

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,800

Open access books available

122,000

International authors and editors

135M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.

For more information visit www.intechopen.com



Hot Start 7-Deaza-dGTP Improves Sanger Dideoxy Sequencing Data of GC-Rich Targets

Sabrina Shore, Elena Hidalgo Ashrafi and Natasha Paul
TriLink BioTechnologies, Inc.
USA

1. Introduction

DNA sequencing has developed substantially over the years into a more cost-effective and accurate technique for scientific advancement in medical diagnostics, forensics, systematics, and genomics. Sanger dideoxy sequencing is currently one of the most established and popular sequencing methods (Sanger et al., 1977). Over the years, sequencing methods have become automated, faster and more specific and now allow sequencing of difficult and unknown regions of DNA. Several protocols have been developed over the years which include new dye chemistries, use of modified nucleotide analogs, use of additives in the sequencing reaction, and variations to the sequence cycling parameters (Prober et al., 1987; Kieleczawa, 2006). These modified protocols allow for sequencing through difficult regions of DNA and may be applied in the pre-sequencing PCR step or in the actual sequencing reaction itself. Despite these improvements, there continues to be DNA regions that are problematic to sequence, such as AT, GT, GC-rich regions, regions high in secondary structure, hairpins, homopolymer regions, and regions with repetitive DNA sequence (Kieleczawa, 2006; Frey et al., 2008). These challenging DNA templates often result in ambiguous sequencing data which include false stops, compressions, weak signals, and premature termination of signal. In particular, sequences high in GC content still suffer from several of these problems despite all the advancements.

Templates high in GC content have higher melting temperatures that do not allow for adequate strand separation of the DNA duplex in standard sequencing protocols. The tendency for these sequences to form complex secondary structures, such as hairpins or G-quadruplexes (Simonsson, 2001) can prevent a DNA polymerase from processively replicating an entire stretch of sequence (Weitzmann et al., 1996). The innate secondary structure of GC-rich templates and the strength of the DNA duplex can be obstacles in sequencing reactions as well as in PCR. GC-rich PCR assays are often plagued by mis-priming and inadequate amplicon yield which in turn provide poor quality DNA samples for downstream sequencing reactions (Shore & Paul, 2010). When a DNA template high in GC content is sequenced, base compressions, weak signals from high background noise, and truncated sequencing reads are typical results. Base compressions are due to secondary structure and cause abnormal migration during the electrophoretic separation step. These fragment irregularities in migration often plague the downstream sequence analysis, where the software is unable to discriminate between the fragments (Mutz et al., 2000). As a result,

several advancements have been developed to address the challenge of GC-rich DNA in the pre-sequencing PCR step as well as in the sequencing reaction itself.

To overcome the higher melting temperatures of GC-rich DNA sequences, sequence cycling parameters have been adjusted to use higher temperatures. Additives such as dimethyl sulfoxide (DMSO), betaine, formamide, and glycerol are often added into sequencing reactions to reduce secondary structure and to promote strand separation (Jung et al., 2002). It is well known that the use of modified analogs such as 7-deaza-dGTP (7-deaza-2'-deoxyguanosine-5'-triphosphate) and dITP (2'-deoxyinosine-5'-triphosphate) can be used in the sequencing reaction in place of dGTP or in some combination thereof to reduce secondary structure (Mutz, Svante et al., 2000). 7-deaza-dGTP is a modified dGTP analog which lacks a nitrogen at the seven position of the purine ring. The absence of this nitrogen destabilizes G-quadruplex formation by preventing Hoogsteen base pairing without affecting Watson-Crick base pairing. Alternatively, dITP alters the Watson-Crick base pairing by reducing the number of hydrogen bonds from three to two in a base pair between inosine and cytosine. This reduces the strength of GC-rich duplexes by lowering the melting temperature.

Efforts have also focused on improving the PCR amplification of GC-rich targets to provide a good quality DNA template prior to the sequencing reaction. The addition of 7-deaza-dGTP in the PCR step permanently disrupts secondary structure by incorporation of this modified analog into the amplicon. The pre-sequencing PCR amplification effectively linearizes the DNA template and prepares it for sequencing. Though the optimal ratio of 7-deaza-dGTP to dGTP in a PCR reaction has been thoroughly investigated and determined, this method often requires additional components such as a Hot Start polymerase or additives to be included into the reaction (McConlogue et al., 1988).

A recently developed Hot Start technology named CleanAmp™ employs a thermolabile group on the 3' hydroxyl of a deoxynucleoside triphosphate. The presence of the protecting group blocks low temperature primer extension, which can often be a significant problem in PCR. At higher temperatures, the protecting group is released to allow for incorporation by the DNA polymerase and for more specific amplification of the intended target (Koukhareva & Lebedev, 2009). CleanAmp™ dNTPs are Hot Start versions of standard deoxynucleotide triphosphates (dATP, dCTP, dGTP, and dTTP) while CleanAmp™ 7-deaza-dGTP applies the same concept to a modified nucleoside triphosphate, 7-deaza-dGTP. Previous results have shown that a dNTP mixture containing CleanAmp™ 7-deaza-dGTP provides a significant improvement over standard 7-deaza-dGTP in the amplification of GC-rich targets in PCR assays (Shore & Paul, 2010). CleanAmp™ 7-deaza-dGTP Mix is a formulation which contains the CleanAmp™ versions of dATP, dCTP, dGTP, dTTP, and 7-deaza-dGTP where the 7-deaza-dGTP:dGTP ratio is 3:1. CleanAmp™ 7-deaza-dGTP applies a Hot Start technology to a secondary structure reducing analog that is permanently incorporated into the PCR amplicon. Preliminary sequencing results on PCR amplicons generated using the CleanAmp™ 7-deaza-dGTP Mix have shown an improvement in sequencing reads over a standard (unmodified) 7-deaza-dGTP mix. The significant improvement in data quality when CleanAmp™ 7-deaza-dGTP Mix was compared to an analogous mix containing unmodified versions of the dNTPs instigated further experimental inquiries to identify the optimal mix composition for use prior to sequencing of PCR products.

Herein we will compare the application of a CleanAmp™ 7-deaza-dGTP Mix to a standard 7-deaza-dGTP mix for the PCR amplification of GC-rich targets in preparation for Sanger dideoxy sequencing. We show that CleanAmp™ 7-deaza-dGTP Mix provides an improvement compared to the standard version of a 7-deaza-dGTP mix and provide guidance as to the best ratio of 7-deaza-dGTP to dGTP to use for optimal PCR and downstream sequencing performance. Performance categories that weigh into this decision include measures of PCR performance, such as preliminary amplicon yield and amplicon quality, and measures of sequencing performance including percent of high quality base calls, read length, and pairwise identity. The most crucial metric for determining sequence performance is the percent high quality base calls, which provides a numerical readout of fragment resolution and sequence determination. The effects of amplicon yield and amplicon quality were also assessed to determine if these parameters directly correlate with sequencing quality.

2. Methods

The use of the CleanAmp™ technology in a pre-sequencing PCR amplification step was explored by comparing analogous reactions with standard dNTPs to investigate the effect of Hot Start activation on the PCR step. The ratio of 7-deaza-dGTP to dGTP was thoroughly investigated to determine an optimal mixture that will provide robust PCR yield and accurate Sanger dideoxy sequencing results. The effect of magnesium chloride concentration on amplicon yield and subsequent sequencing results was also investigated for mixtures with high 7-deaza-dGTP ratios, which showed low amplicon yield under normal 1.5 mM magnesium chloride concentrations. Experimental data evaluates five GC-rich targets of varying GC content and amplicon length.

