



## OPEN LNA blockers for improved amplification selectivity

Jaime Prout, Michael Tian, Alicia Palladino, Jason Wright & John F. Thompson

LNA-containing oligonucleotides bind DNA more tightly than standard DNA, so they can interact with targeted sequences and affect multiple processes. When a desired DNA is present at low concentrations relative to nearly identical undesired DNAs, LNAs can block amplification of unwanted DNAs. Using a short rAAV and synthetic DNA sequence as a model, we studied the length, number, and positioning of LNA bases to improve blocker effectiveness. Oligonucleotides 18–24 bases long with LNAs at every other position were most effective. Highly degenerate targets were used to characterize the impact of mismatches on blocking. Mismatches at LNA ends had little impact on blocking activity. Single and double mismatches were tolerated with longer blockers, especially if the mismatches were near LNA ends. Shorter LNAs were more selective, with > 1 mismatch preventing effective blocking. Neither the strand to which a blocker bound nor the distance between the blocker and priming sites greatly impacted blocking efficiency. We used these findings to design blockers of wild-type DNA versus the single-base *ALAT* PiZ allele. Blockers are most specific when the mismatch is located away from the LNA 5' end. Pairs of partially overlapping blockers on opposite strands with a centrally-located mismatch have maximal activity and specificity.

Locked Nucleic Acid (LNA) nucleotides are identical to natural nucleotides except for a methylene bridge spanning the deoxyribose sugar<sup>1,2</sup> which makes them more stable in double-stranded structures and more resistant to degradation<sup>3</sup>. The higher melting temperatures ( $T_m$ s) of oligonucleotides that include LNA bases provide greater specificity and new functions. They have been used successfully as PCR primers/probes<sup>4–6</sup>, as antisense reagents<sup>7–9</sup>, as selective binders for distinguishing single-nucleotide variants<sup>10–18</sup>, as agents for selective capture/degradation<sup>19,20</sup>, and as polymerization/splicing blockers<sup>21–28</sup>. Each of these roles requires that the LNA bind tighter than the corresponding pure DNA, but some functions may also require additional attributes that could be affected by the number and location of LNA bases within the oligonucleotide. Furthermore, proteins may interact with LNAs differently than with standard DNAs, dictating whether LNA or DNA should occupy a particular position.

While LNA blockers have been used in multiple situations, those studies have little commonality to provide insight into preferred designs. Some use chimeric<sup>11–13,15,21,23,26</sup> or pure<sup>10,17,27</sup> LNAs of 16 nt or shorter, while others use chimeric LNAs of 20 nt or longer<sup>16,22,24</sup>. For 20-mers, there are more than one million possible LNA-DNA configurations for each of the more than 1 trillion possible sequences. In addition, functional predictions of how different LNAs will perform as primers, blockers, or in other roles are even less well characterized because additional factors beyond the well-studied  $T_m$ <sup>29–33</sup> may play a critical role in how different LNAs perform<sup>34–36</sup>.

Blocking with LNAs is not the only method that has been used to prevent the amplification of majority DNAs when the detection of minority DNAs is desired. Other nucleic acid analogs like Peptide Nucleic Acids (PNA) selectively block amplification<sup>37,38</sup>. Choice of PNAs versus LNAs for a particular application will depend on the details of the system. LNAs have the advantages of lower cost and the ability to make longer molecules to reduce the number of perfect binding sites in complex genomes. PNAs can bind more strongly and are not susceptible to nucleases. A concern for PNAs can be specificity where the shorter PNAs may bind multiple genomic sites and potentially interfere with other DNAs of interest. In our case, we often wish to amplify DNAs of several kilobases or more arising from multiple genomic regions. Knowing that blockers will not interfere with DNAs elsewhere in the genome can be important. Thus, the more predictable and selective LNAs are preferred in our application though PNAs can be suitable in other situations.

One method for classifying blockers is whether they prevent the annealing of a PCR primer or prevent the elongation of that primer<sup>39</sup>. Annealing blockers are often used when the priming site for the desired amplicon is different by only one or by a small number of bases relative to the undesired amplicon. When few differences are present, shorter LNAs are often used to enhance the difference between perfect and imperfect binding. When the regions available for priming of the desired and unwanted DNAs are identical, annealing blockers cannot

Department of Genomics and Computational Biology, Homology Medicines, Inc., Bedford, MA 01730, USA. email: Thompson.john.f@gmail.com

be used. Instead, sequences specific to the interior of the undesired amplicon must be bound by an elongation blocker. Our aim was to preferentially amplify integrated recombinant Adeno-Associated Virus (rAAV) while preventing the amplification of episomal copies of the transduced rAAV. Episomal DNA contains the same sequences as integrated DNA<sup>40</sup>, so elongation rather than annealing blockers must be used. Because other rAAVs and DNAs of interest may contain different sequences and hence need different amplification conditions, it is helpful to understand how to design effective LNA elongation blockers for different contexts and applications.

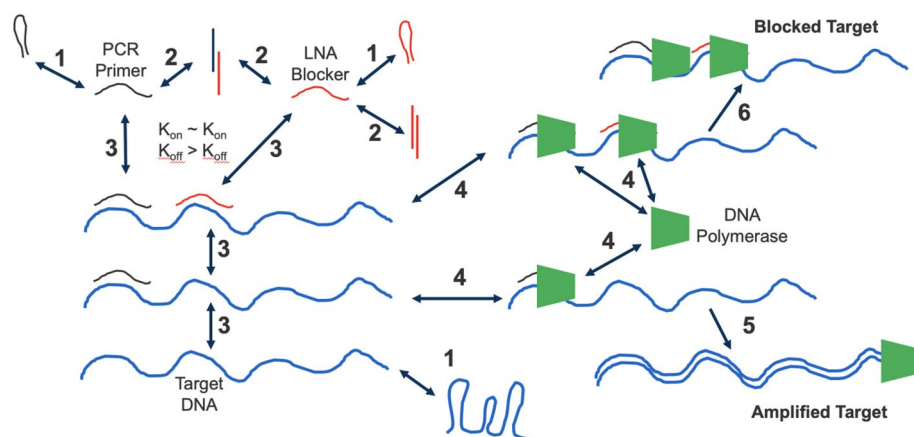
## Results

The process of blocking amplification involves many individual components. As diagrammed in Fig. 1, the priming and blocking oligonucleotides and the target DNA could all potentially participate in unimolecular, bimolecular, and higher-order interactions. During primer/blocker design, sequences are picked that minimize internal secondary structure and the formation of homo/heterodimers. The amplifying polymerase could bind to both oligo/target complexes and either stabilize or destabilize those interactions, with or without the additional complexity of polymerization. In addition to these potential interactions, there is also the likelihood that any component could interact with other DNAs in the mix because the reason for wanting to block the target DNA is to improve the signal relative to other related DNAs present in the complex mixture.

The large number of potential substrates and conditions prevent a complete analysis of all LNA lengths and sequences. For these experiments, we have restricted the total length of the LNA/DNA chimeras to those long enough to bind and provide reasonable specificity within the human genome ( $\geq 16$  nt) while not so long as to allow binding to occur despite the presence of multiple mismatches. For the initial experiments (Table 1), the target DNA was a plasmid containing a sequence derived from a multiple cloning site and episomal rAAV<sup>40</sup>. Subsequent experiments (Table 2) used a synthetic 175mer DNA (Template\_0) with the same targeted sequence flanked by different priming sequences (Fig. 2). Additional experiments used other DNA targets as described in Table S1. Generally, we used DNA polymerase GXL that lacks 5' > 3' exonuclease activity. All LNA-containing oligonucleotides were capped at both the 5' and 3' ends to minimize degradation and prevent extension as a primer. Initially, some LNAs were made with phosphorothioate linkages to prevent degradation. No effect was seen (data not shown), so later LNAs did not include such linkages.

Qualitative assays were run at varying LNA concentrations to identify where to focus with length and the number/positioning of LNAs (Table 1). For these studies, varying numbers of LNAs were situated in the middle, at either end, or throughout the oligonucleotide. The LNAs were added to PCR reactions at 0.1  $\mu$ M or 10  $\mu$ M, and all were targeted to the same DNA site for consistent comparisons. We found that maximum effectiveness occurred when chimeric molecules were  $\geq 20$  nt and half of the positions were substituted with LNAs. Furthermore, an even distribution of LNAs throughout the molecule (LNA\_10) was more effective than LNAs clustered together (3'MCS, 3'MCS\_5', and 3'MCS\_3'), so all later LNAs contained alternating LNA/DNA.

