GC-rich DNA regions are found in >60% of gene promoters in higher eukaryotes (1). Most housekeeping and tumor suppressor genes, as well as ~40% of tissue-specific genes, contain high-GC sequences in their promoter regions (2). Therefore, the study of GC-rich regions is of great interest in biological and clinical investigations. The amplification of genes of interest by PCR is an essential step in the study of their function and regulation, but DNA regions that are >60% GC require special PCR protocols due to their complex and strong secondary structures that resist denaturation and interfere with primer annealing. Secondary structures can also block the advance of DNA polymerase during the extension step, resulting in poor amplification of the desired products or in the amplification of shortened PCR products lacking the sequence containing the secondary structure (3–8).

There are a large number of protocols for amplification of GC-rich DNA. Some methodologies involve protocols such as stepdown PCR, slowdown PCR, or PCR with multiple heat pulses during the extension phase (8,9). Certain organic co-solvents such as DMSO, formamide, and ethylene glycol have been shown to facilitate DNA strand separation by disrupting base pairing, and other agents such as tetraalkylammonium (TAA) salts and betaine have also been used (10–13); however, high concentrations of these additives inhibit DNA polymerase activity (14,15). In a different approach, nucleotide analogs such as 7-deaza-2′-deoxyguanosine (7dGTP), 2′-deoxyinosine 5′-triphosphate (dITP), and N4-methyl-2′-deoxycytidine 5′-triphosphate (N4me-dCTP) have been reported as being useful by destabilizing G:C base pairs (9,16).

GC-rich DNA regions were PCR-amplified with Taq DNA polymerase using either the canonical set of deoxynucleoside triphosphates or mixtures in which the dCTP had been partially or completely replaced by its N4-methylated analog, N4-methyl-2′-deoxycytidine 5′-triphosphate (N4me-dCTP). In the case of a particularly GC-rich region (78.9% GC), the PCR mixtures containing N4me-dCTP produced the expected amplicon in high yield, while mixtures containing the canonical set of nucleotides produced numerous alternative amplicons. For another GC-rich DNA region (80.6% GC), the target amplicon was only generated by re-amplifying a gel-purified sample of the original amplicon with N4me-dCTP–containing PCR mixtures. In a direct PCR comparison on a highly GC-rich template, mixtures containing N4me-dCTP clearly performed better than did solutions containing the canonical set of nucleotides mixed with various organic additives (DMSO, betaine, or ethylene glycol) that have been reported to resolve or alleviate problems caused by secondary structures in the DNA. This nucleotide analog was also tested in PCR amplification of DNA regions with intermediate GC content, producing the expected amplicon in each case with a melting temperature (Tm) clearly below the Tm of the same amplicon synthesized exclusively with the canonical bases.

METHOD SUMMARY

Here, we report use of the nucleotide analog N4-methyl-2′-deoxycytidine 5′-triphosphate to partially or totally replace dCTP in PCR amplification of high-GC content DNA regions resistant to amplification using the canonical set of nucleotides.
human genes (Table 1) using the PrimerQuest program from Integrated DNA Technologies (IDT, Coralville, IA). Primers were selected that contained the fewest possible guanines in their sequence in order to minimize the reduction of the primer hybridization temperature for the DNA template containing N\textsuperscript{4}meC (19). Primers were also evaluated for their tendency to form homodimers, heterodimers, and hairpin structures using the OligoAnalyzer program (IDT). Finally, primers were evaluated with primer-BLAST (www.ncbi.nlm.nih.gov/tools/primer-blast/) to ensure their uniqueness within the human genome.

**PCR experiments**

All PCR experiments were performed in a LightCycler 1.5 thermocycler (Roche, Basel, Switzerland) in borosilicate glass capillaries (Roche). PCR was carried out in a total volume of 10 µL. The amplification mixtures contained 30 ng of human genomic DNA (Roche); 0.5 µM of each primer (IDT); PCR buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3) (Sigma-Aldrich, St. Louis, MO); 200 µM of each deoxyribonucleotide (dTTP, dGTP, dTTP, and dCTP) (Sigma-Aldrich) or 200 µM N\textsuperscript{4}me-dCTP (TriLink Biotechnologies, San Diego, CA) instead of dCTP in the case of modified amplicons; 1.5 U HotStart Taq DNA polymerase (Sigma-Aldrich); 2.5 mM MgCl\textsubscript{2} (Sigma-Aldrich); and SYBR Green I (Invitrogen, Carlsbad, CA) at a final concentration of 0.4X.