2.1 PCR

The following five GC-rich targets were investigated: ACE (60% GC-rich), BRAF (74%), GNAQ (79%), GNAS (84%) (Frey, Bachmann et al., 2008) and B4GN4 (78%) (Zhang et al., 2009). PCR reactions were set up in a 50 µL total volume and consisted of 1x PCR buffer (20 mM Tris pH 8.4, 50 mM KCl; Invitrogen), 1.5 mM MgCl₂ (Invitrogen), 2.0 U *Taq* DNA polymerase (Invitrogen), 0.2 µM primers (TriLink BioTechnologies), 0.2 mM d(A,C,T), X % dGTP, and Y % 7-deaza-dGTP (TriLink Biotechnologies) where X and Y are varying percentages of 0.2 mM (X/Y = 30/70, 25/75, 20/80, 10/90, 0/100). Nucleotide mixes consisted of all standard dNTPs (New England Biolabs) or all CleanAmp™ dNTPs (TriLink BioTechnologies) and 10 ng Human Genomic DNA (Promega) as template. For PCR with the 90 and 100% 7-deaza-dGTP nucleotide mixtures, different concentrations of magnesium chloride were used: 1.5, 2, or 2.5 mM. Five replicates for each condition were prepared, each in a single, thin-walled 200-µL tube and placed in a BioRad Tetrad 2 thermal cycler with a thermal cycling protocol 95 °C (10 min); [95 °C (40 s), X °C (1 s), 72 °C (1 min)] 35x; 72 °C (7 min) where X (annealing temperature) varied according to the target (ACE 70 °C; BRAF 67 °C; B4GN4 57 °C; GNAQ 66 °C; GNAS 64 °C). The GNAQ and GNAS targets went through 40 PCR cycles. After thermal cycling, all 5 reactions were combined, and 15 µL of product was run on a 2% agarose E-gel (Invitrogen) to determine preliminary amplicon yield before PCR cleanup. The five combined PCR reactions were then cleaned up to remove excess primers and dNTPs using the Qiagen PCR Cleanup kit (Qiagen). After cleanup, 5 µL of

purified samples were run on an agarose gel to visualize the final product to be sequenced. Image J software was used to integrate amplicon and off-target bands on the E-gels both before PCR cleanup and after (Abramoff, 2004).

2.2 Sanger dideoxy Sequencing

PCR products generated with 70, 75, and 80% mixtures of 7-deaza-dGTP that were submitted for sequencing were amplified using 1.5 mM MgCl₂. The following 90 and 100% samples which were submitted for sequencing required additional magnesium chloride in the PCR step: ACE (100% - 2 mM MgCl₂ for standard and CleanAmp™), B4GN4 (90% - a mixture of the 1.5 and 2 mM MgCl₂ for standard dNTPs and 100% - 2 mM MgCl₂ for standard and CleanAmp™), GNAQ (100% - 2 mM MgCl₂ for standard and CleanAmp™), GNAS (90 and 100% - 2 mM MgCl₂ for standard and CleanAmp™). Those samples not mentioned contained the standard 1.5 mM MgCl₂. Each target was amplified in five individual PCR experiments and analyzed by sequencing. Samples were submitted to Eton Bioscience, Inc. (San Diego, CA) where they were quantified using a nanodrop and submitted for Sanger dideoxy sequencing with the Big Dye Terminator v3.1 (Applied Biosystems) kit used for cycle sequencing reactions. A specified amount for each PCR product was used during the sequencing reaction, depending on the number of base pairs in the amplicon: ACE (156 bp, 10 ng DNA), BRAF (185 bp, 12 ng DNA), B4GN4 (720 bp, 28 ng DNA), GNAQ (642 bp, 27 ng DNA), GNAS (242 bp, 15 ng DNA). All sequencing results were analyzed with the Genious Pro v5.5.2 sequencing software (Drummond AJ, 2011).

2.3 Data analysis

Several categories of results were examined including preliminary amplicon yield, amplicon quality, pairwise identity, read length of target, and percent of high quality base calls. All values were averaged for the five independent PCR or sequencing runs and statistical analysis was done using Graphpad Prism Software (San Diego, CA). A two tailed t-test was used to test the null hypothesis that CleanAmp™ and standard dNTPs yield the same results in each 7-deaza-dGTP mixture. A one-way ANOVA test was used to test the null hypothesis that all 7-deaza-dGTP compositions perform equally within a group (CleanAmp™ or standard dNTPs). The Tukey-Kramer post test was also done in addition to the one-way ANOVA to compare all levels of substitution with one another (Prism). Statistical probability values are indicated on graphs and tables.

3. Results

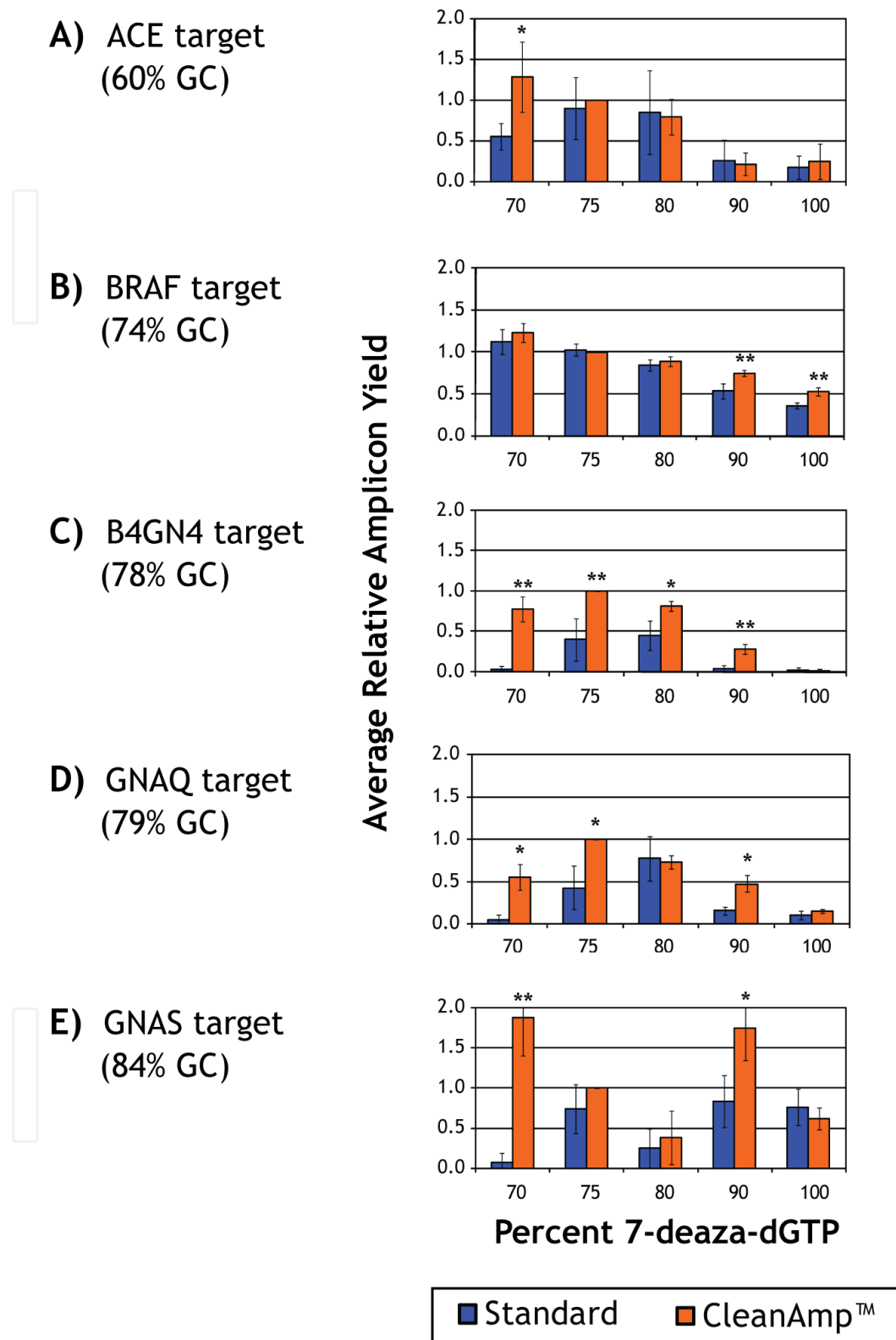
Several experimental factors were considered to rank the performance of the PCR amplification and downstream sequencing for five different GC-rich targets with 60 to 84 % GC content. These targets were amplified using five nucleotide mixtures containing different percentages of 7-deaza-dGTP, where the nucleotide mix used in the PCR step was either unmodified (standard) or Hot Start (CleanAmp™). Experimental performance criteria sought to identify the influence of the PCR conditions on amplicon formation (yield and quality) and on Sanger sequencing data quality (percent of high quality base calls, read length, and pairwise identity) (Kieleczawa et al., 2009).

3.1 PCR specificity and yield

The first factor was assessment of the PCR yield of reactions with CleanAmp™ (or hot start dNTPs) and standard dNTPs to determine if one would provide a superior result. The graphs in Figure 1 show this comparison, where each of the five targets was PCR amplified with five different 7-deaza-dGTP blends, ranging from 70 to 100% substitution of dGTP with 7-deaza-dGTP. The preliminary amplicon yield is presented as the average relative amplicon yield for each target, where the raw data values for target band gel densitometry in each experiment were normalized to the 75% CleanAmp™ 7-deaza-dGTP mixture. Amplicon yield was quantified prior to PCR cleanup and reflects approximately how much DNA yield was generated from an independent PCR reaction. While many reactions generated sufficient amplicon for downstream sequencing from a single set-up, all reactions contained five replicates to ensure that the amplicon yield after PCR cleanup was sufficient for sequencing. It was also noted that reactions using a higher percentage of 7-deaza-dGTP often required an increase in magnesium chloride concentration for sufficient amplicon yield. Therefore, the 90 and 100% 7-deaza-dGTP brews were prepared with 1.5, 2, and 2.5 mM MgCl₂ to ensure that enough product was formed. From the samples with varying MgCl₂, the PCR product with the highest yield and least off-target per run was submitted for sequencing. Preliminary experiments indicated that altered magnesium chloride concentration in the PCR step did not significantly affect sequencing reads, so this variable was eliminated when analyzing results.