Based on the results in Table 1, a new set of 18–30 nt blockers was synthesized with alternating DNA/LNA nucleotides (Fig. 2, Table 2, and Table S1). LNAs were made with sequences identical to the top (forward, F) strand and the bottom (reverse, R) strand of the 175 nt synthetic Template\_0 (Fig. 2). In most cases, pairs of oligos were made, varying with respect to whether the first LNA was placed at position 1 or position 2. Oligonucleotides with names ending in 1 started with the first nucleotide as LNA, while oligonucleotides with names ending in 2 started with the second nucleotide as LNA with DNA/LNA alternating thereafter. All chimeras with LNAs starting with the second position had higher predicted  $T_m$ s than the equivalent molecules starting with LNA in the first



**Figure 1.** Potential interactions among intended reactants during PCR blockage. Potential intra- and inter-molecular interactions are shown for the PCR primer (black), the LNA blocker (red), the target DNA (blue), and the DNA polymerase (green). The DNAs/LNA can undergo folding into interfering secondary structures (1). The primer/blocker can potentially interact in a non-productive complex (2) with each other or in a productive or non-productive complex with the target DNA (3). LNAs have a slower off-rate versus DNAs<sup>46</sup>. The DNA polymerase can then bind the primer/blocker/target complex (4) and either extend the primer to generate an Amplified Target (5) or, if polymerization is halted by the LNA when the LNA is bound to the target DNA, a Blocked Target (6).

Name	Blocking at:		Pred Tm	Length	# LNAs	Position within binding site																				
	10 $\mu$ M	0.1 $\mu$ M				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
3'MCS	High	Low	69.6	20	10		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
3'MCS_5'	Low	None	65.8	20	9		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
3'MCS_3'	Low	None	65.3	20	9		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
DNA	None	None	51.4	20	0		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
All_LNA	Mod	None	81.0	20	20		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
LNA_10	High	High	66.9	20	10		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
LNA_7	Low	None	63.8	20	7		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
LNA_5	Low	None	58.2	20	5		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
LNA_4	Low	None	58.4	20	4		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
LNA_2	None	None	51.7	20	2		C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
3'MCS_16	Low	None	67.3	16	10					G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	
3'MCS_18	Low	None	66.6	18	10				T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A
3'MCS_22	Mod	None	70.6	22	10	C	C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
						A																				
3'MCS_24	Mod	None	68.8	24	10	C	C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
						A	T	A																		
3'MCS_30	High	High	70.4	30	10	C	C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
						A	T	A	A	G	T	A	G	C												
3'MCS_40	High	High	74.4	40	10	C	C	T	G	C	A	G	G	T	C	T	A	G	A	T	A	C	G	T	A	G
						A	A	T	A	A	G	T	A	G	C	A	T	G	G	C	G	G	G	T	T	

**Table 1.** Qualitative assessment of LNA/DNA blocking efficiency. Positions with LNAs are in white, positions with DNAs are in gray, and positions with no nucleotides are blacked out. Using a plasmid as the target DNA, different LNA/DNA chimera concentrations were added to the PCR mix at either 0.1  $\mu$ M or 10  $\mu$ M, and the product was analyzed via quantitative capillary electrophoresis. Blocking was assessed visually to determine relative efficiency.  $T_m$ s were calculated using default conditions at <https://www.idtdna.com/calc/analyzer>.

position (Table S1). Individual LNAs were used to block amplification at 1  $\mu$ M, with their effectiveness measured using quantitative capillary electrophoresis relative to samples with no LNA added (Fig. S1, Table 2). Effectiveness varied significantly among oligonucleotides. In most cases (10 of 12 tested), blockers with LNAs starting in the second/even positions were more effective than the same blockers with LNAs in the first/odd positions. Nearly half (15 of 34) of the blockers tested with primers F1/R1 yielded >90% blocking, and another 6 blocked >80%.

The initial results with the F1/R1 primer pair suggested that blockers separated on the same strand from the primers by 20–40 nt did not work as well as those positioned further away. The length of DNA gaps between the primers and blockers was changed to determine whether this affected blocking efficiency. Primer R1 was replaced with primers R3A and R3B that were 61 and 43 nt closer to the blocker binding sites to generate altered primer/blocker spacing. Blocker effectiveness did not change when the new primers were 20–40 nt away from the blocker, suggesting that, if there is a gap distance effect, it is not strong. Similarly, moving the primer by 4 bp to change helical orientation on the forward strand (primer F1 versus F3) had minimal effect on blocking (Table 2).

While the blocking efficiency of LNAs is a critical feature, some applications require specificity so that non-targeted sequences are minimally affected. To assess how target mismatches affect blocking efficiency, Template\_D was synthesized with a blocker binding region containing multiple partially degenerate positions located at the site bound by blockers 18\_F2, 18\_R2, 20\_F2, and 20\_R2 (Fig. S2). Template\_D had a different sequence relative to Template\_0 beyond the 20 nt identical region bound by these blockers (discussed in Fig. S2 legend). The degeneracy introduced by using a mixture of nucleotides during synthesis allows the creation of a highly complex pool of DNA molecules with a wide range of variants that can be used as a substrate for determining the effect of DNA mismatches on LNA blocking. This pool contained potentially over 1 trillion different DNA molecules with randomly situated variants within the degenerate region. Because the synthetic nucleotide pool favored the reference sequence (79%), molecules averaged only 4 variants each. In addition to the 18- and 20-mers that could be tested on both Template\_0 and Template\_D, we also wanted to test Template\_D with a longer blocker. A new pair of 24 nt LNAs was made (24D\_F2 and 24D\_R2). Even though the 24D blockers were 24 nt long, they covered only 21 degenerate positions because three positions bound by 24D consist of 100% reference sequence.

When there is a constant level of degeneracy in DNA, the frequency of variants within a given length can be predicted mathematically using classic combination/permutation equations (<https://en.wikipedia.org/wiki/Combination>). When R is the proportion of correct reference sequence and L is the length of the degeneracy, the frequency of a perfect reference sequence (no variation) is  $R^L$ . For DNA synthesized to be 79% Reference/21% Mismatch, the percent of reads matching the reference sequence perfectly is predicted to be 1.4% for an 18mer and falls to 0.7% for a 21mer. For predicting the number of variant sites in all molecules, the expected frequency for any given number of mismatches (MM) in the target region of length L can be calculated from the equation:

$$\% \text{ reads with mismatches (MM)} = 100 \times R^{L-MM} \times (1-R)^{MM} \times \frac{L!}{MM! \times (L-MM)!}$$

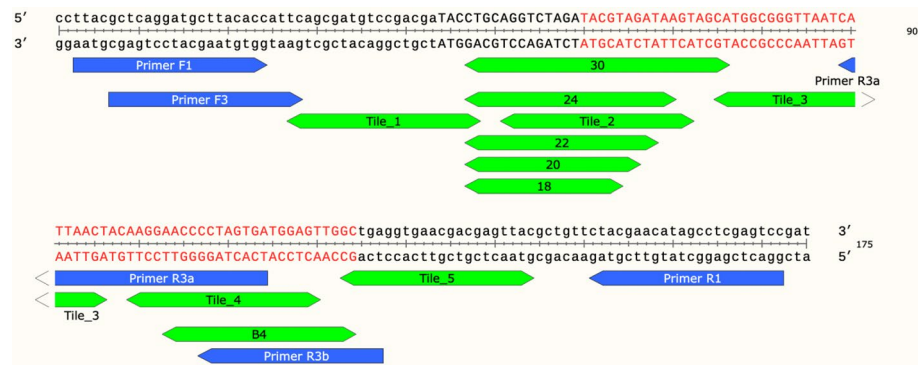
For example, to determine the frequency of molecules with 3 mismatches in a 20mer that was 79% reference sequence, one would calculate  $100 \times (0.79)^{17} \times (0.21)^3 \times (20!)/((3!) \times (17!))$  yielding 19.2%, meaning that 19.2%

Name	Oligo start	Oligo end	Primers F1/R1		Primers F3/R1		Primers F1/R3A		Primers F1/R3B	
			Primer/blocker gap	% Blocking	Primer/blocker gap	% Blocking	Primer/blocker gap	% Blocking	Primer/blocker gap	% Blocking
Tile_1F1	27	48	2	98	-2	99				
18_F1	47	64	22	73						
18_F2	47	64	22	89						
20_F1	47	66	22	68						
20_F2	47	66	22	83						
22_F1	47	68	22	58						
22_F2	47	68	22	95						
24_F1	47	70	22	91						
24_F2	47	70	22	97						
30_F1	47	76	22	90						
30_F2	47	76	22	94						
Tile_2F1	51	72	26	27	22	25				
Tile_3F1	75	96	50	82	46	83				
Tile_4F1	99	120	74	87	74	88				
B4_F1	103	124	78	98						
B4_F2	103	124	78	99						
Tile_5F1	123	144	98	90	94	91				
Name	Oligo start	Oligo end	Primers F1/R1		Primers F3/R1		Primers F1/R3A		Primers F1/R3B	
			Primer/blocker gap	% Blocking	Primer/blocker gap	% Blocking	Primer/blocker gap	% Blocking	Primer/blocker gap	% Blocking
Tile_5R1	144	123	6	96	6	96				
B4_R1	124	103	26	35						
B4_R2	124	103	26	49						
Tile_4R1	120	99	30	15	30	16	-31	10	-13	16
Tile_3R1	96	75	54	68	54	88	-7	96	11	87
30_R1	76	47	74	59			13	99	31	99
30_R2	76	47	74	68						
Tile_2R1	72	51	78	46	78	50	17	87	35	72
24_R1	70	47	80	93			19	96	37	92
24_R2	70	47	80	98						
22_R1	68	47	82	80			21	94	39	91
22_R2	68	47	82	95						
20_R1	66	47	84	78			23	92	41	89
20_R2	66	47	84	86						
18_R1	64	47	86	94			25	81	43	62
18_R2	64	47	86	74						
Tile_1R1	48	27	102	94	102	91	41	98	59	95

