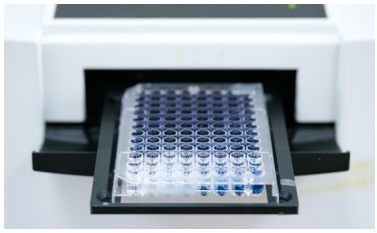
Microsynth

Troubleshooting Guide:
Understanding Your Sequencing Results

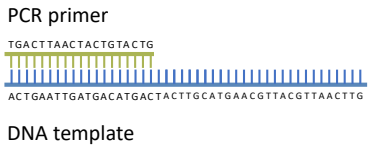
Dear Valued Customer

Microsynth always strives to get the most out of your DNA sequencing samples. Optimized processing steps, proven protocols, and over 25 years of experience in the area of Sanger sequencing usually result in long read lengths and clean sequences. Nevertheless, some factors can affect the successful generation of a high-quality sequence. However, before we review our compilation of the most common problems experienced in ABI Sanger sequencing, including example chromatogram images and descriptions of possible causes and solutions, we would like to share four critical tips with you. When considering and mastering these basics, it is highly likely that you will obtain the desired sequencing result.



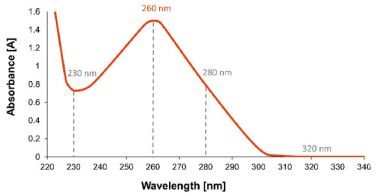
01: Pay Attention to the Optimal DNA Sample and Primer Concentrations

- Provide 40-100 ng/µl for plasmids or 1.5 ng/µl per 100 bases for PCR products (for volumes, please see our user guides for the desired service).
- Provide 10 µM primer in a separate 1.5 ml tube or premix 3 µl of 20 µM primer with your template.



02: Use Well-designed Sequencing Primers

- Primer length should be approximately 20 bases (± 2)
- G/C content should be around 50%
- Further parameters can be viewed in the primer design guidelines.



03: Start with Pure DNA Template

- Use commercial isolation kits
- Re-suspend or elute DNA in water or Tris-HCl. Avoid EDTA (TE buffer)
- Determine purity by A260/A280 and A260/A230. The optimal range is 1.8 to 2.0 for both.

04: Use our Ecoli NightSeq® Service for Plasmid Sequencing

- Instead of isolating the plasmid yourself, you send us part of your *E. coli* colony or 5 µl of your bacterial solution in our Ecoli NightSeq® tubes or plates. This means you only need to take responsibility for the sequencing primers.
- It isn't a problem if you only have very low amounts of plasmid DNA. Simply provide us with 1 µl of your plasmid DNA.

Index

Good Sequencing Reaction, Reliable Sequence Data

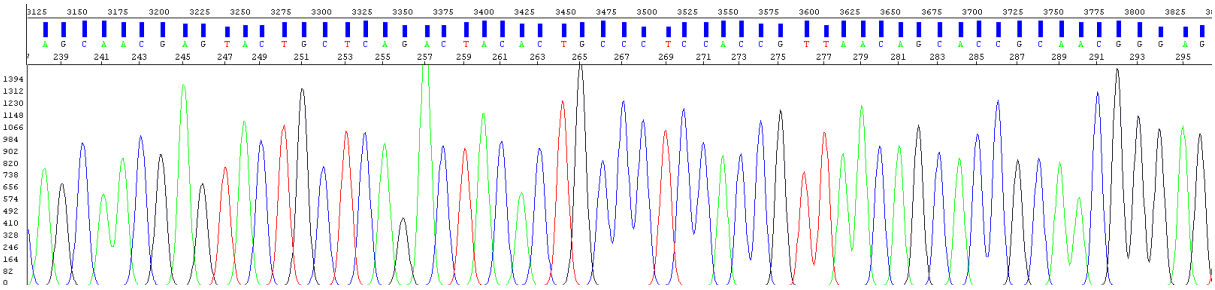


***Figure 1**: Electropherogram of a high-quality DNA sequencing reaction. It shows distinct peaks and no or very low background noise.*

Microsynth provides helpful user guides for all Sanger services that are listed on the Microsynth website in our customer support/download area. These will give you detailed descriptions of how to obtain reliable sequencing results.

Understanding and solving failed sequencing reactions and/or problematic sequencing data requires a thorough analysis of the corresponding sequencing chromatograms. Please see the following pages for a more comprehensive discussion of major sequencing failures and the related troubleshooting actions!

You will learn how to interpret your Sanger results. Once you have received the Sanger results, some questions that may arise can be answered in this document. The chromatograms show different phenomena, possible reasons are listed, and hints for making improvements are given.