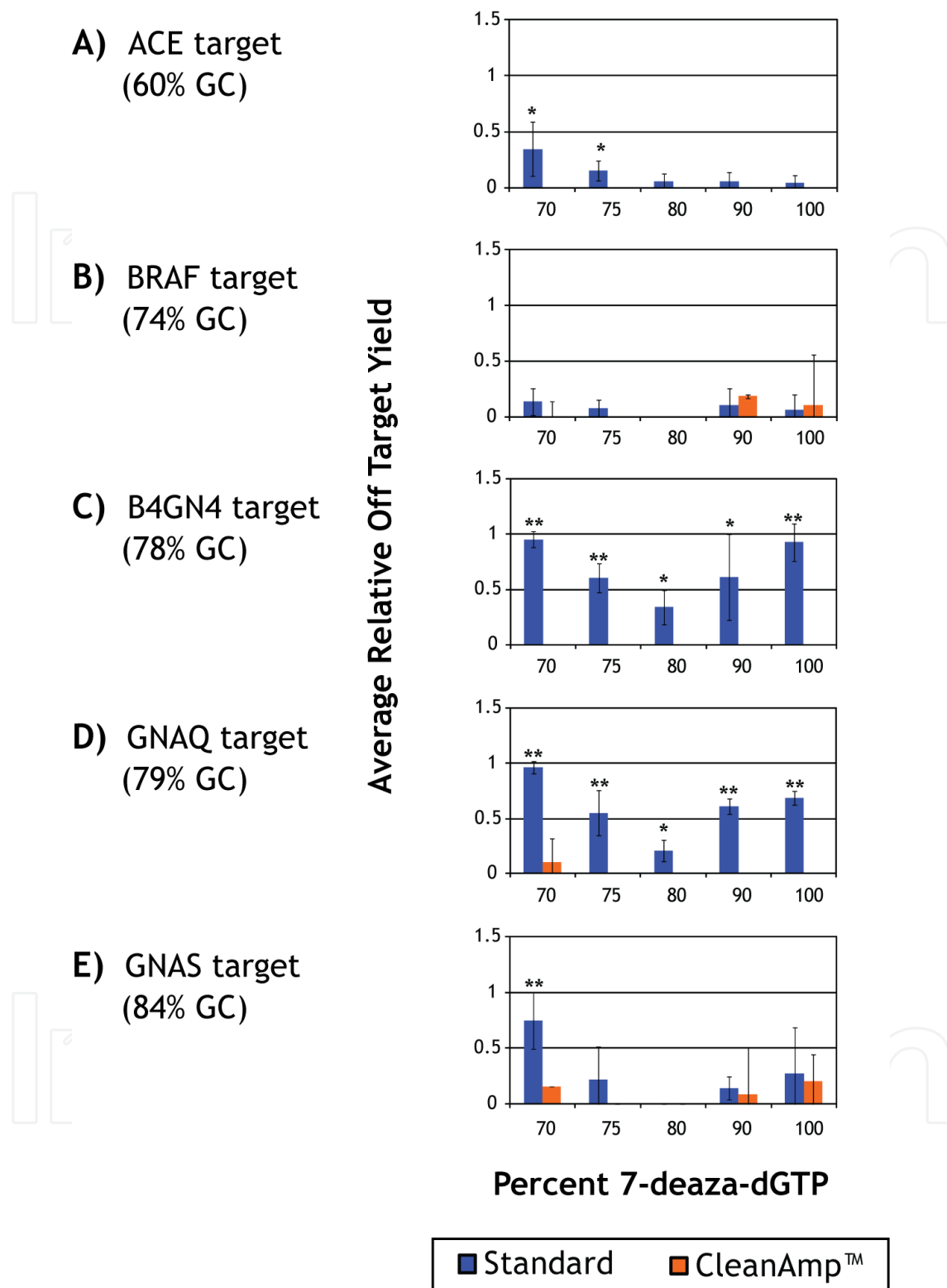
When the amplicon yield of a reaction containing standard dNTPs was compared to an analogous reaction using CleanAmp™ dNTPs for a given percentage of 7-deaza-dGTP, twelve out of twenty five reactions showed a significant improvement in amplicon yield with CleanAmp™ dNTPs (Figure 1). While there were not many statistically significant differences for the lower GC-rich targets ACE (60%) and BRAF (74%), the three highest GC-rich targets: B4GN4 (78%), GNAQ (79%), and GNAS (84%) showed an improved amplicon yield using CleanAmp™ dNTPs over standard dNTPs for several 7-deaza-dGTP compositions. In the case of B4GN4, CleanAmp™ dominates in all the 7-deaza-dGTP mixtures except 100%, where very little amplicon was formed with either type of dNTPs. Furthermore, the addition of magnesium chloride had little effect on amplicon yield for this target producing very low yields prior to PCR cleanup.

The second factor was the evaluation of several 7-deaza-dGTP nucleotide blends (from 70 to 100%) across each group (CleanAmp™ or standard dNTPs) for a given target to determine an ideal amount of 7-deaza-dGTP for optimal PCR yield. In addition to the graphs shown in Figure 1, Table 1 includes the raw data averages and numerical standard deviations for this figure. Also featured in this table are the results for statistical comparison of each of the 7-deaza-dGTP compositions with one another across each of the two groups: standard or CleanAmp™ dNTPs. The colored boxes in the tables represent which of the five 7-deaza-dGTP mixtures fared the best with each type of dNTP. The results of this analysis indicate that there was not just one composition that dominated over the others for all five targets but there were some noteworthy trends. First, it was common that the 90 or 100% compositions produced less or comparable amplicon yield than any of the lower percentage compositions despite the additional MgCl₂ used. Second, although several different 7-deaza-dGTP blends gave comparable amplicon yields for a given target, the blend that was optimal was not consistent from one target to the next. Results were more well-defined for the CleanAmp™ group than for standard dNTPs. For CleanAmp™ dNTPs, a 75% 7-deaza-dGTP mix produced the most



Amplicon yield for each target was normalized to the 75% CleanAmp™ 7-deaza dGTP mixture, and all values were analyzed with a two tailed t-test where (* $p < 0.05$; ** $p < 0.01$).

Fig. 1. Average Relative Amplicon Yield for five GC-rich targets (A-E) amplified using a dNTP mixture with 70 to 100% 7-deaza-dGTP.



Off-Target (primer dimer and mis-priming) yield for each target was quantified relative to the desired amplicon and analyzed with a two tailed t-test where probability values (* $p < 0.05$; ** $p < 0.01$) specify statistically significant values.

Fig. 2. Average Relative Off-Target Yield for five GC-rich targets (A-E) amplified using a dNTP mixture with 70 to 100% 7-deaza-dGTP.

amplicon for two targets, B4GN4 and GNAQ. A 70% 7-deaza-dGTP mix provided the highest yield for BRAF and gave higher yields than at least three other mixtures for ACE and GNAS. On the other hand, standard dNTP mixtures showed few obvious trends and varied considerably from one target to the next. For example, the B4GN4 target had comparable amplicon yields for the 75 and 80% mixtures, which provided higher yields than the remaining mixtures, while for the GNAS target, the 75, 90, and 100% yields were comparable and had greater yields than the 70 and 80% mixtures. Overall, one single 7-deaza-dGTP composition could not be identified for highest amplicon yield. In addition to amplicon yield another important factor in PCR product preparation for sequencing is amplicon quality.