**Table 2.** Blocking efficiency. Positions of LNA blockers versus Template\_0 are listed. More detailed information on each blocker can be found in Table S1. Four sets of primers (shown in Fig. 2) were used with Template\_0 and the indicated blockers added at 1  $\mu\text{M}$ . The length of the gap between the elongating primer and the blocker is shown, as well as the efficiency of blocking for each primer set. Blockers identical to the top strand (F) are in the top half of the table, and blockers identical to the bottom strand (R) are in the bottom half of the table.

of all 20mers should have exactly three mismatches somewhere within their sequence. The predicted read frequencies as a function of the number of mismatches for 18, 20, and 21 degeneracies are shown in Fig. 3A. The frequencies for the 24D LNAs were calculated with 21 degeneracies because three template positions bound by the LNAs were reference, not degenerate. To compare the predicted frequency with the actual frequency, the target DNA pool was amplified with primers that bound the constant primer region outside of the degenerate regions. After barcoding, the amplified samples were sequenced using a MiSeq and the frequency of variant positions compared to the predicted results. The mismatch frequency for the samples with no LNA blocking closely mirrors the calculated values for 79% reference/21% mismatch (Fig. 3A and Table S2). Values were calculated only up to 10 mismatches because, with these synthetic parameters, more than 99% of DNA molecules have 10 variants or fewer.

To determine the impact of mismatches on blocking, amplifications of the complex pool of DNA sequences that make up Template\_D were performed with either no LNA or individual LNAs at 1  $\mu\text{M}$ . The primers used to



**Figure 2.** Template\_0 target DNA and primer/LNA binding locations. The target oligonucleotide to be blocked (Template\_0) includes a left primer region (1–43), a multiple cloning site (MCS) with SbfI, XbaI, and SnaBI sites (44–59), AAV2 sequence (4495–4558 in <http://www.ncbi.nlm.nih.gov/nucore/AF043303.1>, 60–124, shown in red) and a right primer region (125–175). Two forward primer sequences on the left, F1 and F3, and three reverse primers on the right, R1, R3A, and R3B (blue arrows) were used to amplify the sequence with or without LNA blockers. Each LNA blocker (green double arrows) was made in four versions, two corresponding to the top strand (F, forward) and two corresponding to the bottom strand (R, reverse). The pair of F LNAs and the pair of R LNAs were synthesized starting with LNAs at either position 1 or position 2. These are listed individually in Table 2 and Table S1.

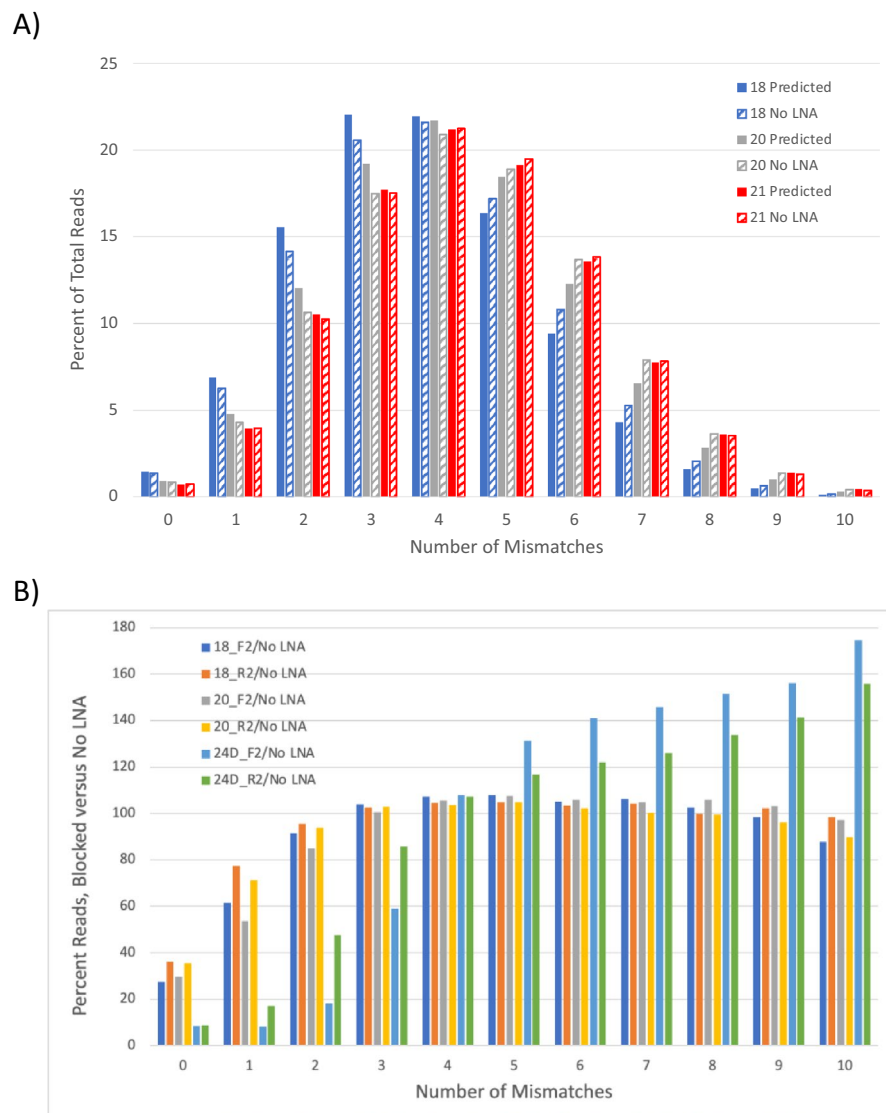
amplify Template\_D each bound identical, non-degenerate regions of the DNAs while the blockers had trillions of different sequences to which they could bind. The ability of LNAs to bind certain sequences caused blocking of those DNAs while others amplified normally. If variants in the target pool are differentially affected by LNA blockers, the read distribution should change accordingly. When blocked samples were examined (Fig. 3B), all 18 and 20 nt blockers had 27–36% as many reads with no mismatches as the unblocked samples, while the 24 nt blockers had only 8.4–8.6% as many reads with no mismatches as unblocked samples, showing that specific blocking had occurred, and the amount of blocking was length dependent. The read frequency for each mismatch was very similar when 18\_F2/18\_R2 and 20\_F2/20\_R2 were compared, indicating that the strand being blocked does not matter with this target. There is a greater deviation between 24D\_F2 and 24D\_R2, but this may be due to the asymmetric location of degenerate positions in the two blockers. 24D\_R2 has three fixed target positions near its 5' end, while the fixed positions for 24D\_F2 are near its 3' end.

For the 18 and 20 nt blockers, there were 54–77% as many reads with 1 mismatch relative to no LNA. There is only a slight difference in reads (85–96% as many reads relative to no LNA) with 2 mismatches relative to no LNA, indicating that any mismatches lead to poor blocking at these LNA lengths. All blocking activity was lost with three or more mismatches for 18 and 20 nt blockers. With the 24 nt blockers, there was significant blocking with 0, 1, and 2 mismatches. There was lower but measurable blocking (59–86% as many reads relative to no LNA) with 3 mismatches. This indicates that the specificity of blocking is dependent on the length of the blocker and longer LNAs are capable of blocking to some degree, even in the presence of up to three suitably placed mismatches.

