5902–5910 *Nucleic Acids Research, 2018, Vol. 46, No. 12* *Published online 24 May 2018*
doi: 10.1093/nar/gky392

Nucleoside analogs to manage sequence divergence in nucleic acid amplification and SNP detection

Zunyi Yang[1,2,†], Hyo-Joong Kim[2,†], Jennifer T. Le[1,2], Chris McLendon[1,2], Kevin M. Bradley[1,2], Myong-Sang Kim[2], Daniel Hutter[1,2], Shuichi Hoshika[1,2], Ozlem Yaren[1] and Steven A. Benner[1,2,*]

[1]Foundation for Applied Molecular Evolution (FfAME), 13709 Progress Boulevard, Box 7, Alachua, FL 32615, USA and [2]Firebird Biomolecular Sciences LLC, 13709 Progress Blvd, Box 17, Alachua, FL 32615, USA

Received February 22, 2018; Revised March 26, 2018; Editorial Decision March 29, 2018; Accepted May 02, 2018

ABSTRACT

Described here are the synthesis, enzymology and some applications of a purine nucleoside analog (H) designed to have two tautomeric forms, one complementary to thymidine (T), the other complementary to cytidine (C). The performance of H is compared by various metrics to performances of other 'biversal' analogs that similarly rely on tautomerism to complement both pyrimidines. These include (i) the thermodynamic stability of duplexes that pair these biversals with various standard nucleotides, (ii) the ability of the biversals to support polymerase chain reaction (PCR), (iii) the ability of primers containing biversals to equally amplify targets having polymorphisms in the primer binding site, and (iv) the ability of ligation-based assays to exploit the biversals to detect medically relevant single nucleotide polymorphisms (SNPs) in sequences flanked by medically irrelevant polymorphisms. One advantage of H over the widely used inosine 'universal base' and 'mixed sequence' probes is seen in ligation-based assays to detect SNPs. The need to detect medically relevant SNPs within ambiguous sequences is especially important when probing RNA viruses, which rapidly mutate to create drug resistance, but also suffer neutral drift, the second obstructing simple methods to detect the first. Thus, H is being developed to detect variants of viruses that are rapidly mutating.

INTRODUCTION

The genetic variability of rapidly evolving viruses, such as HIV, can eventually cause established assays to underestimate the prevalence of the virus in a population, the load of the virus in a patient, or even the very presence of the virus in a patient. Mutations in regions that bind to primers and probes lower their affinities and could eventually prevent their binding entirely (1–3). Related to our long-standing interest to develop assays to detect pathogenic viruses with genetic variability (4,5), we were interested in evaluating various strategies to overcome these problems.

One strategy simply creates multiple primers that perfectly match all conceivable variants in a population (6), either by explicit synthesis of multiple primers, or by using mixtures of phosphoramidites when adding a nucleotide at an ambiguous site (7). However, each level of added diversity dilutes the perfectly matched primer, creates opportunities for off-target amplification, and reduces assay sensitivity. Further, even with these strategies, medically relevant drug-resistance mutations embedded in a segment of a target that contains irrelevant variability can rapidly become difficult to detect (3).

An alternative strategy to manage target sequence diversity uses a 'universal nucleobase', a synthetic nucleotide in a primer that can pair with all natural nucleobases at its binding site (8,9). A large body of literature covers many clever designs for universal nucleobases (10–12). In practice, however, inosine is common as a universal base, notwithstanding some problematic aspects of its behavior (13,14). Because of these, no single analog can be considered to be truly universal (15,16).

True universality is, however, not always needed (or even desired) for biological applications. First, if too many universal nucleotides are added to a probe, the resulting molecule may have too many places to bind off-target in a complex DNA sequence background. Further, when seeking to manage a naturally divergent sequence, evolutionary analyses often find that transition mutations (replacing a purine by a purine, or a pyrimidine by a pyrimidine) are more common than transversions (replacing a purine by a

*To whom correspondence should be addressed. Tel: +1 386 418 8085; Fax: +1 386 418 8856; Email: manuscripts@ffame.org; Homepage: http://ffame.org and http://firebirdbio.com
†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.
Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

pyrimidine, or a pyrimidine by a purine) in coding and non-coding regions (17,18). In these cases, biversality (binding to either purine, or to either pyrimidine, but not to both purines and pyrimidines) may be preferred over universality, managing the more common ambiguity without creating the excessive off-target binding.

Tautomerism is one strategy to create biversality. For example, Brown and his coworkers made candidate biversals by exploiting the tautomerization of oximes (19). Their pyrimidine biversal used the oxime-alkoxyamine tautomerization of N6-methoxycytidine (20); here, the oxime tautomer is hydrogen-bond complementary to A, while the enamine tautomer is complementary to G (Figure 1). Their corresponding purine biversal exploited oxime-alkoxyamine tautomerization in a purine ring with N6-methoxydiaminopurine (21,22); the oxime tautomer is hydrogen-bond complementary to C, while the alkoxyamine tautomer is complementary to T (Figure 2).

In Brown's pyrimidine biversal, the methoxy group is 'tied back' as shown in $\mathbf{P_C}$ and $\mathbf{P_T}$ (right in Figure 1). Without the tie-back, rotation of the methoxy group can obstruct biversal **P** pairing with G and A (left in Figure 1). Because of the nitrogen at position 7, purines themselves do not offer any ready position to tie back the exocyclic N-O unit (Figure 2, left). Thus, in the classical purine biversal systems, the melting temperatures are lower, presumably because the methoxy group in a '*cis*' conformation interferes with hydrogen bonding to the Watson–Crick complement.

We approached the purine biversal problem by synthesizing a tricyclic analog based on a 7-deazapurine, whose carbon allows for a point of attachment of the ring. This should have electronic properties similar to the conformationally flexible oxime. Brown and his coworkers attempted to make a purine analog, but without reporting success (23,24).

Here, we report successful synthesis of a tricyclic purine biversal with the alkoxyl substituent 'tied back'. We then report biophysical data that compare binding of oligonucleotides containing various biversal analogs to complementary and semi-complementary oligonucleotides. We follow with results from polymerase chain reaction (PCR) amplification supported by primers containing these biversals, and ligation architectures to detect single nucleotide polymorphisms (SNPs) using probes having various numbers of these biversals.

MATERIALS AND METHODS

Oligonucleotide

Oligonucleotides were synthesized in 1.0 μmol scale using the recommended protocol for an ABI 394 DNA synthesizer, with extended coupling time (10 min) for the tricyclic purine biversal **H**. All natural (A, G, C and T) ultramild phosphoramidites were from Glen Research. Oligonucleotides were cleaved from solid support and deprotected by treatment with 0.05 M K_2CO_3 in MeOH (0.2 ml, 55°C overnight). The deprotection mixture was diluted with 1 M TEAA (0.4 ml), desalted by Sep-Pak C18 cartridge (Waters) and purified by ion exchange high-performance liquid chromatography (HPLC, Dionex DNAPac PA-100 22 × 250 mm Prep column, eluent A = 25 mM NaOH + 0.1 M NaCl, eluent B = 25 mM NaOH + 1 M NaCl, from 10 to 50% B in 30 min, flow rate 10 ml/min). The appropriate fractions were collected, neutralized by aqueous acetic acid, desalted by Sep-PaK and lyophilized. The purity of each oligonucleotide was analyzed by ion exchange HPLC (traces in Supplementary Data). DNA with only natural nucleobases was obtained from Integrated DNA Technologies (IDT), Coralville, IA, USA.