Target	% 7-deaza-dGTP	Average Relative Amplicon Yield (Standard deviation)		Average Relative Off-Target (Standard deviation)	
		Standard	CleanAmp™	Standard	CleanAmp™
ACE (60% GC; 156 bp)	70	0.55 (0.17)	1.28 (0.43)	0.35 (0.24)	0 (0.0)
	75	0.90 (0.38)	1.00 (0.0)	0.15 (0.09)	0 (0.0)
	80	0.85 (0.51)	0.79 (0.22)	0.05 (0.07)	0 (0.0)
	90	0.26 (0.25)	0.21 (0.14)	0.06 (0.08)	0 (0.0)
	100	0.17 (0.15)	0.24 (0.22)	0.04 (0.07)	0 (0.0)
BRAF (74% GC; 185 bp)	70	1.13 (0.15)	1.23 (0.11)	0.14 (0.12)	0 (0.0)
	75	1.03 (0.07)	1.00 (0.0)	0.08 (0.07)	0 (0.0)
	80	0.85 (0.07)	0.89 (0.06)	0 (0.0)	0 (0.0)
	90	0.54 (0.09)	0.75 (0.04)	0.11 (0.15)	0.19 (0.42)
	100	0.37 (0.04)	0.53 (0.05)	0.06 (0.14)	0.11 (0.24)
B4GN4 (78% GC; 720 bp)	70	0.03 (0.04)	0.78 (0.15)	0.95 (0.07)	0 (0.0)
	75	0.41 (0.26)	1.00 (0.0)	0.60 (0.13)	0 (0.0)
	80	0.45 (0.18)	0.82 (0.06)	0.34 (0.16)	0 (0.0)
	90	0.04 (0.04)	0.28 (0.06)	0.61 (0.39)	0 (0.0)
	100	0.02 (0.03)	0.02 (0.02)	0.92 (0.17)	0 (0.0)
GNAQ (79% GC; 642 bp)	70	0.04 (0.06)	0.56 (0.15)	0.96 (0.05)	0.10 (0.22)
	75	0.43 (0.26)	1.00 (0.0)	0.55 (0.20)	0 (0.0)
	80	0.77 (0.26)	0.73 (0.08)	0.21 (0.10)	0 (0.0)
	90	0.15 (0.05)	0.47 (0.10)	0.61 (0.07)	0 (0.0)
	100	0.10 (0.05)	0.14 (0.02)	0.69 (0.06)	0 (0.0)
GNAS (84% GC; 242 bp)	70	0.07 (0.12)	1.87 (0.47)	0.75 (0.25)	0.15 (0.14)
	75	0.74 (0.30)	1.00 (0.0)	0.22 (0.30)	0 (0.0)
	80	0.25 (0.24)	0.38 (0.33)	0 (0.0)	0 (0.0)
	90	0.83 (0.33)	1.74 (0.40)	0.14 (0.11)	0.09 (0.02)
	100	0.76 (0.23)	0.62 (0.14)	0.27 (0.41)	0.20 (0.45)

Amplicon yields were integrated and normalized relative to the 75% CleanAmp™ 7-deaza-dGTP mixture. Off-target amplification yields are the fraction of off-target product formed relative the desired amplicon as determined by gel densitometry. The five sets of 7-deaza-dGTP compositions in each group (standard or CleanAmp™) were analyzed by a one-way ANOVA and Tukey-Kramer post test where ($p < 0.05$; $p < 0.01$; $p < 0.001$) for a given percentage of 7-deaza-dGTP. Boxes outlined in color represent means that give statistically significant values.

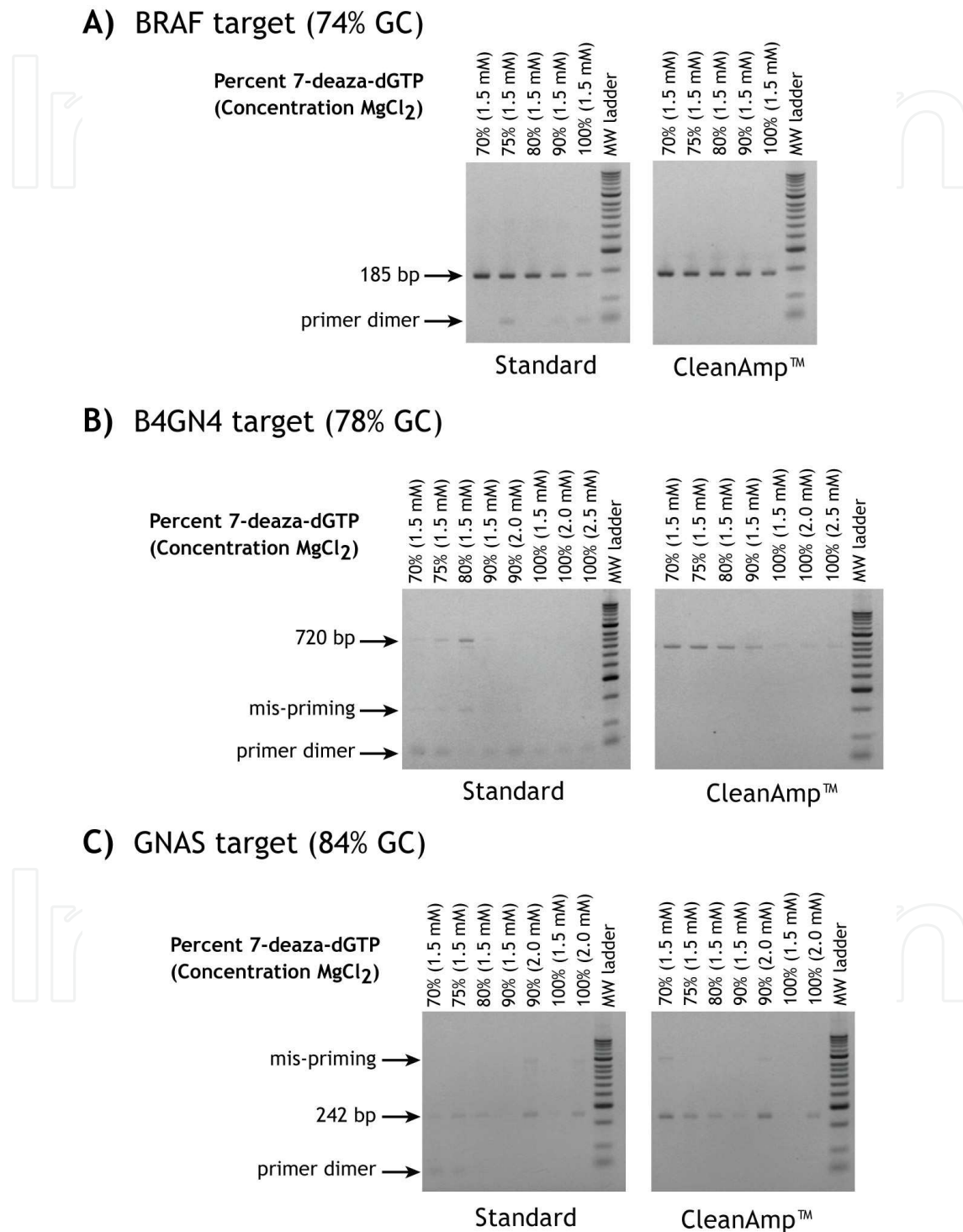
Table 1. Relative Average Amplicon Yield and Off-Target Integration Data

Amplicon quality indicates the purity of the sample that is being sent for sequencing. A high quality sample should contain only the DNA target to be sequenced and be free of any contaminants, excess primers, excess dNTPs, or off-target products which might interfere with the sequencing reaction. In this study, the PCR products went through a commonly used PCR cleanup process that rids the samples of excess dNTPs and primers but does not remove off-target products that were generated during the PCR. Generation of a high quality PCR product with no off-target would eliminate the more laborious step of gel purification prior to sequencing. Therefore, in this chapter the amplicon quality was assessed by integrating the amount of average relative off-target products in the sample after the five PCR replicates were pooled, cleaned, and concentrated. Amplicon quality is represented graphically as the fraction of off-target (mis-priming and primer dimer) generated relative to the amplicon (Figure 2, Table 1), so the lower this value, the higher the sample quality will be.

Thirteen out of twenty five reactions showed significant reduction in off-target when CleanAmp™ dNTPs were used (Figure 2). The ACE off-target products consisted entirely of primer dimer while the other four targets were prone to a combination of primer dimer and mis-priming side products (Figure 3). The two lowest percent GC-rich targets, ACE and BRAF, produced the lowest amount of off-target and highest amplicon quality, with comparable performance between CleanAmp™ and standard dNTPs. The other three target reactions formed substantially more off-target, especially when amplified with standard dNTPs. Figure 2 results show that amplicon quality is highest in most cases when CleanAmp™ mixtures are used. For the B4GN4 and GNAQ targets, reactions with standard dNTPs produced significantly more off-target products than CleanAmp™ dNTPs regardless of the percent 7-deaza-dGTP. For all five amplicons, the reactions using 75 and 80% 7-deaza-dGTP produced no primer dimer or mis-priming when CleanAmp™ was used. As was the case for amplicon yield, there was not just one composition of 7-deaza-dGTP that produced the highest quality DNA product. However it was evident that the use of CleanAmp™ dNTPs reduced the amount of off-target compared to standard dNTPs, providing an amplicon with a higher chance of successfully being sequenced.