To assess the relative importance of different positions within the LNAs for blocking specificity, the locations of mismatches within all reads with a single mismatch were examined. The positional frequency of single mismatches in the No LNA sample relative to all reads in the sample is uniform as each mismatch has no effect on amplification (Table S2). The frequency of single mismatch reads by position relative to all reads in 18 and 20 nt blocked samples varies significantly and is shown in Fig. 4A. The frequencies are given as a percentage relative to the values observed in the unblocked sample. Because effective blocking was observed with both one and two mismatches with 24D\_F2 and 24D\_R2, the positional effects for both one and two mismatches are shown for those blockers (Fig. 4B). If all positions are equally important for blocking, all blocked samples would have the same percentage across the length of the LNA. Indeed, that is the observation with the 24D\_F2 LNA with a single mismatch: it does not matter where the mismatch occurs, there is 90–95% blocking, independent of position. For all other conditions with one or two mismatches, there is a positional dependence on blocker effectiveness. Positions very close to either end of the blocked region have lower relative read frequencies than the more centrally located positions. The lower read frequency for mismatches near the blocker ends means templates with those mismatches are blocked despite the mismatch, so they are less represented among all reads. If selectivity of blocking is desired, significant blocking in the presence of mismatches is not a good thing. Though the position of the mismatch matters, the identity of the base substitution at a given position is generally irrelevant. At most positions, all mismatches have similar blocking effects (Table S3). There are some differences but no obvious trends.

An additional common feature among blockers that did not perform well was the presence of an A on the same strand adjacent to the 5' end of the blocker binding site. When the base was an A, the average blocking effectiveness was 42% (n = 5), while it was 84% for C (n = 16), 93% for G (n = 4), and 81% for T (n = 9). To test the same blockers with a different adjacent sequence, two new templates were made (Template\_A1 and Template\_A2, Figs. S3 and S4). Many LNA blockers shown in Table 2 had overlapping binding sites, so making a base change adjacent to one binding site created a mismatch within other binding sites. Only the subset of the original blockers that maintained perfect binding sites were tested on the new templates. For the blockers tested,

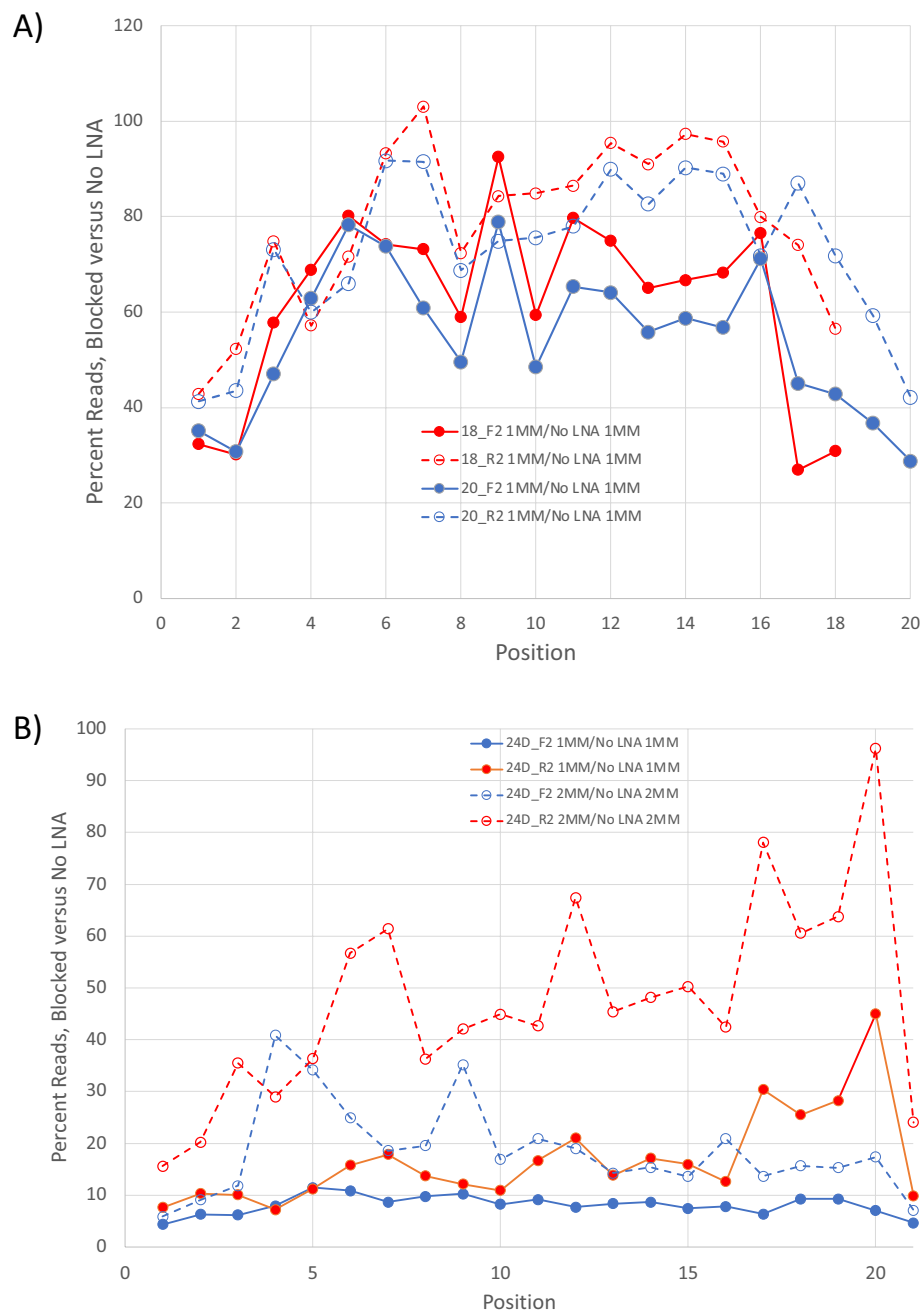




**Figure 3.** Predicted and observed mismatch read frequencies. **(A)** The predicted mismatch frequency (solid fill) and observed variant read frequency for unblocked samples (hatched fill) as a percent of total reads normalized by sample for 0–10 variants is shown for degenerate DNA lengths of 18 (blue), 20 (gray), and 21 (red) nt. These predictions were made using the combination equation in the text where  $R=0.79$ ,  $L=18$  or 20 or 21 and  $MM=0-10$ . >99% of all molecules with these degeneracy lengths and 79% reference sequence should have 10 or fewer variants so higher variant values are not shown. Even though the 24D blockers are 24 nt long, a length of 21 is used for its calculations because only 21 of the 24 positions on the target DNA are degenerate. The calculated frequencies are nearly identical to the observed variant frequencies in DNA reads when the amplification is not blocked by LNAs (Table S2). **(B)** For forward (F2) or reverse (R2) blockers covering 18, 20, or 21 degenerate positions, the percent of reads for each sample versus no LNA present is shown for each mismatch value. The number of mismatches allowed while still retaining blocking activity varies by length.

all preceding non-As were changed to A, and all preceding As were changed to another base (Table 3, Figs. S3 and S4). The eleven changes from C/G/T to A resulted in an average decrease in efficiency of 21% while the three changes from A to G/T resulted in an average increase in efficiency of 15%. In addition to Template\_A1 and Template\_A2 which have uniformly altered sequences, it is also possible to use Template\_D to address the question of adjacent sequence. Two of the blockers used, 18\_R2 and 24D\_R2, have a degenerate position adjacent to the 5' end of their binding site on that target DNA. Thus, the frequency of each base prior to their binding sites can be examined for differences. There do not appear to be significant base-specific effects at those sites with Template\_D; so, if adjacent bases have an impact, the effect is small, may be more complex than just a single base, or may be evident only in certain conditions.

A key component of the overall amplification and blocking process is the DNA polymerase. The experiments described thus far were carried out with GXL polymerase, which has 3' > 5' exonuclease activity, but no 5' > 3' exonuclease activity as some other polymerases have. Two additional polymerases were tested in parallel against



**Figure 4.** Blocking as function of position. **(A)** The relative read frequency for amplification with blockers versus amplification with no LNAs is shown as a function of position. In panel A, the frequency of reads with single mismatches is shown for 18 (red) and 20 (blue) nt LNA blockers. Blockers identical to the forward strand are shown with solid lines/symbols, and blockers identical to the reverse strand are shown with dashed lines/open symbols. The overall shape for all forward primers and for all reverse primers is similar. **(B)** LNAs blockers that cover 21 degenerate positions are shown with data for reads for one mismatch (1MM, solid lines/symbols) and for reads with two mismatches (2MM, dashed lines/open symbols). Blockers identical to the forward strand are shown in blue, and the reverse strand in red. Blocking in the presence of mismatches is not necessarily desirable as it indicates a lack of selectivity.

a subset of LNA blockers carried out in conditions and temperatures recommended by the suppliers. As shown in Table 4, the two polymerases which lack 5' > 3' exonuclease activity, GXL and Q5, had similar blocking profiles. The polymerase that does exhibit 5' > 3' exonuclease activity, GoTaq, behaves differently. Template\_U has no LNA binding sites, but it is blocked to the same extent as Template\_O with 6 of the 8 blockers tested. There may be a small amount of specific blocking with the two best blockers, 30\_R1 and 30\_R2. Literature results for blocking of Taq have produced discrepant results<sup>11,41</sup>. Those experiments were run at different temperatures, so

