

ARTICLES

Fundamentals of Sequencing of Difficult Templates—An Overview

Jan Kieleczawa

Wyeth Research, Biological Technologies Department, Cambridge, MA

Despite enormous progress, the DNA sequencing of difficult regions, whether they are part of complex genomes or individual constructs, still presents a significant challenge and may require many trials, brute-force, or the intervention of a very experienced sequencer (and sometimes all of the above). Very early on, it was realized that sequencing of various types of difficult templates requires distinct treatments. To sequence through GC-rich regions, the addition of DMSO, NP-40/Tween-20 detergents, or the mix (4:1) of BD3.0:dGTP3.0 was sometimes helpful. To get through long poly-A/T tails, sometimes one would be successful using tailored poly-A/T (V/B) N primers or primers that spanned part of pre-tail and tail regions. A few years ago Invitrogen introduced a set of sequencing additives that proved to be useful for many different types of difficult templates.

KEY WORDS: difficult templates, heat denaturation, DNA sequencing.

In this paper, we present data using a modified ABI DNA sequencing protocol that in combination with Invitrogen's and others' reagents allows for a clean read through many complex DNA regions. The major part of this modification is the incorporation of a 5-min controlled-heat denaturation step of the template in the low-salt buffer and in the presence or absence of various additives. A collection of over twenty difficult templates was tested under several conditions, and in 7 out of 22 cases no data at all were obtained using a standard ABI-like sequencing protocol. By adding just a heat-denaturation step we were able to get 300–800 good-quality bases.

The last three decades have witnessed amazing progress in the development of many tools and techniques in molecular biology, one of which is DNA sequencing. Since its humble beginnings,^{1,2} when it still took a considerable amount of skill and effort to sequence even a relatively small molecule, DNA sequencing is now almost as routine as, for example, gene cloning or polymerase chain reaction (PCR) technology. Quite often, the only barrier to performing any sequencing experiment is the cost of a sequencing instrument itself (which ranges from

below \$100,000 to about \$500,000, depending on manufacturer and required throughput). In the simplest incarnation of sequencing, all one needs to do is to combine a few components (DNA, primer, water, additives, if any, and detection mix), perform cycle sequencing, purify, and run on a detection instrument. If all is carried out correctly, one can expect up to 1000 readable bases/reaction. This is true most of the time, but only when the starting DNA template is of good quality, sufficient quantity, and does not contain regions that are considered difficult. For the purposes of this paper, we consider a template (or just some part of it) to be difficult when it cannot be sequenced using a standard ABI-like DNA-sequencing protocol.³

The difficult templates can be classified into the following categories:

(a) *GC rich*. There is no agreement on a lower threshold percentage of GC at which a template is considered difficult, although a template (or at least a 100- to 150-base-long part) with >60–65% GC content could reasonably be considered difficult.^{4–7}

(b) *Containing various repeats*. These include: di- and trinucleotide, direct, inverted and Alu repeats. Examples are: AG, CA, CT, GT, AGG, ACC, CCG, CCT, CTT, GCC, GGA, CCCTTT, and any other combination of these repeats.^{7,8} From a sequencing point of view, long stretches (>40) containing only two nucleotides that do not necessarily form repeats also are difficult to sequence (J. Kieleczawa, unpublished observations).

ADDRESS CORRESPONDENCE AND REPRINT REQUESTS TO: Dr. Jan Kieleczawa, Wyeth Research, Biological Technologies Department, 87 Cambridgepark Drive, Cambridge, MA 02140 (phone: 617-665-8318; fax: 617-665-8350; email: jkieleczawa@wyeth.com).

(c) *Containing hairpin structures.* Technically these structures consist of two inverted repeats, separated by at least three nucleotides. They warrant classification in a separate category due to the increasing importance of si/shRNA-based research that quite often requires sequencing through such structures.^{7,9–13}

Strong hairpin structures are also an integral part of some vectors, e.g., the pDONR/pDEST series from Invitrogen (www.invitrogen.com) or the inverted terminal repeat sequences in adeno-associated viruses.

(d) *Containing long homopolymer stretches.* Most typical are poly-A/T tails resulting from constructs obtained through reverse-transcriptase (RT) PCR amplification of mRNAs. The length of such tails varies from around 20 to over 100 A/Ts. The poly-G/C stretches not only contribute to the overall GC richness but also may form strong hairpins and other complex secondary structures.

(e) *Containing sequence motifs causing band compressions.*¹⁴ These are predominantly 5'-YGN_{1–2}AR motifs, where Y and R are pyrimidine and purine residues, respectively, and N can be any base.

One of the complicating factors in sequencing many different categories of difficult templates is the increasingly apparent realization that there may not be a one-method-fits-all solution and that each category requires a separate approach (or even set of approaches). A number of papers describe modified sequencing protocols. However, they appear to apply only to specific types of difficult templates at best.^{4–7,15–19}

On the other hand, a single approach that incorporates a 5-min heat-denaturation step (with or without some other additives) was successfully applied to many different categories of difficult templates.^{7,13,20}

In this paper, we review both general and specific approaches to sequencing many different kinds of difficult templates, starting with the description of the heat-denaturation step and its impact on the quality of sequencing data.

