Nucleobase analogs for degenerate hybridization devised through conformational pairing analysis

Michelle L. Abraham, Maria Albalos, Toumy Guettouche, Michel J. Friesenhahn, and Thomas R. Battersby
Siemens Medical Solutions Diagnostics, Berkeley, CA, USA

BioTechniques 43:617-624 (November 2007)
doi 10.2144/000112602

A conformational pairing analysis was used to devise nucleobase analogs capable of forming nonselective and energetically favorable base pairs opposite either the purine or the pyrimidine constituents of nucleic acids. 5-methylisocytosine and isoguanine were conceived as a degenerate pyrimidine and a degenerate purine, respectively. Data from previous DNA duplex melting experiments verified that the analogs can act as degenerate nucleobases as hypothesized. Isoguanine also formed unusually stable base pairs with guanine. A quantitative PCR assay yielding equivalent results across hepatitis C virus (HCV) subtypes was created with this system, despite the use of a single probe targeted to a polymorphic region. Amplification curves using probes with 5-methylisocytosine or isoguanine opposite appropriate ambiguous target positions exhibited more signal than curves from similar probes containing common degenerate nucleobase hypoxanthine.

INTRODUCTION

In degenerate hybridization a single nucleic acid sequence hybridizes nonselectively to target sequences containing polymorphic sites. Degenerate probes are created from otherwise specific sequences with nucleobase analogs acting as degenerate bases opposite ambiguous target sites. Interaction of an analog and intended target nucleobases should be nonselective and should not be markedly destabilizing relative to Watson-Crick interaction. Degenerate probes offer the advantage of reducing the number and total amount of probes required to target ambiguous sites. Fewer probes simplify assay optimization, reduce background signal, and potentially reduce assay costs. A degenerate site need not be opposite a primary polymorphic site interrogated by a hybridization probe; it may simply allow targeting of regions with incidental sequence variation.

A significant challenge of designing a degenerate base is to compensate for the different hydrogen bonding patterns presented by the nucleobase constituents of nucleic acids (A, C, G, and T/U). Previous strategies for nonselective pairing with two or more of the nucleobases include tautomerization (1), bond rotation to present different hydrogen bonding faces (2,3), or by dispensing with hydrogen bonding altogether (4–6). Here we employ a systematic conformational pairing analysis of nucleobase analogs to devise degenerate bases. Using this approach, 5-methylisocytosine (F) and isoguanine (J) were conceived as degenerate nucleobases that minimally destabilize nucleic acid duplexes opposite purine (R) and pyrimidine (Y) ambiguous sites, respectively.

MATERIALS AND METHODS

Oligodeoxyribonucleotides

A 27-mer forward and 27-mer reverse primer pair, targeted to sequence regions conserved across all targets examined, was used. Primers were obtained from Operon Biotechnologies (Huntsville, AL, USA) with reversed-phase high-performance liquid chromatography (HPLC) purification. Probes 1–8 were 5′-labeled with 6-fluorescein and 3′-labeled with Black Hole Quencher 1. Probes 1 and 2, 4–6, and 8 were obtained from Biosearch Technologies (Novato, CA, USA) with anion exchange and reversed-phase HPLC purification. Probes 3 and 7 containing F or J were synthesized on a 394 DNA/RNA Synthesizer (Applied Biosystems, Foster City, CA, USA). Phosphoramidites were purchased from Glen Research (Sterling, VA, USA). Solid support with Black Hole Quencher was purchased from Biosearch Technologies. The oligodeoxy nucleotides (ODNs) were purified by reversed-phase HPLC using the Wave System (Transgenomic, Omaha, NE, USA). ODN mass was verified as being within 0.5% of the expected mass by matrix-assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry using a Voyager-DE (Applied Biosystems). ODN purity was ≥93% by electrophoresis in a polyacrylamide-filled capillary on a 30CE instrument (Agilent Technologies, Santa Clara, CA, USA).

Probes against sense strand of PCR product were: probe 1, 5′-TTTGCAGA CCAACACTACTGGCT-3′; probe 2, 5′-TTTGCACCCACGCTAC TCGGCT-3′; probe 3, 5′-TTTCGCCA CCCAACJCCTACTGGCT-3′; probe 4, 5′-TTTCGCCCACCACCTAC TCGGCT-3′. Probes against antisense strand of PCR product were: probe 5, 5′-GCCGATAGTGTTGGGTTTCG GAAA-3′; probe 6, 5′-GCCGAGTA GCGTTGGGTTGGGCAA-3′; probe 7, 5′-GCCGATAGTGTTGGGTTGGGTTGGGCAA-3′; probe 8, 5′-AGCCG AGTAGTTGGGTTTGGGCAA-3′. Bold nucleobases are opposite target polymorphic sites.