Name	Template 0	Sequence change	Template A1	Difference A1	Template A2	Difference A2
A > N						
30_R1	0.47	77 A > T	0.92	0.45		
Tile_4R1	0.31	121 A > G			0.16	-0.15
B4_R1	0.21	125 A > G	0.37	0.16		
N > A						
18_F1	0.73	46 C > A	0.82	0.09		
Tile_1R1	0.94	49 C > A			0.60	-0.33
Tile_2F1	0.27	50 C > A			0.27	0.00
Tile_4F1	0.87	98 C > A			0.37	-0.50
Tile_5R1	0.96	145 C > A			0.85	-0.11
Tile_3F1	0.82	74 G > A			0.50	-0.32
B4_F1	0.98	102 G > A	0.61	-0.37		
Tile_5F1	0.90	122 G > A			0.44	-0.47
Tile_1F1	0.98	26 T > A			0.80	-0.18
Tile_2R1	0.46	73 T > A			0.50	0.03
Tile_3R1	0.68	97 T > A			0.52	-0.16

**Table 3.** Impact of preceding base. Fourteen blockers were tested on two templates, the original Template\_0 and either Template\_A1 or Template\_A2. Both Template\_A1 (Fig. S3) and Template\_A2 (Fig. S4) have altered sequence adjacent to selected blocker binding sites. Below, the first three blockers involved a change from A to either T or G, while the last eleven blockers involved a change from C, G, or T to A. The initial blocking efficiency with Template\_0 is shown in the second column and the second blocking efficiency is shown in the 4th or 6th column with all LNAs added at 1  $\mu$ M.

Polymerase >	GXL—68C		Q5—72C		GoTaq—68C		GoTaq—60C	
Template >	T_O (%)	T_U (%)	T_O (%)	T_U (%)	T_O (%)	T_U (%)	T_O (%)	T_U (%)
20_F1	72	35	64	23	26	37	37	0
20_F2	73	22	66	26	43	43	47	-4
20_R1	69	24	32	27	40	39	47	-3
20_R2	92	23	84	28	40	32	51	-10
30_F1	92	40	89	29	44	40	38	-6
30_F2	95	53	90	24	41	40	34	2
30_R1	98	41	97	22	53	42	37	-14
30_R2	95	38	99	25	48	37	44	-9

**Table 4.** LNA blocking as a function of DNA polymerase. Eight blockers with moderate to high blocking effectiveness were evaluated with three different polymerases using the indicated extension temperatures. The manufacturers' suggested buffer and cycling conditions were used for each polymerase except for when GoTaq was used with a 60 °C extension. Blocking was attempted with both Template\_0 that contains a specific binding site and Template\_U which did not contain any site recognizable by these blockers. Both templates were amplified using primers F1 and R1 with LNA added to 1  $\mu$ M.

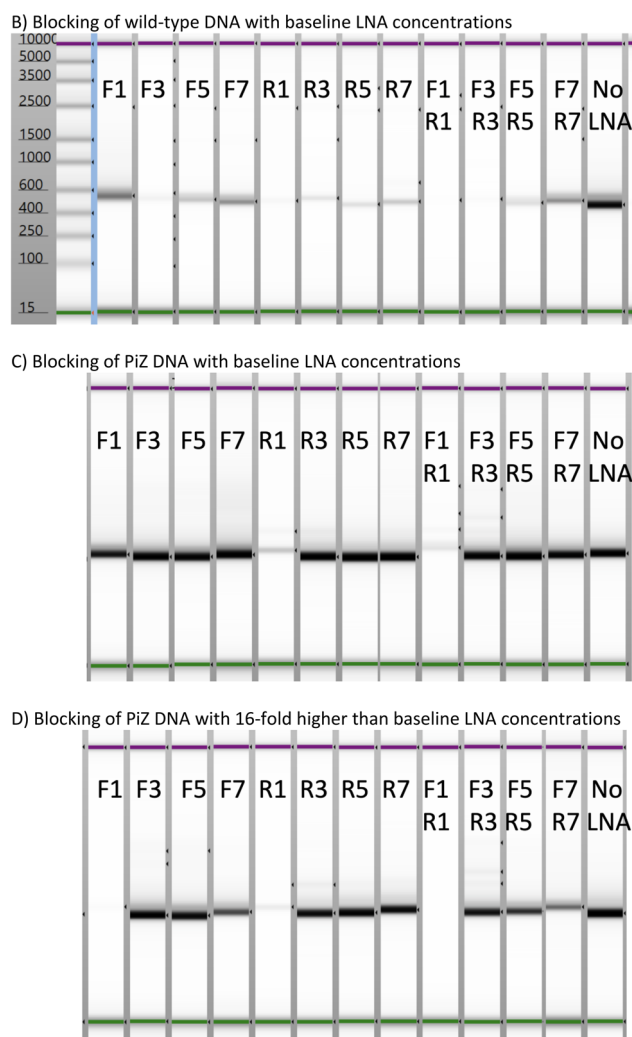
we also tested GoTaq with an extension temperature of 60 °C where LNA blocking was reported. We find specific blocking at that temperature though not as much as found with the other polymerases.

To see if these design properties could be extended to another target of biological interest, LNA chimeras were designed to the wild-type allele of *A1AT* to see whether its amplification could be blocked specifically relative to the single base PiZ mutation which is linked to COPD and other respiratory diseases<sup>42</sup>. Four pairs of LNAs matching the wild-type sequence were made (Fig. 5A). These LNA pairs placed the base corresponding to the PiZ mutation at the 5' end and at the third, fifth, or seventh position in from the 5' end. This was done for both DNA strands (Forward and Reverse). All LNAs started at the second position and alternated through the molecule. Five of the eight were 16 nt long but the three with the lowest  $T_m$ s were extended one additional base to make the  $T_m$ s more similar (Table S1). The region around the PiZ allele was amplified from human genomic DNA using primers listed in Table S1 to generate a 646 bp product. LNAs were titrated versus wild-type DNA to find concentrations of each that would yield extensive but not complete blocking. These concentrations varied significantly and were independent of  $T_m$ . The best *A1AT* LNA blocker, A1AT\_R1\_2, was 30 times more potent than the worst blocker, A1AT\_F1\_2. Based on this, genomic DNA with either wild-type or PiZ allele was amplified with the eight individual LNAs as well as with each pair of corresponding F/R LNAs. As shown in Fig. 5, concentrations that yielded significant blocking of wild-type *A1AT* had no effect on PiZ *A1AT* except



## A) LNAs synthesized for A1AT

Chr14:94378625-94378594  
 A1AT\_F1-2  
 5' GTGCTGACCATCGACRAGAAAGGGACTGAAGC 3'  
 3' CACGACTGGTAGCTGYTCTTCCCTGACTTCG 5'  
 A1AT\_R1-2  
 A1AT\_F3-2  
 5' GTGCTGACCATCGACRAGAAAGGGACTGAAGC 3'  
 3' CACGACTGGTAGCTGYTCTTCCCTGACTTCG 5'  
 A1AT\_R3-2  
 A1AT\_F5-2  
 5' GTGCTGACCATCGACRAGAAAGGGACTGAAGC 3'  
 3' CACGACTGGTAGCTGYTCTTCCCTGACTTCG 5'  
 A1AT\_R5-2  
 A1AT\_F7-2  
 5' GTGCTGACCATCGACRAGAAAGGGACTGAAGC 3'  
 3' CACGACTGGTAGCTGYTCTTCCCTGACTTCG 5'  
 A1AT\_R7-2



**Figure 5.** Blocking *A1AT*. (A) The sequence surrounding the PiZ allele in *A1AT* is shown (central base, R/Y). G/C is the wild-type sequence and A/T is the PiZ allele. LNAs are named with an F if identical to the top strand and R if identical to the bottom strand. The number adjacent to F/R indicates where the PiZ allele is in the LNA. Individual LNAs and how they overlap are shaded. Because each LNA varies by two in length, the alternating LNAs are positioned identically in these series. Based on potency in initial studies at 1 and 10  $\mu\text{M}$ , concentrations were titrated to obtain good but not complete blocking at the lowest concentrations used. The lowest concentration for A1AT\_R1-2 was 0.1  $\mu\text{M}$ , followed by A1AT\_F3-2, A1AT\_R3-2, and A1AT\_R5-2 at 0.2  $\mu\text{M}$ , A1AT\_F5-2 and A1AT\_R7-2 at 0.5  $\mu\text{M}$ , A1AT\_F7-2 at 1  $\mu\text{M}$ , and A1AT\_F1-2 at 3.3  $\mu\text{M}$ . (B) Blocking of wild-type DNA with LNA concentrations listed in (A). (C) Blocking of PiZ DNA with LNA concentrations listed in (A). (D) Blocking of PiZ DNA with LNA concentrations 16-fold higher than listed in (A). Data is not shown for the 16 $\times$  wild-type DNA because all LNAs completely blocked amplification.

for A1AT\_F1-2 and A1AT\_R1\_2 which blocked the PiZ DNA nearly as well as wild-type despite the mismatch at the 5' end for both. The combination pairs all blocked wild-type DNA better than the individual blockers. When LNA concentrations were increased 16-fold, five LNAs blocked PiZ DNA amplification less than 30% (A1AT\_F3-2, A1AT\_F5-2, A1AT\_R3\_2, A1AT\_R5\_2, and A1AT\_R7\_2) while A1AT\_F1-2 and A1AT\_R1\_2 blocked it more than 95%. More effective blocking of PiZ DNA was also observed with all combinations except for A1AT\_F7\_2/ A1AT\_R7\_2 which was similar to the individual blockers. This blocker pair overlaps each other by 13 nt so it is likely they will bind each other in solution, potentially affecting their ability to bind to both wild-type and PiZ DNA.