MATERIALS AND METHODS

All materials and methods presented in this review were extensively described in earlier publications^{7,13,20,27,29,31} and will be not repeated here unless there is a notable deviation. However, for convenience, we will define standard and modified sequencing protocols. *Standard sequencing protocol* refers to a basic protocol recommended by Applied Biosystems technical literature, as described, for example, in reference 3, with the exception that the final volume is 10 μ L. Briefly, DNA, primer, water, and dye terminator mix are combined, and cycled 25 times: 96°C/10 sec, 50°C/5 sec, 60°C/4 min. In the modified

sequencing protocol, we combine DNA, primer, and 10 mM Tris (pH 8.0), heat-denature samples for 5 min at 98°C, and then add dye-terminator mix. If additives are used, they are included in the heat-denaturation step.

CONTROLLED HEAT DENATURATION OF PLASMIDS

In the early days, DNA sequencing was dominated by non-thermostable polymerases, primarily the T7-based Sequenase,^{21–23} and to sequence any double-stranded template, the first step needed to be strand separation for the efficient annealing of the primer. There were many protocols designed for strand separation, all including heating (22°C to 100°C) in the presence or absence of 0.1 to 0.3 N NaOH followed by neutralization, precipitation, and re-suspension in the desired solution.^{24–27} These protocols were time-consuming, cumbersome, and insufficiently reliable for routine use. Instead, a simple heat-denaturation step was suggested in low-salt buffers and at elevated temperatures, to convert supercoiled plasmid DNA efficiently to a single-stranded form amenable for sequencing.²⁰ The denaturation can be carried out in water, but it occasionally produces additional bands that effectively reduce the amount of template available for sequencing (J. Kieleczawa, unpublished observation). The time needed to convert DNA effectively from supercoiled form to single-stranded (ss) form depends on the size of the plasmid; the bigger the plasmid, the shorter time needed for this conversion. It takes 7.5 min (10 mM Tris-Cl, pH 8.0 buffer, 98°C) to convert 75% of pGem3zf (3.2 kbp) from supercoiled to ss form. (The 75% conversion level is selected to avoid potential degradation of DNA.) For any plasmid bigger than pGem3zf, one needs to subtract 1 min per multiple of 2.5 kbp from 7.5 min to achieve a similar level of denaturation. In fact, there is a linear relationship between size of a plasmid, at least in the 3- to 20-kbp size range, and time needed for efficient conversion to a form suitable for sequencing.²⁰ As a rule of thumb, the time needed to achieve a similar level of denaturation in water should be halved compared to the time in 10 mM Tris-Cl buffer. Furthermore, the linear relationship between the size of the plasmid and the time needed for conversion holds true only for plasmids that do not contain any difficult regions. For example, templates rich in GC or with CTT require 30 and 20 min, respectively, for 75% conversion (J. Kieleczawa, unpublished observation). Surprisingly, plasmids with long poly-A/T tracts (70–80 bases) also require up to 20 min for effective conversion to ss form under similar conditions. In the presence of 2 mM MgCl₂ (the final concentration of Mg ions under optimal cycle sequencing conditions using ABI's dye-terminator chemistry³), there is no, or very little, conversion of supercoiled

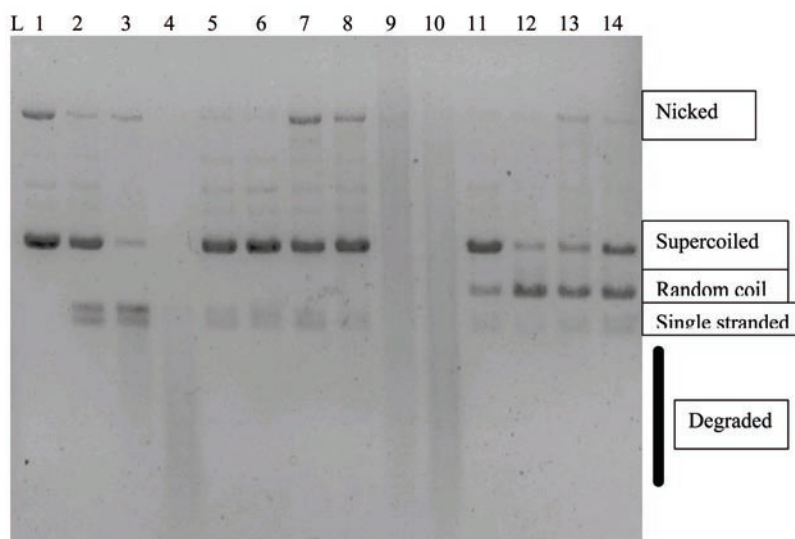


FIGURE 1

Denaturation of pGem3zf under different conditions. Two hundred nanograms of pGem3zf was denatured using heat-denaturation conditions described in this paper and in several published papers. (1) control pGem3zf; (2) 2.5 min heat-denaturation (HD); (3) 7.5 min HD; (4) 15 min HD (DNA in lanes 1–4 was in 10 mM Tris-Cl, 0.01 mM EDTA (pH 8.0) = TEsl. The DNA in lanes 5–14 was also in TEsl but had additional components): (5) in the presence of 2 mM MgCl₂, 7.5 min HD; (6) in the presence of BigDye terminator v3.0/4X final dilution with MgCl₂ at 2 mM, 7.5 min HD; (7) in the presence of 2 mM MgCl₂ followed by 40 sequencing cycles [(10 sec/96°C)(5 sec/50°C)(2 min/60°C)]; (8) as in lane 6 but followed by cycling as in lane 7; (9) 7.5 min HD, add 2 mM MgCl₂ followed by cycling as in lane 7; (10) 7.5 min HD, add dye-terminator as in lane 6, followed by cycling as in lane 7; (11) in the presence of 0.1 N NaOH incubated at room temperature for 5 min, neutralized with 0.1 N HCl;²⁵ (12) in the presence of 0.2 N NaOH incubated at room temperature for 5 min, neutralized with 0.2 N HCl;²⁶ (13) in the presence of 0.2 N NaOH incubated at 85°C for 5 min, neutralized with 0.2 N HCl;²⁶ (14) in the presence of 0.1 N NaOH incubated for 3 min at 100°C, neutralized with 0.1 N HCl.²⁷