Melting temperature analysis

Oligonucleotides were combined in equimolar concentrations (1 μM) in buffer (0.1 M NaCl, 0.02 M sodium cacodylate pH 7.0). Absorbance of each solution was monitored at 260 nm over a temperature range from 20°C to 85°C with a ramp rate of 0.5°C/min. Melting curves were measured on a Shimadzu UV spectrophotometer UV-1800 with a thermal Peltier cell block and temperature probe.

Sanger sequencing of the PCR products generated with 3H- and 4H-containing primers

Wild-type (WT) DNA template was amplified by standard primers, 3H- or 4H-containing primers (1 μM of each primer) in 1× AmpliTaq Gold reaction buffer (15 mM Tris–HCl, 50 mM KCl, pH 8.3 at 25°C), $MgCl_2$ (3 mM), dNTPs (each 0.2 mM) and 2.5 units of Hot Start AmpliTaq Gold DNA polymerase (5 U/μl, ABI, total 25 μl). PCR: initially at 95°C for 10 min; followed by 35 cycles (95°C for 10 s, 50°C for 30 s, 72°C for 30 s); finally 72°C for 5 min. Upon completion, samples (10 μl) were mixed with 6× agarose loading dye (2 μl, Promega) and analyzed on 3% agarose gel.

PCR products (2 μl) were cloned (pCRTM4-TOPO®) using TOPO TA Cloning Kit (Invitrogen) and transformed into One Shot® TOP10 competent cells following manufacturer's instruction. Blue-white screens gave 32 colonies that were sequenced (BioBasic). The sequencing results are shown in Supplementary Table S1 and Figure S1.

Comparison of the amplification efficiency of 3H-, 4H-, 4K-, and 4I-containing primers using real-time PCR

WT templates (3×10^6 and 3×10^4 copies of WT-90) and divergent templates (3×10^6 and 3×10^4 copies of Div4-90) were obtained from IDT and amplified in 1× AmpliTaq Gold buffer (15 mM Tris–HCl, 50 mM KCl, pH 8.3 at 25°C), with $MgCl_2$ (3 mM), dNTPs (each 0.2 mM), Hot Start AmpliTaq Gold DNA polymerase (2.5 units, 5 U/μl, ABI), and EvaGreen® dye (0.2 × Biotium) using five types of primers (standard, 3**H**-, 4**H**-, 4**K**-, and 4**I**-containing primers, Table 2), respectively. PCRs and no template control (NTC) were performed in replicates under identical conditions: 95°C for 8 min, followed by 40 cycles (95°C for 10 s, 50°C for 30 s, 72°C for 30 s) in the Roche LightCycler® 480 real-time PCR system. LightCycler® 480 software was used to calculate threshold cycle (C_t) and T_m. Upon completion of PCR, each sample (10 μl) was

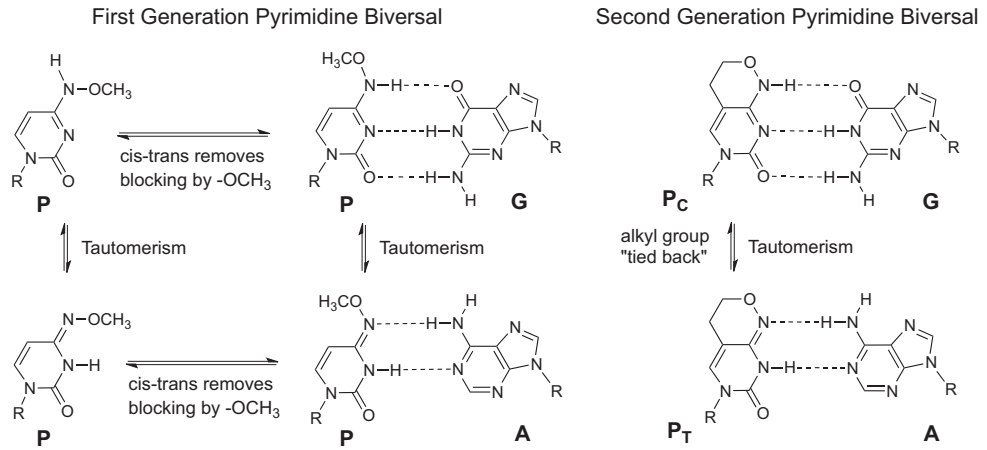


Figure 1. Two candidate pyrimidine biversal nucleoside analogs that exploit oxime-alkoxylamine tautomerism. A second generation biversal nucleoside ($\mathbf{P_C}$ and $\mathbf{P_T}$, right) showed better pairing to G/A than the first generation biversal nucleoside (**P**, left), presumably because the appended CH_3O- group in the first species can swing around to a '*cis*' conformation that may obstruct nucleobase pairing.

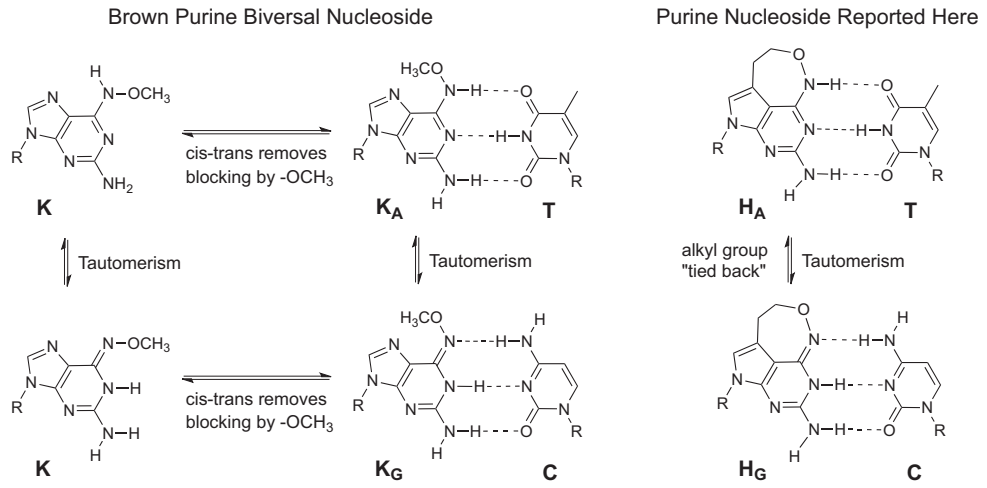


Figure 2. Two candidate purine biversal nucleosides. In the purine biversal nucleoside reported by Brown (**K**, left), the appended CH_3O- group can swing to a '*cis*' conformation that may obstruct base pairing with either T or C. In the new candidate purine biversal reported here (**H**, right), this cannot happen. However, the additional ring may influence the tautomer ratio, based on geometries of substituents attached to double bonds.

mixed with 6× agarose loading dye (2 μl, Promega) and analyzed on agarose gel (3%) electrophoresis (Figure 3; Supplementary Figures S2 and S3a).

General protocol for the ligation assay

A mixture (15 μL) of ^{32}P-labeled acceptor probe (100 nM), donor probe (5′-phosphorylated, 100 nM), and target (100 nM) in 1× ligation buffer was first heated at 90°C for 1 min in a Bio-Rad Thermal Cycler, then, cooled to 45°C with a ramp rate of 0.2 °C/second (unless stated otherwise). 5 μL of Taq DNA Ligase (40 U/assay) in 1× ligation buffer was added to the above mixture at 45 °C. Each ligation assay (20 μL) was incubated at 45°C for 5, 20, or 60 min. Aliquots (7 μl) were quenched with 7 μL of loading dye (95% of formamide with 10 mM ethylenediaminetetraacetic acid) and resolved by polyacrylamide gel electrophoresis (20% PAGE, 7 M urea). (Figures 4 and 5; Supplementary Figures S4a and S4b).