Background Knowledge:

- Sample Requirements
- Primer Design Guidelines
- Ecoli NightSeq® User Guide

No Sequencing Reaction, Unreliable Sequence Data

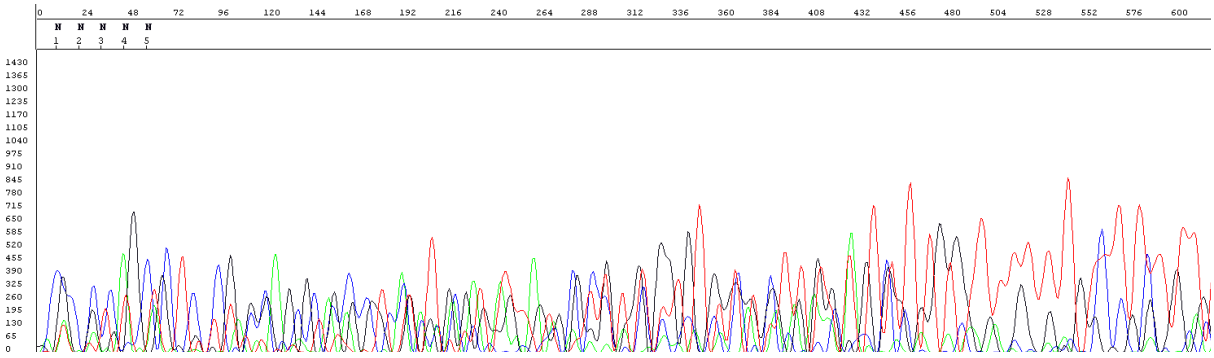


Figure 2: *No sequence information could be deduced.*

No sequence information could be deduced, which means that the base calling failed and only NNNNN has been annotated. Signal strength is very low (just background noise).

Possible Cause	Action
Primer binding site not available	• Check whether the standard primer fits your vector. • Double-check your primer design. • Double-check your construct (e.g. restriction enzyme digests).
Inappropriate amounts of DNA and / or primer	• Double-check the DNA concentration of your sample and/or primer (please be aware: photometrical measurements can often lead to an overestimation of DNA amounts!) • Keep the required DNA concentration ranges.
Presence of inhibitory contaminants	• Use commercial isolation kits. • Purify the sample through columns to remove contaminants like salts, EDTA, alcohol, proteins, RNA, detergents, cesium and phenol. • Dilute excessive highly concentrated DNA (even excessively high amounts of DNA inhibit the sequencing reactions.) • Use our Ecoli NightSeq service for plasmids.

Tip: Microsynth's Primer Search Tool

- Enables the easy searching of standard primer binding sites on your template. Simply paste your reference sequence or vector sequence into the online application on our website.

In-depth Knowledge:

The last 5 -7 bases of the 3' end of your sequencing primer are the most critical. They need to fit 100 %.

Superimposed Signals from Start to End

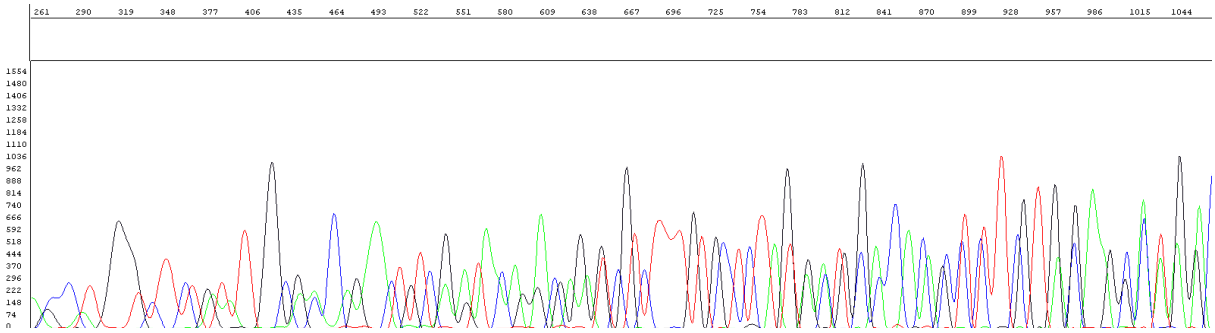


Figure 3: *Superimposed signals from the beginning until the end of the sequence.*

Possible Cause	Action
The average signal strength is too low	• Double-check the average signal strength. Average low signal strength (<200) could be linked to background signals (-> see no sequencing reaction section).
Primer binding site is not unique	• Double-check if the primer binding site is unique. Multiple primer binding causes superimposed signals.

Solution: Find a primer that binds uniquely to your sample. If sequences are very repetitive, try to find a primer with one (several would be better) bases at the 3' end that are specific to the target sequence and thus act as a sort of "anchor".

In-depth Knowledge:

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, which leads to double peaks that are not congruently superposed, but instead 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.

Possible Cause	Action
Unspecific PCR amplification while generating the PCR product	• Avoid unspecific PCR amplification while generating the PCR product. Either one or both PCR primers bind to more than one position on the template DNA leading to multiple, mixed PCR products.

Tips: if mixed PCR products of similar or even identical length cannot be separated properly in the gel:

- Enhance the specificity of PCR amplification by optimizing PCR conditions
- Redesign of the PCR primers
- Subclone PCR products before sequencing
- If possible, nested sequencing primers can be used to sequence just the PCR product of interest

In-depth Knowledge:

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 in size between the PCR products, the target template should be purified from the gel.
Avoid a situation whereby one of the PCR primers works as a forward and reverse primer while generating the PCR product. 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	Action
Residual primers and/or dNTPs present in the PCR sample	• Purify PCR products from PCR primers and dNTPs.