to ss form either during prolonged heat denaturation at 98°C or during cycle sequencing (Figure 1). It appears that the primary part that is transformed to a form amenable to sequencing is the nicked form; hence, one needs to use more DNA than necessary to compensate for the partial conversion. Table 1 shows the sequence data for different NaOH-induced plasmid denaturation protocols and the denaturation conditions recommended in this work. Based on the data presented in references 7, 20, and 28, there are the following advantages to incorporation of the heat denaturation step into a sequencing protocol:

(a) The conversion of supercoiled DNA to ss form and destruction of any residual DNase activity is very effective.

(b) The amount of DNA needed for optimal results is three- to fivefold lower than that recommended by a standard ABI protocol.

(c) At a given amount of DNA, the fluorescent signal strength is approximately threefold higher than that of samples without a heat-denaturation step (which is directly related to point “a” above).

(d) Read length ($Q \geq 20$) is increased by 72.3 ± 20 (~10% more) for standard (nondifficult) templates.

(e) Very effective when sequencing many types of difficult templates. In all cases, clearer data are generated with this step, and in 7/22 cases the heat-denaturation method succeeded in generating readable data where standard methods failed (see Table 2). **[AU: Table 1 has not been cited. Please indicate where it belongs.]**

(f) First correctly called base is 3–5 nucleotides closer to the 3' end of the primer (for M13 primer with pGem3zf DNA).

(g) Fewer sequencing errors (4.4 errors/1000 bases) are detected than with a standard protocol.

The kinetics of reversal from ss form back to supercoil-like form is extremely slow (measured in days) both in the presence or absence of 2 mM MgCl₂ and at temperatures ranging from 0 (ice) to 22°C (Kieleczawa, unpublished observation). So, even if the assembly of sequencing reactions takes a few hours, as it could in a factory-like sequencing pipeline, there is no need to take any special precautions after the initial heat denaturation (performed with DNA, primer, low-salt buffer, or water) and before the addition of dye-terminator mix. Furthermore, as pointed out by Kieleczawa and Wu,²⁹ the dye-terminator mix can be stored for several days at room temperature

TABLE 1

Comparison of Read Lengths and Signal Strengths for Varying DNA Concentrations and Different Denaturation Conditions

Denaturation Conditions	This Work/H ₂ O ^a		This Work/TEs ^b		Ref. 25	Ref. 26 ^c	Ref. 26 ^d	Ref. 27	
	-HD	+HD	-HD	+HD					
25 ng DNA	RL	489.5±87.3	744.3±53.8	488.0±84.7	724.7±71.0	268.6±62.1	284.1±124.5	223.6±92.2	0
	Q ≥ 20 SS	13.0±3.7	32.1±6.1	12.3±2.5	27.2±6.0	8.8±2.2	10.6±3.8	8.6±2.0	5.8±0.5
50 ng DNA	RL	585.4±177.7	764.2±44.1	625.4±106.0	784.9±59.1	577.4±44.3	409.9±59.0	436.7±93.0	339.0±90.9
	Q ≥ 20 SS	15.6±5.2	47.0±8.6	16.9±3.4	45.8±19.8	12.6±2.5	10.5±2.2	9.4±1.8	7.6±1.2
200 ng DNA	RL	837.7±25.1	843.6±22.2	724.2±71.5	866.4±29.0	792.0±39.8	663.3±113.7	621.2±76.8	712.8±41.4
	Q ≥ 20 SS	59.8±9.8	119.2±14.1	41.7±7.4	123.3±20.1	65.1±2.2	43.1±9.7	29.8±8.2	26.1±5.9

Varying amounts of pGem3zf plasmid DNA (data are the average of four samples for each condition) were pretreated differently before subjecting them to cycle sequencing. Processed sequencing reactions were run on an ABI3100 equipped with an 80-cm capillary array as recommended by the manufacturer. The data were evaluated in terms of Q ≥ 20 read length (RL) and signal strength (SS is in fluorescent units as produced by an ABI instrument).

^aDNA was resuspended in water and sequenced using standard ABI protocol (-HD = no heat denaturation), or samples were heat denatured (+HD) for 2.5 min at 98°C before cycle sequencing.

^bDNA was resuspended in 10 mM Tris/0.01 mM EDTA (pH 8.0) and sequenced using standard ABI protocol (-HD), or samples were heat denatured for 7.5 min at 98°C before cycle sequencing.

^cDNA was denatured at room temperature (22°C) for 5 min in the presence of 0.2 N NaOH. Following this treatment, an equal volume of 0.2 N HCl was added to neutralize this solution. The mixtures were supplemented with 40 μL of water, 6 μL of 3 M sodium acetate (pH 6.0), and 160 μL 95% ethanol, and were centrifuged for 30 min at 3400 g. Supernatants were discarded, and precipitates were washed twice with 200 μL of 70% ethanol. Finally, samples were dried for 15 min at 65°C and resuspended in 6 μL of water, 1 μL 5 μM M13 reverse primer, and 3 μL of twofold diluted BigDye terminator v3.0, and electrophoresed as described in reference 20.