Although it is important to generate a high quality PCR product with adequate yield, other measures of the sequencing results, such as read length, pairwise identity, and percent high quality base scores should also be considered. Read length is the number of bases that were called for a given target. Optimally, this value should match the length of the reference sequence, provided it does not exceed the ~1000 base pair limits of the current Sanger dideoxy sequencing technology (Slater & Drouin, 1992; Kieleczawa, Adam et al., 2009). For some challenging targets shorter than these upper limits, the read lengths can often become truncated due to complex secondary structures or regions of DNA that the polymerase can not read through such as GC-rich regions. Results in Table 2 indicate that the average read lengths for the five GC-rich targets were comparable between CleanAmp™ and standard dNTPs in nearly every case. These results were not surprising since these five targets, which were chosen mainly for GC content are not long enough to accurately assess the impact on read length. Therefore the effect of GC content on read length remains yet to be determined in assays with longer amplicons. B4GN4, the longest target with 720 base pairs, varied the most in read length based on standard deviations, suggesting that this was one of the most difficult targets out of the group and just beginning to approach the length

threshold where sequencing becomes more of a challenge. For these targets, no one 7-deaza-dGTP mixture fared better than the rest, as all of the average read length values were statistically comparable to one other. In addition to the read length of the sequence, it is critical to identify the correct bases within a sequence.



Gel images show the variability among the 3 different GC-rich targets in amplicon yield and off-target products (primer dimer and mis-priming).

Fig. 3. Agarose gel images of BRAF, B4GN4, and GNAS amplicons post PCR cleanup

3.2 Sanger dideoxy sequencing data quality

The pairwise identity in an alignment of two sequences is the percentage of shared identical bases (Drummond AJ, 2011). In this chapter, all sequences were known, so all experimental sequencing read-outs were aligned to the appropriate GenBank reference sequences (Dennis A. Benson, 2005). If the read-outs matched exactly, they would have 100% pairwise identity to the reference sequence. For ACE and BRAF, the sequencing data matched with the reference sequences nearly 100% of the time for both standard and CleanAmp™ dNTPs in all 7-deaza-dGTP compositions. For the three targets with higher GC content, results showed that standard and CleanAmp™ dNTPs pairwise identity values approached 100% with no statistically significant differences for most 7-deaza-dGTP mix compositions. However, at 70% 7-deaza-dGTP, three targets amplified with CleanAmp™ dNTPs had higher pairwise identity values than standard dNTPs. Another noteworthy outlier was B4GN4, which was prone to the highest level of off-target formation. Although no one 7-deaza-dGTP composition improved results over any of the others in the group, reactions employing CleanAmp™ dNTPs provided a higher pairwise identity to the B4GN4 reference sequences at lower 7-deaza-dGTP substitutions. While read length and correct base calls are important parameters in determining the quality of sequencing data, the most critical parameter is the percentage of high quality base calls or HQ percentage.

HQ percentage differs from pairwise identity, as it is a measure of the confidence by which the sequencing software can determine a sequence (Drummond AJ, 2011). Often, the pairwise identity may be a 100% match to the reference sequence but have very low confidence values for each base call within the sequence. The confidence in base calling becomes more important when sequencing unknown regions of DNA where there is no reference sequence. In these cases, the resultant sequence can only be trusted if the confidence of the sequencing software is high enough. Phred scores, or quality scores are a widely accepted measure of the quality of DNA sequences. Phred scores are numerical estimates of error probability for a given base (Ewing & Green, 1998). Sequencing softwares each have their own scale for base scoring, and in these studies, the percent of high quality base calls in a sequence read-out (HQ%) is defined to be the percent of bases that have a quality score (phred score) greater than 40. The highest score is a one in a million (10^{-6}) probability of a calling error where a middle score (20-40) represents a probability of only a one in a thousand (10^{-3}) (Drummond AJ, 2011). The data presented herein includes the HQ percentages for each sample (Table 2 and Figure 4).

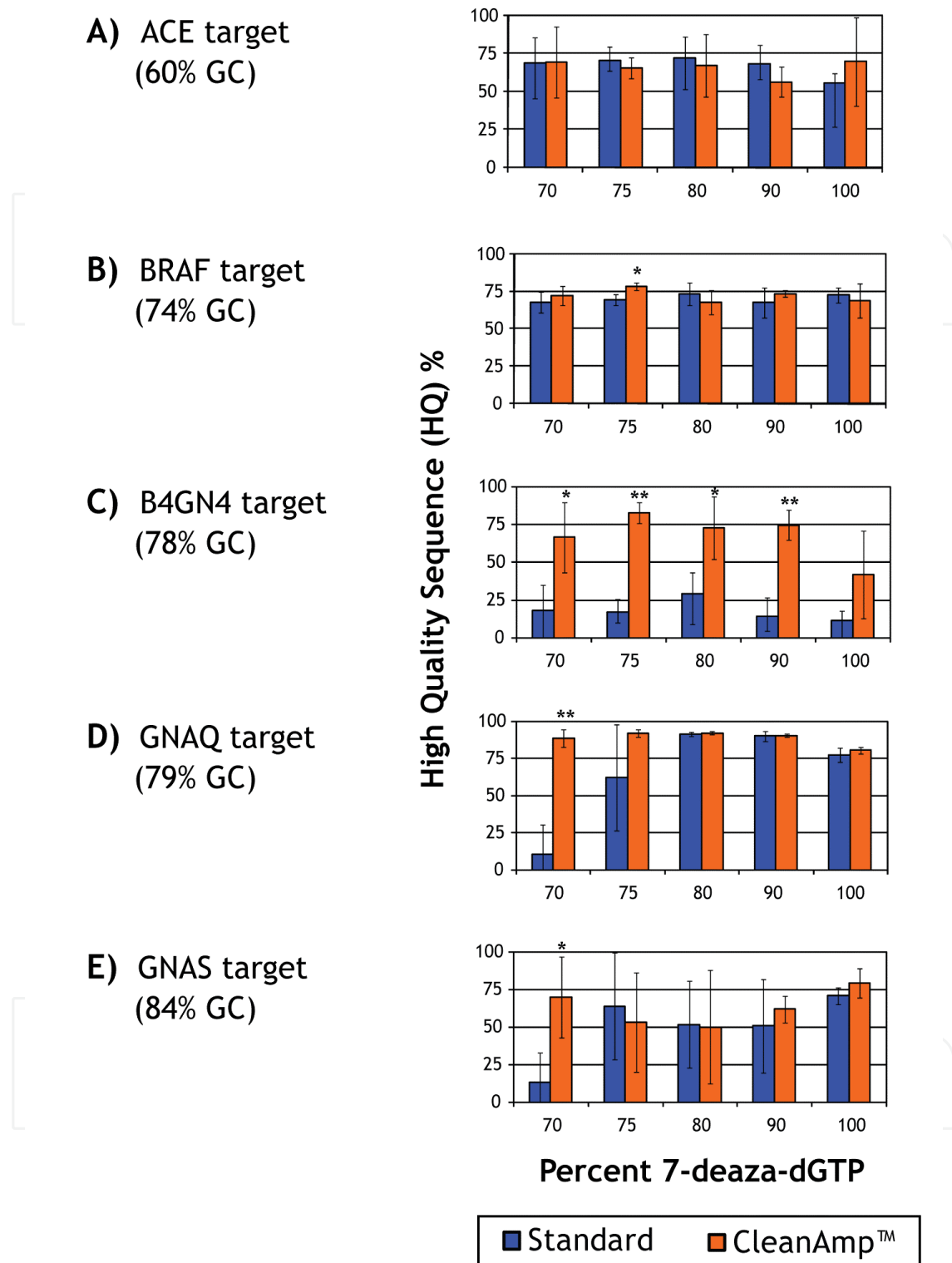
The HQ percent scores for the ACE and BRAF sequencing data showed minimal differences between CleanAmp™ and standard dNTPs for a given 7-deaza-dGTP mix composition. For B4GN4, GNAQ, and GNAS, the sequencing of amplicons generated with CleanAmp™ dNTPs showed significantly improved HQ scores at 70% 7-deaza-dGTP relative to analogous reactions with standard dNTPs which did not reach a value of 50%. CleanAmp™ also yielded higher HQ scores ($p < 0.05$) in four out of the five mixtures for the B4GN4 target. When looking at the compositions of 7-deaza-dGTP across each group of CleanAmp™ or standard dNTPs, all HQ percent values for each 7-deaza blend are comparable for low GC content targets. For targets with greater than 75% GC content, there are some noticeable differences in the lowest (70%) and the highest (100%) substitutions of 7-deaza-dGTP. However, there were no statistically significant differences between HQ values in the middle mix compositions of 75, 80, and 90% (Table 2). Although a specific optimal percentage of 7-deaza-dGTP could not be identified, results indicate that 75, 80, and 90% blends provided the best results across a wide

range of targets. Furthermore, CleanAmp™ was found to improve high quality base calls for seven out of the twenty five reactions which included certain challenging targets and lower 7-deaza-dGTP compositions for higher GC-rich species.