## Discussion

PCR has been used to routinely amplify rare DNA sequences from complex mixtures. When the desired DNA has unique sequence characteristics, high degrees of amplification are readily achieved. When the targeted sequence is overwhelmed by an excess of highly similar sequences, selective amplification may not be as straightforward, and detecting specific rare DNAs in a mix of nearly identical molecules can be challenging. For example, when seeking to identify rAAV DNAs that have integrated into the host cell genome, there can be a high level of episomal DNA, up to thousands of copies per cell, obscuring less frequent genomic integration events. Sherman et al.<sup>24</sup> used 27–32 nt blockers with 9 LNAs to address a similar issue with retroviral and lentiviral genomic integrations. These viral integration studies encountered a less severe problem due to the absence of high copy episomal DNA. Similar scenarios, such as complex metagenomic samples, somatic mutations, multiple paralogs, repetitive sequences, or pseudogenes, can also be complicated by DNA ratios for which the desired amplification is challenging. In the rAAV scenario, unique sequences are present in the non-integrated segments of the episomal virus relative to the integrated version, providing wide latitude in sequence choice. In other situations, there may be only one or a small number of changes, limiting the flexibility in blocker choice.

Based on thermodynamic studies<sup>29–33</sup>, all the oligonucleotides with LNA substitutions we studied should stably bind the target DNAs of interest. While LNA binding is a minimal requirement, the positioning and number of LNAs may or may not affect function, so relying on binding alone to predict function would be ill-advised. A 50% fractional LNA content performed best in initial experiments, especially when LNAs were located throughout the blocker rather than clustered in one region. Our initial observations suggesting that alternating LNA substitutions are functionally superior is consistent with other studies that indicate special LNA binding properties with such an arrangement<sup>43,44</sup>. Blocking activity was generally better when the LNAs started at the second position rather than the first.

To allow analysis of trillions of different targets, a highly degenerate pool of DNAs was synthesized for blocking studies. For DNA with 18–21 degenerate positions with the potential for all four bases at each position, this was equivalent to testing up to 4 trillion different targets simultaneously. Because the target DNA was synthesized to favor the reference sequence, molecules averaged four variants each, but could have anywhere from 0 to 21 variant positions. It is unlikely that most molecules with > 10 variants are represented because the pool was designed to test DNAs with fewer variants. We found that the exact sequence of mismatches did not generally have an effect at any blocked position. However, the position of the mismatch on the template does matter. Reads with mismatches very close to either end of the blocker were found far less often than more centrally located mismatches, indicating that mismatches at the ends still allowed less specific blocking, while more interior mismatches prevented blocking. If blocking of both perfect and slightly imperfect matches is not an issue, this positional effect will not matter.

However, when the intent is to distinguish between two DNA sequences that differ by only a single base, the positional effect is critical. Placing the mismatch away from the blocker ends is necessary if there is to be good blocking with a perfect match and no blocking with a single mismatch. We used the PiZ single base mutation in the *A1AT* gene as an example for evaluating the generalizability of these findings. Four pairs of LNAs that placed the mismatch position at the 5' end or 3, 5, or 7 nt into the LNA confirmed that internal LNA positions are superior to the 5' end with respect to specificity. While all LNAs block the wild-type allele, the 5' end mismatch also blocks PiZ allele DNA with nearly the same effectiveness. When LNAs binding to both strands are used, improved blocking is observed. Of the LNAs we tested, the combination of A1AT\_F7-2 and A1AT\_R7-2 had the greatest overlap between them, 13 nt. Even with this high degree of overlap that could lead to stable binding to each other, blocking of the targeted allele remained strong though with slightly reduced specificity toward the mutation. The most effective and specific blockers had the mismatch located 3 or 5 nt from the 5' end.

Previously, there have been reports that polymerases with 5' > 3' exonuclease activity eliminate LNA blocking activity<sup>11</sup>, while others have found no effect<sup>41</sup>. In our hands, Taq polymerase with 5' > 3' exonuclease activity is blocked non-specifically by LNAs when amplification is carried out at the recommended extension temperature. Amplification is inhibited 26–53% (Table 4) when there are specific target binding sites present (Template\_0) and 32–48% when there are no LNA binding sites present (Template\_U). There may be some specific blocking with the best blockers, 30\_R1 and 30R\_2. GXL and Q5 polymerases, which both lack 5' > 3' exonuclease activity, are non-specifically inhibited to a lesser degree and have much higher specific inhibition when there is a target binding site. If the amplification conditions that produced disparate results in the literature are compared, the most notable difference is the extension temperature during amplification. When primers are extended at 72 °C<sup>11</sup>, polymerases with 5' > 3' exo activity exhibit no blocking while extension at 60 °C<sup>41</sup> retains blocking. We can mimic both results by adjusting the Taq extension temperature. Previous examination of Taq and the Stoffel analog, which lacks 5' > 3' exo activity, showed marked differences in temperature dependence of single versus double-stranded template replication<sup>45</sup> that may explain why 5' > 3' exo activity matters in some situations but not others. Other polymerase-specific properties like fidelity and processivity were not tested but could also be

envisioned to affect blocking. Thus, if certain polymerases are required for a specific application, it is necessary to test them to ensure that they can be effectively blocked by the desired LNAs.

The factors identified here are not guarantees of LNA blocker success, but they provide a guide when designing amplification blockers. The widely different concentrations of the *AIAT* LNAs required to achieve similar blocking on overlapping sequences highlight the remaining design unpredictability. Blockers of 18–24 nt with alternating LNAs beginning in the second position are a reasonable starting point when selectivity is not critical. If high selectivity for single base changes is needed, shorter oligos should be used (16–18 nt), while longer ones (24 nt) can be used if one or two mismatches are acceptable. For more complete blocking, both DNA strands should be targeted so that neither undesired strand would be synthesized. If only one or a few sequence differences can be exploited and selectivity is needed, it would be best to place mismatches away from the LNA ends.

## Materials and methods

**PCR conditions and blocking efficiency.** PCR reactions were carried out with GXL polymerase (Takara), Q5 polymerase (NEB), or GoTaq polymerase (Promega) with the addition of specified LNAs to the listed final concentrations. The target DNA used for Table 1 results was 0.001 ng plasmid DNA<sup>40</sup>. The target DNA used for *AIAT* in Fig. 5 was Promega human genomic DNA for the wild-type allele or DNA from a human cell line containing the PiZ allele. The target DNA used for all other experiments was synthetic DNA templates (0.001 ng) whose sequences are listed in Table S1. Most experiments were carried out with the DNA polymerase GXL and cycling parameters of 18 cycles of 98 °C for 10 s, 55 °C for 15 s, and 68 °C for 30 s. In Table 4, other polymerases were also used with cycling parameters for Q5 of initial denaturation of 98 °C for 30 s followed by 18 cycles of 98 °C for 10 s, 55 °C for 15 s, and 72 °C for 30 s and a final extension at 72 °C for 2 min. Cycling parameters for GoTaq consist of initial denaturation of 95 °C for 2 min followed by 18 cycles of 95 °C for 30 s, 55 °C for 15 s, and 60 °C or 68 °C for 30 s and a final extension at 60 °C or 68 °C for 5 min. All experiments included controls with no LNA added. Primers, LNAs, and non-degenerate templates were ordered from Integrated DNA Technologies (IDT), and the degenerate template (Template\_D) was ordered from Genewiz/Azenta. All template, LNA, and primer sequences are provided in Table S1. The Template\_0 sequence was derived from a plasmid which was used as the template DNA for the reactions in Table 1. That sequence included a synthetic multicloning region and a region of AAV DNA. PCRs with LNAs, primers, and templates were tested as described in Tables 1 and 2. For the Table 2 experiments, at least 3 replicates were performed for each LNA/Template combination. From each PCR reaction, 1 µL was run using quantitative capillary electrophoresis according to the recommended protocol for D1000 tapes (Agilent) (Fig. S1).