Hepatitis C Virus Transcripts

Hepatitis C virus (HCV) transcripts were produced from viral samples of subtypes 1a, 1b, 2a, and 2b and quantified by phosphate analysis, as described previously (7).

Reverse Transcription PCR with HCV Transcripts

Reverse transcription PCR (RT-PCR) was conducted in 96-well microplates in 25-µL volumes. Individual
Short Technical Reports

wells contained 1 μL OneStep RT-PCR enzyme mix (Qiagen, Valencia, CA, USA), 5.0 mM MgCl₂, 0.3 mM each deoxynucleoside triphosphate, and 30 nM ROX reference dye (Stratagene, La Jolla, CA, USA) in OneStep RT-PCR buffer. Each reaction contained 400 nM each primer and 125 nM probe. Target levels of 1 × 10³, 1 × 10⁴, and 1 × 10⁵ copies of HCV transcript and no template controls were run in triplicate. Amplification was performed on a Stratagene Mx3000P with 50°C reverse transcription incubation for 40 min; 95°C Tₐq polymerase activation for 15 min; and 40 cycles of 95°C for 15 s; 67°C for 1 min; and 72°C for 30 s. MxPro software determined threshold cycles (Cₚ) employing amplification-based threshold (set at 5%), adaptive baseline, and moving average algorithms.

Transcripts of subtypes 1a and 2a were targets on one microplate with probes 1–4 and on another microplate with probes 5–8. These experiments were repeated with transcripts of subtypes 1b and 2b using probes 1–3 on one microplate and probes 5–7 on another microplate.

RESULTS AND DISCUSSION

Conformational Pairing Analysis

Conformational analysis was limited to analogs with carbon/nitrogen ring systems isosteric to natural nucleosides and bearing only the types of functionality found in the natural nucleobases. These analogs are referred to herein as purines or pyrimidines. In pairs formed from such analogs, determination of the number of hydrogen bonds (H-bonds) and potential repulsive interactions in accessible pairing conformations allows prediction, to a first approximation, of the relative stability of most pairs (8). However, relative stability of some purine-purine pairs is not straightforward because of accessible syn- and anti-purine conformations in base pairing. Further, pyrimidine-pyrimidine pairs in duplex DNA are usually significantly destabilizing versus purine-pyrimidine pairs and are correspondingly less desirable as degenerate bases. From these considerations, we

<table>
<thead>
<tr>
<th>Nucleobasea</th>
<th>Mean Tₘ Opposite T</th>
<th>Mean Tₘ Opposite C</th>
<th>Mean Tₘ Differenceb</th>
<th>Tₘ Rangec</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>54.4</td>
<td>42.1</td>
<td>12.3</td>
<td>14.1</td>
</tr>
<tr>
<td>G</td>
<td>49.4</td>
<td>59.0</td>
<td>9.6</td>
<td>11.6</td>
</tr>
<tr>
<td>I</td>
<td>49.7d</td>
<td>53.7</td>
<td>4.0</td>
<td>4.9</td>
</tr>
<tr>
<td>J</td>
<td>52.0</td>
<td>51.6</td>
<td>0.4</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Oligodeoxynucleotides 5′-CACN-ACTTTCTCT-3′ and 5′-GGAGAAAGTN2GT-3′ were used. Each strand was 1.6 μM in 0.1 M NaCl, 0.04 M sodium citrate, pH 7.9. Original data for individual duplexes appear in Reference 8. Tₘ, melting temperature.

aMean Tₘs include designated nucleobase in N₁ and N₂ positions.
bAbsolute value of the difference between the mean Tₘ opposite T and the mean Tₘ opposite C.
cHighest Tₘs minus lowest Tₘ for the designated nucleobase opposite T and C. This provides a measure of Tₘ variation due to sequence context when compared with the corresponding “Mean Tₘ Difference.”
dOnly one of the two possible Iₘs was measured.
eAbsolute value of the difference between the mean Tₘ opposite A and the mean Tₘ opposite G.
fHighest Tₘs minus lowest Tₘ for the designated nucleobase opposite A and G.
focused on purine-pyrimidine interaction to devise nucleobase analogs hybridizing nonselectively to relatively common R or Y polymorphic sites with minimal duplex destabilization.

