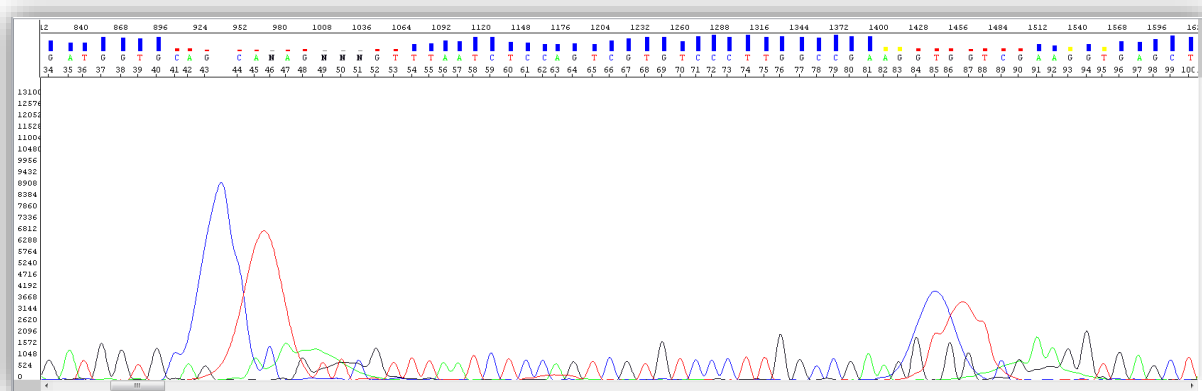
Figure 8-1: Dyeblobs in an otherwise background free sequence

Dye blobs appear as broad, undefined peaks of a single color. The actual peaks can be seen below the dye blob but are somewhat distorted and difficult to read. Dye blobs usually occur quite early in the sequence, approximately at position 70.

Cause: Dye blobs are caused by non-incorporated dye terminator molecules that have not been removed by the purification process and are then loaded with the sequencing products onto the sequencer.
>
> **Solution:** Dye blobs are usually observed in samples with low signal strength. Samples with a weak signal are caused by either 1) not enough DNA so that fewer template for the amplification and labeling reaction is present, hence after finishing the sequencing protocol, a larger proportion of unincorporated dye molecules remains; or 2) contaminating substances that inhibit the sequencing reaction. It is speculated that certain contaminants preferentially bind to the dye molecules. However, we have also observed that certain customer samples, partly depending on the type of template preparation, have a higher tendency of producing dye blobs, regardless of signal strength.

Case 9: Spikes
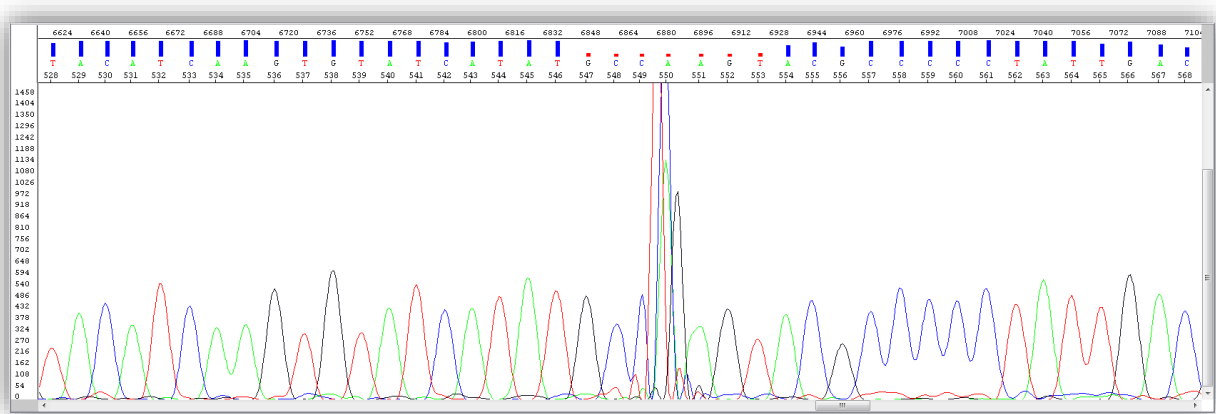
Figure 9-1: Sequencing electropherogram depicting a "Spike"

Spikes are multicolored, fused large peaks within the sequence that normally cover only one or two nucleotides.

Cause: Spikes are caused within the DNA sequencer either by tiny air bubbles in the polymer solution or small, polymerized polymer particles which flow accidentally into the capillaries. In samples with low signal strength they are more pronounced than in samples with high signal strength.

Solution: In general spikes no longer appear with reinjection of the same sample, so that the problem is solved by re-injecting the sample.