**Evaluating primer-LNA binding position.** To determine if the distance from the primer to the LNA binding site was important, a forward primer, F3, was designed for use in conjunction with primer R1 and tested with tiling LNAs 1–5 directed to either strand with Template\_0. Additionally, two reverse primers, R3a and R3b, were designed and tested in conjunction with primer F1 and tested with LNAs 18\_R1, 20\_R1, 22\_R1, 24\_R1, 30\_R1, and tiling LNAs 1-4R with Template\_0.

**Evaluating the base prior to LNA binding region.** To investigate whether the base just prior to the LNA binding region was important, two DNAs, Template\_A1 (Fig. S3) and Template\_A2 (Fig. S4), were synthesized with a series of single base changes relative to Template\_0. With primers F1 and R1 and using conditions stated above, LNAs 18\_F1, 30\_R1, B4\_F1, and B4\_R1 were evaluated with Template\_A1 and LNAs Tile\_1F, Tile\_1R, Tile\_2F, Tile\_3F, Tile\_3R, Tile\_4F, Tile\_4R, Tile\_5F, and Tile\_5R were evaluated with Template\_A2.

**Quantitative capillary electrophoresis data analysis.** For experiments where Agilent quantitative capillary electrophoresis was used to measure the signal, the concentration was recorded for the band of interest for each sample. Replicates were averaged, and percent blocking was calculated with the formula

$$\% \text{ Blocking} = \left( 1 - \frac{[\text{average LNA signal}]}{[\text{average no LNA signal}]} \right) \times 100$$

**Degenerate template sequencing.** LNAs 18\_F1, 18\_F2, 18\_R1, 18\_R2, 20\_F1, 20\_F2, 20\_R1, and 20\_R2, 24D\_F2, 24D\_R2, and No LNA were evaluated using Template\_D where the LNA binding region was synthesized with a degenerate sequence with 79% reference base/7% each non-reference base to determine the role of mismatches in the target sequence. PCR was performed as above, except primers 66Set1 and R1-2 were used. These primers included sequences compatible with Illumina sequencing. Replicates were pooled and purified using Monarch PCR and DNA Cleanup Kit (NEB). A barcoding PCR was performed with 1 µL of each purified PCR1 using GXL polymerase with the following conditions, 98 °C for 2 min, 25 cycles of 98 °C for 10 s, 55 °C for 15 s, 68 °C for 30 s, and a final extension at 68 °C for 3 min. Barcoded products were purified again, and 1 µL run via Agilent Quantitative Capillary Electrophoresis according to the recommended protocol for D1000 tapes to check for clean products and calculate molarity. Products were then pooled equimolarly, and 50% PhiX was added for library complexity. The DNA pool was melted, diluted, and loaded onto an Illumina MiSeq following the standard V2 300 Cycle kit protocol.

**Sequencing data analysis.** Demultiplexed Fastq files were collected from the instrument, and fastQC was performed to QC the data. Reads were aligned to the reference sequence with Bowtie, and the resulting bam file was filtered to remove secondary and supplementary alignments and unmapped, and poor-quality reads. Next,

for each mapped read, each base aligning to the degenerate part of the template was extracted and added to a pivot table, with each row consisting of a readID and the columns containing corresponding individual bases by position. Each base was labeled as R for reference, N for not called, or the base was left unchanged if it did not match the reference. For each readID, the number of mismatches (bases that were not R or N) was calculated in the region targeted by the LNA. The dataset was then separated by the number of mismatches. For each dataset, the number of reference and non-reference bases were counted per position, allowing base frequencies at each position to be calculated.

### Data availability

Sequence data is available at SRA with the accession number PRJNA910530. All other data can be found in the manuscript or supplementary files.

Received: 3 January 2023; Accepted: 20 March 2023

Published online: 24 March 2023

### References

- Obika, S. *et al.* Stability and structural features of the duplexes containing nucleoside analogues with a fixed N-type conformation, 2'-O,4'-C-methyleneribonucleosides. *Tetrahedron Lett.* **39**, 5401–5404. [https://doi.org/10.1016/S0040-4039\(98\)01084-3](https://doi.org/10.1016/S0040-4039(98)01084-3) (1998).
- Singh, S. K., Koshkin, A. A., Wengel, J. & Nielsen, P. LNA (locked nucleic acids): Synthesis and high-affinity nucleic acid recognition. *Chem. Commun.*, 455–456. <https://doi.org/10.1039/A708608C> (1998).
- Frieden, M., Hansen, H. F. & Koch, T. Nuclease stability of LNA oligonucleotides and LNA–DNA chimeras. *Nucleosides Nucleotides Nucleic Acids* **22**, 1041–1043. <https://doi.org/10.1081/NCN-120022731> (2003).
- Ballantyne, K. N., van Oorschot, R. A. & Mitchell, R. J. Locked nucleic acids in PCR primers increase sensitivity and performance. *Genomics* **91**, 301–305. <https://doi.org/10.1016/j.ygeno.2007.10.016> (2008).
- Latorra, D., Arar, K. & Hurley, J. M. Design considerations and effects of LNA in PCR primers. *Mol. Cell. Probes* **17**, 253–259. [https://doi.org/10.1016/S0890-8508\(03\)00062-8](https://doi.org/10.1016/S0890-8508(03)00062-8) (2003).
- Wadle, S., Lehnert, M., Rubenwolf, S., Zengerle, R. & von Stetten, F. Real-time PCR probe optimization using design of experiments approach. *Biomol. Detect. Quantif.* **7**, 1–8. <https://doi.org/10.1016/j.bdq.2015.12.002> (2016).
- Kurreck, J., Wyszko, E., Gillen, C. & Erdmann, V. A. Design of antisense oligonucleotides stabilized by locked nucleic acids. *Nucleic Acids Res.* **30**, 1911–1918. <https://doi.org/10.1093/nar/30.9.1911> (2002).
- Braasch, D. A., Liu, Y. & Corey, D. R. Antisense inhibition of gene expression in cells by oligonucleotides incorporating locked nucleic acids: Effect of mRNA target sequence and chimera design. *Nucleic Acids Res.* **30**, 5160–5167. <https://doi.org/10.1093/nar/gkf651> (2002).
- Straarup, E. M. *et al.* Short locked nucleic acid antisense oligonucleotides potentially reduce apolipoprotein B mRNA and serum cholesterol in mice and non-human primates. *Nucleic Acids Res.* **38**, 7100–7111. <https://doi.org/10.1093/nar/gkq457> (2010).
- Albitar, A. Z., Ma, W. & Albitar, M. Wild-type blocking PCR combined with direct sequencing as a highly sensitive method for detection of low-frequency somatic mutations. *J. Vis. Exp.* <https://doi.org/10.3791/55130> (2017).
- Dominguez, P. L. & Kolodney, M. S. Wild-type blocking polymerase chain reaction for detection of single nucleotide minority mutations from clinical specimens. *Oncogene* **24**, 6830–6834. <https://doi.org/10.1038/sj.onc.1208832> (2005).
- Peng, J. *et al.* Wildtype blocking pcr coupled with internal competitive amplified fragment improved the detection of rare mutation of KRAS. *Mol. Med. Rep.* **16**, 2726–2732. <https://doi.org/10.3892/mmr.2017.6883> (2017).
- Yang, H. *et al.* A tailored LNA clamping design principle: Efficient, economized, specific and ultrasensitive for the detection of point mutations. *Biotechnol. J.* **16**, e2100233. <https://doi.org/10.1002/biot.202100233> (2021).
- You, Y., Moreira, B. G., Behlke, M. A. & Owczarzy, R. Design of LNA probes that improve mismatch discrimination. *Nucleic Acids Res.* **34**, e60. <https://doi.org/10.1093/nar/gkl175> (2006).
- Johnson, M. P., Haupt, L. M. & Griffiths, L. R. Locked nucleic acid (LNA) single nucleotide polymorphism (SNP) genotype analysis and validation using real-time PCR. *Nucleic Acids Res.* **32**, e55. <https://doi.org/10.1093/nar/gnh046> (2004).
- Ang, D. *et al.* Novel method for PIK3CA mutation analysis: Locked nucleic acid–PCR sequencing. *J. Mol. Diagn.* **15**, 312–318. <https://doi.org/10.1016/j.jmoldx.2012.12.005> (2013).
- Denys, B., El Housni, H., Nollet, F., Verhasselt, B. & Philippe, J. A real-time polymerase chain reaction assay for rapid, sensitive, and specific quantification of the JAK2V617F mutation using a locked nucleic acid-modified oligonucleotide. *J. Mol. Diagn.* **12**, 512–519. <https://doi.org/10.2353/jmoldx.2010.090137> (2010).
- Ishige, T., Itoga, S. & Matsushita, K. Locked nucleic acid technology for highly sensitive detection of somatic mutations in cancer. *Adv. Clin. Chem.* **83**, 53–72. <https://doi.org/10.1016/bs.acc.2017.10.002> (2018).
- Dolinsek, J., Dorninger, C., Lagkouravdos, I., Wagner, M. & Daims, H. Depletion of unwanted nucleic acid templates by selective cleavage: LNAzymes, catalytically active oligonucleotides containing locked nucleic acids, open a new window for detecting rare microbial community members. *Appl. Environ. Microbiol.* **79**, 1534–1544. <https://doi.org/10.1128/AEM.03392-12> (2013).
- Hermans, N. *et al.* Toehold-enhanced LNA probes for selective pull down and single-molecule analysis of native chromatin. *Sci. Rep.* **7**, 16721. <https://doi.org/10.1038/s41598-017-16864-7> (2017).
- Hummelshoj, L., Ryder, L. P., Madsen, H. O. & Poulsen, L. K. Locked nucleic acid inhibits amplification of contaminating DNA in real-time PCR. *Biotechniques* **38**, 605–610. <https://doi.org/10.2144/05384RR01> (2005).
- Nie, H., Evans, A. A., London, W. T., Block, T. M. & Ren, X. D. Ultrasensitive quantification of hepatitis B virus A1762T/G1764A mutant by a SimpleProbe PCR using a wild-type-selective PCR blocker and a primer-blocker-probe partial-overlap approach. *J. Clin. Microbiol.* **49**, 2440–2448. <https://doi.org/10.1128/JCM.02472-10> (2011).
- Oldenburg, R. P., Liu, M. S. & Kolodney, M. S. Selective amplification of rare mutations using locked nucleic acid oligonucleotides that competitively inhibit primer binding to wild-type DNA. *J. Invest. Dermatol.* **128**, 398–402. <https://doi.org/10.1038/sj.jid.5700920> (2008).
- Sherman, E. *et al.* INSPIRED: A pipeline for quantitative analysis of sites of new DNA integration in cellular genomes. *Mol. Ther. Methods Clin. Dev.* **4**, 39–49. <https://doi.org/10.1016/j.omtm.2016.11.002> (2017).
- Vestheim, H. & Jarman, S. N. Blocking primers to enhance PCR amplification of rare sequences in mixed samples—A case study on prey DNA in Antarctic krill stomachs. *Front. Zool.* **5**, 12. <https://doi.org/10.1186/1742-9994-5-12> (2008).
- Taya, Y. *et al.* Applications of blocker nucleic acids and non-metazoan PCR improves the discovery of the eukaryotic microbiome in ticks. *Microorganisms*. <https://doi.org/10.3390/microorganisms9051051> (2021).
- Everaert, C. V. J. *et al.* Blocking unwanted sequences by target specific high-affinity binding oligonucleotides during transcriptome library preparation. *bioRxiv*. <https://doi.org/10.1101/2022.03.11.483910v1> (2022).
- Liu, C., Qi, R. J., Jiang, J. Z., Zhang, M. Q. & Wang, J. Y. Development of a blocking primer to inhibit the pcr amplification of the 18S rDNA sequences of *Litopenaeus vannamei* and its efficacy in *Crassostrea hongkongensis*. *Front. Microbiol.* **10**, 830. <https://doi.org/10.3389/fmicb.2019.00830> (2019).