^dAs described under footnote c, but denaturation was at 85°C.

The denaturation conditions from reference 25: DNA was denatured for 5 min at room temperature in the presence of 0.1 N NaOH and neutralized with 0.1 N HCl. Remaining steps were as described above.

The denaturation conditions from reference 27: DNA was denatured for 3 min at 100°C in the presence of 0.1 N NaOH followed by neutralization with 0.1 N HCl. Remaining steps were as described above.

(or even at 37°C) with no adverse influence on the quality of the resulting sequences. Hence, all assembly and pre-treatment steps can be safely performed at room temperature, with the only necessary precaution being control of evaporation of the sequencing mix to avoid excessive imbalance in the dNTPs and Mg ion concentrations.

See Figures 2, 3, and 4 for examples of the effect of heat denaturation on the sequence quality.

EXAMPLES OF SEQUENCING OF VARIOUS DIFFICULT TEMPLATES

In this section we briefly describe possible modifications that present the best chance(s) to sequence through specific types of difficult templates.

(a) *Sequencing of GC-rich templates.* This category of templates is arguably the most commonly encountered in a typical sequencing laboratory (DSRG survey, 2005 data, unpublished). Most often, the solution of choice is to add DMSO to the final concentration of 2.5–5%, and it seems to be effective in templates with up to 60–72% GC content.^{4–6} On the other hand, the inclusion of a 5-min heat-denaturation step is even more effective⁷ and, of course, does not require the addition of DMSO. In our DNA sequencing core laboratory, we have developed protocols (built into our laboratory information management system—LIMS) that automatically switch between various chemistries depending on the GC content of a particular region in a template. When GC content is ≤70%, just heat denaturation alone is sufficient to sequence a major-

TABLE 2

Sequencing Difficult Templates Using Various Protocols

DNA Type	Sequencing Method → DNA Characteristics ↓	Method No.							
		1	2	3	4	5	6	7	8
1	60% GC	834	868	818	862	895 ACG	884 Bet	511	280
2	63% GC	208	572	NT	121	673 C	635 GC	552	549
3	66% GC	538	847	867	854	894 AG	858 Bet	728	635
4	67% GC	0	989	NT	222	1087 C	1025 GC	177	945
5	75% GC	0	660	NT	960	974 F	940 Bet	747	713
6	80% GC	661	920	914	920	940 ACG	933 Bet	842	848
7	86% GC	0	504	NT	928	967 C	988 GC	895	675
8	94% GC	0	277	NT	270	304 A	302 Bet	318	420
9	CT repeats	569	886	107	618	950 C	815 Bet	845	927
10	GA repeats	363	761	168	416	900 ACG	825 Bet	580	765
11	CTT/TTTCCC repeats	222	277	NT	193	914 CBD	898 Bet	477	NT
12	TCC/GCC repeats	0	830	NT	853	929 C	893 Bet	855	NT
13	CTT repeats	401	431	NT	411	540 BCE	466 Bet	455	426
14	CCCCNN/TTCC repeats	390	609	503	805	1000 AFG	1000 Bet	954	890
15	GGGGNN/AAGG repeats	378	991	873	1008	1011 ACG	995 Bet	882	886
16	Alu repeat	0	118	NT	107	262 BDE	258 GC	820	NT
17	19 Gs homopolymer	706	743	809	834	860 AC	NT	712	NT
18	19 Cs homopolymer	487	537	403	409	856 AC	NT	471	NT
19	23 Gs homopolymer	283	404	407	418	606 A	NT	605	NT
20	28 base ITR hairpin*	410	439	NT	NT	462 C	NT	NT	NT
21	RNAi plasmid forward primer	0	590	NT	601	710 ACD	885 Bet	811	811
22	RNAi plasmid reverse primer	566	820	NT	750	806 CD	806 GC	963	955

Twenty-two different difficult DNAs were sequenced using a variety of protocols. Method 1 is equivalent to the standard ABI protocol.³ Method 2 is equivalent to protocol described in references 7, 13, 20 (includes 5 min heat-denaturation step). Method 3 is like protocol 1 but in the presence of 5% DMSO. Method 4 is like protocol 2, but DMSO was included during heat-denaturation step. Method 5 is like protocol 2, but one of seven reagents from Invitrogen's (Carlsbad, CA) sequence enhancer Rx set was present during denaturation step (only data for the longest reads are reported here). Method 6 is like method 2, but included some other additive (GC melt is from BD Biosciences, San Jose, CA; Betaine was from Sigma-Aldrich, St. Louis, MO). Method 7 was like method 2, but BigDye V3.1 was substituted with dGTP V3.0. Method 8 was like method 2, but BigDye V3.1 was substituted with 4:1 (v/v) mix of BigDye V3.1 and dGTP V3.0 as suggested by G. Grills (personal communication). All data are expressed in Q ≥ 20 values as described earlier.⁷

*For these two templates, the only currently available method of reading through is to apply a two-step protocol as described in reference 7.

ity of difficult regions. When the GC content is between 70 and 80%, we add reagent A (from a panel of seven Rx reagents sold by Invitrogen, Inc., Carlsbad, CA) or 1 M betaine (purchased from Sigma-Aldrich, St. Louis, MO). For templates with GC content from 80 to 90%, we use dGTP V3.0 chemistry in conjunction with reagent A or C.