Solution: order PCR purifications by Microsynth

In-depth Knowledge:

It is essential to optimize the PCR reaction to ensure primers and dNTPs are not added in excess and that they are almost completely consumed 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	Action
Primers with a high melting point	• Avoid sequencing primers with a high melting point. Redesign the primers.

Solution: use the software Oligo Calculator provided on our website to calculate melting points

In-depth Knowledge:

Primers with a very high melting point (Tm> 65 ° C) are not recommended for sequencing. Those primers are often GC rich or very long and these two factors increase the ability of the primer to form secondary structures and/or dimers. We recommend that primers are designed 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)

Superimposed Signals occur within the Sequence

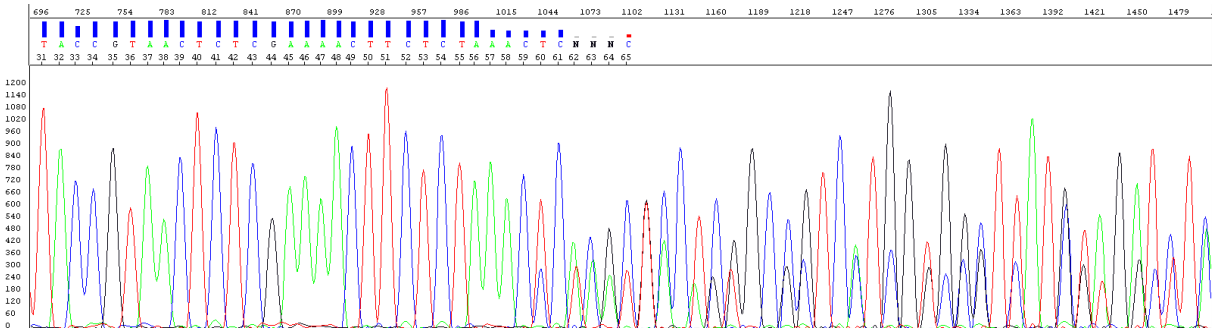


Figure 4 *Superimposed signals follow a region of well resolved peaks.*

Possible Cause	Action
A mixture of plasmids in the sample	• Pick a single colony for plasmid preparation.
A mixture of PCR-products due to insertions/deletions	• For instance, if two PCR products differ in one or several deletions or insertions, then superimposed signals occur after this indel.
Insert consists of direct repeats	• If the insert consists of direct repeats then 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 onwards.

Tip: order untrimmed data to receive entire sequence information

- For CRISPR/Cas sequencing, in particular, you need all sequence information to analyze using TIDE

In-depth Knowledge:

In case of doubt, you should re-streak your bacterial colonies onto a new agar plate to make sure that only single colonies are picked. After plasmid isolation, it should be checked again by DNA restriction analysis and subsequent gel electrophoresis to determine whether the expected inserts are present or not.
A plasmid preparation that contains more than one plasmid (for example two vectors with different inserts or one vector with insert and one without) typically shows a readable sequence with well-resolved peaks at the beginning (resulting from the plasmid backbone), followed by a region with superimposed signals. In rare cases, this can also be a result of unstable inserts that mutate during the growth of the host strain, e.g. when the insert is toxic or it causes other negative selection pressures.

Superimposed Signals - Smaller Peaks Ahead

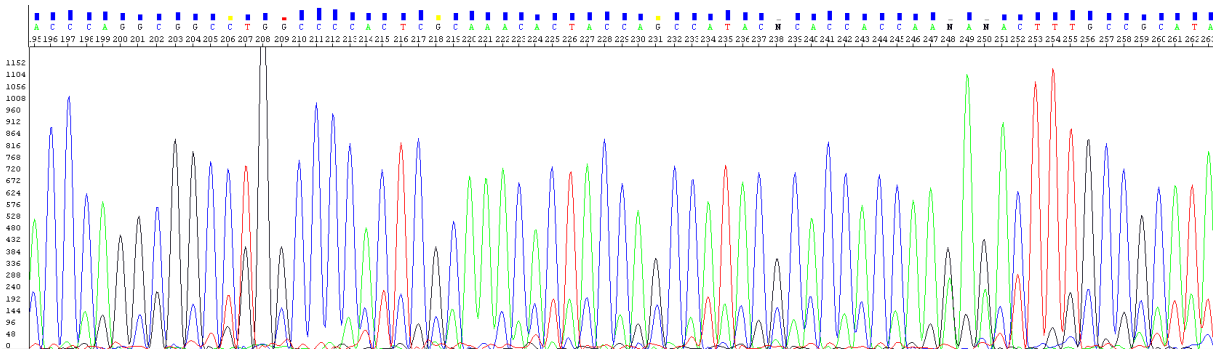


Figure 5: *Smaller peaks show the same base as the next real peak/bases.*

Possible Cause	Action
Improperly purified or degraded sequencing primer	• Insufficient purification of primers during the synthesis process lead to a certain amount of remaining n-1 primers that are truncated by one base at the 5' end. When the template is sequenced, these remaining n-1 primers bind to the template DNA and lead to labeled DNA fragments that are precisely truncated by this one missing base. • Inappropriate storage conditions of primers.