Case 10: Chromatogram Shows Widened Peaks
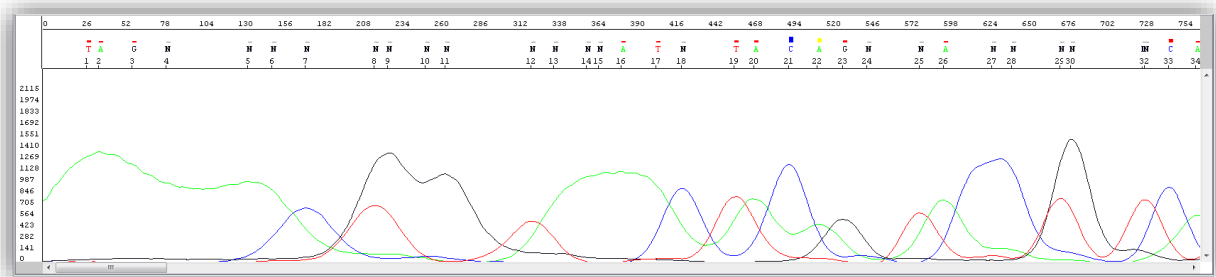
Figure 10-1: wide peaks caused by a blocked capillary, salts or anionic contamination.

The sequence is characterized by bands that gradually widen and thus can no longer be resolved. The first peak to broaden may already occur at position 1 (as is the case in the above figure), but sometimes the widening of the peaks appears only later in the sequence.

Cause: For this problem, there may be several reasons: (1) an air bubble or a block in the capillary, (2) excess salts, (3) anionic contaminant of a "kit-purified" plasmid or (4) DNA concentration is way too high.

Solution: For cause (1) a reinjection is the solution. If the cause is more likely in the items (2) or (3) the reaction can (a) be diluted and re-injected, (b) repeated with less template or (c) repeated after dialysis of the template. Cause (4) makes resequencing using less template necessary.

Case 11: Homopolymer Regions
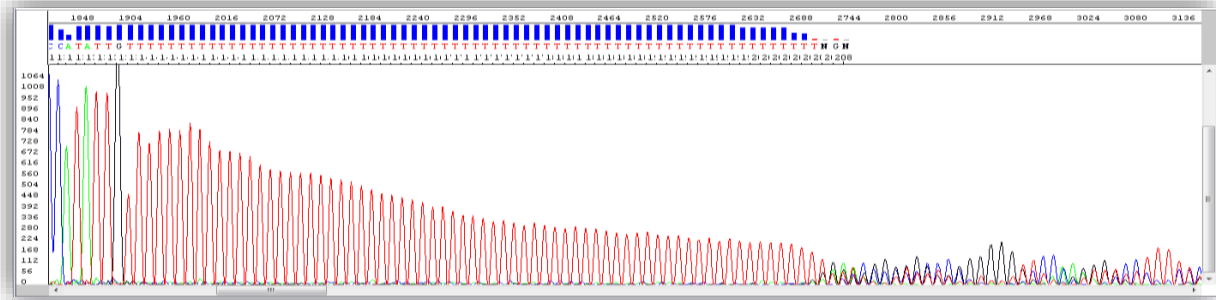
Figure 11-1: Deterioration of data quality after a polyT Region
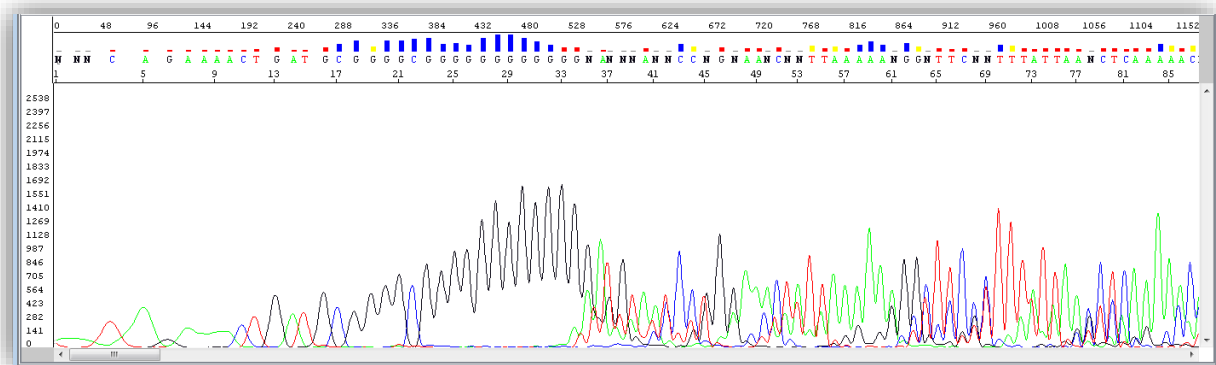
Figure 11-2: Deterioration of data quality after a polyG region.