It is worth remembering that dGTP Big Dye occasionally produces band compression and needs to be corrected using standard BigDye chemistry.¹³ Above 90% GC content, we (through our LIMS) apply the two-step process that includes PCR amplification of the fragment in question, using 7-deaza-dGTP instead of dGTP, followed by

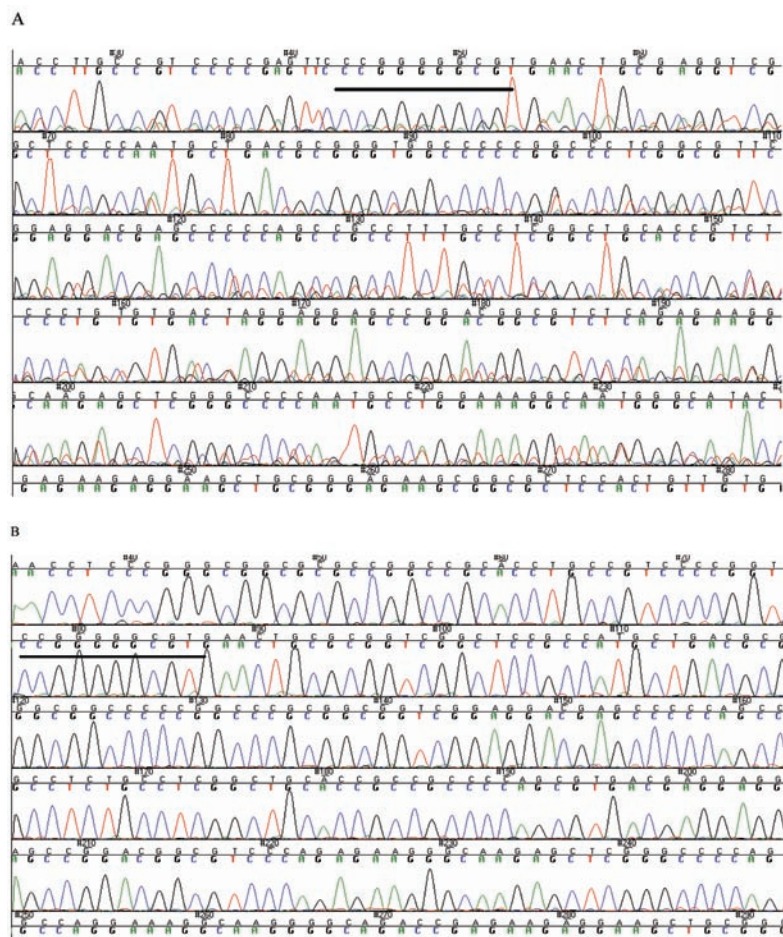


FIGURE 2

Chromatograms of GC-rich template (about 67%) sequenced using a standard (**A**) and modified (**B**) protocol. The *solid bar* indicates identical sequence regions.

purification and sequencing using standard dye-terminator chemistry. A similar two-step approach, and, so far, the only successful one, is also used for sequencing through the 28-bp hairpin structure in the ITR sequence in adeno-associated virus. A few attempts to use Tween-20 and NP-40 detergents to help with sequencing of these types of templates were unsuccessful.

See Table 2 for examples of sequencing through a few GC-rich templates.

(b) *Sequencing regions containing various repeats.* The most effective way to deal with many different kinds of repeats is to apply a heat-denaturation step. More specifically, for AG/GA repeats, heat denaturation and reagent A are a very effective combination. For GCCCCT/TCCT and similar repeats, we use dGTP chemistry with reagent A. We generally recommend, in addition to the heat-denaturation step, reagent A or betaine for repeats that contain Gs and reagent C for repeats that contain Cs. See Table 2 for examples of sequencing through some repeat structures.

(c) *Sequencing of regions with hairpin structures.* The detailed protocol varies depending on the length of the hairpin. The heat denaturation step alone is sufficient to sequence

through hairpins that are ≤ 15 bases long.¹³ On the other hand, Esposito et al. had to introduce a special blocking agent to overcome problems with sequencing fragments cloned in pDONR221.¹⁶ For the standard, 19-bp hairpin structures that are currently commonly used in siRNA experiments,^{9–12} we recommend the addition of reagent C (or A) to successfully read correct sequences. Only occasionally is dGTP chemistry necessary—for example, when GC pairs are concentrated at one end of the stem (Mostafa Ait-Zahra and Jan Kieleczawa, unpublished observation). For longer hairpin structures—for example, in the 28-bp ITR region—the two-step protocol (PCR in the presence of 7-deaza-GTP followed by standard sequencing) has to be used to sequence through such strong hairpin structures. See Table 2 for examples of sequencing through hairpin structures.

(d) *Sequencing through long homopolymer stretches.* In many ways sequencing through homopolymer stretches is the most challenging. Heat denaturation alone is sufficient to read through G or C homopolymers of 15 bases or less (though addition of reagent A or betaine improves the quality of data). However, if G or C stretches are longer