To evaluate the usefulness of N\textsuperscript{4}me-dCTP in the amplification of highly GC-rich target sequences, we selected regions of the human ARX and GNAS1 genes (78.9% GC and 80.6% GC, respectively).

For the comparison between additives, we chose to amplify the ARX gene. Additives were used in the following concentrations: 2.5%–10% DMSO (Sigma-Aldrich), 0.5–2 M betaine (Sigma-Aldrich), and 0.5–1.5 M ethylene glycol (Fermont, Mexico City, Mexico).

Other less GC-rich segments were also amplified with N\textsuperscript{4}me-dCTP in order to determine the T\textsubscript{m} reduction of the amplicons when N\textsuperscript{4}meC was incorporated.

**Cycling protocols**

The standard thermal program used was: 95°C for 1 min; followed by 45 cycles of 95°C for 30 s; 62°C, 64°C, or 66°C for 30 s; and 72°C for 60 s; with a final extension at 72°C for 4 min. The slowdown program (9) used a heating rate of 2.5°C s\textsuperscript{-1} and a cooling rate of 1.5°C s\textsuperscript{-1}, with temperatures of 95°C for 1 min, then 37 cycles of 94°C for 30 s, annealing for 30 s at the temperatures detailed below, and 72°C for 1 min, followed by a final extension at 72°C for 4 min. The annealing temperature was set at 70°C for the first 2 cycles and then reduced by 1°C every second cycle until 60°C was reached (22 total cycles for 11 steps of decreasing annealing temperature); this was followed by 15 additional cycles with annealing at 60°C.

**PCR product characterization**

PCR products were evaluated by amplicon melting temperature (T\textsubscript{m}), which was obtained from the melting curves (Supplementary Material) provided by the LightCycler 1.5 computer automatically after thermocycling. Furthermore, all PCR products were analyzed by agarose gel electrophoresis: 7–µL aliquots of the PCR mixtures were applied in loading buffer (0.25% xylene cyanol, 0.25% bromophenol blue, 30% glycerol) to 2% agarose gels (Invitrogen) in 1X TBE buffer (Invitrogen) and electrophoresed in a subcell GT electrophoresis chamber (Bio-Rad, Hercules, CA) at 85 V for 60 min. The gels were stained in aqueous ethidium bromide (0.2 µg/mL); rinsed with water, and photographed on a transilluminator (320 nm) (Model LMS-20E; Ultra-Violet Products Ltd., Cambridge, UK). For size standards, a 100-bp or 50-bp DNA marker ladder

---

**Materials and methods**

**Primer design**

Primer pairs were designed to amplify GC-rich segments located in various DNA regions, using the normal complement of nucleotides in the PCR mixture or replacing dCTP with N\textsuperscript{4}me-dCTP, to determine if there is any advantage gained by using this analog. In addition, the effects of some of the above-mentioned additives on PCR were also compared. It was found that the use of N\textsuperscript{4}me-dCTP affords substantial benefits in the amplification of DNA sequences with high GC content and that these benefits exceed those attainable with commonly used co-solvents and additives.

Table 1. Main characteristics of the primers designed for this study.