A systematic analysis of purine-pyrimidine interaction was conducted (see Supplementary Figure S1 available online at www.BioTechniques.com). A brief summary follows. Analogs bearing H-bonding functionality able to form a maximum number of H-bonds, and no repulsive interactions (such as ionized nucleobases or amines not involved in H-bonding) in purine-pyrimidine interaction, will minimally destabilize duplex DNA (Figure 1A). Watson-Crick pairing is the only feasible way for purine-pyrimidine pairs to form three H-bonds, so no H-bonding pattern on an analog can form three H-bonds with two or more nucleobases. There are three energetically feasible pairing conformations within nucleic acid duplexes of purine-pyrimidine pairs joined by two H-bonds (8). Slightly less stable than Watson-Crick pairs joined by 3 H-bonds are Watson-Crick pairs joined by 2 H-bonds, in which at least one partner lacks a third H-bonding functional group and there are no repulsive interactions (as in the A-T pair). Finally, slightly less stable than these Watson-Crick pairs are wobble pairing (as in the G-T pair) and a conformation sometimes called reverse wobble pairing, not possible with the natural nucleobase pairs. Among these conformations, it is only possible to form pairs of comparable stability with analogs capable of forming a wobble pair with one nucleobase and a reverse wobble pair with another nucleobase. Commercially available analogs F and J bear the only patterns of H-bonding functionality meeting this description (Figure 1, B and C).

Oligodeoxynucleotide Duplex Melting

ODN melting experiments monitoring absorbance of a 13-mer duplex have been previously conducted with a large number of purine and pyrimidine analogs at a single position within an otherwise identical duplex sequence (8). Two melting temperatures (T<sub>ms</sub>) can be measured for each nucleobase pair (except for homopairs) within the duplex. Of particular relevance are data generated by pairing the standard nucleobases with J, F, or hypoxanthine (I). Purine I (called inosine as the nucleoside) has been used for many years as a universal nucleobase, intended to hybridize nonselectively to all of the standard nucleobases (9). J and F as a degenerate purine and pyrimidine can be directly compared with these data to I in identical roles.

The performance of J and F in these experiments demonstrates that these analogs function quite well as degenerate bases (Table 1). T<sub>ms</sub> of duplexes containing J and F were relatively high and less variable than T<sub>ms</sub> for other nucleobases, including I, in identical pairings. Interestingly, T<sub>ms</sub> for duplexes with J opposite G (52.2°C and 52.7°C) were very close to T<sub>ms</sub> for duplexes with J opposite the natural pyrimidines (8). This result suggests that J functions not only as a degenerate nucleobase against pyrimidines C and T, but also against purine G.

Quantitative Reverse Transcription PCR with Degenerate Probes

A potential use of degenerate probes containing J and F was illustrated in a model quantitative RT-PCR assay for HCV. This assay resembles diagnostic
assays intended for measuring viral load in clinical laboratories (10), in which equivalent quantification across subtype samples is desired. The probe binding regions of four HCV transcript targets in these experiments contained two genotype-specific polymorphic sites (Figure 2A). Genotype 1 transcripts had a common sequence motif in the probe binding region, and genotype 2 transcripts had a different motif with two isolated single base differences. A single probe was used in each reaction with one of four assay formats: (i) a probe matching the genotype 1 motif, (ii) a probe matching the genotype 2 motif, (iii) a degenerate probe containing either F or J, or (iv) a degenerate probe containing I.

Designing probes specific to the sense and antisense strands of PCR products allowed assessment in separate assays of either F or J as a degenerate nucleobase in the probe.

Predictably, a matched probe always outperformed a probe containing two standard base mismatches in both the sense and antisense probe sets. The C_T was substantially delayed for targets that had standard base mismatches with the probe. Any single probe with only standard nucleobases was not suitable for quantification across genotypes 1 and 2 (Figure 2, B and C).