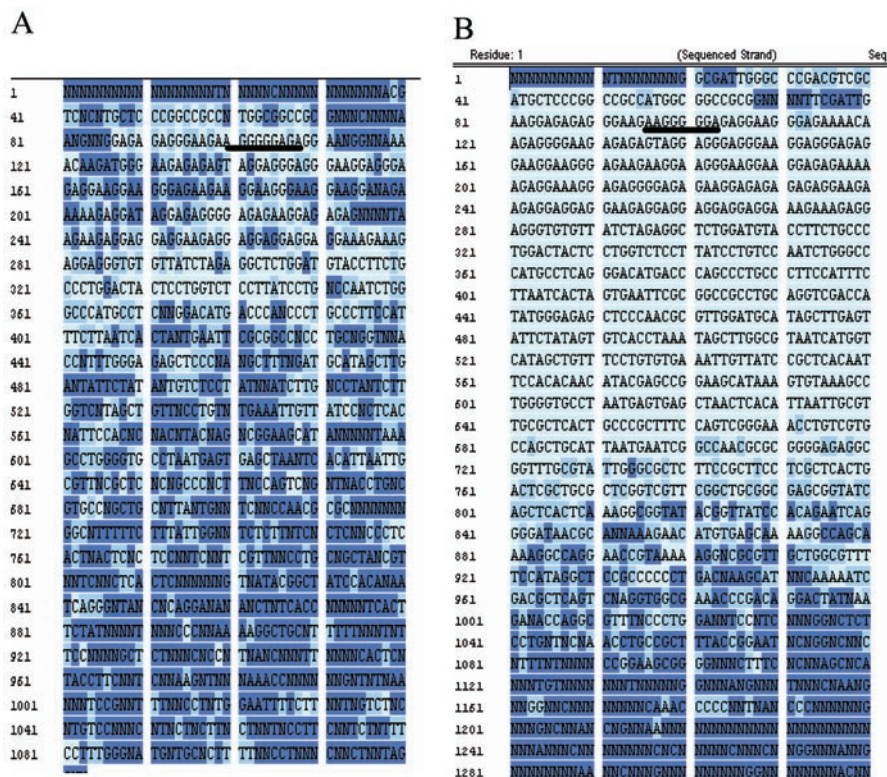


FIGURE 4

An example of a template with 59-bases GA non-repeat stretch sequenced using a standard (A) and modified (B) protocol. These views represent quality scores as displayed by Sequencher program. The light blue color indicates bases with higher Q quality and darker blue color indicates bases with lower Q values. The black bar shows the region with the same sequence.

studies³¹ we prepared twelve different difficult DNAs using six conventional DNA preparation methods and a Templphi-based protocol to evaluate whether there is any advantage using one kit over another. Two well-established commercial DNA kits (from EdgeBiosystems and Qiagen) seem to be most consistent in producing high-quality DNAs leading to the longest reads,³¹ as shown in Table 3. The Templphi produces branched and somewhat unstructured molecules^{32,33} that, we were expecting, would be easier to read through, but unfortunately this turned out to be incorrect. The technical literature from Amersham³⁴ (currently part of GE Healthcare) claims that certain types of difficult templates, prepared using Templphi, are easier to sequence compared to those isolated with conventional methods. However, this claim needs to be independently verified.

THE WEB AS A RESOURCE FOR INFORMATION ON SEQUENCING OF DIFFICULT TEMPLATES

In the era of Yahoo and Google, it is only natural that one would be tempted to scan the Web for information/advice regarding sequencing of difficult templates. Typ-

ing, e.g., “difficult DNA templates” into Google’s search engine results in about 341,000 hits (211,000 in Yahoo’s search engine), which leads to about 1000 sites as of January 2006. Rarely, though, does one get any advice beyond “DMSO or glycerol” for GC-rich samples, or some unspecified proprietary treatment in a few commercial sites. Quite unusual is also the fact that almost all sites cite the same/similar amounts of DNA needed for optimal read length, concentrations of interfering agents, etc., without providing references to any published or unpublished data. An example of such unsubstantiated advice is the recommended amount of PCR fragment needed for optimal read length: use 10 ng of DNA/100 bases. In our experiments²⁸ we demonstrate that even for a PCR fragment of 1250 base pairs, 1 ng of DNA is fully sufficient for optimal read length. In fact, we carried out similar experiments on PCR fragments of about 500 and 900 bases (with the amount of DNA/reaction varied from 0.1 ng to over 500 ng) and the results did not support the 10 g/100 bases claim (J. Kieleczawa, unpublished observation); in all cases, just 1 ng of DNA gave the optimal read length.

TABLE 3

The Performance of Six Classical Plasmid Preparation Methods

DNA preparation method	DNA Sequencing Method No.							
	Method 1		Method 2		Method 3		Method 4	
	No.	%	No.	%	No.	%	No.	%
EdgeBiosystems	7	58.3	4	33.3	4	33.3	5	38.5
Eppendorf	0	0	0	0	0	0	0	0
Marligen	2	16.7	0	0	0	0	1	7.7
Promega	0	0	0	0	0	0	0	0
Sigma	0	0	0	0	5	41.7	4	30.8
Qiagen	3	25.0	8	67.3	3	25.0	3	23.0

Twelve different difficult templates were prepared using six commercial kits. Data shows how many times (No and %) given method produced DNA that resulted in longest reads (measured with $Q \geq 20$ values) using four sequencing methods. Method #1 is equivalent to a standard ABI-like protocol. Method #2 included heat denaturation only. Methods #3 and 4 were the two best methods for a given DNA template and included heat denaturation and one of additives (see reference 7 for details). The twelve templates used in this experiment is a subset of templates shown in Table 2.

Explanations for the table. No. = number of times the method gave the best results.

Therefore, a reader (PC user) seeking fast Web advice concerning any aspect of DNA sequencing needs to be highly skeptical and selective, as the advice one gets may not necessarily be helpful (although in most cases it will not hurt either).