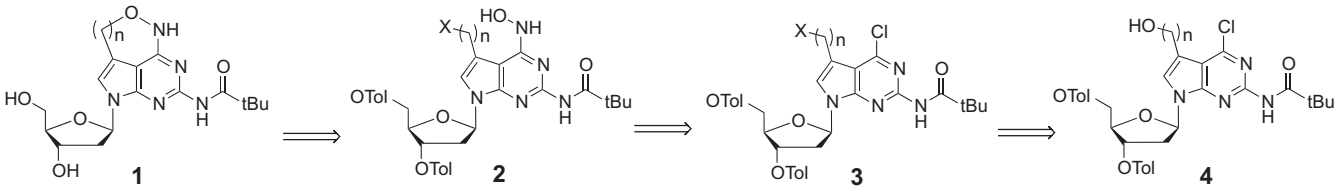
RESULTS AND DISCUSSION

Nucleoside and oligonucleotide synthesis

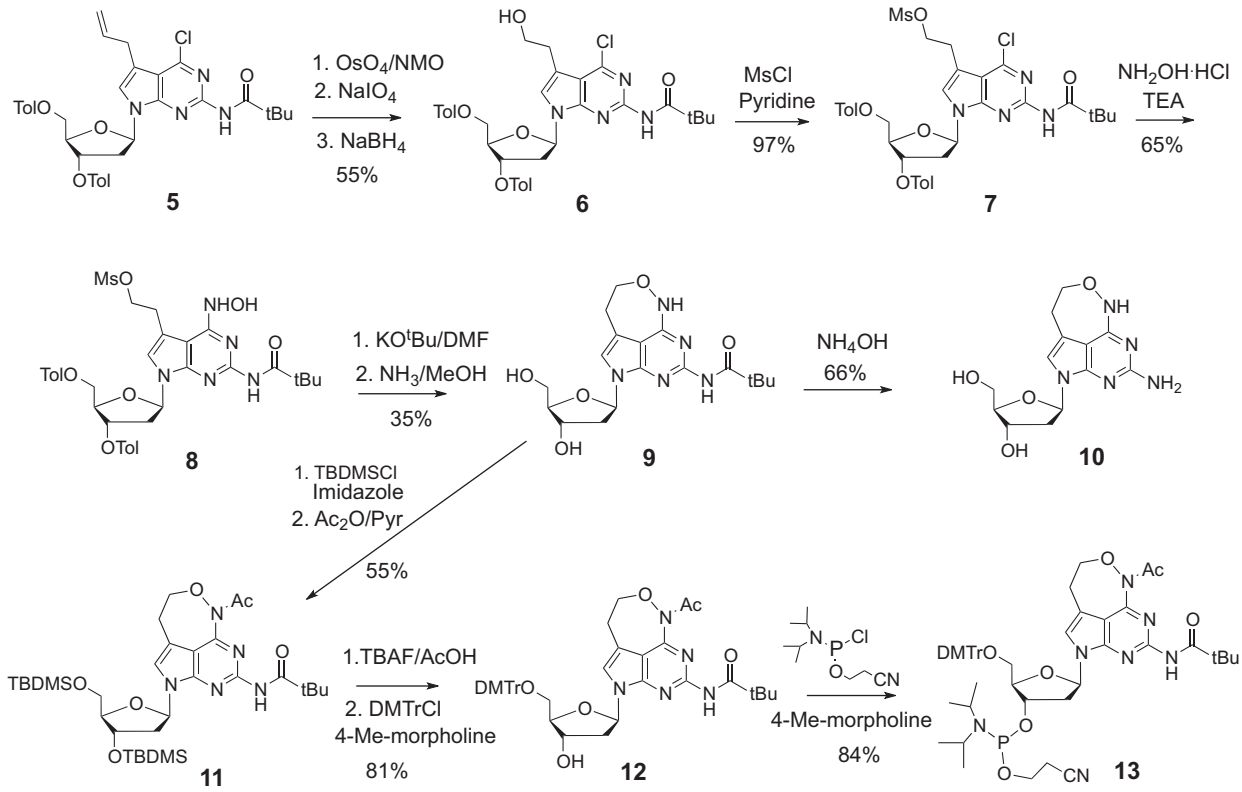
The synthesis plan sought the tricycle **1** via intramolecular closure of **2**, made by displacement of chloride by hydroxylamine. That ring could have either one ($n = 1$) or two ($n = 2$) CH_2 units. In either, the leaving group for compound **3** might arise from an alcohol in **4**, available from a vinyl unit if $n = 1$ and from an allyl unit if $n = 2$ (Scheme 1). Both were made, but only the tosylate and mesylate with $n = 2$ proved to be satisfactory.

Thus, a new synthesis began by converting **5** (Scheme 2) to an alcohol that was mesylated to give **7** (25). The chloride in **7** was displaced by hydroxylamine followed by cyclization of **8** with potassium tert-butoxide in DMF; the reaction failed when THF was used as solvent.

The product from cyclization, after being purified by chromatography, was treated with methanolic NH_3 at 25°C to give analytically pure **9**, confirmed by ^{1}H and ^{13}C nuclear



Scheme 1. Retrosynthetic strategy for biversal tricyclic purine nucleoside.



Scheme 2. Synthesis of biversal tricyclic purine nucleoside **10** and phosphoramidite **13**.

magnetic resonance (NMR) and high-resolution mass spectrometry (HRMS). Removal of the pivaloyl group in NH_3 at 55°C gave deprotected tricyclic nucleoside **10**, confirmed by NMR and HRMS. The UV of **10** shows two maxima (270, 301 nm), assigned to the imine and alkoxylamine tautomers.

Protected phosphoramidite **13** made from **9** (Scheme 2) was used for solid phase DNA synthesis. Deprotection in concentrated NH4OH at 55°C overnight gave side-reactions. Successful deprotection was achieved by K_2CO_3 (0.05 M) in MeOH (55°C overnight) to give oligonucleotides with 2–4 tricyclic purine nucleosides (**H**) easily resolved by HPLC (Tables 1–4).

Thermal stability

For hybridization, complementary oligos with all possible nucleotides at the biversal pairing site were synthesized. For comparison, oligos containing first-generation purine biversal nucleobase (**K**, Glen Research), **H**, **I**, and natural A and G in the same position were synthesized (Table 1).

The first set of denaturation experiments (Table 1) reproduce Brown's results with his first generation purine biversal **K**. Consistent with his reports, **K** showed biversality and specificity. The T_m values of duplexes containing two or three **K**:T pairs were only slightly higher than those with **K**:C pairs. The values of duplexes with purine:purine **K**:G and **K**:A mismatches were lower; the ΔT_m was ∼2°C per substitution. T_m values of **K**:C and **K**:T were much lower than the values of duplexes containing the natural pairs (G:C and A:T), also consistent with Brown's reports.