Solution: order a new primer

In-depth Knowledge:

Through proper storage the hydrolysis of a primer can be slowed down. We recommend storing primers as stocks in 10mM Tris pH 8.0 at -20°C and to prepare fresh dilutions of stocks before sequencing.

High Background / Noisy Signals

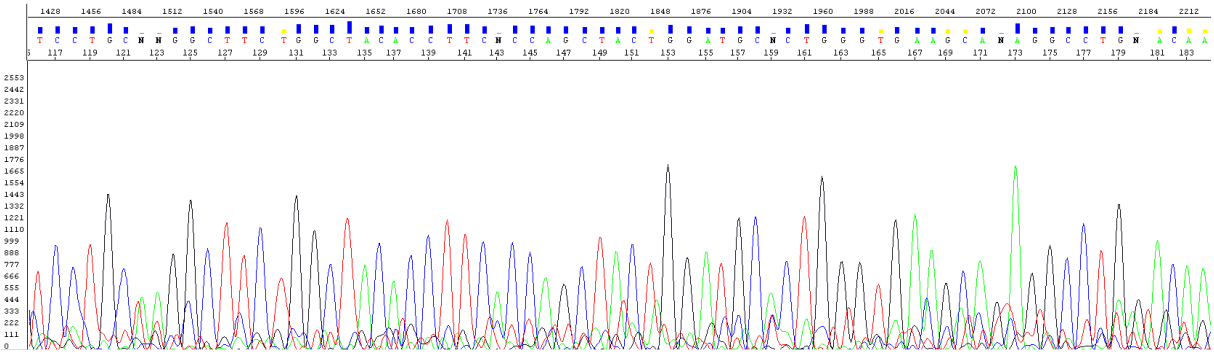


Figure 6: *Sequence electropherogram depicting peaks within a noisy background.*

"Noisy" sequences are easily recognizable by multiple peaks, a large number of "Ns" (if your sequencing preferences are set to "N for unclear bases"), and in average low signal strength.

Possible Cause	Action
Wrong DNA concentration	• Double-check the DNA concentration. The DNA concentration should be approximately 80 ng/µl for plasmids and 1.5 ng/µl per 100 bases for PCR products.
Inhibitory contaminants, such as salts or phenol	• Purify your DNA via column-based protocols. • Try to prepare your template DNA again to obtain a better sequencing result. The cycle sequencing reaction is very sensitive to the presence of certain contaminants, some of which are capable of inhibiting polymerase activity either partially or even completely.
Degraded DNA due to the activity of nucleases, repeated freeze/thawing, excessive exposure to UV light or bisulfite treatment.	• Prepare new samples.
Inefficient primer binding (low Tm, degenerated primers, mismatches)	• Double-check primer properties. • Design new sequencing primers.

In-depth Knowledge:

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 base calling software will attempt to compensate for this by raising all signals to detectable levels. As a result, the background signals are enhanced as well, which results 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 it will not be detected in samples with higher signal strength.

Nuclease contaminations within template DNA or repeated freezing and thawing events can lead to degradation over time. Even small amounts of nucleases can have strong effects on DNA integrity, 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 DNA need to be purified from an agarose gel, please keep in mind that long exposure to UV light can lead to single-strand breaks, degradation, and mutations. You should minimize the exposure time and UV intensity to counteract the degradation of DNA. Or even better, avoid UV exposure: if you load a small volume of your template DNA next to the preparative section of your gel, you can just expose this part of the gel to UV light, mark the band of interest by cutting it out, and then cut your preparative part next to it after switching off the UV light. If the DNA is treated with bisulfite for methylation experiments, you should make sure that you avoid long incubation periods at high temperatures. This is because substantial amounts of DNA can be destroyed in the process.

The Tm of a primer is defined as being the temperature at which 50% of the oligonucleotide and its complementary partner form a duplex. The Tm of an oligo can be roughly determined by the following equation: Tm = 2°C (A+T) + 4°C (G+C)
In our standard cycle sequencing protocol, annealing takes place at 50°C. If the Tm of your sequencing primer is much lower than 50°C, the hybridization to your template is decreased. Consequently, a significantly smaller amount of elongated fragments are generated. In such instances, please increase the Tm of your primer to approximately 52°C-58°C, for example by attaching additional bases to the 3' or 5' end. In addition, degenerated primers and primers with mismatches hybridize with lower efficiency due to reduced affinity between primer and template. If mismatches or degenerated bases are present directly at or close to the 3' terminus of your primer, it is very likely that sequencing fails or gives a low-quality result. Microsynth does not recommend primers with degenerated bases for DNA sequencing.