SUMMARY

In this paper, we have reviewed past and current methods to sequence through many types of difficult templates. Although a number of solutions were suggested over the last few years, they all seem to be quite specific to a particular type of difficult template. The possible exception is the heat-denaturation modification, which appears to be more broadly applicable for several different difficult templates and in combination with several additives gives the best chance to obtain good-quality data. Data presented in Table 2 show that for 32% of difficult templates, adding just a heat-denaturation step was the only possible way to obtain any sequencing data. For 22% of templates, the improvement was on the order of 5–10% in read length. Increase of 10–50% in read length was evident for another 14% of templates, and for the remaining 32% the increase in read length was on the order of 50–175%. It is apparent that more experiments need to be carried out to develop more general rules, and it is entirely possible that, in the

interest of time, any given very difficult template may need to be sequenced using a few different chemistries in parallel. One of the potential venues to speed the development of general rules for sequencing of difficult templates is to organize a bank of well-characterized DNAs and involve the broad community in applying a range of technologies. The model of such community effort is well established—e.g., the DNA Sequencing Research Group (DSRG) conducted a study on the effect of DMSO on sequencing of some GC-rich templates.^{5,6} In collaboration with bioinformaticians, who now have extensive computational tools, it should be possible to look for a more detailed correlation between sequence patterns, especially right before and after difficult regions, and the success of a given chemistry. It is also conceivable that one of the big sequencing centers, NIH or some other entity, will establish a special unit that will investigate in greater detail the sequencing of all types of difficult templates, and the art will become science.

As we approach an era of \$1000-a-genome sequencing, one still-unanswered question is how these new technologies will deal with more complex regions. It is possible that difficult regions will not present any challenges, as most of these methods rely on assembly of very short overlapping fragments. Hopefully, researchers who are

TABLE 4

Comparison of Read Lengths between DNAs Prepared Using Either TempliPhi or Qiagen/Marligen Methods

DNA No.	Characteristics	Preparation Method	DNA Sequencing Method No.		
			1	2	3
1	CCT/TTTCCC	TempliPhi	340	515	594
		Qiagen	375	495	647
2	TCC/GCC	TempliPhi	635	640	626
		Qiagen	599	659	642
3	CTT/CCCT/CCTT	TempliPhi	0	0	0
		Marligen	337	458	454
4	Alu-repeats	TempliPhi	128	147	292
		Qiagen	214	240	259
5	90% GC	TempliPhi	111	278	285
		Marligen	0	292	327
6	88% GC	TempliPhi	143	165	181
		Qiagen	313	502	587
7	75% GC	TempliPhi	148	610	567
		Qiagen	565	642	600
8	63% GC	TempliPhi	524	591	596
		Qiagen	556	575	596
9	63% GC + strong hairpin	TempliPhi	<100	<100	<100
		Qiagen	<100	358	100
10	54% GC	TempliPhi	634	663	634
		Qiagen	640	671	643
11	shRNA hairpin 1	TempliPhi	497	487	501
		Qiagen	609	663	653
12	shRNA hairpin 2	TempliPhi	355	421	571
		Qiagen	270	343	526

The TempliPhi-derived templates were prepared as described in reference 35. Method #1 is equivalent to a standard ABI-like protocol. Method #2 included heat denaturation only. Method #3 was the best protocol for a given template and included heat denaturation and one of the additives (see reference 7 for details). The templates used in this experiment are a subset of templates shown in Table 2.

directly involved in the development of new technologies will be able to answer such questions relatively soon.

ACKNOWLEDGMENTS

I wish to thank Dr. Laird Bloom of Wyeth Research for critical review of this manuscript and many valuable suggestions. The support and encouragement by the management of the Biological Technologies department of Wyeth Research is also greatly appreciated. I would like to thank Jones and Bartlett Publishers, Sudbury MA, for permission to reprint table 1-1 from (J. Kieleczawa, ed) *DNA Sequencing: Optimizing the Process and Analysis*, 2005, and tables 1-1 and 1-4 from (J. Kieleczawa, ed) *DNA Sequencing: Optimizing the Preparation and Cleanup*, 2006.

REFERENCES

1. Maxam AM, Gilbert W. A new method of sequencing DNA. *Proc Natl Acad Sci USA* 1977;74:560–564.
2. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463–5467.
3. BigDye Terminator v3.1 Cycle Sequencing Kit. Protocol. 2002. Part number 4337035 Rev. A. Applied Biosystems. Foster City, CA.
4. Adams PS, Dolejsi MK, Hardin S, Mische S, Nanthakamur B, Riethman H, et al. DNA sequencing of a moderately difficult template: Evaluation of the results from a *Thermus thermophilus* unknown test sample. *BioTechniques* 1996;21:678.
5. Adams PS, Dolejsi MK, Grills G, Hardin S, McMinimy DL, Rush J, et al. Effects of DMSO, thermocycling and editing on a