H was biophysically better by some metrics. Thus, the T_m of duplexes containing **H**:T pairs were almost the same as those of duplexes containing A:T pairs, which were higher than the same duplexes with **K**:T pairs by ∼2.3°C per substitution. The **H**:C pair had T_ms similar to those with **K**:C pairs. However, context variability was large. In duplexes with two central X:Y pairs, the T_m with **H**:C pairs was 3°C higher than with **K**:C pairs, while in duplex with two end X:Y pairs, the T_m of **H**:C pairs was 2°C lower than the T_m

Table 1. Melting temperatures of duplexes modified with purine analogs

5′-GAG TCT CGA CA**X** AG**X** TCC CAG AGG-3′ 3′-CTC AGA GCT GT**Y** TC**Y** AGG GTC TCC-5′							
X:Y	T_m	X:Y	T_m	X:Y	T_m	X:Y	T_m
H:C	63	**K**:C	60	**I**:C	67	G:C	72
H:T	68	**K**:T	62	**I**:T	61	G:T	60
H:G	63	**K**:G	57	**I**:G	61	A:C	55
H:A	59	**K**:A	58	**I**:A	65	A:T	67
5′-GA**X** TCT CGA CAG AGA TCC CA**X** AGG-3′ 3′-CT**Y** AGA GCT GTC TCT AGG GT**Y** TCC-5′							
H:C	58	**K**:C	60	**I**:C	65	G:C	69
H:T	65	**K**:T	61	**I**:T	58	G:T	57
H:G	60	**K**:G	57	**I**:G	59	A:C	56
H:A	59	**K**:A	57	**I**:A	61	A:T	64
5′-GAG TCT CG**X** CAG AG**X** TCC CA**X** AGG-3′ 3′-CTC AGA GC**Y** GTC TC**Y** AGG GT**Y** TCC-5′							
H:C	56	**K**:C	57	**I**:C	66	G:C	71
H:T	65	**K**:T	59	**I**:T	57	G:T	55
H:G	58	**K**:G	48	**I**:G	57	A:C	51
H:A	54	**K**:A	53	**I**:A	64	A:T	67

A total of 1 μM of equimolar oligonucleotides were combined in buffer containing 100 mM of NaCl, 20 mM of sodium cacodylate (pH 7). The heating and cooling cycle was performed between 20°C to 85°C with 0.5°C/min gradient.

Table 2. Primers and targets used in PCR

Name	Sequence	
Std Primer	5`-TGGGA**A**GTTCA**A**TAAGG**A**ATACC**A**CATC-3`	3`-G**A**CCT**A**CATCC**A**CTACGTAT**A**AAAAGT-5`
3H Primer	5`-TGGGAAGTTCA**H**TAAGG**H**ATACC**H**CATC-3`	3`-G**H**CCT**H**CATCC**H**CTACGTAT**H**AAAAGT-5`
4H Primer	5`-TGGGA**H**GTTCA**H**TAAGG**H**ATACC**H**CATC-3`	3`-G**H**CCT**H**CATCC**H**CTACGTAT**H**AAAAGT-5`
4K Primer	5`-TGGGA**K**GTTCA**K**TAAGG**K**ATACC**K**CATC-3`	3`-G**K**CCT**K**CATCC**K**CTACGTAT**K**AAAAGT-5`
4I Primer	5`-TGGGA**I**GTTCA**I**TAAGG**I**ATACC**I**CATC-3`	3`-G**I**CCT**I**CATCC**I**CTACGTAT**I**AAAAGT-5`
WT-90 Target	5`-TGGGA**a**GTTCA**a**TAAGG**a**ATACC**a**CATC ---------- C**t**GGA**t**GTAGG**t**GATGCATA**t**TTTTCA-3`	
Div-90 Target	5`-TGGGA**g**GTTCA**g**TAAGG**g**ATACC**g**CATC ---------- C**c**GGA**c**GTAGG**c**GATGCATA**c**TTTTCA-3`	

Dashed line means the middle of the amplicon, not written out.

Table 3. Ligation probes with biversal purine

Name	Donor Probe (3`-5`)	Acceptor Probe (3`-5`)
Standard Probe	CTTTT**A**ATC**A**TCT**A**AAGTCT	CTTGAATT**A**TTCTCTTG**A**GTTCTGAAGACCCT
K-Probe	CTTTTAATCATCT**K**AA**K**TCT	CTT**K**AATT**K**TTCTCTT**K**AGTTCTGAAGACCCT
	CTTTT**K**ATC**K**TCT**K**AA**K**TCT	
H-Probe	CTTTTAATCATCT**H**AA**H**TCT	CTT**H**AATT**H**TTCTCTT**H**AGTTCTGAAGACCCT
	CTTTT**H**ATC**H**TCT**H**AA**H**TCT	
I-Probe	CTTTT**I**ATC**I**TCT**I**AA**I**TCT	CTT**I**AATT**I**TTCTCTT**I**AGTTCTGAAGACCCT
Wt-g Target	5`-GAAAA**T**TAG**T**AGA**T**TT**C**AGA	g AA**C**TTAA**T**AAGAGAA**C**TCAAGACTTCTGGGA-3`
Div2 Target	5`-GAAAA**T**TAG**T**AGA**C**TT**T**AGA	g AA**T**TTAA**C**AAGAGAA**C**TCAAGACTTCTGGGA-3`
Div3 Target	5`-GAAAA**T**TAG**C**AGA**C**TT**T**AGA	g AA**T**TTAA**C**AAGAGAA**T**TCAAGACTTCTGGGA-3`

of **K**:C pairs. With three substitutions pairing with C, no preference of **H** over **K** was seen (Table 1).

H displayed less biversality in these studies. The difference in T_ms of **K** paired with T versus C was only ∼0.7°C per substitution, T being the preferred partner for **K**. In contrast, the preference of **H** for T over C was ∼3°C per substitution. This implies that **H** resembles adenine more than guanine. This parallels its enzymatic properties (see below); **H** is more 'enamine' than 'oxime'.

Table 4. Ligation probes with biversal purine for SNP detection

Name	Donor Probe (3`-5`)	Acceptor Probe (3`-5`)
Standard	TCTTTTTTA**G**TCATT**G**TC**A**T	G**A**CCT**A**CATCC**A**CTACGTAT**A**AAAAGT
K- Probe	TCTTTTTTA**K**TCATT**K**TC**K**T	G**K**CCT**K**CATCC**K**CTACGTAT**K**AAAAGT
H- Probe	TCTTTTTTA**H**TCATT**H**TC**H**T	G**H**CCT**H**CATCC**H**CTACGTAT**H**AAAAGT
Wt-C Target	5-AGAAAAAAT**C**AGTAA**C**AG**T**A	C**T**GGA**T**GTAGG**T**GATGCATA**T**TTTTCA-3`
Mut-T Target	5-AGAAAAAAT**C**AGTAA**C**AG**T**A	T**T**GGA**T**GTAGG**T**GATGCATA**T**TTTTCA-3`
Div-C Target	5-AGAAAAAAT**T**AGTAA**C**AG**C**A	C**C**GGA**T**GTAGG**C**GATGCATA**T**TTTTCA-3`
Div-T Target	5-AGAAAAAAT**T**AGTAA**C**AG**C**A	T**C**GGA**T**GTAGG**C**GATGCATA**T**TTTTCA-3`

This can be rationalized. Double bonds force attached atoms to lie in a plane. In the enamine tautomer complementary to T, the moving double bond lies in a 6-ring ($\mathbf{H_A}$, Figure 2, right), which is naturally planar. However, in the oxime tautomer complementary to C ($\mathbf{H_G}$, Figure 2, right), this double bond is part of a 7-ring where its substituents (including oxygen) strain to be planar (Figure 2). These geometric factors may account for the preference of **H** for T, compared with a smaller preference of **K**.