Tip: use Ecoli NightSeq® for difficult to isolate and low quantity plasmids

Truncated Sequence (due to demanding structures)

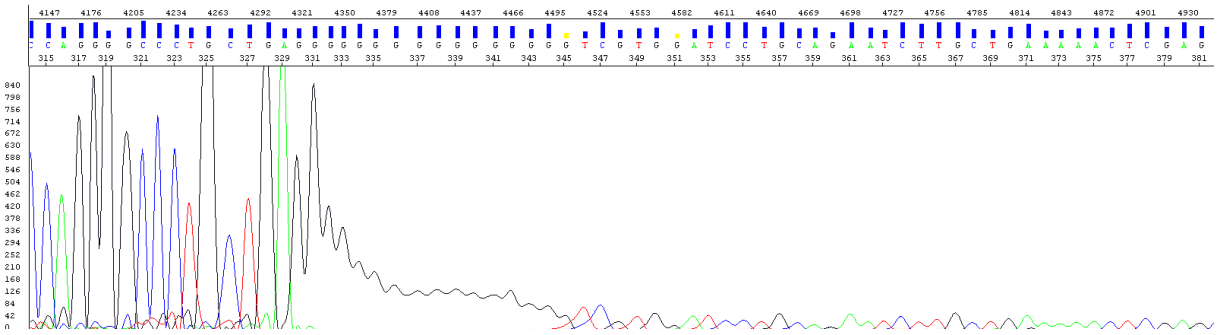


Figure 7: *Repeats of C or G cause drastic drops in signal intensity. However, the sequence remains readable.*

Possible Cause	Action
C or G repeats	• Sequence from the opposite side towards this repetitive stretch. • Send the plasmid sample and primer separately. • Standard sequencing at Microsynth is optimized for GC-rich sequences as well as hairpin structures. If challenging structures cause severe signal drops your reaction will be repeated by applying a special treatment.

Truncated Sequence (due to hairpin structures)

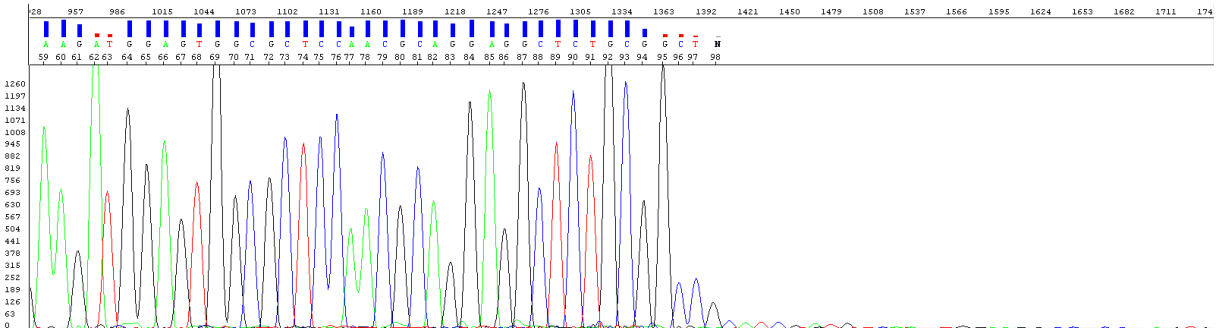


Figure 8: *Hairpin structure causes a severe signal drop immediately at the beginning of the secondary structure.*

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

Possible Cause	Action
Hairpin structures	• Use different sequencing primers. • Place the sequencing primer as close as possible to the hairpin area to promote its unwinding. • Sequencing of the complementary strand can sometimes provide much better results. • Send plasmid sample and primer separately. • Standard sequencing at Microsynth is optimized for GC-rich sequences as well as hairpin structures. If challenging structures cause severe signal drops your reaction will be repeated by applying a special treatment.

Tip: Linearize the plasmids with restriction enzymes

- Demanding structures are more accessible in linearized DNA

In-depth Knowledge:

GC rich DNA, and to a lesser extent AT rich DNA, shows a strong tendency to form secondary structures. This is 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 either interfere with the function of the polymerase or inhibit it completely. The enzyme dissolves from the template and sequencing is terminated. Hairpin structures may not melt during a standard sequencing protocol, meaning 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 partially dissolved, a sharp drop in signal intensity after an initial series of clear signals can be seen. With the latest BigDye Terminator cycle sequencing chemistry (v3.1), many GC-associated problems have been greatly improved. However, to date, not all challenges have been overcome.