- template with 72% GC rich area: Results from the 2nd Annual ABRF sequencing survey demonstrate that editing is the major factor in improving sequencing accuracy. Ninth International Genome Sequencing and Analysis Conference. *Microb Comp Genomics* 1997;2:198.
6. Adams PS, Dolejsi MK, Grills G, McMinimy D, Morrison P, Rush J, et al. An analysis of techniques used to improve the accuracy of automated DNA sequencing of a GC-rich template: Results from the 2nd Annual ABRF DNA Sequence Research Group study. Available at: <http://www.abrf.org>; search for 2nd ABRF DNA Sequence Research Group Study. 1999.
 7. Kieleczawa J. Sequencing of difficult DNA templates. In Kieleczawa J (ed), *DNA Sequencing: Optimizing the Process and Analysis*. Jones and Bartlett Publishers, Sudbury MA, 2005, pp. 27–34.
 8. Zhao X, Haqqi T, Yadav SP. Sequencing telomeric DNA templates with short tandem repeats using dye terminator cycle sequencing. *J Biomol Tech* 2000;11:111–121.
 9. Sharp PA. RNAi and double-strand RNA. *Genes and Dev*. 1999;13:139–141.
 10. Hammond SM, Caudy AA, Hannon GJ. Posttranslational gene silencing by double-stranded RNA. *Nat Rev Genet* 2001;2:110–119.
 11. Brummelkamp TR, Bernards R, Agami R. A system for stable expression of short interfering RNAs in mammalian cells. *Science* 2002;296:550–553.
 12. Elbashir SM, Harborth J, Weber K, Tuschl T. Analysis of gene function in somatic mammalian cells using small interfering RNAs. *Methods* 2002;26:199–213.
 13. Kieleczawa J. Simple modifications of the standard DNA sequencing protocol allow for sequencing through siRNA hairpins and other repeats. *J Biomol Tech* 2005;16:220–223.
 14. Yamakawa H, Nakajima D, Ohara O. Identification of sequence motifs causing band compressions on human cDNA sequencing. *DNA Res* 1996;3:81–86.
 15. Ducat DC, Herrera FJ, Triezenberg SJ. Overcoming obstacles in DNA sequencing of expression plasmids for short interfering RNAs. *BioTechniques* 2003;34:1140–1144.
 16. Esposito D, Gillette W, Hartley JL. Blocking oligonucleotides improve sequencing through inverted repeats. *BioTechniques* 2003;35:914–920.
 17. Gerstner A, Sasvari-Szekely M, Kalasz H, Guttman A. Sequencing difficult templates using membrane-mediated loading with hot sample application. *BioTechniques* 2000;28:628–630.
 18. Langan JE, Rowbottom L, Liloglou T, Field JK, Risk JM. Sequencing of difficult templates containing poly (A/T) tracts: Closure of sequencing gaps. *BioTechniques* 2002;33:276–280.
 19. Thomas MG, Hesse, SA, McKie AT, Farzaneh F. Sequencing of cDNA using anchored oligo dT primers. *Nucleic Acid Res* 1993;21:3915–3916.
 20. Kieleczawa J. Controlled heat-denaturation of DNA plasmids. In Kieleczawa J (ed), *DNA Sequencing: Optimizing the Process and Analysis*. Jones and Bartlett Publishers, Sudbury MA, 2005, pp. 1–10.
 21. Tabor S, Richardson CC. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc Natl Acad Sci USA* 1987;84:4767–4771.
 22. Tabor S, Richardson CC. DNA sequence analysis with a modified bacteriophage T7 DNA polymerase: Effect of pyrophosphorolysis and metal ions. *J Biol Chem* 1990;265:8322–8328.
 23. Tabor S, Richardson CC. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerases I family is critical for distinguishing between deoxy and dideoxyribonucleotides. *Proc Natl Acad Sci USA* 1995;92:6339–6343.
 24. Haltiner M, Kempe T, Tjian R. A novel strategy for constructing point mutations. *Nucleic Acid Res* 1985;13:1015–1025.
 25. Hattori M, Sasaki Y. Dideoxy sequencing method using denatured plasmid templates. *Anal Biochem* 1986;152:232–238.
 26. Toneguzzo F, Glynn S, Levi E, Mjolsness S, Hayday A. Use of chemically modified T7 DNA polymerase for manual and automated sequencing of supercoiled DNA. *BioTechniques* 1988;6:460–469.
 27. Yie Y, Wei Z, Tien P. A simplified and reliable protocol for plasmid DNA sequencing: Fast miniprep and denaturation. *Nucleic Acid Res* 1993;21:361.
 28. Kieleczawa J, Bajson K. Evaluation of methods for cleanup of DNA sequencing reactions. In Kieleczawa J (ed), *DNA Sequencing II: Optimizing the Preparation and Clean Up*. Jones and Bartlett Publishers, Sudbury MA, 2006, pp. 219–240.
 29. Kieleczawa J, Wu P. Prolonged storage of plasmid DNAs under different conditions: Effects on plasmid integrity, spectral characteristics, and DNA sequence quality. In Kieleczawa J (ed), *DNA Sequencing II: Optimizing the Preparation and Clean Up*. Jones and Bartlett Publishers, Sudbury MA, 2006, pp. 259–274.
 30. Flick PK. Plasmid preparation methods for DNA sequencing. In Kieleczawa J (ed), *DNA Sequencing: Optimizing the Process and Analysis*. Jones and Bartlett Publishers, Sudbury MA, 2005, pp. 99–115.
 31. Kieleczawa J, Li T, Wu P. Preparation of difficult DNA templates using seven different commercial methods. In Kieleczawa J (ed), *DNA Sequencing II: Optimizing the Preparation and Clean Up*. Jones and Bartlett Publishers, Sudbury MA, 2006, pp. 1–14.
 32. Dean RB, Nelson JR, Giesler TL, Lasken RS. Rapid amplification of plasmid and phage DNA using phi29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* 2001;11:1095–1099.
 33. Nelson JR, Cai YC, Giesler TL, Farchaus JW, Sundaram ST, Ortiz-Rivera M, et al. 2002. TempliPhi: Phi29 DNA polymerase-based rolling circle amplification of templates for DNA sequencing. *BioTechniques* 2002;32:S44–S47.
 34. New tool for resolving difficult templates. 2004. GE Healthcare. Piscataway, N.J.