However, **H** in some T_m studies behaved more like a 'universal' base than a 'biversal' base. Thus, T_m values of duplexes with two or three **H**:G mismatches were similar to those with two or three **H**:C matches. The T_m of the **H**:A mismatch was lower than the T_m of **H**:C by only ∼0.7°C per pair. In summary, **H**:T (by 2.4°C per pair) > **H**:G (by 0.6°C per pair) > **H**:C (by 0.7°C per pair) > **H**:A. Further, duplexes with two or three **H**:G mismatches were more stable than **K**:G mismatches (∼2.7°C per pair); and the **H**:A pair was more stable than **K**:A (∼0.6°C per pair).

The thermal stability of duplexes where **H** is paired with all natural bases was further compared to the stability of duplexes having inosine (**I**) at the corresponding sites. Overall, **I**:C matches have the same stability as **H**:T matches, duplexes containing two or three **I**:A pairs were more stable than **H**:A mismatches (∼2.6°C per pair), and the **I**:G pair was less stable than **H**:G (∼0.6°C per pair). In summary **I**:C (by 1.1°C per pair) > **I**:A (by 2°C per pair) > **I**:T ≈ **I**:G.

Biversal H supports PCR

We then asked whether DNA containing **H** could prime on DNA/RNA templates. Results with **K** were consistent with results from Brown (26). Primers targeted the consensus segments of the reverse transcriptase gene of HIV-1B. Biversal **H**'s or **K**'s replaced three or four As in primers at sites known to have high divergence. The WT sequence was taken from the consensus. The divergent target had transition mutations of A to G and T to C in forward and reverse primer binding regions (Table 2).

Amplicons from PCR were resolved by agarose gel electrophoresis (Supplementary Figure S1). Both 3**H**- and 4**H**-primers gave clean, full-length amplicons at rates similar to rates of amplicon generation with perfectly matched standard primers. This showed that **H**-containing primers effi-

ciently hybridize with the WT targets, consistent with the biophysical studies. Further, these results showed that Taq DNA polymerase can use biversal **H**-containing DNA as a template, incorporating a standard dNTP opposite **H** to support the PCR (Supplementary Figure S1).

PCR products from 3**H**- and 4**H**-primers were cloned and sequenced (Supplementary Table S1). Results showed that biversal **H** directed incorporation of dTTP 95% of the time, and dCTP 5% of the time. In contrast, Brown's biversal **K** directed incorporation of dTTP 87% the time and dCTP 13% of the time (26). The statistics are not large enough to draw conclusions about the influence of context on selectivity.

Management of sequence divergence in PCR amplification

To evaluate the ability of biversal-containing primers to amplify targets with mutations, a target that had diverged at four sites, replacing A's by G's in one and T's by C's in another, was probed in two concentrations ('Hi' = 3×10^6 copies; 'Lo' = 3×10^4 copies); a consensus 'WT' target (3×10^6 or 3×10^4 copies of WT-90) was added to a PCR reaction mixture. Standard primers with four A:C mismatches were used to illustrate the value of the biversal primers (Table 2). Amplicons were resolved by electrophoresis (Figure 3A).

For standard primers, signal threshold was passed after 6.7 cycles (Figure 3C); the target with four A:C mismatches required 12.1 cycles at high target concentrations (Div4-Hi, Figure 3C). With target at lower concentrations, the mismatched primer was also less efficient. This showed clearly the value of biversality.

With both **H** and **K** at high and low target concentrations, amplification rates were essentially identical (Figure 3B and C). Paralleling the higher affinity of **K** for T over C, the C_t values were 6.3 and 6.8 at high target concentrations, and 14.6 and 15.0 at low target concentrations (4K, Figure 3C). The ΔC_t here is 0.4–0.5. Interestingly, the greater preference of **H** for T versus C was not manifested in the amplification efficiencies. The C_t values were 6.6 and 7.2 at high target with three **H**'s, and 6.4 and 7.1 with four **H**'s. The corresponding numbers at low target with three **H**'s were 14.8 and 15.4, versus 14.7 and 15.4 with four **H**'s (3H and 4H, Figure 3C).

Also interesting, the fluorescence of EvaGreen bound to products from **H**- and **K**-containing primers is 30–50% of that from standard primers (Figure 3B). However, the PCR amplicons from each assay showed almost the same intensity by agarose gel/ethidium bromide analysis (Figure 3A). This suggests that EvaGreen intercalates less well in a duplex containing biversals, is intrinsically less fluorescent upon binding, or has its fluorescence quenched.

Amplification curves from **K**-containing primers showed abnormal shapes (Figure 3B). This might be due to incorporation of both dTTP and dCTP opposite **K**, producing amplicon mixtures with **K** variously paired with T or C. In contrast, since **H** directs incorporation of primarily dTTP, later amplifications with the **H**-containing primer may be better. This is consistent with amplification curves, more sigmoidal with **H** than **K** (Figure 3B).

Primers containing **H** were next compared with primers containing inosine (**I**) (Figure 3D), relative to standard primers. All PCR conditions were the same (Figure 3B and C), except that primer concentrations were reduced to 0.5 μM and polymerase concentrations were reduced to 0.05 unit/μl. Further, uracil-DNA glycosylase (UDG, NEB) and a mixture of dUTP/dTTP were added into PCR to prevent carry-over contamination.

Amplification of all targets was slower (Figure 3D), reflecting lower concentrations of primers and polymerase. For **H**-primers, C_t values were 7.8 and 10.3 cycles (WT-Hi and Div4-Hi, Figure 3D). C_t values were worse (16.8 and 19.0 cycles) for WT and divergent targets (WT-Lo and Div4-Lo, Figure 3D). ΔC_t values were 2.5 and 2.2 for WT and divergent targets at high and low concentrations (4H, Figure 3D).

For **I**-primers, C_t values for the divergent target (**I** pairs C) were 9.1 and 14.6 at high and low target concentrations (Figure 3D and Supplementary Figure S3a). However, C_t values for WT target (**I** pairs T) were 12.2 and 22.9 cycles, with ΔC_t values of 3.1 and 8.3 at high and low concentrations (4**I**, Figure 3D). This corresponds to the T_m preference of **I** with C. For standard primers, ΔC_t between matched and mismatched targets were 11.2 and 7.3 cycles. Overall, the **H**-containing primers had smaller ΔC_t differences between WT and divergent targets than standard and **I**-containing primers (Figure 3D), showing an advantage of **H**.

To test the combination of biversal **H** and **P** in primers to manage sequence divergence, standard and **H** and **P**-primers were tested to amplify WT (HIV-1C consensus) and divergent targets (four or six mutations in primer binding segments). As expected, **H** and **P**-containing primers can amplify all targets and provide more uniform amplification than the standard primers (Supplementary Figure S3b).

In addition, T_ms of amplicons from 1 μM of Std-, 3**H**-, 4**H**- and 4**K**-containing primers were 79.2, 79.4, 79.6 and 75.0°C. T_ms of the amplicons from 0.5 μM of Std-, 4**H**- and 4**I**-containing primers and dUTP/dTTP mixture were 78.4, 78.6 and 77.3°C. The T_ms of amplicons from standard primers and **H**-primers were almost identical, and higher than amplicons from **I**- and **K**-containing primers. T_ms of these purine analog-containing amplicons also agreed with the melting temperature studies in Table 1.

Management of sequence divergence in SNP analysis

We then asked if the biversal-containing probes worked in ligation assays to detect medically significant SNPs with mutations nearby. This is especially a problem when seeking drug resistance in rapidly evolving RNA viruses. For example, the SNP in HIV reverse transcriptase that confers resistance to common drugs lies between sites that rapidly diverge in the coding region. These uninteresting SNPs confound efforts to detect the relevant SNPs in standard architectures that use standard nucleobases.