Truncated Sequence (due to repetitive sequence patterns)

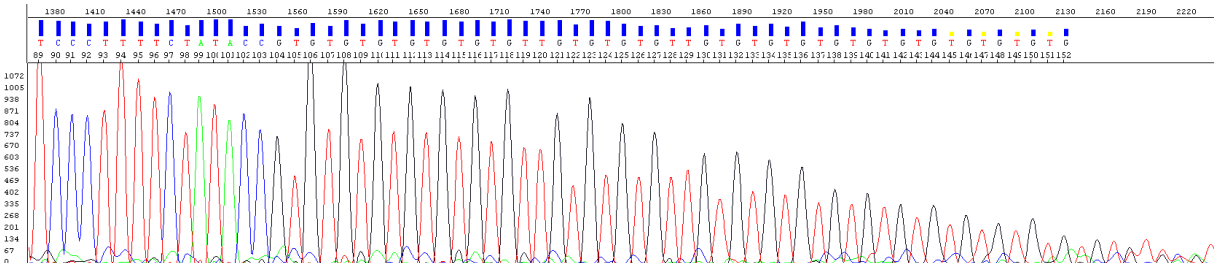


Figure 9: *Repetitive GT-signals leading to a gradual decrease of signal intensity.*

Typically, repetitive areas (usually mono- or dinucleotide repeats) lead to strong signal intensities at the beginning of the sequence, which will decline rapidly until the signals become undetectable.

Possible Cause	Action
Formation of secondary structures due to the repeats	• The nucleotide composition and the size of a repetitive region have a major influence on whether 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 considerably improve the ability to sequence templates that were previously classified as "difficult".

Dye Blobs

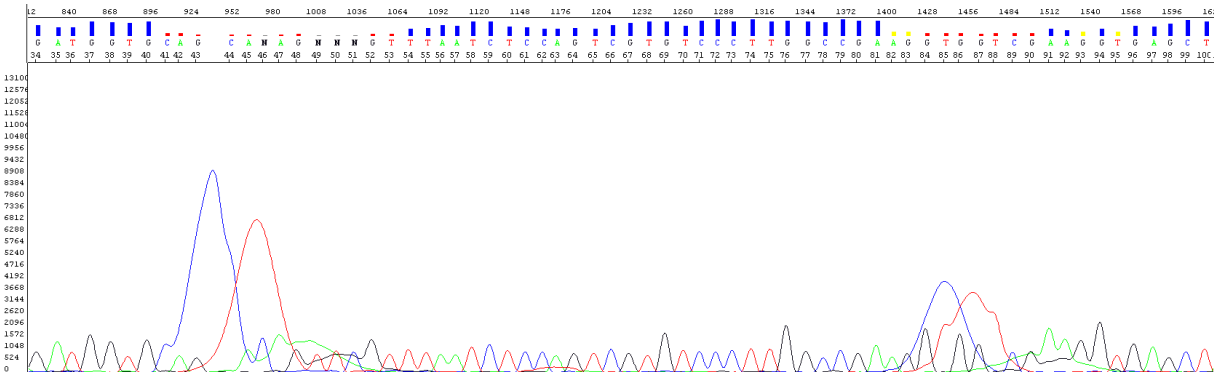


Figure 10: *Dye blobs in an otherwise background free sequence.*

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 (in this example at positions 40 and 80).

Cause: sequencing artifact. Dye blobs are caused by non-incorporated dye terminator molecules that have not been removed during the purification process and consequently co-migrate during capillary electrophoresis.

Solution: request a reload or repetition of the reaction (if your DNA concentration is in our required range).

In-depth Knowledge:

Dye blobs are usually observed within sequencing reactions 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 to produce dye blobs, regardless of signal strength.

Spikes

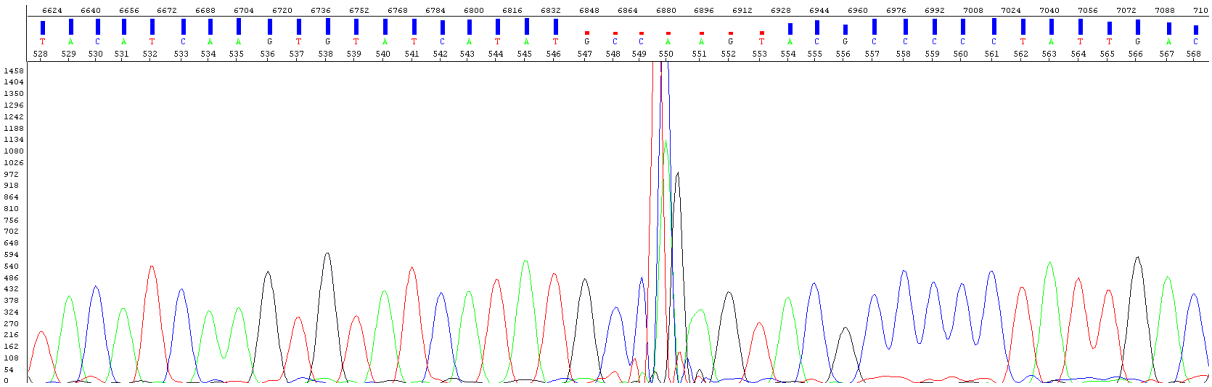


Figure 11: *Sequencing electropherogram depicting a "Spike"*

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

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

Solution: request a reload or repetition of the reaction (this will usually be done automatically)

In-depth Knowledge:

In general, spikes no longer appear after the reinjection of your sample. Therefore, this problem can be solved quickly.