29. Bruylants, G., Bocconcelli, M., Snoussi, K. & Bartik, K. Comparison of the thermodynamics and base-pair dynamics of a full LNA:DNA duplex and of the isosequential DNA:DNA duplex. *Biochemistry* **48**, 8473–8482. <https://doi.org/10.1021/bi900615z> (2009).
30. Hughesman, C., Fakhfakh, K., Bidshahri, R., Lund, H. L. & Haynes, C. A new general model for predicting melting thermodynamics of complementary and mismatched B-form duplexes containing locked nucleic acids: Application to probe design for digital PCR detection of somatic mutations. *Biochemistry* **54**, 1338–1352. <https://doi.org/10.1021/bi500905b> (2015).
31. Owczarzy, R., You, Y., Groth, C. L. & Tataurov, A. V. Stability and mismatch discrimination of locked nucleic acid-DNA duplexes. *Biochemistry* **50**, 9352–9367. <https://doi.org/10.1021/bi200904e> (2011).
32. McTigue, P. M., Peterson, R. J. & Kahn, J. D. Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochemistry* **43**, 5388–5405. <https://doi.org/10.1021/bi035976d> (2004).
33. Kaur, H., Arora, A., Wengel, J. & Maiti, S. Thermodynamic, counterion, and hydration effects for the incorporation of locked nucleic acid nucleotides into DNA duplexes. *Biochemistry* **45**, 7347–7355. <https://doi.org/10.1021/bi060307w> (2006).
34. Di Giusto, D. A. & King, G. C. Strong positional preference in the interaction of LNA oligonucleotides with DNA polymerase and proofreading exonuclease activities: Implications for genotyping assays. *Nucleic Acids Res.* **32**, e32. <https://doi.org/10.1093/nar/gnh036> (2004).
35. Levin, J. D., Fiala, D., Samala, M. F., Kahn, J. D. & Peterson, R. J. Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers. *Nucleic Acids Res.* **34**, e142. <https://doi.org/10.1093/nar/gkl756> (2006).
36. Hagedorn, P. H. *et al.* Locked nucleic acid: Modality, diversity, and drug discovery. *Drug Discov. Today* **23**, 101–114. <https://doi.org/10.1016/j.drudis.2017.09.018> (2018).
37. Brodyagin, N., Katkevics, M., Kotikam, V., Ryan, C. A. & Rozners, E. Chemical approaches to discover the full potential of peptide nucleic acids in biomedical applications. *Beilstein J. Org. Chem.* **17**, 1641–1688. <https://doi.org/10.3762/bjoc.17.116> (2021).
38. Fouz, M. F. & Appella, D. H. PNA clamping in nucleic acid amplification protocols to detect single nucleotide mutations related to cancer. *Molecules* <https://doi.org/10.3390/molecules25040786> (2020).
39. Vestheim, H., Deagle, B. E. & Jarman, S. N. Application of blocking oligonucleotides to improve signal-to-noise ratio in a PCR. *Methods Mol. Biol.* **687**, 265–274. [https://doi.org/10.1007/978-1-60761-944-4\\_19](https://doi.org/10.1007/978-1-60761-944-4_19) (2011).
40. Chen, H. M. *et al.* Molecular characterization of precise in vivo targeted gene integration in human cells using AAVHSC15. *PLoS One* **15**, e0233373. <https://doi.org/10.1371/journal.pone.0233373> (2020).
41. Sidon, P. *et al.* Combined locked nucleic acid and molecular beacon technologies for sensitive detection of the JAK2V617F somatic single-base sequence variant. *Clin. Chem.* **52**, 1436–1438. <https://doi.org/10.1373/clinchem.2006.066886> (2006).
42. Janciauskiene, S. & Welte, T. Well-known and less well-known functions of alpha-1 antitrypsin. Its role in chronic obstructive pulmonary disease and other disease developments. *Ann. Am. Thorac. Soc.* **13**(Suppl 4), S280–S288. <https://doi.org/10.1513/AnnalsATS.201507-468KV> (2016).
43. Pabon-Martinez, Y. V. *et al.* LNA effects on DNA binding and conformation: From single strand to duplex and triplex structures. *Sci. Rep.* **7**, 11043. <https://doi.org/10.1038/s41598-017-09147-8> (2017).
44. Xu, Y. *et al.* The ability of locked nucleic acid oligonucleotides to pre-structure the double helix: A molecular simulation and binding study. *PLoS One* **14**, e0211651. <https://doi.org/10.1371/journal.pone.0211651> (2019).
45. Lawyer, F. C. *et al.* High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl.* **2**, 275–287. <https://doi.org/10.1101/gr.2.4.275> (1993).
46. Christensen, U., Jacobsen, N., Rajwansi, V. K., Wengel, J. & Koch, T. Stopped-flow kinetics of locked nucleic acid (LNA)-oligonucleotide duplex formation: Studies of LNA-DNA and DNA-DNA interactions. *Biochem. J.* **354**, 481–484. <https://doi.org/10.1042/0264-6021:3540481> (2001).

## Author contributions

J.P., A.P., and M.T. carried out experimental work and assisted in experimental design, data analysis, and writing of the manuscript. J.W. assisted in experimental design and writing of the manuscript. J.E.T. assisted in experimental design, data analysis, and writing of the manuscript.

## Funding

This work was funded by Homology Medicines Inc.

## Competing interests

All authors are current or former employees of Homology Medicines.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-31871-7>.

**Correspondence** and requests for materials should be addressed to J.E.T.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023