Ligation probes were designed with different numbers of **H** and **K** biversals (Table 3). WT target (WT-g) was chosen from the consensus sequence of the RT gene of HIV-1B; the divergent targets (Div2 and Div3) have either two or three mutations near the ligation site. Here, the medically irrelevant SNPs replaced C by T.

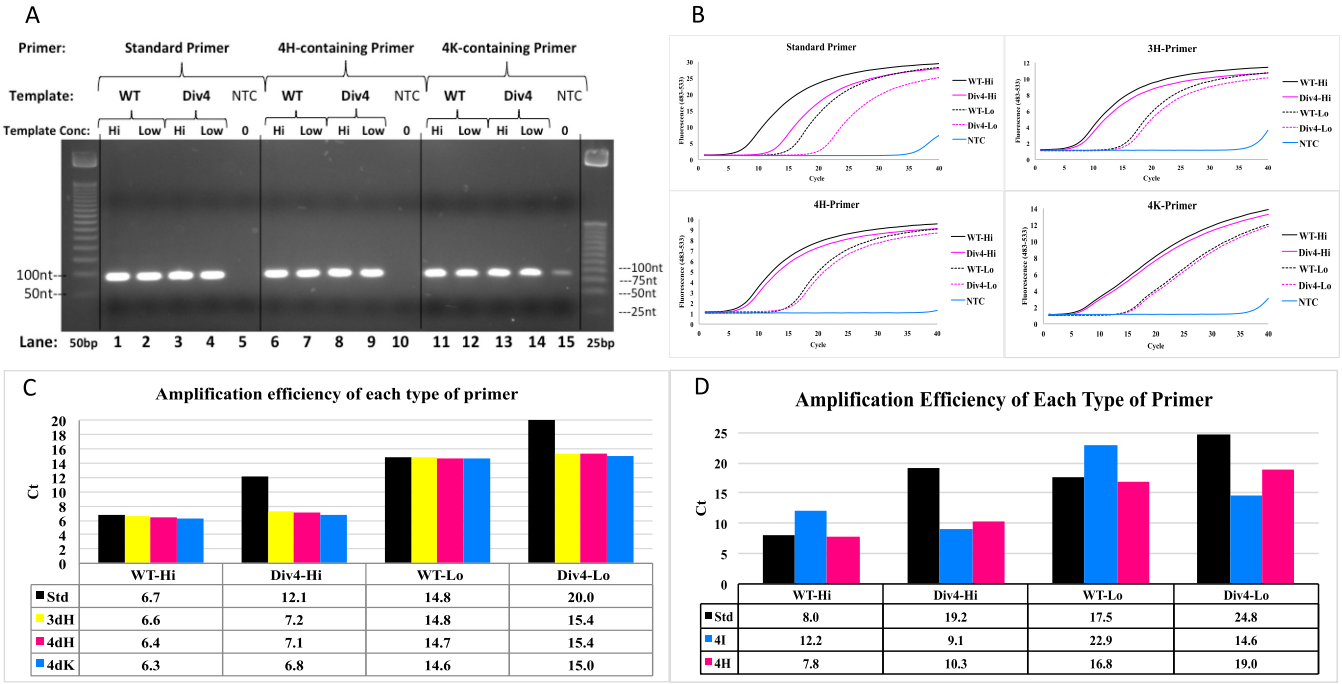


Figure 3. (**A**) PCR assays using standard primer or biversal 4H- and 4K-containing primers. Lanes 1, 6 and 11: 3×10^6 copies of WT target. Lanes 2, 7 and 12: 3×10^4 copies of WT target. Lanes 3, 8 and 13: 3×10^6 copies of divergent target (Div4). Lanes 4, 9 and 14: 3×10^4 copies of divergent target (Div4). Lanes 5, 10 and 15: no template control (NTC). The 50 bp (left) and 25 bp (right) DNA ladder. Staining is with ethidium bromide. (**B**) Real-time PCR curves show the amplification of standard primer, 3H-, 4H- and 4K-primer targeting on WT or divergent (Div4) consensus targets. Amplification efficiency was monitored based on the fluorescence of EvaGreen® dye at two target concentrations, 3×10^6 copies (Hi) or 3×10^4 copies (Lo) per assay. NTC indicates no template control. All curves represent three replicates. (**C**) Performance of purine biversal (3H, 4H and 4K) primers by C_t. Black: amplification of standard primer (Std) on WT and divergent (Div4) targets at high (Hi) or low (Lo) concentrations. Yellow: amplification of 3H-primer on WT and Div4 targets at both target concentrations. Magenta and Cyan bars indicate performances of 4H- and 4K-primers (all primers were 1 μM). (**D**) Performance evaluation of purine analog (4I and 4H) primers by C_t. Black bars: Standard primer (Std) on WT and divergent (Div4) targets at higher (Hi) or lower (Lo) target concentration. Cyan: The amplification of 4I-primer on WT and Div4 targets at Hi or Lo target concentration. Magenta bars: The performance of 4H-primers. All primers were 0.5 μM.

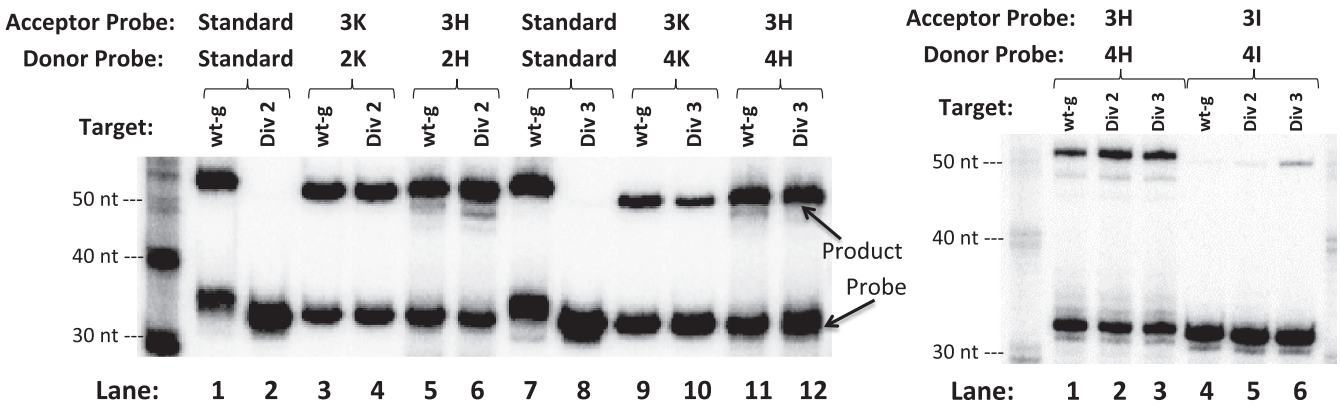


Figure 4. Ligation of standard probes and purine analog (K-, H- and I-containing) probes on wild-type (WT-g) or divergent (Div2 and Div3) targets. Acceptor probes were 5′-^{32}P labeled, ligation assays were resolved on PAGE gel (20% with 7 M urea). Lanes 1, 2, 7 and 8 (left gel): Standard probes were ligated on wild-type (WT-g) or divergent targets (Div2 and Div3). Lanes 3, 4, 9 and 10 (left gel): Biversal K-probes were used. Lanes 5, 6, 11, 12 (left gel) and 1, 2 and 3 (right gel): Biversal H-probes were used. Lanes 4, 5 and 6 (right gel): I-containing probes were used.