Poor Resolution

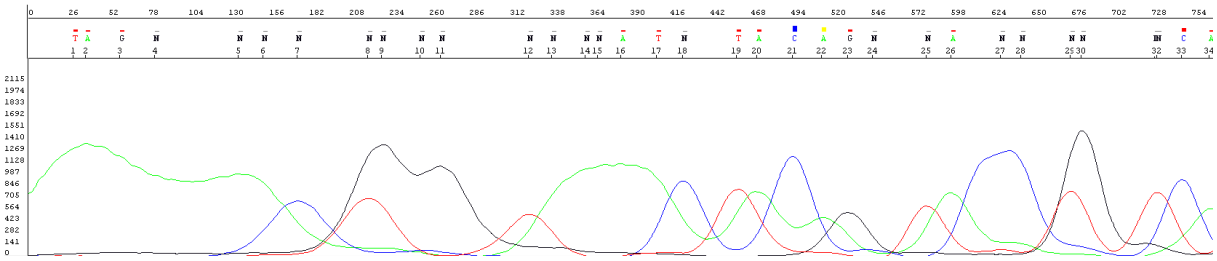


Figure 12: *Wide peaks caused by overloaded or blocked capillaries, salts or anionic contamination.*

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.

Possible Cause	Action
DNA concentration	• Double-check the DNA concentrations. • Send the DNA concentrations in the required range, not too highly concentrated DNA.
Capillary problem	• Wait for the result of the repetition. • If the poor resolution is caused by a capillary problem your reaction will be automatically repeated.

In-depth Knowledge:

There could be several reasons for this problem: (1) an air bubble or a block in the capillary, (2) excess salts, (3) anionic contaminant of a "kit-purified" plasmid, or (4) the DNA concentration is far too high.
For cause (1), a reinjection is the solution. If the cause is more likely to fall within items (2) or (3) then the reaction can (a) be diluted and reinjected, (b) repeated with less template, or (c) repeated after dialysis of the template. Cause (4) makes resequencing using less template necessary.

Homopolymer Regions

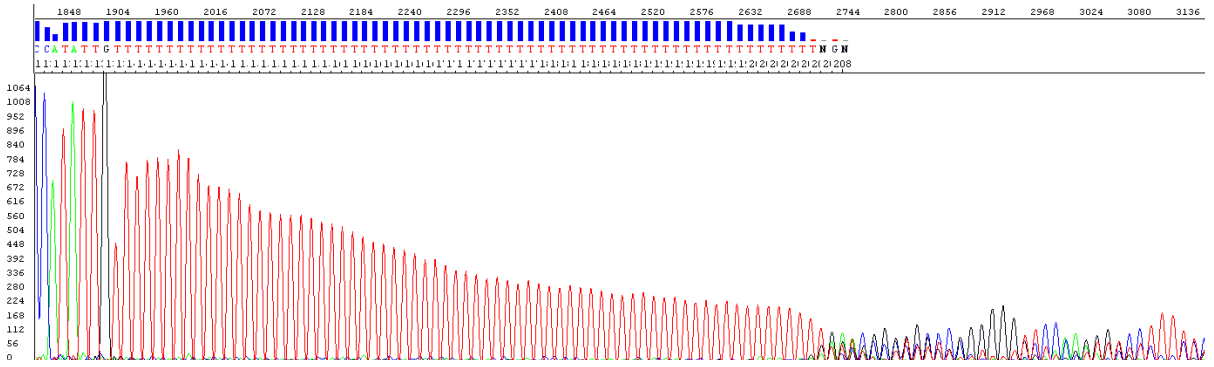


Figure 13-1: *Decline of data quality after a polyT region.*

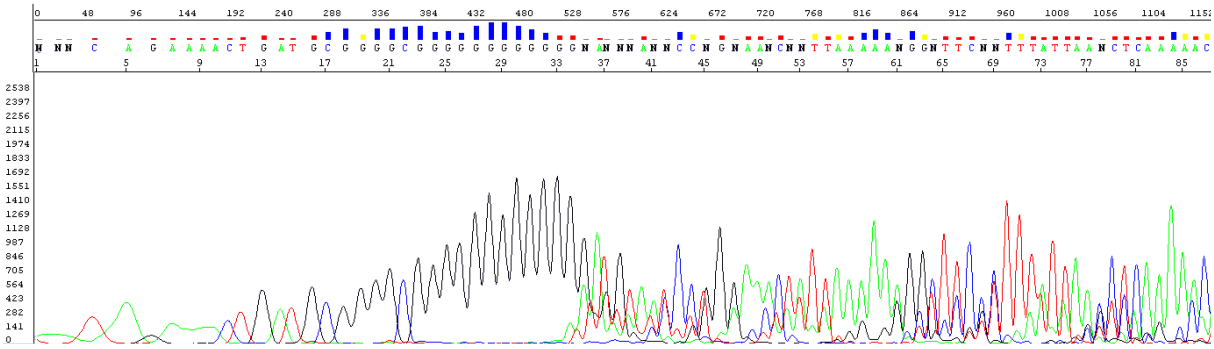


Figure 13-2: *Decline of data quality after a polyG region.*

The decline of data quality after a homopolymer region.