Target	% 7-deaza-dGTP	Average HQ% (Standard deviation)		Average Pairwise Identity (Standard deviation)		Average Read Length (Standard deviation)	
		Standard	CleanAmp™	Standard	CleanAmp™	Standard	CleanAmp™
ACE (60%)	70	69.2 (8.6)	69.5 (4.2)	99.6 (0.5)	99.8 (0.4)	108 (0)	108 (0)
	75	70.7 (6.4)	65.9 (5.4)	98.7 (1.9)	99.8 (0.4)	108 (0)	108 (0)
	80	72.0 (5.7)	67.2 (7.4)	100.0 (0)	100.0 (0)	108 (0)	108 (0)
	90	68.3 (7.0)	56.5 (25.5)	100.0 (0)	100.0 (0)	108 (0)	108 (0)
	100	56.0 (28.0)	69.9 (4.4)	99.8 (0.4)	99.8 (0.4)	108 (0)	108 (0)
BRAF (74%)	70	67.8 (7.1)	72.3 (6.3)	97.2 (2.6)	99.4 (0.7)	154.8 (2.5)	154.4 (1.1)
	75	69.7 (3.5)	78.3 (2.5)	98.2 (1.5)	99.4 (0.7)	153.8 (0.8)	153.6 (0.9)
	80	73.5 (7.6)	67.9 (7.9)	99.2 (0.8)	99.8 (0.3)	154.0 (1.4)	154.4 (0.9)
	90	67.7 (10.0)	73.6 (2.1)	98.3 (0.6)	99.3 (0.7)	154.4 (1.8)	154.0 (1.0)
	100	72.7 (5.1)	68.9 (11.5)	98.8 (1.3)	98.5 (1.4)	154.2 (1.5)	154.4 (3.4)
B4GN4 (78%)	70	18.1 (16.6)	66.3 (23.5)	55.3 (4.7)	97.1 (3.2)**	362.6 (204.1)	680.0 (2.6)*
	75	16.7 (8.9)	82.4 (6.9)	79.9 (10.9)	99.3 (0.2)	567.6 (95.4)	678.0 (1.4)
	80	29.2 (14.0)	72.6 (20.6)	86.2 (10.3)	99.0 (0.5)	667.2 (41.6)	672.8 (10.6)
	90	14.2 (12.2)	74.5 (9.9)	68.4 (10.7)	98.7 (1.2)**	535.6 (141.5)	666.4 (27.8)
	100	11.6 (6.0)	41.6 (29.0)	70.0 (16.9)	94.3 (6.6)	634.8 (158.7)	579.8 (195.4)
GNAQ (79%)	70	10.4 (19.8)	88.6 (6.0)	75.7 (17.0)	99.4 (0.7)*	508.8 (67.5)	597.6 (5.4)
	75	62.2 (35.8)	91.6 (2.6)	93.2 (10.6)	99.6 (0.2)	507.0 (204.7)	600.2 (0.5)
	80	91.3 (1.6)	92.0 (1.1)	99.5 (0.2)	99.5 (0.2)	600.0 (0)	600.8 (0.8)
	90	90.0 (3.2)	90.2 (1.1)	99.6 (0.1)	99.4 (0.1)	597.2 (5.7)	601.0 (0)
	100	77.3 (4.7)	80.4 (2.0)	99.3 (0.6)	99.5 (0.2)	600.4 (1.1)	601.0 (0.7)
GNAS (84%)	70	13.0 (19.5)	69.6 (27.1)	65.3 (17.0)	98.3 (2.8)*	162.0 (54.9)	193.2 (0.5)
	75	63.9 (35.8)	52.9 (33.2)	89.4 (20.8)	98.3 (2.8)	175.0 (25.4)	193.2 (0.5)
	80	51.6 (29.0)	49.9 (37.8)	89.6 (22.1)	90.8 (18.3)	164.4 (64.0)	195.2 (5.5)
	90	50.7 (31.2)	61.9 (8.9)	91.2 (17.4)	99.4 (0.2)	194.6 (3.6)	193.0 (0)
	100	70.7 (5.6)	79.3 (9.7)	99.0 (0.5)	99.2 (0.5)	193.2 (0.4)	188.2 (10.7)

The five sets of 7-deaza-dGTP compositions in each group (standard or CleanAmp™) were analyzed by a one-way ANOVA and Tukey-Kramer post test where probability values shown with highlighted boxes ($p < 0.05$; $p < 0.01$; $p < 0.001$) find the means statistically significant for HQ values, pairwise identity, and read length. In addition, Pairwise Identity and Read Length comparison of CleanAmp™ to standard dNTPs in individual 7-deaza-dGTP compositions, which is not shown by bar graph, are indicated by stars where (* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$) represent values that are statistically significant.

Table 2. Average HQ, Pairwise Identity, and Read Length Data Averages



The percentage of bases in a sequencing read out which had a high quality score of 40 or higher determined a HQ percent value. These values were averaged after alignments to each reference sequence and analyzed with a two tailed t-test where probability values (* $p < 0.05$; ** $p < 0.01$) find the means statistically significant.

Fig. 4. Average HQ Percentages from sequencing read-outs for all five GC-rich targets (A-E) amplified with 7-deaza-dGTP mixes of different compositions.

In summary, these studies have investigated both the percent of 7-deaza-dGTP substitution and the influence of standard and CleanAmp™ versions of the nucleotide mix on PCR and Sanger sequencing performance. When different metrics such as amplicon yield, amplicon quality, HQ percentage and sequencing chromatogram quality are considered, there are notable instances where reactions employing CleanAmp™ dNTPs have either comparable performance or statistically significant improvements in performance. To better understand how these parameters interplay with one another, a more detailed analysis will be presented in the Conclusion section.

4. Conclusion

Both standard and CleanAmp™ dNTPs can effectively generate PCR amplicons with GC-rich sequence when amplified in combination with 7-deaza-dGTP. The use of this nucleotide analog effectively linearizes the DNA sequences in preparation for Sanger dideoxy sequencing by destabilizing secondary structures such as G-quadruplexes. This allows the DNA fragments to migrate more predictably through the polyacrylamide gel and reduces the possibility of ambiguous base calls in sequencing results (Motz, Svante et al., 2000). CleanAmp™ dNTPs also offer the added benefit of reduced off-target amplification due to Hot Start activation in the PCR assay. Therefore the effect of Hot Start PCR activation in conjunction with the extent of 7-deaza-dGTP substitution was investigated to determine its potential benefit.