Three **H** or **K** biversals were put in the acceptor probes, and two or four biversals were put in the donor probes, to pair with sites holding mutations in the targets. The efficiency of ligations was compared with standard probe ligations on WT target (Figure 4). As seen in Figure 4 (left, lanes 1 and 7), standard probes gave ∼50% ligation product with the perfectly matched WT target (WT-g). In contrast, the same probes produced no ligation product at all with divergent targets (Div2 and Div3) that had two or three nearby medically irrelevant mismatches (Figure 4, left, lanes 2 and 8).

Here, biversality implemented via tautomerism has a large and potentially useful impact. For 2**K**-donor probes containing two **K**'s to match medically irrelevant SNPs,

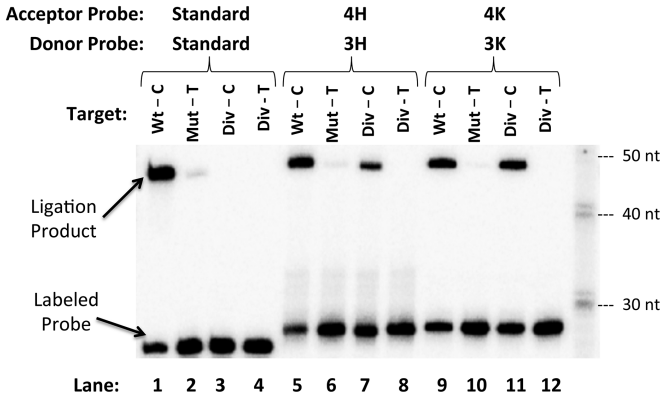


Figure 5. SNP detection by ligation of standard and biversal H- and K-probes on wild-type (WT-C) or mutant (Mut-T, Div-C, Div-T) targets. Acceptor probes 5′-^{32}P labeled; products resolved on PAGE gel (20%, 7 M urea). Lanes 1, 2, 3 and 4: Ligation using standard probes on WT or mutation targets. Lanes 5, 6, 7 and 8: with **H**-probes. Lanes 9, 10, 11 and 12: with **K**-probes. A 10-bp DNA ladder (right).

~50% ligation was seen for both targets WT-g and Div2. With the 4**K**-donor probe, ligation products decreased to ~30% for WT target and further for the divergent target (Figure 4, lanes 9 and 10). The decreased ligation may be caused by the lower T_m of **K**:C and **K**:T pairs relative to the A:T pair; alternatively, the freely rotating OMe group of biversal **K** may disturb the double helix structure of the ligation probe and target, a disturbance recognized by the ligase enzyme. For **H**-containing probes, results were markedly better. Ligation products were produced for all targets at 50% levels. Further, the medically irrelevant mutations in the divergent targets had no significant impact on ligation efficiency for the 4**H**-containing probes (Figure 4, left, lanes 5, 6, 11 and 12).

Next, **H** in the ligation probes was replaced with inosine (**I**). Inosine-containing probes failed to ligate on the wild-type (WT-g) and divergent (Div2) targets (Figure 4, right, lanes 4 and 5), and made negligible product on Div3 (Figure 4, right, lane 6). In contrast, **H**-probes successfully produced products for all targets (Figure 4, right, lanes 1–3). This shows another advantage of **H** over **I**.

To detect SNPs in variable targets in a ligation assay, standard probes can detect only perfectly matched targets (Table 4, Wt-C, Figure 5, lane 1). As expected, standard probes generated no ligation products with targets having single C to T mutations directly at the ligation sites (Mut-T, Div-T, Figure 5, lanes 2 and 4). However, they also failed with a matching at the exact ligation site if that site had flanking mismatches (Div-C) (Figure 5, lane 3).

This failure was not seen with probes containing **H** and **K**. These not only ligate when held by perfectly matched targets (WT-C, Figure 5, lanes 5 and 9), but also with mutations flanking that site (Div-C, Figure 5, lanes 7 and 11). **H**- and **K**-containing probes also differentiate a medically relevant C to T change at the ligation site, regardless of the presence of nearby medically irrelevant SNPs (Mut-T, Div-T, Figure 5, lanes 6, 8, 10 and 12).

We also asked whether the second-generation biversal pyrimidine (**Figure 1**) could allow medically relevant SNPs to be detected using ligation probes with **P** at the positions where A←→G transition mutations were found (Supplementary Table S2). **P**-containing probes generated equal amount of product for all targets (a-WT, Div5 and Div6) (Supplementary Figure S4a). This shows, as with **K** and **H**, that **P**-containing probes accommodate one or two mutations in regions flanking an interesting site. Standard probes failed to give ligation product if two mismatches (T:G) were present in the flanking region with two A to G mutations (Supplementary Figure S4a, lane 3). Likewise, probes where **I** replaces **P** also failed to give products when two transition mutations (A to G) were in the flanking region and caused two **I**:G mismatches (data not shown). For the SNP detection, biversal **P**-containing probes performed nearly identically as the standard probes to differentiate medically relevant SNP changes (Supplementary Figure S4b).

CONCLUSION

The 'concept' behind Brown's biversal pyrimidine nucleobase is based on the addition of an oxygen atom to the exocyclic amino group of base to shift the imine-enamine tautomeric equilibrium (Figures 1 and 2) (27). Thermal denaturation experiment of DNA with Brown's pyrimidine biversal nucleosides showed that the 'tied-back' bicyclic version was much better than N6-methoxycytidine, perhaps for steric reasons. Brown, however, was unable to make a tied-back version of the purine biversal.

The successful synthesis of the 'tied back' tricyclic purine oxime reported here allows, for the first time, extensive evaluation of the 'oxime tautomerism' strategy to create biversality. Overall, the strategy is successful. **H**-containing primers perform as well as perfectly matched standard primers in PCR. They work better than standard primers that are mismatched against mutated targets (by approximately five cycles with multiple mismatches). This **H** allows more uniform amplification of perfectly matched targets and mutated targets in an evolved virus.

However, by thermodynamic measurements, the added ring in **H** appears to perturb its tautomeric ratio in favor of the **H**:T match over the **H**:C match. This perturbation is reflected in enzymatic pairing with template **H**, making it behave more like A. This may be useful in PCR amplification, as the preference for **H**:T matches allows ambiguity in targets to be 'resolved' in amplicons in favor of T.

Primers containing **H** performed competitively with the universal base inosine (**I**) in many contexts. However, **H** performed better in ligation assays. We expect this to be its principal application over inosine. **H** may also be adapted to Taqman-like assays (28), where small numbers of mismatches have large impacts on the outcome.

Single oligonucleotide probes that match ambiguous sites with 'universal bases' or 'biversal bases' could also replace mixed primers. They might also replace consensus primers that overlook the divergence in the initial PCR priming step, as in PANDAA (Pan Degenerate Amplification and Adaptation) to detect HIV drug-resistance (29). Biversal base-containing primers and probes can also be used in isothermal amplification and other genotyping assays to detect medically relevant SNPs at sites flanked by irrelevant variation.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors thank the National Institutes of Health for funding support.

FUNDING

National Institute of Allergy and Infectious Diseases of the National Institutes of Health [R21AI133567]; National Institute of General Medical Sciences of the National Institutes of Health [R41GM115130, R42GM115130]. Funding for open access charge: National Institute of Allergy and Infectious Diseases of the National Institutes of Health [R21AI133567].