Possible Cause	Action
"Slides" (slippage) of DNA polymerase on the template during elongation due to homopolymer region	• Sequence from the opposite side.
	• Sequence with alternative primers. • Use the polyT A standard primer. • Design a follow-on primer, 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 that can result in an unambiguous sequence.
	• Sub-clone PCR products before sequencing.

In-depth Knowledge:

This effect is usually very strong in PCR products or other linear templates, but it can also occur in plasmids, especially when the polyA region of cDNA is sequenced.
It is believed that these difficulties are caused by the "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.

PCR Products with Overdriven Signals

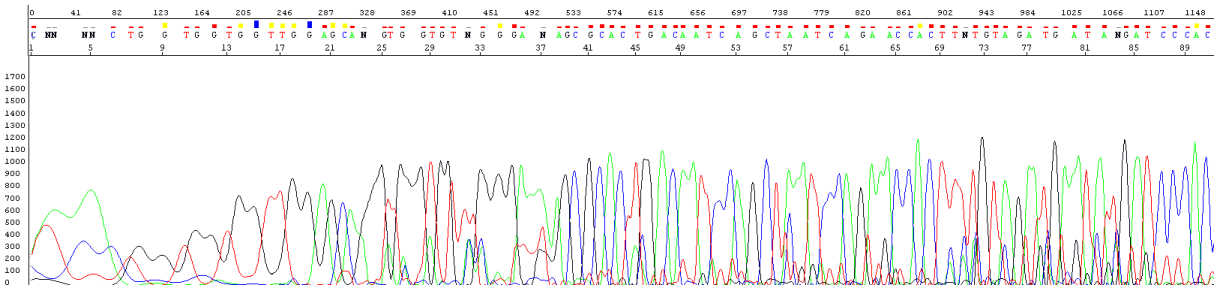


Figure 14: *Sequencing electropherogram depicting a PCR product with heavily overdriven signals leading to completely distorted peaks.*

Possible Cause	Action
DNA concentration is too high	• Send lower concentrated DNA (within our required range) • Dilute your samples and please send them again.

In-depth Knowledge:

DNA sequencers generate clean signals in a wide range of intensities. However, if the upper-resolution limit is exceeded, distortion is generated, as can happen with an overdriven amplifier.
The DNA concentration is far too high. In particular, short PCR products, where many short fragments are generated, are prone to overdriven signals and thus to distorted peaks.

Sequencing Options & Preferences

In general, data is analyzed by base callers that have various parameter adjustment options.
It is possible to adapt the parameters according to your own needs. Therefore, the most important keywords are explained:

Base calling is the process of assigning bases to chromatogram peaks via sequencing analysis software. The de facto standard software for the analysis of sequencing data is the KB base caller from Applied Biosystems (ABI), which is known to consistently produce quality outputs. However, in many cases, the quality and length of sequence reads can be further improved through the use of third-party software.

"Optimized" base calling: by analyzing the important parameters of the initial sequencing output, Microsynth Seqlab can evaluate whether the application of additional software tools will be beneficial or not. If all requirements for additional software use are fulfilled, you will receive longer sequence reads with augmented quality values. Otherwise, sequencing results generated via the KB base caller ("Conservative" base calling) are provided.

"Conservative" base calling: as the name suggests, the ABI KB base caller is more conservative in converting chromatogram peaks to bases. This software is superior in its ability to detect ambiguous base positions and it should be your preferred choice if your aim is mutational analysis or sequencing of short PCR fragments (up to 800 bp).

IUPAC Basecalling: this setting is similar to the option "N for Unclear Bases" with the difference being that the full range of the IUPAC code table is used to further distinguish between ambiguous positions (R, Y, K, M, S, W, B, D, H, V, and N). The assignment of IUPAC bases is particularly helpful if you are looking for heterozygous mutations.

N for unclear bases: the base calling process usually assigns one of the four-letter codes A, C, G, or T to peak positions in the chromatogram. With this setting, you can decide whether your sequence should contain N for ambiguous bases*, or if the analysis software should make an assignment in any case.

***Quality values** (these are often referred to as QV or Phred values) are calculated for each position and they are used for the assignment process.

Sequence trimming: automated DNA sequencers occasionally produce poor-quality reads. This is particularly the case near the sequencing primer site and towards the end of longer sequence runs. If you choose sequence trimming, low-quality regions (misleading data) will be removed from the ends of your sequences.

Tip: for CRISPR/Cas samples, it is suitable to use untrimmed data.

Important Info: We're Here to Help!

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

Microsynth AG
Schützenstrasse 15
9436 Balgach
Switzerland
Phone: +41 71 726 10 04
Email: sanger.support@microsynth.ch

Microsynth Seqlab GmbH
Maschmühlenweg 36
37081 Göttingen
Germany
Phone: +49 551 37 000 15/17
Email: info@microsynth.seqlab.de

Microsynth Austria GmbH
Leberstrasse 20
1110 Wien
Austria
Phone: +41 71 726 10 04
Email: sanger.support@microsynth.ch