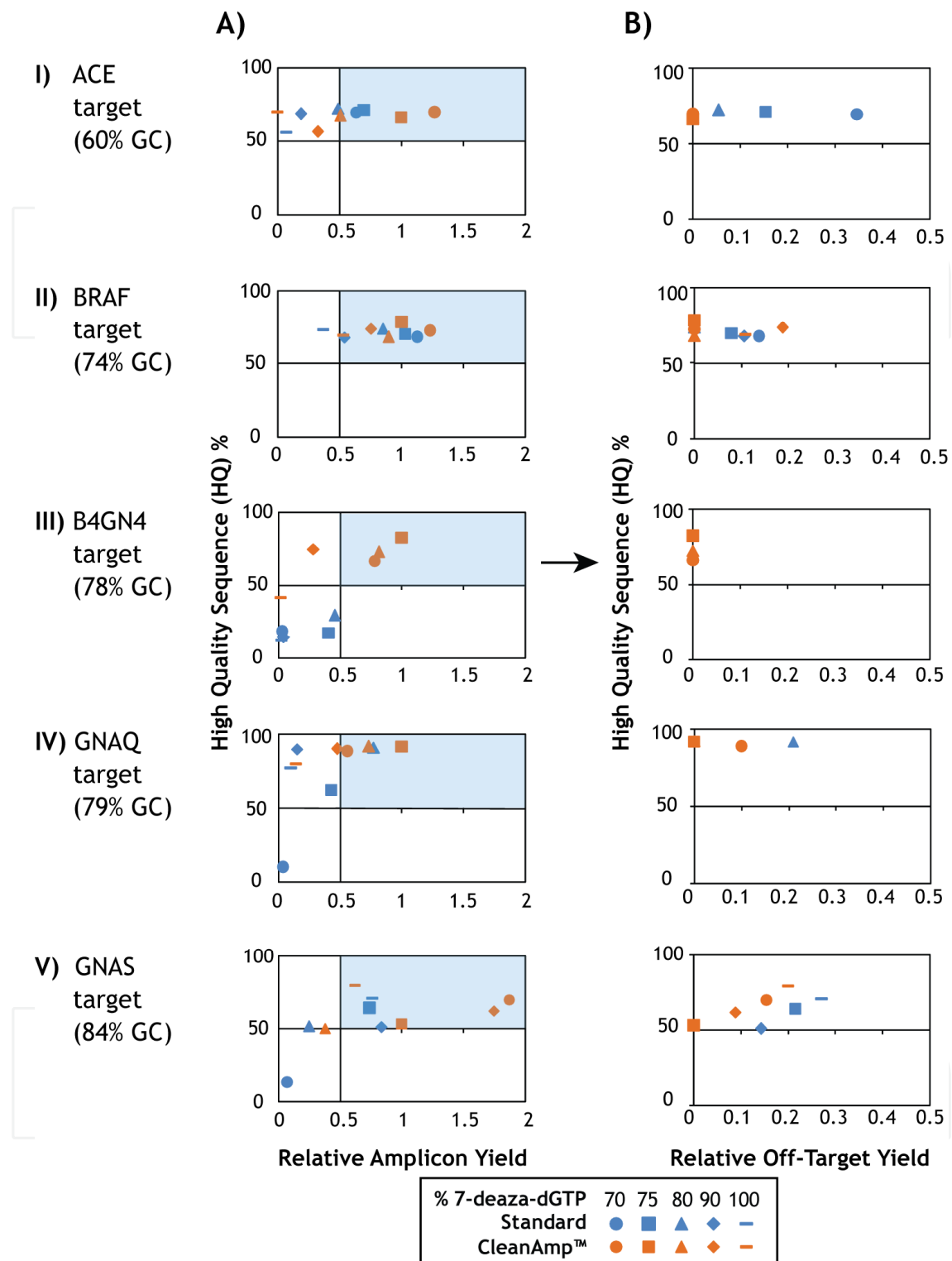
For the targets with lower GC content (less than 75%), CleanAmp™ dNTPs helped to reduce off-target product formation at the PCR step but gave comparable results to standard dNTPs in all other categories. For the three highest GC-rich targets, CleanAmp™ improved amplicon yield and amplicon quality with several different 7-deaza-dGTP compositions, indicating that the Hot Start activation is a much-needed benefit. In one case, B4GN4, CleanAmp™ significantly improved amplicon yield, amplicon quality, and percent HQ over standard dNTPs for at least four out of the five 7-deaza-dGTP mixtures (from 70-100% 7-deaza-dGTP). Quality sequencing results for this target were not achieved with standard dNTPs alone. The use of CleanAmp™ dNTPs at the PCR stage also improved amplicon yield, pairwise identity, and percent HQ with the 70% 7-deaza-dGTP composition. However, reactions with an analogous mixture of standard dNTPs were not as successful, indicating that standard dNTP mixtures may require a higher percentage of 7-deaza-dGTP. Although the categories of pairwise identity and read length showed minimal differences when it came to using standard or CleanAmp™ dNTPs, CleanAmp™ dNTPs improved the PCR assay and down stream sequencing results over standard dNTPs in DNA targets with GC content higher than 75%.

After analyzing the results in five categories individually, it was determined that three of the categories, amplicon yield, amplicon quality, and percent HQ, were most affected by the variables being investigated. Therefore the influence of these categories on one another was more thoroughly studied to discern the most optimal percent 7-deaza-dGTP mixture. Figure 5A(I to V) shows scatter plots of the percent HQ and amplicon yield for all variables being tested. The shaded portion of the plot highlights dNTP mixtures that reached a threshold of at least 50% relative amplicon yield and 50% high quality bases called. The dNTP compositions that were found in this region of the scatter plot were identified and re-plotted in a scatter plot of HQ percent versus amplicon quality (Figure 5B (I to V)). Optimal compositions for Figure 5B lie highest on the plots for HQ scores and furthest to the left for the least amount of off-target formed or highest amplicon quality.

Several of the different CleanAmp™ dNTP mixtures met the threshold requirements for amplicon yield, were high in amplicon quality and yielded high HQ scores. While many of the standard dNTP mixtures also have adequate amplicon yield, the amplicon quality suffered for several targets. From the scatter plot analysis in Figure 5, the 75% CleanAmp™ mixture provided adequate amplicon yield, best amplicon quality and highest HQ scores for 4 out of the 5 targets. For the 75% mixture of CleanAmp™ dNTPs in the GNAS target adequate yield and high amplicon quality were evident but the HQ scores were lower for this composition. The lesser correlation of GNAS to the other samples may be due to its higher GC content (84%) and the higher concentrations of magnesium chloride, which was needed for adequate amplicon yield. For targets with higher GC than 80% composition, the data indicated that complete substitution of 7-deaza-dGTP may be necessary. If more replicates are pooled to produce enough amplicon at a lower magnesium concentration then it is likely that the off-target products will decrease and base call quality will still remain high. If standard dNTPs are used in the PCR step, an 80% 7-deaza-dGTP mixture was the optimal composition. However amplicon quality can sometimes still be affected at this composition with standard dNTPs, which may require a more laborious gel purification step to remove off-target products prior to sequencing.

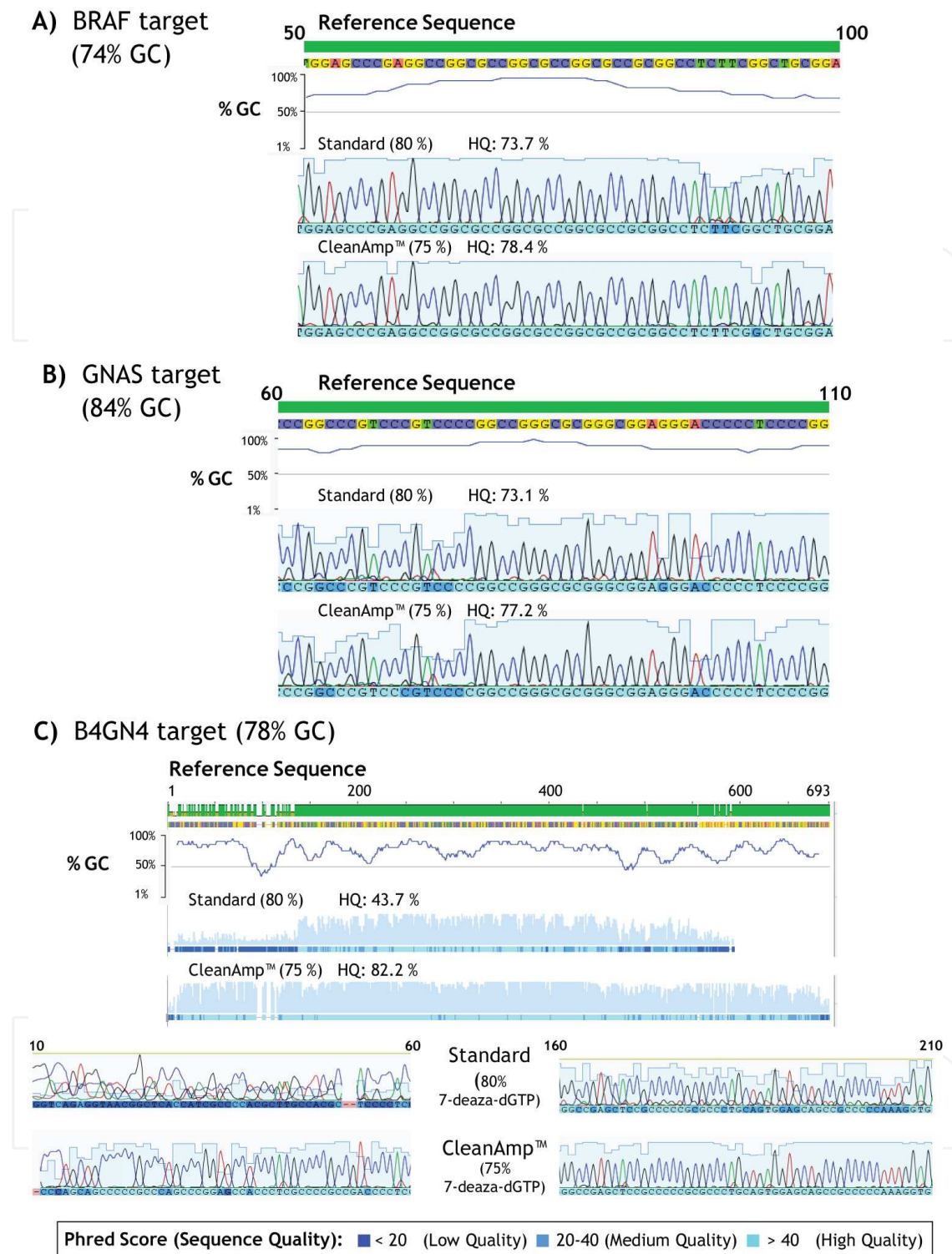
In addition to the numerical metrics from the sequencing data of the five targets, the sequencing chromatograms were studied. In Figure 6, representative chromatograms for the regions with high GC content for the BRAF, GNAS, and B4GN4 targets are presented. Comparisons show the optimal compositions of CleanAmp™ with 75% 7-deaza-dGTP and standard with 80% 7-deaza-dGTP. For BRAF (74% GC), a region from 50-100 bp is shown. Since the analogous HQ percentages are similar for templates with standard and CleanAmp™ dNTPs, it was not surprising that the chromatograms are similar in this region. Similarly for GNAS, the HQ percentages were comparable for both CleanAmp™ and standard dNTP samples, with a modest improvement in chromatogram shape and base call confidence for the CleanAmp™ dNTP target. For B4GN4, there were significant differences in the HQ percentages between standard (HQ: 43.7%) and CleanAmp™ (HQ: 82.2%). The sequencing trace for standard dNTPs died out at 600 bp, while the trace for CleanAmp™ dNTPs persists to the end of the target (~700 bp). Two representative regions of sequence are shown (10-60 bp and 160-210 bp), where reactions with CleanAmp™ dNTPs had strong performance for both regions, and reactions with standard dNTPs had poor performance in the early part of the read, culminating in stronger performance mid-sequence.