Conflict of interest statement. The authors declare the following competing financial interests. S.A.B. owns intellectual property in this area. Patents have been submitted for certain of the technologies reported here.

REFERENCES

1. Korn,K., Weissbrich,B., Henke-Gendo,C., Heim,A., Jauer,C.M., Taylor,N. and Eberle,J. (2009) Single-point mutations causing more than 100-fold underestimation of human immunodeficiency virus type 1 (HIV-1) load with the Cobas TaqMan HIV-1 real-time PCR assay. *J. Clin. Microbiol.*, **47**, 1238–1240.
2. Drexler,J.F., de Souza Luna,L.K., Pedroso,C., Pedral-Sampaio,D.B., Queiroz,A.T., Brites,C., Netto,E.M. and Drosten,C. (2007) Rates of and reasons for failure of commercial human immunodeficiency virus type 1 viral load assays in Brazil. *J. Clin. Microbiol.*, **45**, 2061–2063.
3. Clutter,D. S., Rojas Sanchez,P., Rhee,S. Y. and Shafer,R. W. (2016) Genetic variability of HIV-1 for drug resistance assay development. *Viruses*, **8**, 48.
4. Glushakova,L.G., Bradley,A., Bradley,K.M., Alto,B.W., Hoshika,S., Hutter,D., Sharma,N., Yang,Z., Kim,M.-J. and Benner,S.A. (2015) High-throughput multiplexed xMAP Luminex array panel for detection of twenty two medically important mosquito-borne arboviruses based on innovations in synthetic biology. *J. Virol. Methods*, **214**, 60–74.
5. Yaren,O., Alto,B.W., Gangodkar,P.V., Ranade,S.R., Patil,K.N., Bradley,K.M., Yang,Z.Y., Phadke,N. and Benner,S.A. (2017) Point of sampling detection of Zika virus within a multiplexed kit capable of detecting dengue and chikungunya. *BMC Infect. Dis.*, **17**, 293.
6. De Bel,A., Marissens,D., Debaisieux,L., Liesnard,C., Van den Wijngaert,S., Lauwers,S. and Pierard,D. (2010) Correction of Underquantification of human immunodeficiency virus Type 1 Load with the Second Version of the Roche COBAS AmpliPrep/COBAS TaqMan Assay. *J. Clin. Microbiol*, **48**, 1337–1342.
7. Young,T.P., Napolitano,L.A., Paquet,A.C., Parkin,N.T., Fransen,S., Trinh,R., Haddad,M., Hackett,J. Jr and Cloherty,G.A. (2013) Minimal sequence variability of the region of HIV-1 integrase targeted by the Abbott RealTime HIV-1 viral load assay in clinical specimens with reduced susceptibility to raltegravir. *J. Virol. Methods*, **193**, 693–696.
8. Liang,F., Liu,Y.-Z. and Zhang,P. (2013) Universal base analogues and their applications in DNA sequencing technology. *RSC Adv.*, **3**, 14910–14928.
9. Loakes,D. (2001) The applications of universal DNA base analogues. *Nucleic Acids Res.*, **29**, 2437–2447.
10. Ingale,S.A., Leonard,P., Yang,H.Z. and Seela,F. (2014) 5-Nitroindole oligonucleotides with alkynyl side chains: universal base pairing, triple bond hydration and properties of pyrene 'click' adducts. *Org. Biomol. Chem.*, **12**, 8519–8532.
11. Petrie,K.L. and Joyce,G.F. (2010) Deep sequencing analysis of mutations resulting from the incorporation of dNTP analogs. *Nucleic Acids Res.*, **38**, 8095–8104.
12. Day,J.P., Bergstrom,D., Hammer,R.P. and Barany,F. (1999) Nucleotide analogs facilitate base conversion with 3′ mismatch primers. *Nucleic Acids Res.*, **27**, 1810–1818.
13. Zheng,L.D., Gibbs,M.J. and Rodoni,B.C. (2008) Quantitative PCR measurements of the effects of introducing inosines into primers provides guidelines for improved degenerate primer design. *J. Virol. Methods*, **153**, 97–103.
14. Li,Y.Z., Zhao,P., Zhang,M.D., Zhao,X.Y. and Li,X.Y. (2013) Multistep DNA-Templated synthesis using a universal template. *J. Am. Chem. Soc.*, **135**, 17727–17730.
15. Watkins,N.E. and SantaLucia,J. (2005) Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.*, **33**, 6258–6267.
16. Vorobiev,A.V., Scarr,N.K., Belousov,Y. and Lukhtanov,E.A. (2013) 7-Aminobutynyl-8-aza-7-deazahypoxanthine as a quasi-universal nucleobase. *Nucleosides Nucleotides Nucleic Acids*, **32**, 421–438.
17. Stoltzfus,A. and Norris,R.W. (2016) On the causes of evolutionary transition:transversion bias. *Mol. Biol. Evol.*, **33**, 595–602.
18. Collins,D. W. and Jukes,T. H. (1994) Coding sequences since the human-rodent divergence. *Genomics*, **20**, 386–396.
19. LaPointe,S.M., Wheaton,C.A. and Wetmore,S.D. (2004) The degenerate properties of modified purine and pyrimidine DNA bases: a density functional study. *Chem. Phys. Lett.*, **400**, 487–493.
20. Lin,P.K.T. and Brown,D.M. (1989) Synthesis and duplex stability of oligonucleotides containing cytosine-thymine analogs. *Nucleic Acids Res.*, **17**, 10373–10383.
21. Brown,D.M. and Lin,P.K.T. (1991) Synthesis and duplex stability of oligonucleotides containing adenine guanine analogs. *Carbohydr. Res.*, **216**, 129–139.
22. Kamiya,H., Muratakamiya,N., Lin,P.K.T., Brown,D.M. and Ohtsuka,E. (1994) Nucleotide incorporation opposite degenerate bases by taq DNA-polymerase. *Nucleosides Nucleotides*, **13**, 1483–1492.
23. Williams,D.M., Loakes,D. and Brown,D.M. (1998) Synthesis of a hydrogen-bond-degenerate tricyclic pyrrolopyrimidine nucleoside and of its 5 '-triphosphate. *J. Chem. Soc., Perkin Trans. 1*, **21**, 3565–3570.
24. Williams,D.M., Yakovlev,D.Y. and Brown,D.M. (1997) The synthesis of a tricyclic pyrrolopyrimidine related to N-6-hydroxyadenine. *J. Chem. Soc., Perkin Trans. 1*, **8**, 1171–1178.
25. Angelov,T., Guainazzi,A. and Schärer,O. D. (2009) Generation of DNA interstrand cross-links by post-synthetic reductive amination. *Org. Lett.*, **11**, 661–664.
26. Hill,F., Loakes,D. and Brown,D.M. (1998) Polymerase recognition of synthetic oligodeoxyribonucleotides incorporating degenerate pyrimidine and purine bases. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 4258–4263.
27. LaPointe,S.M., Wheaton,C.A. and Wetmore,S.D. (2004) The degenerate properties of modified purine and pyrimidine DNA bases. *Chem. Phys. Lett.*, **400**, 487–493.
28. Colucci,G. and Knobel,R. (2011) The COBAS (R) TaqMan (R) hepatitis C virus assays: automated systems for accurate and sensitive viral load quantification. *Expert Rev. Mol. Diagn.*, **11**, 793–798.
29. Macleod,I. J., Rowley,C. F. and Essex,M. (2015) Methods of determining polymorphisms. WO 2015/048730.