<table>
<thead>
<tr>
<th>Gene</th>
<th>GenBank</th>
<th>Primer sequence (5’-3’)</th>
<th>T\textsubscript{m} (°C)</th>
<th>Product length (bp)</th>
<th>GC (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GNAS1</td>
<td>AH002748</td>
<td>TCTCTCTCTCTCTCTCCACC GCGGCGCTTATTTGCC</td>
<td>61.6</td>
<td>124</td>
<td>80.6</td>
</tr>
<tr>
<td>ARX</td>
<td>NG_008281</td>
<td>CCGGAGGTCCCTCAAA CTATCTTCTCTCTTGCA</td>
<td>65.5</td>
<td>522</td>
<td>78.9</td>
</tr>
<tr>
<td>BRAF</td>
<td>NG_007873.2</td>
<td>CTGTGTTTAAGATGCGGGCGCTGA AAGGGGAGGGGGGAGGT</td>
<td>72.1</td>
<td>185</td>
<td>74.6</td>
</tr>
<tr>
<td>N5MC1</td>
<td>NT_008046</td>
<td>CCGTCCTCTGAGTCACTACGGT GCCGCTGCTGCAAGGCC</td>
<td>70.3</td>
<td>308</td>
<td>70.8</td>
</tr>
<tr>
<td>TPPI-2</td>
<td>YA044097</td>
<td>CGGTTGTGCGGCTGGGGCTGA AAGGGGAGGGGGGAGGT</td>
<td>63.3</td>
<td>267</td>
<td>67.2</td>
</tr>
<tr>
<td>ACTB</td>
<td>NG_007992.1</td>
<td>GCTAATCTCCTGCTCTATTT GACACCACCTGTACTCTTAT</td>
<td>62.0</td>
<td>220</td>
<td>55.5</td>
</tr>
<tr>
<td>GAPDH</td>
<td>NG_007073.2</td>
<td>CTCTCCTCTCTCTCTCC CCCACACCTCTCTCTCTAGGCT</td>
<td>60.3</td>
<td>195</td>
<td>54.4</td>
</tr>
</tbody>
</table>
Results and discussion

Previously, we demonstrated the capacity of N\(^{\text{me}}\)-dCTP in the reaction mixture to produce specific PCR products with reduced Tm values from DNA regions with moderate GC content (19). Here, GC-rich segments of different human genes were tested by PCR amplification using a reaction mixture containing the canonical set of nucleoside triphosphates at 200 \(\mu\)M each or one in which dCTP was replaced with N\(^{\text{me}}\)-dCTP, either completely (200 \(\mu\)M) or with various proportions of N\(^{\text{me}}\)-dCTP:dCTP (i.e., 10:1 (182 \(\mu\)M:18 \(\mu\)M), 5:1 (167 \(\mu\)M:33 \(\mu\)M), or 3:1 (150 \(\mu\)M: 50 \(\mu\)M)). Figure 1 shows the amplification reactions for the ARX gene using either a canonical nucleotide mix or mixes containing N\(^{\text{me}}\)-dCTP in different proportions.

PCR amplifications conducted with the canonical complement of nucleotides and different thermal profiles produced spurious electrophoretic bands in all cases (Figure 1A). In contrast, amplifications performed with N\(^{\text{me}}\)-dCTP in different proportions produced a single specific band of the expected molecular size (Figure 1B). Similarly, when the dCTP was completely replaced by N\(^{\text{me}}\)-dCTP, the specific band was again obtained (Figure 1B, Lane 7).

Analysis of possible hairpin structures in the ARX gene by the OligoAnalyzer program shows the potential for two extensive hairpin structures formed mainly by G:C base pairs in the central region of the desired amplicon (Supplementary Material), with AG values of -6.49 and -0.84 kcal/mol at 72°C and 50 mM NaCl. These hairpins might block the advance of DNA polymerase during the extension step, resulting in a shorter amplicon due to the enzyme skipping past the hairpin or other amplicons, including longer ones, due to the opportunistic amplification of less difficult to amplify regions of the genome. In contrast, when N\(^{\text{me}}\)C is incorporated by the DNA polymerase during PCR, the resulting DNA product acts as a better template for subsequent PCR amplifications, likely as a result of weaker G:N\(^{\text{me}}\)C base pairs leading to less stable hairpin structures. Upon replacing the G:C pairs in the hairpin-prone regions by A:T pairs (for a rough approximation of the energetics of G:N\(^{\text{me}}\)C pairs, see Reference 17) the OligoAnalyzer program predicts only a very weak hairpin in the sequence of interest, with AG = +0.98 kcal/mol at 72°C and 50 mM NaCl.

The 522-bp PCR product of the ARX gene, obtained using N\(^{\text{me}}\)-dCTP instead of dCTP or with 10:1 N\(^{\text{me}}\)-dCTP:dCTP, was isolated from the gel and sequenced using the reverse primer. This gave the correct sequence over the length of the amplicon, and it aligned with the sequence in the NCBI Nucleotide database when searched with BLAST, confirming that the desired amplicon had indeed been obtained. It should also be noted that the potential hairpin structures did not interfere with sequencing, probably because these structures were relatively destabilized in the template DNA due to the partial or total replacement of the cytosines with N\(^{\text{me}}\)Cs.

We also compared amplification of the GNAS1 gene using the different nucleotide mixes. Using the canonical set of nucleotides and testing different PCR programs, the desired 124-bp amplicon was not obtained; instead, the product mixtures contained a number of unwanted larger amplicons, whose identity and distribution varied with the different PCR conditions.
thermal programs used (Figure 2A). It was also not possible to produce the desired 124-bp amplicon even when the N\textsuperscript{4}me-dCTP nucleotide was used, either with different N\textsuperscript{4}me-dCTP:dCTP input ratios or complete replacement of the canonical nucleotide with the N\textsuperscript{4}-methylated analog. However, the desired product was obtained by first conducting a PCR with the canonical set of nucleotides in 10% DMSO, isolating the desired faint band from the agarose gel using the QIAquick gel extraction kit from QIAGEN (Hilden, Germany), and reamplifying this purified material by PCR with the same pair of primers without DMSO and either mixtures of N\textsuperscript{4}me-dCTP:dCTP or with N\textsuperscript{4}me-dCTP only. This provided satisfactory yields in all cases (Figure 2B). Here, secondary-structure analysis by the OligoAnalyzer program showed one of the two primer hybridization sites to be involved formation of a strong hairpin structure with a neighboring stretch of DNA, which should preclude efficient priming.

Regions with lower GC-content (54.4%–74.6%) from different human genes (BRAF, NSMCE2, TFPI-2, ACTB, GAPDH) were amplified using N\textsuperscript{4}me-dCTP instead of dCTP. All PCR amplifications yielded the target amplicons. Table 2 shows the \( T_m \) values for all fragments amplified in both versions of the PCR mixture, namely, with the canonical set of nucleotides or with complete replacement of dCTP with N\textsuperscript{4}me-dCTP. For the amplified fragments, the \( \Delta T_m \) values obtained for the normal and the modified version of the same amplicon ranged from -8.7°C to -11.3°C; contrary to what might have been expected, these \( T_m \) differences did not correlate with the GC content of the amplicons. Thus, GNAS1, the most GC-rich amplicon (80.6% GC), showed a reduction in \( T_m \) of 9.5°C when cytosine was replaced with N\textsuperscript{4}meC, while the much less GC-rich fragment ACTB (55.5% GC), showed a \( T_m \) reduction of 10.5°C, suggesting that there are other factors besides the fractional GC-content involved in determining the specific melting-temperature differences of the amplicons.

Previous efforts to amplify GC-rich regions have focused on the use of organic compounds such as dimethyl sulfoxide (DMSO), betaine, ethylene glycol, or formamide, among others. These approaches have generally required optimization of the additive concentration and sometimes even the combination of two or more additives to achieve satisfactory amplification. To evaluate N\textsuperscript{4}me-dCTP in comparison with PCR additives, the 522-bp region from the ARX gene was amplified using N\textsuperscript{4}me-dCTP or using the canonical complement of nucleotides with added DMSO, betaine, or ethylene glycol. Gel electrophoresis (Figure 3, Lane 2)
illustrates the difficulty of PCR amplification for this highly GC-rich region (78.9% GC) when neither the nucleotide analog nor any additive is used; at least 3 nonspecific products are evident. The use of N4me-dCTP results in amplification of the correct target, and reduces non-specific amplification (Figure 3, Lane 1). Using the canonical set of nucleotides with 2.5%–10% DMSO, as commonly reported (5,21,22), there is a reduction in the number of nonspecific products, but in no case is the 522-bp fragment obtained as the sole amplification product (Figure 3, Lanes 3–6). DMSO concentrations >10% were not tested, as they are reported to inhibit the DNA polymerase (15). A similar result was observed with betaine (Figure 3, Lanes 7–10), which was evaluated at concentrations of 0.5 M, 1 M, or 1.5 M ethylene glycol; Lanes M: 100-bp DNA marker. Figures 3. Gel electrophoresis of the products of PCR amplification of the 522-bp ARX gene region with N4me-dCTP or with different additives. Lane 1: PCR amplification with N4me-dCTP replacing dCTP, without additives; Lane 2: amplification with dATP/dGTP/dTTP/dCTP. Lanes 3–6: amplification with dATP/dGTP/dTTP/dCTP, with 2.5%, 5%, 7.5%, or 10% DMSO. Lanes 7–10: amplification with dATP/dGTP/dTTP/dCTP, with 0.5 M, 1.5 M, or 2 M betaine; Lanes 11–13: amplification with dATP/dGTP/dTTP/dCTP, with 0.5 M, 1 M, or 1.5 M ethylene glycol; Lanes M: 100-bp DNA marker.

Make a Great Library.

DNA size selection is a key aspect of many NGS workflows. The PippinHT not only automates this step, but improves the quality of results. For short-read sequencing or long-range genomics, make your library great.

<table>
<thead>
<tr>
<th>Method</th>
<th>Length Range</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>dbDNA collection</td>
<td>200–10,000 bp</td>
<td>ddRAD-seq, paired-end sequencing</td>
</tr>
<tr>
<td>Flaired End Sequencing</td>
<td>100–5,000 bp</td>
<td>paired-end sequencing, HiSeq, Illumina</td>
</tr>
<tr>
<td>mRNA Library Isolation</td>
<td>200–1,000 bp</td>
<td>ChIP-seq, ChIP-Seq, long-read sequencing</td>
</tr>
<tr>
<td>ddRAD-seq</td>
<td>50–1,000 bp</td>
<td>ddRAD-seq, paired-end sequencing</td>
</tr>
<tr>
<td>Single Molecule Sequencing</td>
<td>5–5,000 bp</td>
<td>paired-end sequencing, HiSeq, Illumina</td>
</tr>
<tr>
<td>Long-Range Genomics</td>
<td>100–10,000 bp</td>
<td>paired-end sequencing, HiSeq, Illumina</td>
</tr>
</tbody>
</table>

PippinHT DNA Size Selection System
- Precise and Reproducible
- Up to 24 Samples per Run

www.sagescience.com
@SageSci
ethylene glycol were 82.7°C, 81.5°C, and 83.9°C, respectively. This shows a greater ability of the N^6-meC nucleotide to reduce the stability of the DNA duplex, compared with the additives used at the highest concentrations recommended. Thus, N^6-me-dCTP is more effective than the various additives tested, both in a functional sense, allowing PCR amplification of difficult DNA sequences, and in a physical sense, giving a larger reduction in the Tm values.

The amplicons containing N^6-meC can be used for sequencing or for hybridization experiments (20). As to their use in cloning, this will depend on the particular restriction enzymes used in the cloning scheme, as the majority of restriction enzymes are inhibited by N^6-methylation of cytosines in their recognition sequence (24,25). In any event, performing PCR with N^6-me-dCTP will be beneficial for hybridizations involving highly GC-rich regions. Regarding the amplification fidelity of N^6-meC-containing DNA, sequencing data showed the expected sequence over hundreds of nucleotides. It should be noted that this does not address potential analog-mediated mutations occurring at lower but significant rates (1:100 or 1:1000 bp), which would have to be examined using cloning experiments. However, the fact that N^6-meC occurs routinely in the DNA of certain microorganisms (26–30) argues for relatively low mutation rates associated with incorporation of this nucleotide analog, at least for the polymerases used by these organisms.

Author contributions
C.R.F.J. performed the experiments. C.R.F.J. and R.C.P. designed the experiments and wrote the initial manuscript. E.G.J. and A.A. analyzed the data and contributed to the discussion and the final text. All authors read and approved the final manuscript.

Acknowledgments
This work was supported by the Secretaría de Investigación y Posgrado of the Instituto Politécnico Nacional (Grant SIP 20131256) and the Consejo Nacional de Ciencia y Tecnología-CONACYT (Grant Ciencia Básica 61322). C.R.F.J. was the recipient of a graduate-student stipend from CONACYT.

Competing interests
The authors declare no competing interests.

References


Received 21 March 2016; accepted 06 July 2016.

Address correspondence to Reynaldo C. Pless, CIICATA, Instituto Politécnico Nacional, Querétaro, Querétaro, Mexico. E-mail: rcppless@yahoo.com

To purchase reprints of this article, contact: biotechniques@fosterprinting.com