The two images show the worsening of the sequence quality for a polyT area (Figure 11-1) and a polyG area (Figure 11-2). The sequence up to the polynucleotide stretch can be very good and clear, but the last base of repeats and all subsequent peaks may have a wavy, "stuttering" pattern of double signals, which can no longer be interpreted unambiguously. In PCR products or other linear templates this effect is usually very strong, but it can also occur in plasmids, especially when the polyA region of cDNA is sequenced.

> **Cause:** It is believed that these difficulties are caused by "slipping" of the growing strand during elongation relative to the template strand ("strand slippage") as the homopolymer region no longer properly hybridizes to the template strand. This produces fragments of different lengths within the same sequence after the homopolymer region.
>
> **Solutions:** For sequencing cloned DNA containing a homopolymer region, there are several options that can lead to a successful sequence. Sequencing of the complementary strand can be more successful, particularly if a polyG region is present. A polyC region can often be sequenced easier.
>
> Oligo dT (15-20T) and dA (15-20A) primers, which have either a defined or wobble base (A / T, G or C) at the 3'-end, hybridize at the end of a homopolymer region and can result in an unambiguous sequence.
>
> It is advisable to clone PCR products containing homopolymers in order to obtain proper sequence data.

Case 12: PCR Products with Overdriven Signals
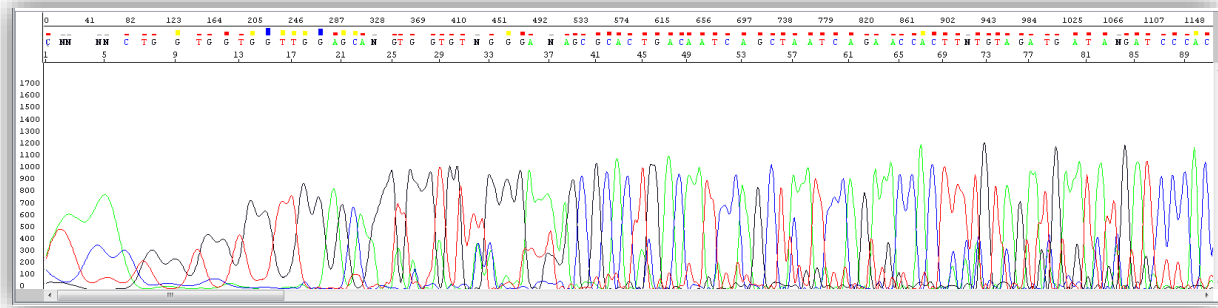
Figure 12-1: Sequencing electropherogram depicting PCR product with heavily overdriven signals leading to completely distorted peaks.
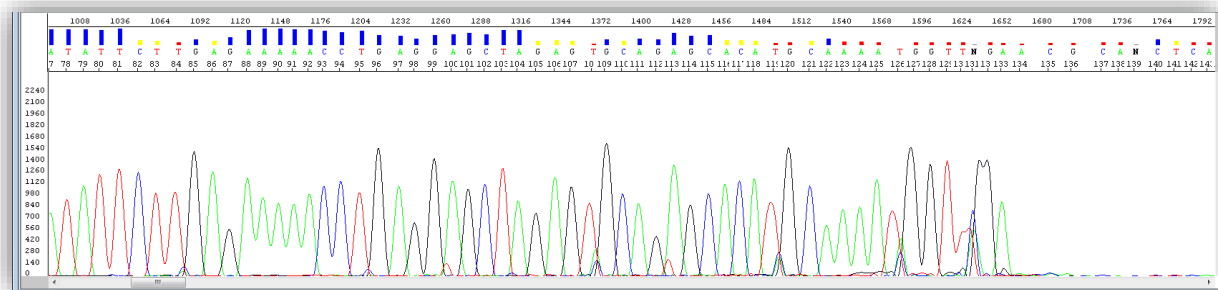
Figure 12-2: Sequencing electropherogram depicting PCR product with slightly overdriven signals. Small signals in the underground show up, indicating that the matrix calculation, which separates the individual fluorescent colors, is overdriven.

DNA sequencers generate clean signals in a wide range of intensities. But if the upper resolution limit is exceeded, distortion is generated, just as you know it from an overdriven amplifier.

> **Cause:** DNA concentration is far too high. Especially short PCR products, where a large number of short fragments are generated, are prone to overdriven signals and thus distorted peaks.

> **Solution:** Please use the amounts specified by Microsynth when sequencing PCR products (1.5 ng/µl per 100 bases).

Important Remark: We're Here to Help!

If you have any further questions regarding the results of your sequencing data or you have problems with their interpretation, please give us a call immediately. Our experienced staff members will assist you. This way problems can be clarified quickly and with the right advice sequencing results can be improved easily in many cases. In order to find your sequences quickly, it is helpful if you can specify the order number or a barcode number.