In contrast, degenerate probes containing either F, J, or I produced much closer C_Ts for targets of genotype 1 or 2 (Figure 2, D and E). This is not surprising considering the same nucleobase pairs are present, albeit in different sequence contexts, in the degenerate probes interacting with targets of genotypes 1 or 2. C_Ts obtained with genotype 1 and 2 transcripts were slightly less variable with probes 3 and 7 containing F or J than for probes 4 and 8 containing I. More significantly, while all probes containing analogs caused delays in C_Ts relative to assays with matched probe and transcript, probes 4 and 8 caused larger delays than probes 3 and 7. Similarly, fluorescent signal normalized against reference dye ROX (dR_n) was always lower with probes containing analogs than with perfectly matched probes, but probes 3 and 7 always had a higher dR_n than probes 4 and 8 (Supplementary Figure S4). It may be reasonably expected that changes in C_T and dR_n in this system would be influenced by the particular mismatch nearest the 5’ end of the probe, and it is interesting to note that probes 3 and 7 performed at least slightly better than probes 4 and 8 in all cases.

Utility of F and J as Degenerate Nucleobases

Analogs F and J, commercially available components of an additional base pair, performed as a degenerate pyrimidine and a degenerate purine, respectively, demonstrating the potential of a simple conformational analysis in predicting nucleobase pairing stability. J further functioned as a degenerate base opposite B (C, T, and G) ambiguous sites. Accounting for this additional degeneracy, J has established tautomeric ambiguity and can adopt an N_7-H tautomer (11,12) opposite purine G to form an unusually stable Watson-Crick purine-purine pair. There are advantages and disadvantages to the additional degeneracy. J cannot be used to specifically target only Y positions, because the interaction of J with G will not allow discrimination of Y from G positions. However, when used in the probes in the HCV-PCR assay, the additional degeneracy of J is irrelevant. In some cases it can be advantageous because J can hybridize with a variety of polymorphisms; J will be nonselective opposite all possible polymorphisms except those containing A.

Nucleobases J and F can increase specificity of nucleic acid hybridization when introduced as a third base pair (13), yet paradoxically the individual analogs produce decreased hybridization selectivity opposite the natural nucleobases. Although the analogs form relatively stable “mismatches” with two or more standard nucleobases, a significant energy gap remains between these mismatches and the F-J pair. The F-J pair generally stabilizes a duplex even more than the canonical pairs (14,15), maintaining this gap.

Analogs F and J have advantages compared with common alternatives stemming from the conformational analysis used to devise the degenerate analogs. Because of comparable H-bond count and pairing geometries, the analogs pair less selectively than common degenerate nucleobase hypoxanthine opposite common Y and R ambiguous positions. F and J also performed better than I in our example quantitative RT-PCR assay. Additionally, because H-bonding is maximized, F and J are less destabilizing to duplexes as degenerate nucleobases than conventional analogs with good nonselectivity opposite nucleobases, such as 3-nitropyrrrole and 5-nitroindole (16). F and J can function in polymerase replication with only triphosphates of the standard nucleobases. Misincorporations of standard nucleobases opposite F or J in a template strand are tolerated if a polymerase does not frequently encounter analog positions. Thus, it may be possible to equivalently detect targets with ambiguous sites in primer binding regions in quantitative PCR assays using a primer containing F or J. However, the analogs probably cannot be used to introduce ambiguous sites in polymerase replication. Isolated J and F positions in oligonucleotide templates have been demonstrated to direct incorporation of T and A nucleotides, respectively, by thermophilic Family A polymerases commonly used in PCR amplification (17). Further work with a variety of polymerases is required to determine the scope of the utility of primers containing J and F, especially considering the often idiosyncratic behavior of individual polymerases with nucleobase analogs.

Finally, this work confirms that the regular purine-pyrimidine pairing observed among the standard nucleobase constituents of nucleic acids extends generally to a larger group of isosteric nucleobase analogs across a range of pairing conformations. Conformational pairing preferences of these isosteric analogs are not notably different from the natural nucleobases, implying that other characteristics of the nucleobases such as chemical stability, lack of tautomeric interconversion, biosynthetic availability, and suitability for enzymatic replication are also relevant to explaining the selection of nucleobases in the genetic code.
COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES


Received 2 April 2007; accepted 11 September 2007.

Address correspondence to Thomas R. Battersby, Siemens Medical Solutions Diagnostics, P.O. Box 2466, Berkeley, CA 94702, USA. e-mail: thomas.battersby@siemens.com

To purchase reprints of this article, contact: Reprints@BioTechniques.com