Overall, these studies represent a thorough investigation of both the effect of 7-deaza-dGTP substitution and the use of a Hot Start PCR technology on PCR amplification and downstream sequencing performance. Though the differences were subtle for the extent of 7-deaza-dGTP substitution when individual parameters were analyzed, the advantages of using CleanAmp™ over standard versions of the nucleotide mix, were more pronounced. Upon a more detailed analysis of amplicon yield, specificity and downstream sequencing quality, optimal nucleotide compositions were revealed. Future studies may include exploration of more targets greater than 80% in GC composition, longer templates, and the incorporation of CleanAmp™ 7-deaza-dGTP into the sequencing step.



In column A, HQ and amplicon yield values that lie in the blue shaded boxes (top right) met HQ and amplicon yield threshold values and were then re-plotted with scatter plots in column B (HQ versus relative off-target yield). Values that lie furthest to the left and highest on the plots in column B are the most optimal mixtures for PCR and downstream sequencing.

Fig. 5. A) HQ-Amplicon Yield and B) HQ-relative off-target yield Scatter Plots for all five GC-rich targets (I-V).



Base call quality is represented by the height of the blue shading behind the peaks which correlates with the base color (shade of blue) in the sequence. Light blue shaded bases (A, C, G, or T) are of high quality and have scores higher than 40. CleanAmp™ dNTPs show improved read length, high quality base scores, and reduced base compressions.

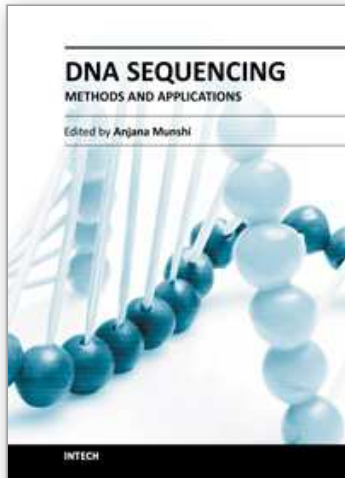
Fig. 6. Sequencing chromatograms of high GC-rich regions of A) BRAF, B) GNAS, and C) B4GN4 comparing standard 7-deaza-dGTP at 80% and CleanAmp™ 7-deaza-dGTP at 75% 7-deaza-dGTP substitution.

5. Acknowledgment

I would like to thank G. Zon, V. Boyd, A. McCaffrey, and R. Hogrefe for careful input and editing of this manuscript

6. References

- Abramoff, M.D., Magalhaes, P.J., Ram, S.J. (2004). "Image Processing with ImageJ." *Biophotonics International* 11(7): 36-42.
- Dennis A. Benson, I.K.-M., David J. Lipman, James Ostell, and David L. Wheeler (2005). "GenBank." *Nucleic Acids Res* 33(Database issue): D34-8.
- Drummond AJ, A.B., Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2011). Geneious Pro v5.4, Available from <http://www.geneious.com/>.
- Ewing, B. & Green, P. (1998). "Base-calling of automated sequencer traces using phred. II. Error probabilities." *Genome Res* 8(3): 186-94.
- Frey, U.H., Bachmann, H.S., et al. (2008). "PCR-amplification of GC-rich regions: 'slowdown PCR'." *Nat Protoc* 3(8): 1312-7.
- Jung, A., Ruckert, S., et al. (2002). "7-Deaza-2'-deoxyguanosine allows PCR and sequencing reactions from CpG islands." *Mol Pathol* 55(1): 55-7.
- Kieleczawa, J. (2006). "Fundamentals of sequencing of difficult templates--an overview." *J Biomol Tech* 17(3): 207-17.
- Kieleczawa, J., Adam, D., et al. (2009). "Identification of optimal protocols for sequencing difficult templates: results of the 2008 ABRF DNA Sequencing Research Group difficult template study 2008." *J Biomol Tech* 20(2): 116-27.
- Koukhareva, I. & Lebedev, A. (2009). "3'-Protected 2'-deoxynucleoside 5'-triphosphates as a tool for heat-triggered activation of polymerase chain reaction." *Anal Chem* 81(12): 4955-62.
- McConlogue, L., Brow, M.A., et al. (1988). "Structure-independent DNA amplification by PCR using 7-deaza-2'-deoxyguanosine." *Nucleic Acids Res* 16(20): 9869.
- Motz, M., Svante, P., et al. (2000). "Improved Cycle Sequencing of GC-Rich Templates by a Combination of Nucleotide Analogs." *BioTechniques* 29(2): 268-270.
- Prism Graphpad Prism Software. San Diego, CA USA.
- Prober, J.M., Trainor, G.L., et al. (1987). "A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides." *Science* 238(4825): 336-41.
- Sanger, F., Nicklen, S., et al. (1977). "DNA sequencing with chain-terminating inhibitors." *Proc Natl Acad Sci U S A* 74(12): 5463-7.
- Shore, S. & Paul, N. (2010). "Robust PCR amplification of GC-rich targets with Hot Start 7-deaza-dGTP." *BioTechniques* 49(5): 841-843.
- Simonsson, T. (2001). "G-quadruplex DNA structures--variations on a theme." *Biol Chem* 382(4): 621-8.
- Slater, G.W. & Drouin, G. (1992). "Why can we not sequence thousands of DNA bases on a polyacrylamide gel?" *Electrophoresis* 13(8): 574-82.
- Weitzmann, M.N., Woodford, K.J., et al. (1996). "The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation." *J Biol Chem* 271(34): 20958-64.
- Zhang, Z., Yang, X., et al. (2009). "Enhanced amplification of GC-rich DNA with two organic reagents." *BioTechniques* 47(3): 775-9.



DNA Sequencing - Methods and Applications

Edited by Dr. Anjana Munshi

ISBN 978-953-51-0564-0

Hard cover, 174 pages

Publisher InTech

Published online 20, April, 2012

Published in print edition April, 2012

This book illustrates methods of DNA sequencing and its application in plant, animal and medical sciences. It has two distinct sections. The one includes 2 chapters devoted to the DNA sequencing methods and the second includes 6 chapters focusing on various applications of this technology. The content of the articles presented in the book is guided by the knowledge and experience of the contributing authors. This book is intended to serve as an important resource and review to the researchers in the field of DNA sequencing.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Sabrina Shore, Elena Hidalgo Ashrafi and Natasha Paul (2012). Hot Start 7-Deaza-dGTP Improves Sanger Dideoxy Sequencing Data of GC-Rich Targets, DNA Sequencing - Methods and Applications, Dr. Anjana Munshi (Ed.), ISBN: 978-953-51-0564-0, InTech, Available from: <http://www.intechopen.com/books/dna-sequencing-methods-and-applications/hot-start-7-deaza-dgtp-improves-sanger-dideoxy-sequencing-data-of-gc-rich-targets>

INTECH
open science | open minds

InTech Europe

University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China

Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the [Creative Commons Attribution 3.0 License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen