Role of Pseudoisocytidine Tautomerization in Triplex-Forming Oligonucleotides

Hartono, Yossa Dwi; Pabon-Martinez, Y. Vladimir; Uyar, Arzu; Wengel, Jesper; Lundin, Karin E.; Zain, Rula; Smith, C. I.Edvard; Nilsson, Lennart; Villa, Alessandra

Published in:
ACS Omega

DOI:
10.1021/acsomega.7b00347

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Citation for pulished version (APA):
Hartono, Y. D., Pabon-Martinez, Y. V., Uyar, A., Wengel, J., Lundin, K. E., Zain, R., ... Villa, A. (2017). Role of Pseudoisocytidine Tautomerization in Triplex-Forming Oligonucleotides: In Silico and in Vitro Studies. ACS Omega, 2(5), 2165-2177. DOI: 10.1021/acsomega.7b00347

ACS OMEGA

Article

http://pubs.acs.org/journal/acsodf

Role of Pseudoisocytidine Tautomerization in Triplex-Forming Oligonucleotides: In Silico and in Vitro Studies

Yossa Dwi Hartono,[†,‡] Y. Vladimir Pabon-Martinez,[§] Arzu Uyar,[†,#] Jesper Wengel,[∥] Karin E. Lundin,[§] Rula Zain,[§,⊥] C. I. Edvard Smith,[§] Lennart Nilsson,[†] and Alessandra Villa*[,†]

[†]Department of Biosciences and Nutrition, Karolinska Institutet, SE-141 83 Huddinge, Sweden

[‡]Division of Structural Biology and Biochemistry, School of Biological Sciences, Nanyang Technological University, 60 Nanyang Drive, 637551, Singapore

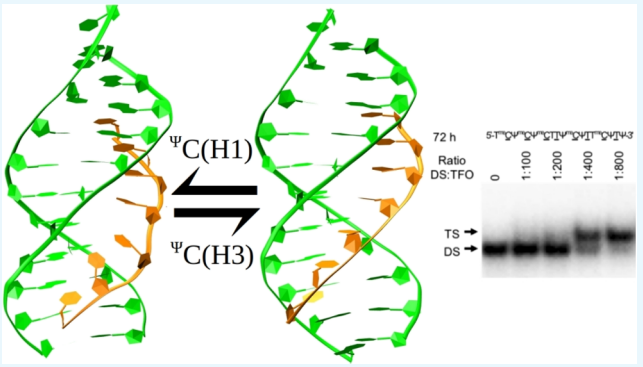
[§]Department of Laboratory Medicine, Clinical Research Center, Karolinska Institutet, SE-141 86 Huddinge, Sweden

[∥]Department of Physics, Chemistry and Pharmacy, Nucleic Acid Center, University of Southern Denmark, 5230 Odense M, Denmark

[⊥]Department of Clinical Genetics, Centre for Rare Diseases, Karolinska University Hospital, SE-171 76 Stockholm, Sweden

Supporting Information

ABSTRACT: Pseudoisocytidine ($^{\Psi}$C) is a synthetic cytidine analogue that can target DNA duplex to form parallel triplex at neutral pH. Pseudoisocytidine has mainly two tautomers, of which only one is favorable for triplex formation. In this study, we investigated the effect of sequence on $^{\Psi}$C tautomerization using λ-dynamics simulation, which takes into account transitions between states. We also performed in vitro binding experiments with sequences containing $^{\Psi}$C and furthermore characterized the structure of the formed triplex using molecular dynamics simulation. We found that the neighboring methylated or protonated cytidine promotes the formation of the favorable tautomer, whereas the neighboring thymine or locked nucleic acid has a poor effect, and consecutive $^{\Psi}$C has a negative influence. The deleterious effect of consecutive $^{\Psi}$C in a triplex formation was confirmed using in vitro binding experiments. Our findings contribute to improving the design of $^{\Psi}$C-containing triplex-forming oligonucleotides directed to target G-rich DNA sequences.



INTRODUCTION

The formation of DNA triple helices plays a key role in cellular processes[1] such as regulation of replication and transcription,[2−5] chromosome folding,[6] stabilization of telomeres, and recombination.[7] Triplex-forming oligonucleotides (TFOs) have been used in many biotechnological and biomedical applications that make use of their ability to target the major groove of a DNA duplex. Examples are isolation of specific DNA sequences (triplex affinity capture),[8−10] detection and capture of polymerase chain reaction products,[11] detection of DNA mutation,[12] and site-directed mutagenesis.[13,14] Triplexes can form in different ways: with purine (antiparallel orientation) or pyrimidine (parallel orientation) motifs.[15] In a parallel triplex, T•A−T and C^{+}•G−C base triads are formed ("−" refers to a Watson−Crick base pair and "•" refers to a Hoogsteen base pair).[16]

Triple-helix target sites in the human genome are abundant, especially in promoter regions.[17,18] TFOs targeting G-rich sequences are of biological importance because such regions are frequently present in promoters, which are potential targets for regulating transcription as an antigene strategy.[18] However, the formation of parallel triplexes is not favorable at physiological pH because it requires the protonation of cytosine (pK_a 4.1).[19] This limits the therapeutic application of TFOs as antigene strategy to regulate transcription, specifically when targeting G-rich sequences.[20,21]

Pseudoisocytidine ($^{\Psi}$C) is an artificial pyrimidine analogue that is derived from pseudouridine.[22] It has at least two relevant tautomers, $^{\Psi}$C(H1) and $^{\Psi}$C(H3), corresponding to the presence of a proton at N1 and N3, respectively (Figure 1). Tautomer $^{\Psi}$C(H1) has the hydrogen bond donor/acceptor set for the Watson−Crick hydrogen bonding scheme to guanine, whereas tautomer $^{\Psi}$C(H3) has the same set of hydrogen bond donors/acceptors as a protonated cytidine, which is favorable for Hoogsteen hydrogen bonding to guanine.[23] Tautomer $^{\Psi}$C(H3) is, thus, desirable in a TFO as a substitute for C to target G-rich sequences forming a pyrimidine-motif triplex. The substitution of cytidine by analogues such as $^{\Psi}$C is one strategy to target G-rich sequences at physiological pH.[22,24,25]

Received: March 23, 2017
Accepted: May 2, 2017
Published: May 17, 2017

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

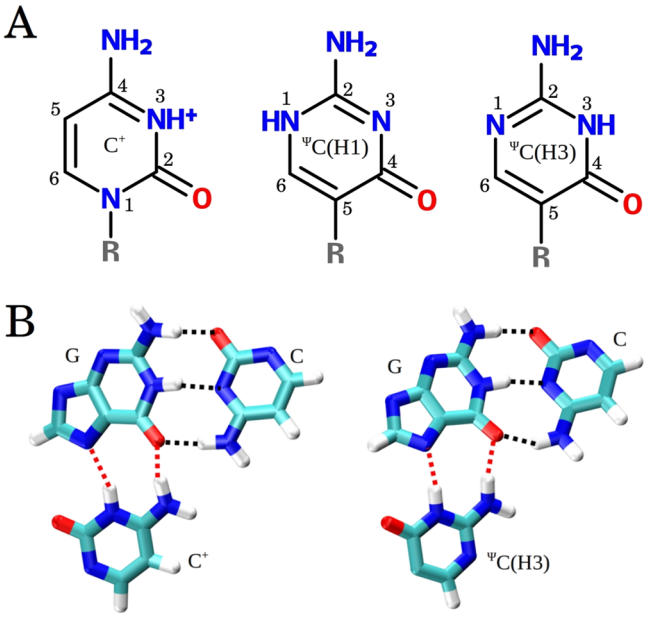


Figure 1. Cytidine and pseudoisocytidine. (A) Structures of protonated cytidine (C^+), pseudoisocytidine tautomers, $^{\Psi}C(H1)$ and $^{\Psi}C(H3)$, with the corresponding atom numbers. R corresponds to the sugar position. (B) Base triad configurations, C^+•G−C and $^{\Psi}C(H3)$ •G−C, with Watson−Crick (black) and Hoogsteen (red) hydrogen bonds.

In this study, we aim to understand the ways of optimizing the design of intermolecular TFOs targeting G-rich sequences by using $^{\Psi}C$ as a C analogue. To achieve this, we combined molecular dynamics (MD) simulations with in vitro binding experiments. In particular, we want to understand the effect of the environment (flanking nucleotides and bound/unbound states) on the tautomerization of $^{\Psi}C$. Experimental investigation of tautomerization is challenging because of the structural similarity, fast interconversion, and ambient aqueous condition.[26] Alternatively, molecular simulation methods, primarily λ-dynamics,[27] can be used to describe the change between tautomeric states. MD simulations have previously been successfully used to investigate DNA triple helices both in a parallel and an antiparallel fashion.[28−30] These studies show that the DNA double-helix overslides in the negative direction to increase the major groove and to accommodate the third strand, and the resultant triple helical conformation is somewhere between A- and B-types, with base pairs remaining almost perpendicular to the helical axis.

Electrophoretic mobility shift assay (EMSA) was used to detect in vitro binding of TFOs containing $^{\Psi}C$ under intranuclear conditions. The triplex intercalator, benzoquinoquinoxaline (BQQ), was used to probe for triplex formation.[31,32] Pseudoisocytidines were incorporated both consecutively and nonconsecutively in the TFO sequence and combined for the first time with sugar-modified nucleotides, locked nucleic acids (LNAs). LNA has an oxy-methylene bridge locking the sugar pucker in C3′-endo rather than in C2′-endo as in DNA, which restricts the conformation of the sugar. The inclusion of LNAs in TFOs has been shown to increase triplex stability.[33] As a target sequence, we used an upstream G-rich region of the human trefoil factor (human TFF) gene close to an estrogen response element (ERE), including runs of consecutive guanines.[34] This sequence was previously studied using bisLNAs, which contain a C-rich TFO part and were found to be a poor target at neutral pH.[35]

First, we discuss the result from λ-dynamics simulations for short single-stranded and triplex DNA (trimers and 7-mers) containing $^{\Psi}C$ in various positions and sequence contexts. Then, we compare the observed sequence effect with the result from in vitro binding experiments of six 17-mer TFOs containing $^{\Psi}C$. Finally, we characterize the structure of the observed triplexes using classical MD.

RESULTS AND DISCUSSION

Pyrimidine motif triplexes are unstable at physiological pH because of the need for protonation of C in the third strand. A way to solve this problem is to use C analogues such as $^{\Psi}C$. Because the aim of this study is to identify rules on how to incorporate $^{\Psi}C$ in LNA containing oligonucleotides (ONs) for optimal hybridization under intracellular conditions, we started by performing simulation studies, specifically investigating how the surrounding bases might influence the tautomerization of $^{\Psi}C$.

Nucleoside Tautomerization: Reference State. The exact tautomeric ratio of $^{\Psi}C$ in an aqueous solution is not known. In the crystal structure, the tautomeric ratio $^{\Psi}C(H1)/^{\Psi}C(H3)$ of the isocytosine base is exactly 1:1.[36] ^1H nuclear magnetic resonance (NMR) spectra did not show separate signals for the two $^{\Psi}C$ tautomers in the aqueous solution, but enzymatic incorporation experiments confirmed the existence of the two tautomers in the solution for the deoxyribonucleoside $^{\Psi}C$.[37] The measured pK_{a1} and pK_{a2} at the two protonation sites corresponding to the two tautomeric states are highly similar [pK_{a1} 3.79 and 3.69 and pK_{a2} 9.36 and 9.42, for tautomers $^{\Psi}C(H1)$ and $^{\Psi}C(H3)$, respectively], and in the same study, ab initio calculation (Hartree−Fock) for the methylated $^{\Psi}C$ base indicates that tautomer $^{\Psi}C(H3)$ is favored over tautomer $^{\Psi}C(H1)$ in a vacuum, and less so when the solvent effect is accounted for using the polarizable continuum model.[38]

Taking these into account, we decided to set the tautomeric ratio of the model system, deoxyribonucleoside $^{\Psi}C$, to be 1:1 as the reference, where only the physical end states were considered. All variations observed in the tautomeric ratio should be interpreted as relative to the reference and not as absolute values. Henceforth, we will call this quantity tautomeric propensity to reflect on this point.

To set the tautomeric ratio to 1:1 in the λ-dynamics formulation, we supply a biasing potential exactly equal to the calculated free energy, $\Delta G_{H1\rightarrow H3}$, for the model system in water, 28.9 ± 0.1 kcal/mol. This results in an average population ratio of 1:1 for deoxyribonucleoside $^{\Psi}C$. To guarantee an optimal transition between the two tautomeric states, we calibrate k_{bias} value to yield a high fraction of physical end states in the trajectory and a high frequency of transitions between the two tautomeric states. By calibrating with a set of 1 ns λ-dynamics runs of the model system at various k_{bias} values, we found k_{bias} = 19.5 kcal/mol to result in >80% physical states and >60 transitions/ns, which we judged to be sufficient to ensure good sampling in simulations of length 1−10 ns (Figure 2). In practice, when this k_{bias} value is applied to other systems containing $^{\Psi}C$, the transition rate is 40−50 ns^{-1}, and more than 80% of the population is physical when there is only one $^{\Psi}C$. For two $^{\Psi}Cs$, a fraction of 60−70% is physical, whereas for three $^{\Psi}Cs$, 50−60% is physical. For an illustration of the

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

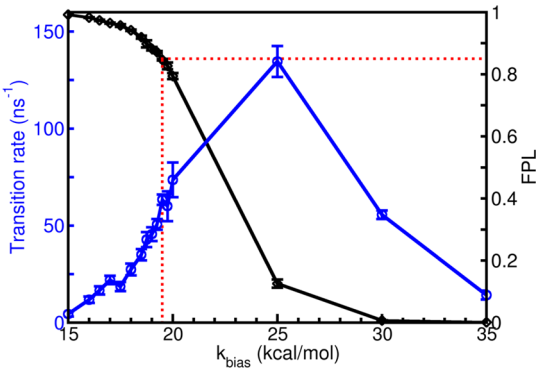


Figure 2. Optimization of k_{bias} value. A k_{bias} value of 19.5 kcal/mol (red) is chosen to maintain the fraction of physical states [fraction physical ligand (FPL), black] above 0.8, while maintaining a moderately high transition rate (blue). The error bar is the standard error of the mean of five independent runs.

fluctuation of λ values at this transition rate and the fraction of physical state, see Figure 3.

Five independent λ-dynamics simulations were performed for nucleoside $^{\Psi}$C (reference compound) and nucleotide $^{\Psi}$C. The addition of a 5′ monophosphate group slightly shifts the tautomeric propensity to favor tautomer $^{\Psi}$C(H1) [from 52 to 41% $^{\Psi}$C(H3)].

Effect of the Neighboring Bases on the Tautomerization of Pseudoisocytidine. We performed λ-dynamics simulations in single-stranded DNA trimers and 7-mers with varying sequences to investigate the effect of neighboring bases on the tautomerization of $^{\Psi}$C in a single strand. Besides DNA bases thymine (T) and cytidine (C) as neighboring residues, we also included protonated and/or 5-methylated C (^{me}C) and restricted sugar moiety (LNA, denoted by underline).

Summarizing Figure 4, we observe that $^{\Psi}$C is 100% $^{\Psi}$C(H3) in the triplex structures, and in the single-stranded ONs, we found the following $^{\Psi}$C(H3) propensities (for the $^{\Psi}$C indicated in **bold**):

- >75%, C$^{+\Psi}$**C**C^{+}, ^{me}C$\underline{}^{\Psi}$**C**T$\underline{}$, T$\underline{}^{\Psi}$**C**^{me}C$\underline{}$, T$\underline{}^{\Psi}$**C**T$\underline{}$, ^{me}C$\underline{}^{\Psi}$**C**^{me}C$\underline{}$, ^{me}C$\underline{}^{\Psi}$**C**^{me}C
- 60−75%, C$^{\Psi}$**C**C, ^{me}C$^{+\Psi}$**C**^{me}C^{+}, C$\underline{}^{+\Psi}$**C**T$\underline{}$, C$\underline{}^{\Psi}$**C**T$\underline{}$, ^{me}C$\underline{}^{+\Psi}$**C**T$\underline{}$, T$\underline{}^{\Psi}$**C**C$\underline{}^{+}$, T$\underline{}^{\Psi}$**C**C$\underline{}$, $^{\Psi}$**C**TT, $^{\Psi}$**C**$^{\Psi}$CT, $^{\Psi}$**C**$^{\Psi}$CT$\underline{}$, $^{\Psi}$**C**$^{\Psi}$C$^{\Psi}$C, TT$^{\Psi}$**C**$^{\Psi}$C$^{\Psi}$CTT, T$^{\Psi}$**C**TTT$^{\Psi}$CT
- 40−59%, T$\underline{}^{\Psi}$**C**^{me}C$\underline{}^{+}$, T$^{\Psi}$**C**T, $^{\Psi}$C$^{\Psi}$**C**T, $^{\Psi}$C$^{\Psi}$**C**T$\underline{}$, $^{\Psi}$C$^{\Psi}$**C**$^{\Psi}$C, TT$^{\Psi}$**C**T$^{\Psi}$CTT, TT$^{\Psi}$**C**TT$^{\Psi}$CT, TT$\underline{}^{\Psi}$**C**^{me}C$\underline{}^{+\Psi}$**C**TT
- 25−39%, TT$^{\Psi}$CTT$^{\Psi}$**C**T, T$^{\Psi}$CTTT$^{\Psi}$**C**T
- <25%, TT$^{\Psi}$**C**, $^{\Psi}$C$^{\Psi}$C$^{\Psi}$**C**, TT$^{\Psi}$C$^{\Psi}$**C**$^{\Psi}$CTT, TT$^{\Psi}$CT$^{\Psi}$CTT, TT$\underline{}^{\Psi}$**C**^{me}C$\underline{}^{+\Psi}$CTT

With a T or C on either side, the propensity slightly shifts to favor tautomer $^{\Psi}$C(H3). Methylated C or protonated C neighbors also shift the propensity to favor tautomer $^{\Psi}$C(H3), but less so when they are both methylated and protonated. Analysis of the base−base interaction energies revealed that $^{\Psi}$C(H3) has favorable electrostatic interactions with a methylated or protonated C neighbor on its 3′-side (Figure S1).

Protonation of an unmethylated LNA-C (C$\underline{}$) neighbor offers little to no improvement in $^{\Psi}$C(H3) propensity, and protonation of ^{me}C$\underline{}$ disfavors tautomer $^{\Psi}$C(H3). We would like to reiterate here that the normal commercial version of LNA-C is always methylated (^{me}C$\underline{}$), and we include unmethylated LNA-C in our computational study to delineate the contributions of methylation and sugar locking.

The 7-mer TT$\underline{}^{\Psi}$C^{me}C$\underline{}^{+\Psi}$CTT, which contains two $^{\Psi}$Cs next to ^{me}C$\underline{}$, disfavors $^{\Psi}$C(H3) compared with TT$^{\Psi}$CT$^{\Psi}$CTT, but notably the position trend is reversed: the 5′ $^{\Psi}$C in TT$^{\Psi}$CT$^{\Psi}$CTT favors $^{\Psi}$C(H3) more, as generally observed in other systems (vide infra), but not in TT$\underline{}^{\Psi}$C^{me}C$\underline{}^{+\Psi}$CTT where the 3′ $^{\Psi}$C favors $^{\Psi}$C(H3). However, by itself, locked sugars in the neighboring LNA residues have a modest to no effect on the tautomeric propensity. There is a modest improvement going from T$^{\Psi}$CT to T$\underline{}^{\Psi}$CT$\underline{}$, but little to none between ^{me}C$^{\Psi}$C^{me}C and ^{me}C$\underline{}^{\Psi}$C^{me}C$\underline{}$ or $^{\Psi}$C$^{\Psi}$CT and $^{\Psi}$C$^{\Psi}$CT$\underline{}$.

When there is more than a single $^{\Psi}$C residue in the system, their tautomeric states do not appear to strongly correlate with each other. In the trimer $^{\Psi}$C$^{\Psi}$CT, when the first $^{\Psi}$C is the $^{\Psi}$C(H1) tautomer, the second $^{\Psi}$C has similar tendencies to be $^{\Psi}$C(H1) or $^{\Psi}$C(H3); and the same is observed when the first $^{\Psi}$C is $^{\Psi}$C(H3) (Table 1). However, when there are more residues in between, the $^{\Psi}$C nearer to the 3′-end always favors tautomer $^{\Psi}$C(H1), as observed in 7-mers TT$^{\Psi}$CT$^{\Psi}$CTT, TT$^{\Psi}$CTT$^{\Psi}$CT, and T$^{\Psi}$CTTT$^{\Psi}$CT. In $^{\Psi}$C$^{\Psi}$C$^{\Psi}$C and TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT, when the first $^{\Psi}$C is $^{\Psi}$C(H3), both the second and third $^{\Psi}$Cs tend to be $^{\Psi}$C(H1).

The position of $^{\Psi}$C in the sequence has a large effect on the tautomeric propensity. When $^{\Psi}$C is at the 5′-end in $^{\Psi}$CTT, the tautomeric propensity is 63% $^{\Psi}$C(H3), whereas when $^{\Psi}$C is at the 3′-end in TT$^{\Psi}$C, it is 24%. This position effect can also be clearly observed in 7-mer T$^{\Psi}$CTTT$^{\Psi}$CT, where both $^{\Psi}$Cs are flanked by T, but their propensities are vastly different [63 and 28% $^{\Psi}$C(H3), respectively]. We observed that when $^{\Psi}$C is positioned toward the 3′-end, it often forms intramolecular

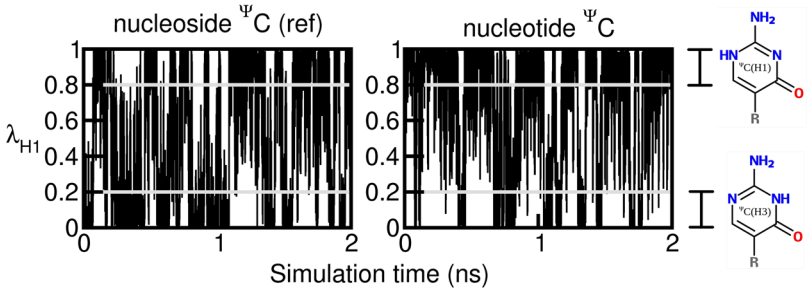


Figure 3. Transition of λ_{H1} values in one run each of nucleoside $^{\Psi}$C (model system used as the reference; henceforth "ref") and nucleotide $^{\Psi}$C. Only $0.8 \leq \lambda_{H1} \leq 1$ is counted as the physical state of tautomer $^{\Psi}$C(H1), whereas $0 \leq \lambda_{H1} \leq 0.2$ would be counted as the physical state of tautomer $^{\Psi}$C(H3). The base structures of the corresponding tautomers are shown. The transition rates are 60 and 40 ns^{-1}, and the fractions of physical states are 83 and 84%.

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

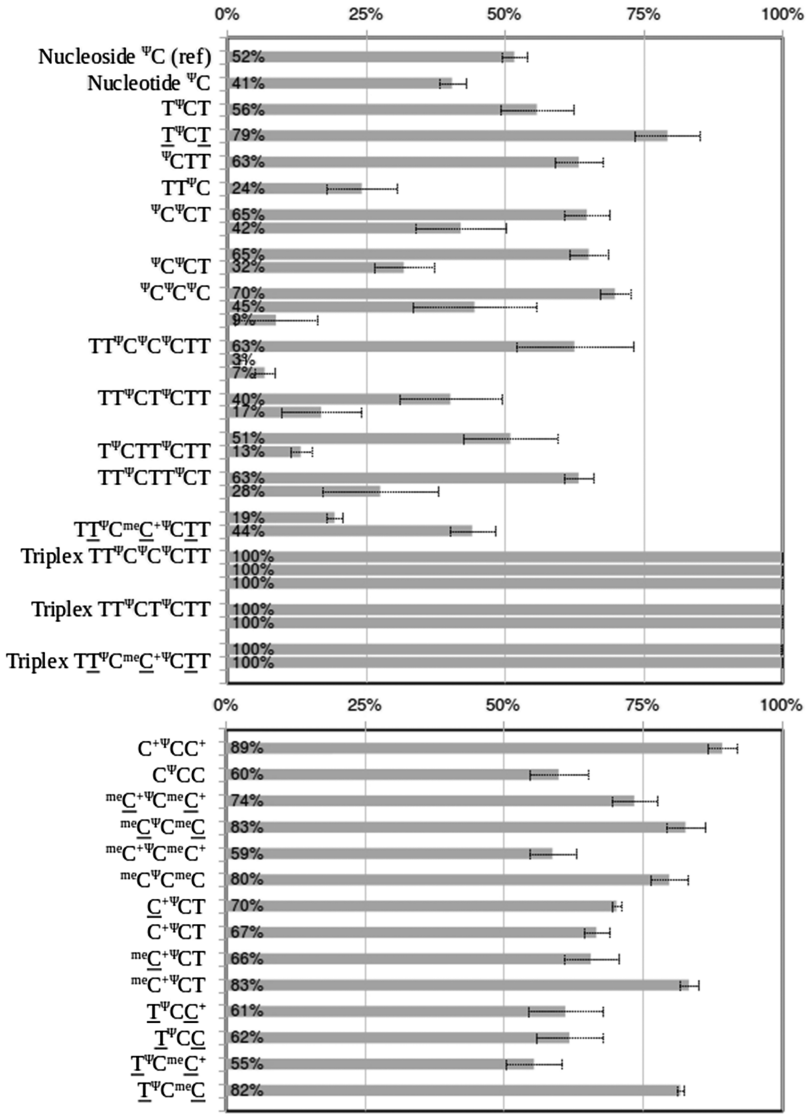


Figure 4. Tautomeric propensity [given in terms of % tautomer ΨC(H3)] of pseudoisocytidine in different systems. ΨC is pseudoisocytidine; meC is 5-methylcytosine; + indicates protonation; underline denotes residues with locked sugar (LNA). When there are multiple ΨCs, data for the ΨC at the nearest 5′-end are presented first. The error bar is the standard error of mean of five independent runs.

hydrogen bonds with the preceding residues. Notably, the hydrogen bonding analyses of trimers and 7-mers show that H1 is much more frequently involved in intramolecular hydrogen bonding compared with N1, N3, and H3, and it is, to a large extent, correlated with the appearance of tautomer ΨC(H1). We select two examples from one run of trimers TΨCT and TTΨC to show such correlation (Figure 5). The position effect can thus be explained in terms of intramolecular hydrogen bonds: when ΨC is positioned toward the 3′-end, it has more available hydrogen bonding partners for H1, favoring the formation of associated tautomer ΨC(H1).

More detailed analyses were undertaken for 7-mer TTΨCΨCΨCTT, which is a fragment of 17-mer TFO5-DNALNAΨC used in the triplex formation experiments. For this 7-mer sequence, we performed another set of conventional MD simulations, fixing the tautomeric states to be the most populated one (combination 311: 59%, Table 1) to exclude artifacts from the dual topology on hydrogen bond and solvent-accessible surface area (SASA) analyses.

The two protonation sites associated with the two tautomers are in similar chemical environments, except for O2 near N3/H3. In 7-mer TTΨCΨCΨCTT (311), residues ΨC4 and ΨC5 have some intramolecular hydrogen bonds involving H1, whereas H3 in ΨC3 has no such hydrogen bonds (Table 2). Notably, the position effect can be observed here: on average, ΨC5 nearer to the 3′-end has its H1 involved in more intramolecular hydrogen bonds than ΨC4, which is nearer to the 5′-end (Table 2).

The SASA for the N1 atom of Ψ in 7-mer TTΨΨΨTT (311) is lower for Ψ4 and Ψ5 than for Ψ3, even when considering the presence of H1 in Ψ4 and Ψ5, whereas for the N3 atom, the SASA is very similar for Ψ3 and Ψ4 even though Ψ3 has H3 present and Ψ4 does not (Figure S2). This is consistent with the observation that Ψ(H1) tends to form intramolecular hydrogen bonds; thus, it tends to be less exposed to the solvent.

Consecutive ΨC lowers ΨC(H3) propensities, except for the first residue at the 5′-end. In ΨCΨCT, the first ΨC has a moderate propensity for ΨC(H3) [65% ΨC(H3)], but the second favors ΨC(H1) instead [42% ΨC(H3)]. The third

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

Table 1. Average Population (in %) of Tautomer Combinations with Standard Error of Mean in Five Independent Runs[a]

2 $^{\Psi}$C								
tautomer combination	$^{\Psi}$C$^{\Psi}$CT	$^{\Psi}$C$^{\Psi}$CT	TT$^{\Psi}$CT$^{\Psi}$CTT	TT$^{\Psi}$CTT$^{\Psi}$CT	T$^{\Psi}$CTTT$^{\Psi}$CT	TT$^{\Psi}$C^{me}C^{+}$^{\Psi}$CTT	triplex TT$^{\Psi}$CT$^{\Psi}$CTT	triplex TT$^{\Psi}$C^{me}C^{+}$^{\Psi}$CTT
11	18 (2)	19 (1)	51 (9)	43 (7)	26 (4)	45 (4)	0 (0)	0 (0)
13	15 (3)	14 (3)	9 (3)	7 (1)	10 (3)	36 (5)	0 (0)	0 (0)
31	39 (7)	48 (5)	33 (9)	43 (6)	46 (6)	12 (2)	0 (0)	0 (0)
33	27 (5)	18 (4)	8 (4)	7 (2)	18 (7)	8 (1)	100 (0)	100 (0)

3 $^{\Psi}$C			
tautomer combination	$^{\Psi}$C$^{\Psi}$C$^{\Psi}$C	TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT	triplex TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT
111	11 (3)	30 (7)	0 (0)
113	1 (1)	3 (1)	0 (0)
131	15 (2)	2 (1)	0 (0)
133	2 (1)	0 (0)	0 (0)
311	40 (9)	59 (7)	0 (0)
313	2 (2)	4 (1)	0 (0)
331	25 (7)	2 (1)	0 (0)
333	4 (3)	0 (0)	100 (0)

[a]Tautomer combination is shown in shorthand; for example, tautomer combination 31 means that the first $^{\Psi}$C is tautomer $^{\Psi}$C(H3) and the second is $^{\Psi}$C(H1).

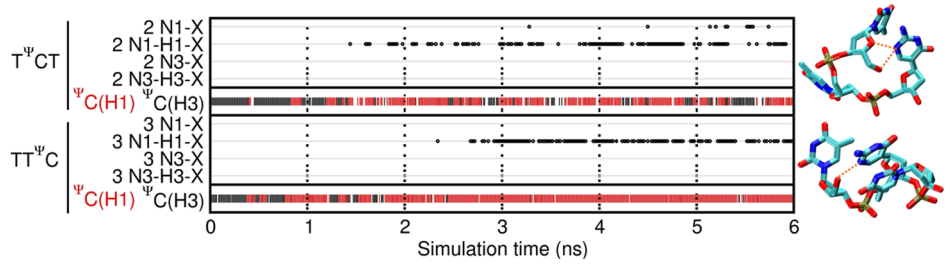


Figure 5. Intramolecular hydrogen bonding and tautomeric states of one run of T$^{\Psi}$CT and TT$^{\Psi}$C. Hydrogen bond label denotes hydrogen bond pairs; for example, (2 N1-X) refers to the intramolecular hydrogen bond involving N1 of residue index 2; X is any intramolecular hydrogen acceptor or donor. Only the physical states of tautomers $^{\Psi}$C(H1) (red) and $^{\Psi}$C(H3) (black) are shown. Snapshots of the two trimers at 4 ns are shown with the intramolecular hydrogen bonds (orange); hydrogen atoms are not shown for clarity.

Table 2. Intramolecular Hydrogen Bond Occupancies of 7-mer TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT Fixed Tautomer 311[a]

	1	2	3	4	5
3 N1-X			0.3		
3 N3-X					
3 N3-H3-X					
4 N1-X					
4 N1-H1-X	0.8		0.1	0.3	
4 N3-X					
5 N1-X					
5 N1-H1-X		0.3	0.4	0.8	0.2
5 N3-X					

[a]Only the hydrogen bonds involving N1, H1, N3, and H3 of the three $^{\Psi}$C residues are shown. The row label is hydrogen bond pairs; for example, (3 N1-X) refers to intramolecular hydrogen bond involving N1 of residue index 3; X is any intramolecular hydrogen acceptor or donor. The column label is the run index of five independent runs. Blank refers to zero hydrogen bond occupancy.

consecutive $^{\Psi}$C has an even more pronounced shift: in $^{\Psi}$C$^{\Psi}$C$^{\Psi}$C, 3′ $^{\Psi}$C has a propensity of only 9% $^{\Psi}$C(H3). Likewise, in 7-mer TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT, the last $^{\Psi}$C has a similar propensity [7% $^{\Psi}$C(H3)], and in addition, the middle $^{\Psi}$C also significantly shifts [3% $^{\Psi}$C(H3)].

The low propensities for $^{\Psi}$C(H3) of consecutive $^{\Psi}$C are deleterious for triplex formation. Not only does the 3′ $^{\Psi}$C have a low propensity for $^{\Psi}$C(H3), but the population of favorable tautomer combinations [all $^{\Psi}$C(H3)] is extremely low. In the trimer $^{\Psi}$C$^{\Psi}$C$^{\Psi}$C, the all-$^{\Psi}$C(H3) (333) population is only 4%; and in 7-mer TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT, it is 0% (Table 1). However, when the 7-mer is in a triplex, the all-$^{\Psi}$C(H3) population becomes 100%. When the middle $^{\Psi}$C in the triplex TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT is substituted with T so that the $^{\Psi}$Cs are no longer consecutive, as in the triplex TT$^{\Psi}$CT$^{\Psi}$CTT, or substituted with ^{me}C^{+} and introducing LNA neighbors, as in triplex T<u>T</u>$^{\Psi}$C^{me}<u>C</u>$^{+}$$^{\Psi}$C<u>T</u><u>T</u>, the all-$^{\Psi}$C(H3) population is, as expected, 100%. This suggests that Hoogsteen hydrogen bonding in the triplex is strong enough to shift the tautomeric propensity to favor $^{\Psi}$C(H3) and confer thermodynamic stability. The low all-$^{\Psi}$C(H3) population in single-stranded systems is of concern because there is only a small amount of the "correct" population, that is, with all-$^{\Psi}$C(H3), that can bind to the duplex, which may result in slow kinetics of binding.

To characterize the behavior of $^{\Psi}$C(H1) in the triplex environment, we performed classical MD simulations of triplex TT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT tautomer combinations 133 and 331. Although $^{\Psi}$C(H3) forms Hoogsteen hydrogen bonds with G, $^{\Psi}$C(H1) partially flips out and interacts with N7 or 5′ phosphate of G instead (Figure 6). The residue $^{\Psi}$C(H1) is not observed to completely flip out, and the triplexes stay mostly stable during 100 ns simulations (Watson−Crick and Hoogsteen hydrogen bonds during the simulations are shown in Figure S3). The

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

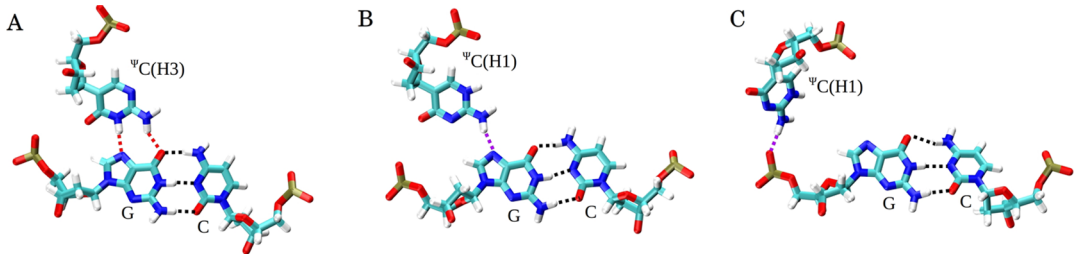


Figure 6. Observed configurations when ΨC is in TFO. (A) Canonical configuration when ΨC(H3) is in TFO with Watson–Crick (black) and Hoogsteen (red) hydrogen bonds. (B,C) ΨC(H1) partially flips out and interacts with N7 or 5′ phosphate of G (purple).

average structures show helical distortions around ΨC(H1) (Figure 7).

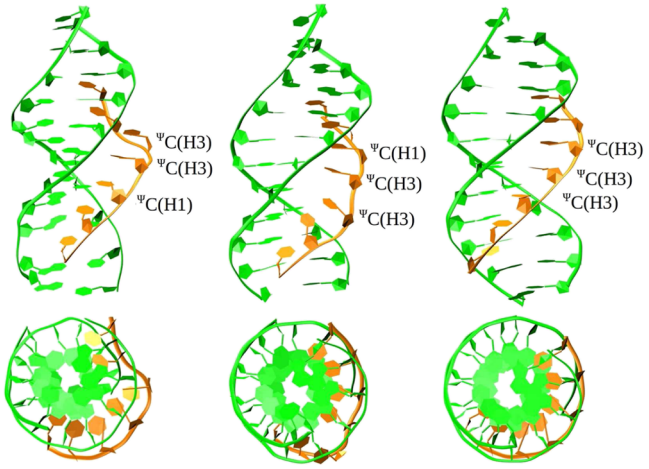


Figure 7. Classical MD simulations of triplexes TTΨCΨCΨCTT tautomer combinations 331, 133, and 333. Average structures from the last 50 ns of the 100 ns simulation are shown in side and top views (duplex in green and TFO in orange).

In summary, from our simulations, we have found that the neighboring residues have different effects on ΨC tautomerization. Methylated or protonated C shifts the tautomeric propensity to favor ΨC(H3); T or LNA neighbors do not affect the tautomerization equilibrium directly; ΨC itself as a neighbor affects the tautomeric propensity to disfavor ΨC(H3), which is not desirable in the context of TFO binding in triplex formation.

Verifying the Effect of Consecutive and Nonconsecutive ΨC in TFOs for in Vitro Binding. Getting a stable TFO formation in vitro requires longer TFO sequences than the 7-mer TFO used in the simulations. To verify the effect of consecutive ΨC-residues, we thus designed TFOs as 17-mers, targeting a region in the human *TFF* gene close to an ERE. This target is a good candidate for in vitro studies of the influence of ΨC in 17-mer TFOs because it contains a majority of Gs, including stretches of consecutive Gs 5′-AGGGG-GAAGGGAAGGAG-3′.[34] We decided to evaluate TFOs with this size because previous in vitro studies performed with 13-mer TFOs containing ΨC bases did not show any TFO binding (unpublished experiments). Each TFO was hybridized with the double-stranded (DS) target for a period of up to 72 h at pH 7.4, and the triplex formation was analyzed using EMSA.

Pseudoisocytidines were located in a consecutive or nonconsecutive manner in the TFOs. Three different stretches of two, three, and five consecutive ΨCs were present in the sequences of TFO1-DNAfullΨC, TFO2-DNAfullΨC-TINA, TFO3-DNALNAfullΨC, and TFO4-DNALNAfullΨC-TINA, where TINA denotes twisted intercalating nucleic acid. One or two thymines (DNA or LNA) were spaced between them, and all of these sequences contained a ΨC at the 3′-end ultimate position.

Initially, the DNA containing TFO1-DNAfullΨC was evaluated. In this sequence, all Cs were substituted by ΨC. After 72 h of incubation, no triplex formation was detected, even in the presence of the triplex-stabilizing BQQ compound

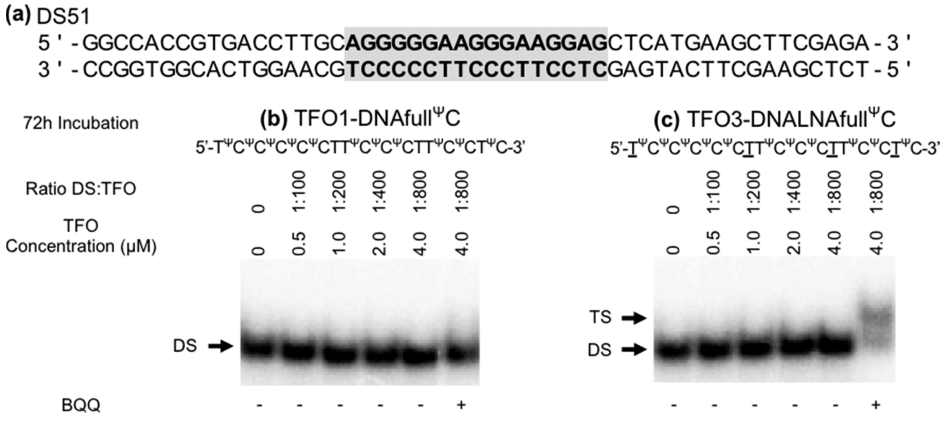


Figure 8. TFO binding of 17-mer TFO sequences containing consecutive ΨC: (a) DS51 and electrophoretic mobility shift profile of DS51 in the presence of (b) TFO1-DNAfullΨC and (c) TFO3-DNALNAfullΨC, both with 11/17 nucleotides being ΨC and in TFO3 4/6 Ts being LNA Ts. Hybridization with TFO, in the absence of and (as indicated only at the highest ratio) in the presence of BQQ, was carried out for 72 h. Triplex structures are detected as slower migrating bands. DNA duplex and triplex complexes are indicated as DS and TS, respectively.

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165–2177

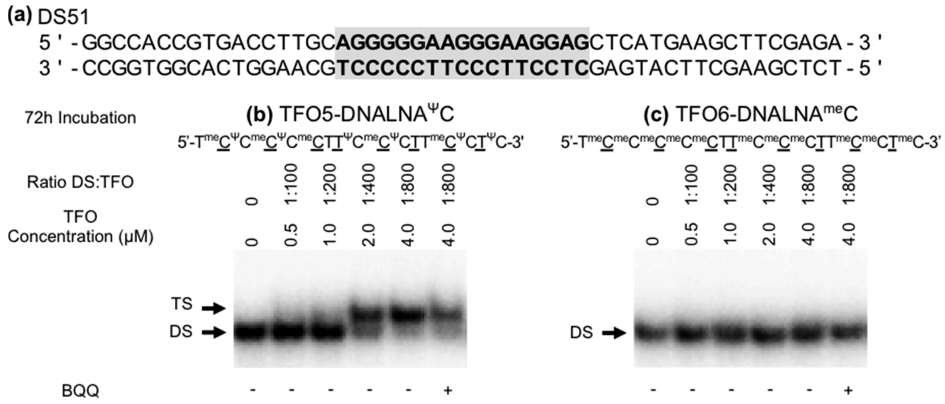


Figure 9. TFO binding of 17-mer TFO sequences containing nonconsecutive $^{\Psi}$C and ^{me}C: (a) DS51 and electrophoretic mobility shift profile of DS51 in the presence of (b) TFO5-DNALNA$^{\Psi}$C and (c) TFO6-DNALNA^{me}C. Hybridization with TFO in the absence (left side) and in the presence (right side) of BQQ was carried out for 72 h. Triplex structures are detected as slower migrating bands. DNA duplex and triplex complexes are indicated as DS and TS, respectively.

(Figure 8). These results confirm our simulations, where we show that $^{\Psi}$C itself as a neighbor affects the triplex formation because of the tautomeric propensity to disfavor $^{\Psi}$C(H3), the desirable tautomer for triplex formation.

Aiming to improve the triplex formation, LNA was included in the TFOs. LNA containing ONs have been shown to improve TFO binding and enhance triplex stability.[33] Thus, a TFO with a similar $^{\Psi}$C distribution as in TFO1-DNAfull$^{\Psi}$C, but including four insertions of LNA T (TFO3-DNALNA-full$^{\Psi}$C), was also evaluated. The presence of LNA combined with $^{\Psi}$C improved the TFO binding but could only be visualized in the presence of BQQ. Moreover, a triplex was only detected at the highest DS/TFO ratio of 1:800, and 100% of triplex formation was never achieved (Figure 8).

To further enhance the TFO binding, a TINA was included at the penultimate 3′-end position of the TFOs. TINA is an intercalator inserted covalently into the TFO[39] and is able to increase the thermal stability of parallel triplexes.[40] The presence of a TINA in the 3′-end of T-rich TFOs has previously been shown to strongly promote the triplex formation at low TFO/DS ratios (Pabon, et al. unpublished result). Thus, TINA was included in the sequence for TFO1 and TFO3 to create TFO2-DNAfull$^{\Psi}$C-TINA and TFO4-DNALNAfull$^{\Psi}$C-TINA, respectively. However, none of these new TFOs showed any improvement compared with the sequences without TINA (Figure S4).

To examine the effect of several consecutive $^{\Psi}$Cs on TFO binding, we designed two ONs with six nonconsecutive $^{\Psi}$Cs. TFO7-DNA$^{\Psi}$C contains three different combinations with $^{\Psi}$C: $^{\Psi}$CCC$^{\Psi}$C, $^{\Psi}$CC$^{\Psi}$C, and $^{\Psi}$CC. Triplex formation was evaluated after 72 h of binding. Our results show that TFO7-DNA$^{\Psi}$C was not able to form a triplex even at the highest concentration of TFO and in the presence of BQQ (Figure S5). The other TFO lacking consecutive $^{\Psi}$Cs, TFO5-DNALNA$^{\Psi}$C, contains the combinations ^{me}C$^{\Psi}$C^{me}C$^{\Psi}$C^{me}C, $^{\Psi}$C^{me}C$^{\Psi}$C, and ^{me}C$^{\Psi}$C and has eight LNA substitutions (three Ts and five ^{me}Cs). At pH 7.4 and at a DS/TFO ratio of 1:400, a shifted band was visible, and at the highest ratio of 1:800, approximately 90% of the triplex formation was achieved (Figure 9). TFO5-DNALNA$^{\Psi}$C was also evaluated at a lower pH (6.0) in a 2-morpholinoethanesulfonic acid (MES) buffer containing the same salt conditions as that of the intranuclear buffer. In comparison with the results at pH 7.4, a shifted band was observed at the DS/TFO ratio of 1:100 in the absence of BQQ. In the presence of BQQ, triplex formation was observed at the DS/TFO ratio of 1:25 (Figure S6). Thus, TFO5-DNALNA$^{\Psi}$C was the only ON showing triplex formation at pH 7.4 under intranuclear salt conditions at a DS/TFO ratio of 1:400 and in the absence of BQQ. This result shows again that LNA improves triplex formation, but it also confirms the conclusion from the simulation experiments that nonconsecutive $^{\Psi}$Cs are the best option to include $^{\Psi}$C in the TFO sequence.

Pyrimidine triplexes formed by the base triplet C^{+}•G−C are pH-dependent. TFOs containing C form stable triplexes under acidic conditions but are in contrast to G- and T-containing TFOs and are less active at physiological pH.[41] Several C analogues have been designed to overcome the requirement of acidic pH; one of them is $^{\Psi}$C. Our TFOs with different combinations of $^{\Psi}$C and another C analogue (^{me}C) and including LNAs address the possibility to target highly C-rich TFOs against sites with several runs of consecutive Gs. Methylated C (^{me}C) has been used to improve pyrimidine TFO binding at neutral pH, forming triplex structures.[42] Here, we also evaluate a TFO6-DNALNA^{me}C that contains ^{me}C instead of $^{\Psi}$C to compare with TFO5-DNALNA$^{\Psi}$C. TFO6-DNALNA^{me}C did not show any triplex formation even at the highest concentration of DS/TFO ratio (1:800) and in the presence of BQQ (Figure 7). Collectively, this shows an enhanced TFO binding when combining LNA with nonconsecutive $^{\Psi}$Cs, also for a TFO targeting a G-rich site.

The observation that TFO6-DNALNA^{me}C did not show any triplex formation agrees with previous studies where triplex formation is disfavored with consecutive C^{+}•G−C triplets[43] because of repulsion between the positive charges from the protonation at N3 of the Hoogsteen C[44] and the competition effect between the Cs in the adjacent C^{+}•G−C.[45]

Pseudoisocytidine has previously been reported to reduce the pH sensitivity in TFOs. Shahid et al. tested different pyrimidine DNA-TFOs against a 21-base target with only a single 4-base C-run,[25] demonstrating that at pH 7.2 and in the presence of 5 mM $MgCl_2$, alternating $^{\Psi}$C with ^{me}C gave the highest triplex stability as determined by the melting experiment and by gel-shift assays at a ratio of 1:500. An 8-mer DNA-TFO containing two $^{\Psi}$Cs was also shown to form a triplex at pH 7.0, whereas the corresponding all-DNA ON containing C or ^{me}C at the same two positions does not.[22] Still, at pH 7.0, the proton concentration is 2.5 times lower than what is found inside of the cell. In intramolecular triplexes, Chin et al. observed

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

Table 3. TFO Sequence and Triplex Formation under Intranuclear Conditions[a]

Name	Evidence of triplex formation in EMSA assay (-BQQ)	Sequence
TFO1-DNAfull$^{\Psi}$C	No	5 ' - T $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C T T $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C T T $^{\Psi}$C $^{\Psi}$C T $^{\Psi}$C - 3 '
TFO3-DNALNAfull$^{\Psi}$C	No	5 ' - T $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C T T $^{\Psi}$C $^{\Psi}$C $^{\Psi}$C T T $^{\Psi}$C $^{\Psi}$C T $^{\Psi}$C - 3 '
TFO5-DNALNA$^{\Psi}$C	Yes	5 ' - T ^{me}C $^{\Psi}$C ^{me}C $^{\Psi}$C ^{me}C T T $^{\Psi}$C ^{me}C $^{\Psi}$C T T ^{me}C $^{\Psi}$C T $^{\Psi}$C - 3 '
TFO6-DNALNA^{me}C	No	5 ' - T ^{me}C ^{me}C ^{me}C ^{me}C ^{me}C T T ^{me}C ^{me}C ^{me}C T T ^{me}C ^{me}C T ^{me}C - 3 '
TFO7-DNA$^{\Psi}$C	No	5 ' - T $^{\Psi}$C C C $^{\Psi}$C T T $^{\Psi}$C C $^{\Psi}$C T T $^{\Psi}$C C T $^{\Psi}$C - 3 '

[a]Triplet sequence in which $^{\Psi}$C gives >75% $^{\Psi}$C(H3) tautomeric propensity in the λ-dynamics simulations are shaded.

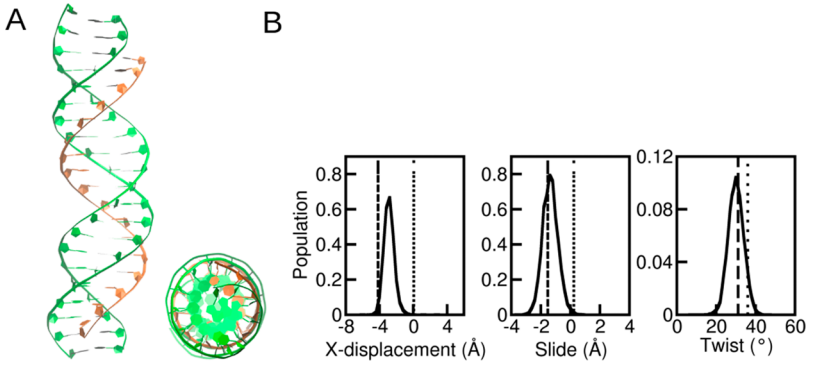


Figure 10. Simulations for triplex-containing TFO5-DNALNA$^{\Psi}$C with classical MD. (A) Average structure from one run (duplex in green and TFO in orange), side and top views. (B) Triplex base pair parameter distributions (excluding 2 residues at either TFO end), from the last 50 ns of four independent 100 ns runs for the triplex containing TFO5. The dashed and dotted lines represent the average values for A- and B-form DNAs, respectively.[51]

stabilization as measured by the melting temperature when three Cs were substituted with $^{\Psi}$C and 2′-*O*-methyl-$^{\Psi}$C.[44−46] Also, it has been shown that $^{\Psi}$C combined with peptide nucleic acid (PNA) at the 3′-end of the TFO in a nonconsecutive manner with every second position containing a T[47,48] can reduce the pH sensitivity. On the basis of NMR experiments, Leitner et al. have demonstrated that for intramolecular TFO-DNA, protonation is disfavored for adjacent C or for C at the end of the triplex,[49] also arguing in favor of our in silico and in vitro results. All of these reports are in line with our conclusion that $^{\Psi}$C in a nonconsecutive manner is the best option for designing DNA/LNA mixmer TFOs containing $^{\Psi}$C. The TFO was able to target the G-rich region under intranuclear conditions when $^{\Psi}$C is flanked by ^{me}C or T, in agreement with simulation results (Table 3). To our knowledge, this is the first time that this is shown in vitro for TFO DNA/LNA containing $^{\Psi}$C.

Two different intercalators have been used in this work: BQQ and TINA. BQQ is a triplex helix-intercalating compound that can bind specifically to and stabilize the triplex structures of purine and pyrimidine motifs.[50] We have also previously used BQQ to confirm and probe for triplex formation using different TFOs (Pabon et al. unpublished result). There we demonstrated that BQQ could also stabilize triplexes formed by LNA-containing TFOs. Here, we can confirm that this is also valid for C-rich TFOs with consecutive $^{\Psi}$C. Also, this is the first study showing that BQQ can stabilize $^{\Psi}$C-containing TFO DNA/LNA. The influence of TINA positioning is not discussed in this work, but we have chosen to locate TINA at the 3′-end position based on previous work (Pabon et al. unpublished result). Surprisingly, the presence of TINA in our $^{\Psi}$C-containing TFOs seems not to increase the rate of triplex formation under the experimental conditions used here.

Structural Characterization of Triplex-Containing TFO5-DNALNA$^{\Psi}$C. To characterize the 3D structure of the triplex containing TFO5-DNALNA$^{\Psi}$C, we performed simulations with classical MD at fixed tautomeric states. We fixed the tautomerization state of $^{\Psi}$C to be $^{\Psi}$C(H3); based on the result in the smaller triplexes, all-$^{\Psi}$C(H3) is favored in the triplex environment. We found that the triplex is stable during the course of the simulation, preserving most of the Watson−Crick and Hoogsteen hydrogen bonds (Figure 10). Out of four independent runs, one has a loss of Hoogsteen hydrogen bond towards the 3′-end, but the other three are similar, where all Hoogsteen hydrogen bonds are preserved most of the time (one such run is shown in Figure S7). The average structure shows no obvious distortion in the triplex structure that might contribute to instability (Figure 10). Upon binding of TFO5, the DNA duplex overslides in the negative direction to accommodate the third strand. The resultant helical structure has slide and twist parameters similar to A-type duplex DNA, but an *x*-displacement value is between those of A- and B-types (Figure 10). Replacing LNA with DNA in the bound TFO does not affect the structural feature of the triplex (Figure S8), supporting that what promotes LNA-containing TFO binding to the DNA duplex is that the TFO is indeed preorganized for major groove binding (Pabon et al. unpublished data). In conclusion, upon binding of $^{\Psi}$C-containing TFO, the triple helical conformation is between A- and B-types with base pairs remaining almost perpendicular to the helical axis, in agreement with what was observed in other DNA duplexes involved in triplex formation.[28]

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

CONCLUSIONS

We have performed λ-dynamics simulations and binding experiments under intranuclear conditions to investigate the ability of pseudoisocytidine to efficiently target the major groove of a DNA duplex and form a triplex structure. In particular, we have investigated the tautomerization of $^{\Psi}$C in different short single-stranded and triplex DNAs. In single strands, we have observed a clear influence of sequence on the tautomeric propensity of $^{\Psi}$C. The predisposition for tautomer $^{\Psi}$C(H3) is higher when the neighboring residues are cytidines (even higher when cytidine is 5-methylated) compared with when they are thymine. When the neighboring residues are $^{\Psi}$C, the propensity of tautomer $^{\Psi}$C(H3) (located between two $^{\Psi}$Cs) is low. Furthermore, the sugar modification LNA on the neighboring residues does not affect the $^{\Psi}$C tautomerization equilibrium directly.

Once the single strand is bound to the targeted duplex, forming a triplex, nearly all $^{\Psi}$Cs are $^{\Psi}$C(H3) tautomers, allowing hydrogen bonding with the Hoogsteen site of the guanine of the double strand, even if the propensity of $^{\Psi}$C(H3) in free, unbound TFOs was very low, as in the case of consecutive runs of $^{\Psi}$Cs. This suggests that Hoogsteen hydrogen bonding in the triplex is strong enough to shift the tautomeric predisposition to favor $^{\Psi}$C(H3) and confer thermodynamic stability.

The in vitro experiment shows that the TFOs having three or more consecutive $^{\Psi}$Cs, such as TFO-DNAfull$^{\Psi}$C and TFO-DNALNAfull$^{\Psi}$C, were unable to form triplexes under intranuclear salt conditions at pH 7.4 (also when ligands promoting triplex formation such as BQQ and TINA were included). Only when nonconsecutive $^{\Psi}$Cs were included in combination with alternating DNA/LNA residues (LNA residues are ^{me}C and T), the 17-mer TFO was able to form a triplex. In the formed triplex, the pseudoisocytidine targeted the Hoogsteen site of the guanine with two hydrogen bonds, and the duplex structure goes under conformation rearrangement with slide and twist parameters similar to A-type, but the *x*-displacement is between those of A- and B-forms.

We conclude, based on the combination of in silico and in vitro studies, that the inclusion of alternating $^{\Psi}$C and the combination with alternating LNA enhances the formation of the evaluated C-rich intermolecular triplexes. Therefore, based on our results, we suggest that when designing DNA/LNA mixmer TFOs containing $^{\Psi}$C, incorporation of $^{\Psi}$C and LNA in a nonconsecutive manner is preferable.

MATERIALS AND METHODS

Simulations. The tautomeric equilibria are influenced by chemical and physical factors, including solvent, ion concentration, and biomolecular environment. An accurate prediction should account for the small energy differences that cause shifts in the tautomeric equilibrium and the need to sample different conformational states accessible to the biomolecule. Methods based on a macroscopic description of the biomolecule and the solvent do not explicitly account for dielectric heterogeneity and response to conformational rearrangement, nor do they take into account the conformational rearrangement.[52] We choose the λ-dynamics approach with an explicit description of the solvent[53] because it addresses these issues by enabling the direct coupling between tautomerization processes and conformational dynamics. Moreover, the accuracy of the method can be improved through a fine calibration of the force field.

Theory. Multisite λ-dynamics[53] is set up for the $^{\Psi}$C residue such that the two tautomeric states $^{\Psi}$C(H1) and $^{\Psi}$C(H3) are described and propagated by continuous variables λ_{H1} and λ_{H3}, respectively. The potential energy function is given by

$$\begin{aligned} U_{tot}(X, \{x\}, \{\lambda\}) = U_{env}(X) &+ \lambda_{H1}(U(X, x_{H1}) \\ &- \Delta G_{H1\rightarrow H3}(\text{model}) \\ &+ \lambda_{H3}(U(X, x_{H3}) + F^{bias}(\lambda_{H1}) \\ &+ F^{bias}(\lambda_{H3}) \end{aligned}$$

where X is the coordinates of the environment atoms, x_{H1} and x_{H3} are the coordinates of atoms in $^{\Psi}$C, corresponding to the tautomers $^{\Psi}$C(H1) and $^{\Psi}$C(H3) respectively, and

$$F^{bias}(\lambda_{Hi}) = \begin{cases} k_{bias}(\lambda_{Hi} - 0.8)^2 & \text{if } \lambda_{Hi} < 0.8 \\ 0 & \text{otherwise} \end{cases}$$

where H*i* refers to the tautomeric state $^{\Psi}$C(H1) or $^{\Psi}$C(H3). λ_{Hi} scales the potential energy of the corresponding tautomer with the constraints

$$0 \le \lambda_{Hi} \le 1 \text{ and } \lambda_{H1} + \lambda_{H3} = 1$$

$\Delta G_{H1\rightarrow H3}$ (model) is the free energy for transforming tautomer $^{\Psi}$C(H1) into $^{\Psi}$C(H3) of the pseudoisocytidine model compound in aqueous solution. This term is included to flatten the potential energy surface such that the two tautomeric states of the nucleoside in solution are equipopulated as the free energy between the two states becomes zero. The model compound structure used in the free energy calculation is the reference state at which the tautomeric ratio $^{\Psi}$C(H1)/$^{\Psi}$C(H3) is 1:1. In the investigated systems, the deviation from this tautomeric ratio would come from the contribution of the environment.

Two harmonic biasing potentials $F^{bias}(\lambda_{H1})$ and $F^{bias}(\lambda_{H3})$ are included to bias the sampling toward the physical end states. In this formulation, $0.8 \le \lambda_{Hi} \le 1$ is considered to be a physical end state. k_{bias} is the force constant of the harmonic potentials. The force constant is equal for both tautomeric states.

Calculation of Free Energy. The free energy, $\Delta G_{H1\rightarrow H3}$(model), was calculated using the Bennett acceptance ratio (BAR) method.[54] The pseudoisocytidine hybrid model compound used in the free energy perturbation calculation is constructed with deoxyribose sugar and 5′ and 3′ hydroxyls and two pyrimidine bases corresponding to the two tautomeric states (CHARMM dual topology); the two bases are maintained within the same volume of space by distance restraints between all pairs of common atoms in the two tautomeric states. The residue is solvated with TIP3P water molecules[55] in a cubic box with 20 Å side length. The CHARMM BLOCK module[56] is used to partition the system into three blocks: (I) environment, (II) tautomer $^{\Psi}$C(H1) base, and (III) tautomer $^{\Psi}$C(H3) base. Interactions between blocks II and III are set to null; interactions between blocks I and II and within block II are scaled with λ; interactions between blocks I and III and within block III are scaled with $1 - \lambda$. Scaling with λ and $1 - \lambda$ is not applied to the bond, angle, and dihedral energy terms. Eight λ values corresponding to alchemical intermediate/end states were used (λ = 0.0, 0.1, 0.2, 0.4, 0.6, 0.8, 0.9, and 1.0), and each window simulation length is 1 ns, with only the last 900 ps used for the BAR

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

Table 4. ON Sequences[a]

name	length (nt)	sequence
TFO1-DNAfull$^{\Psi}$C	17	5′-T$^{\Psi}$C$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$CT$^{\Psi}$C-3′
TFO2-DNAfull$^{\Psi}$C-TINA	17	5′-T$^{\Psi}$C$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$CTP$^{\Psi}$C-3′
TFO3-DNALNAfull$^{\Psi}$C	17	5′-T$^{\Psi}$C$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$CT$^{\Psi}$C-3′
TFO4-DNALNAfull$^{\Psi}$C-TINA	17	5′-T$^{\Psi}$C$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$C$^{\Psi}$CTT$^{\Psi}$C$^{\Psi}$CTP$^{\Psi}$C-3′
TFO5-DNALNA$^{\Psi}$C	17	5′-T^{me}C$^{\Psi}$C^{me}C$^{\Psi}$C^{me}CTT$^{\Psi}$C^{me}C$^{\Psi}$CTT^{me}C$^{\Psi}$CT$^{\Psi}$C-3′
TFO6-DNALNA^{me}C	17	5′-T^{me}C^{me}C^{me}C^{me}C^{me}CTT^{me}C^{me}C^{me}CTT^{me}C^{me}CT^{me}C-3′
TFO7-DNA$^{\Psi}$C	16	5′-T$^{\Psi}$CCC$^{\Psi}$CTT$^{\Psi}$CC$^{\Psi}$CTT$^{\Psi}$CCT$^{\Psi}$C-3′

[a]DNA is indicated in capital letters and LNA is indicated in underlined capital letters; P, p-TINA; ^{me}C, 5-methyl-C; and $^{\Psi}$C, pseudoisocytidine. The commercial version of LNA C is always methylated.

calculation. The CHARMM REPD module[56] is used to run eight alchemical intermediate/end states in parallel and attempt to exchange energies every 1000 steps (Hamiltonian replica exchange). The calculated free energy is 28.9 ± 0.1 kcal/mol.

Calibration of the Biasing Potential Force Constant. The value of k_{bias} is calibrated by performing 1 ns runs of λ-dynamics at various k_{bias} values and observing the fraction of physical end states in the trajectory (FPL, fraction physical ligand) and the frequency of transitions between the two tautomeric states (transition rate, ns^{-1}). An optimal value of k_{bias} was chosen so as to yield a high transition rate and simultaneously maintain a high FPL (above 0.8).[53] The initial k_{bias} values are 15, 20, 25, and 30 kcal/mol. Additional runs are added as needed to determine the optimal value. The value of the optimized k_{bias} is 19.5 kcal/mol.

Simulation Settings. Simulations were performed with CHARMM (version 41a2)[56] with CHARMM36 force field for DNA,[57] TIP3P water,[55] ions,[58] and modified nucleic acids[59,60] (for $^{\Psi}$C). Updated LNA parameters were used (Xu; Nilsson; Villa unpublished result). Initial 7-mer and 17-mer triplex structures were taken from a parallel DNA triplex fiber model from the 3DNA Web server.[61] For both 7-mer and 17-mer triplexes, the duplex is longer than the TFO by 2 residues at either side. The third strand of the triplex is used as the initial structure for the single-strand trimers and 7-mers.

For single-strand monomers, trimers, 7-mers, and 7-mer triplexes, λ-dynamics was used to allow interconversion between the two $^{\Psi}$C tautomers. Several selected systems were also run with the standard MD (fixed tautomer state) to aid analysis—these are run with the same cutoff, settings, and lengths as the λ-dynamics run, but with single topology without λ scaling. For the 17-mer triplex, conventional MD simulation is used and the tautomeric state of $^{\Psi}$C is fixed to tautomer $^{\Psi}$C(H3), the tautomer involved in Hoogsteen hydrogen bonding. For all λ-dynamics simulations, five independent runs were performed. Simulation lengths were chosen so that the standard errors of the mean of the tautomeric propensities from five runs do not exceed 10%. For monomers and trimers, these are 6 ns; 7-mers, 8 ns; and 7-mer triplexes, 40 ns. For MD simulation of the 17-mer triplex, four independent 100 ns runs were performed. Simulation systems and lengths are summarized in Table S1.

The structures were minimized with the steepest descent and adopted-basis Newton−Raphson methods with large position restraints on the heavy atoms. For triplexes, additional distance restraints are added for Watson−Crick and Hoogsteen hydrogen bonds. The systems were solvated in boxes of TIP3P water molecules, with dimensions of 20 × 20 × 20 Å^3 for monomers, 50 × 50 × 50 Å^3 for trimers and 7-mers, 65 × 45 × 45 Å^3 for 7-mer triplexes, and 88 × 42 × 42 Å^3 for 17-mer triplexes. After the addition of sodium ions to neutralize the system, additional sodium and chloride ions are included to reach an ionic concentration of approximately 0.1 M.

λ-dynamics is performed within the CHARMM BLOCK module[56] using the multisite λ-dynamics framework (MSLD).[62] The functional form of λ_{Hi}, $\lambda_{Hi} = \frac{e^{5.5 \sin \theta_{Hi}}}{\sum_i e^{5.5 \sin \theta_{Hi}}}$, was used. Defining λ as a function of θ in this way has been shown to be optimal for sampling and convergence.[63] θ is assigned a fictitious mass of 12 amu·Å^2 (amu = atomic mass unit). The temperature was maintained at 298 K by coupling to Langevin heat bath with a collision frequency of 10 ps^{-1}. λ is saved every 10 steps. The bond, angle, and dihedral energy terms are excluded from scaling by λ so that only geometrically relevant states are sampled. A sampling bias was applied for each tautomeric state with a force constant of 19.5 kcal/mol. A nonbonded list cutoff of 15 Å was used with the electrostatic force switch and van der Waals switch functions between 10 and 12 Å. Simulations were performed in the NVT ensemble with Langevin dynamics with a collision frequency of 10 ps^{-1}. For 7-mer triplexes, distance restraints were applied to Watson−Crick and Hoogsteen hydrogen bonds (distance 2.9 Å and force constant 10 kcal/mol/Å^2) for the base pairs at 5′- and 3′-end positions.

MD simulations for 17-mer triplexes were performed on graphics processing units (GPUs) with CHARMM[56] and the CHARMM/OpenMM[64] interface in the NVT ensemble with Langevin dynamics with a collision frequency of 5 ps^{-1}. A van der Waals force switching function was used between 8 and 9 Å. Particle mesh Ewald was used to treat electrostatic interactions with a nonbonded cutoff of 8 Å and a grid point spacing of 1.0 Å. The distance cutoff in generating the list of pairwise interactions was 17 Å. The temperature was maintained at 298 K by coupling to a Langevin heat bath with a frictional coefficient of 10 ps^{-1}. Distance restraints were applied to Watson−Crick hydrogen bonds (distance 2.9 Å and force constant 10 kcal/mol/Å^2) for the base pair in the 5′- and 3′-ends and all Hoogsteen N7-N3 hydrogen bonds (same distance and force constant). After minimization, the system was equilibrated for 2 ns. During production, the distance restraints on the Hoogsteen hydrogen bonds were released.

In all simulations, the SHAKE algorithm was used to constrain bonds involving hydrogen.[65] A lookup table was used for interactions between water molecules,[66] except for GPU simulations. The leapfrog integrator was used with an integration time-step of 2 fs. The hydrogen bond, interaction energy, and SASA analyses were performed within CHARMM. Triplex base pair step parameter analyses were performed with Curves+.[67]

In Vitro Binding Experiments. *ONs.* Standard methods were used to synthesize the ONs containing pyrimidine

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

Table 5. Target Sequences Used for Experiments[a]

Name	Length (nt)	Sequence
DS48	48	5 ' - G C C A C C G T G A C C T T G C **A G G G G A A G G G A A G G A G** C T C A T G A A G C T T C G A G - 3 '
		3 ' - C G G T G G C A C T G G A A C G **T C C C C T T C C C T T C C T C** G A G T A C T T C G A A G C T C - 5 ' *
DS51	51	5' - G G C C A C C G T G A C C T T G C **A G G G G G A A G G G A A G G A G** C T C A T G A A G C T T C G A G A - 3 '
		3' - C C G G T G G C A C T G G A A C G **T C C C C C T T C C C T T C C T C** G A G T A C T T C G A A G C T C T - 5 ' *

[a]DNA is indicated in capital letters. DS, double-stranded target sequences: DS48 and DS51. The TFO binding site is shown in a gray box with letters in bold. The star (*) indicates the strand which was radiolabelled using [γ-^{32}P] ATP.

analogues and TINA. TFO5 and TFO6 were synthesized at the Nucleic Acid Center at the University of Southern Denmark in Jesper Wengel Laboratory. Mixmer LNA/DNA ONs were synthesized using solid-phase phosphoramidite chemistry on an automated DNA synthesizer on a 1.0 mmol synthesis scale.[68] Purification to at least 85% purity of all modified ONs was performed using reversed-phase high-performance liquid chromatography (RP-HPLC) or ion-exchange HPLC (IE-HPLC), and the composition of all synthesized ONs was verified using matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) analysis recorded using 3-hydroxypicolinic acid as a matrix. TFO1-TFO4 and TFO7 were provided by Anapa Biotech A/S Company from Denmark. DNA target sequences were ordered from Sigma. The ONs and target sequences used here are presented in Tables 4 and 5, respectively. The ON concentrations of stock solutions were confirmed using a NanoDrop spectrophotometer (Thermo Scientific).

ON Hybridization. The double-stranded (DS) target (5.0 nM) was incubated with ONs at different concentrations (0.5, 1.0, 2.0, and 4.0 μM, corresponding to TFO versus DS target ratios of 100, 200, 400, and 800, respectively). Many cytosine-rich ONs can potentially form an intramolecular i-motif[69] and to avoid that, the TFOs were heated before hybridization for 5 min at 65 °C, followed by cooling on ice. Hybridization was performed in an intranuclear buffer (Tris-acetate 50 mM, pH 7.4, 120 mM KCl, 5 mM NaCl, and 0.5 mM MgOAc) and in a total volume of 10 μL at 37 °C for up to 72 h in the absence or presence of BQQ (1 μM).

Preparation of ^{32}P-Labeled dsDNA Target. The pyrimidine strand of the target sequence was labeled using [γ-^{32}P] ATP and T4 polynucleotide kinase (Fermentas) according to the manufacturer's protocol and then purified using a QIAquick Nucleotide Removal Kit (Qiagen). The pyrimidine strand labeled ON was annealed with the unlabeled complementary strand at a 1:1 ratio. The annealing was performed by heating for 5 min at 95 °C, followed by decreasing the temperature to 40 °C at a rate of 1 grade per minute using a thermocycler.

Electrophoretic Mobility Shift Assay (EMSA). DNA complexes were analyzed using nondenaturing polyacrylamide gel electrophoresis 10% (29:1) in Tris-acetate−ethylenediaminetetraacetic acid (EDTA) (TAE) buffer (1×, pH 7.4 supplemented with 0.5 mM MgOAc and 5 mM NaCl). The gels were run at 150 V and 200 mA for 5 h with circulation water cooling and analyzed using a Molecular Imager FX system. The intensity of the gel bands was quantified using the Quantity One software (Bio-Rad). All experiments were repeated three times.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.7b00347.

Electrostatic interaction energy between bases of various trimers from MD simulation, distributions of SASA of $^{\Psi}$C residues, Watson−Crick and Hoogsteen hydrogen bonds during the simulations, TFO binding of 17-mer TFO sequences containing TINA and consecutive $^{\Psi}$C, TFO binding of 16-mer TFO sequences containing no consecutive $^{\Psi}$C, TFO binding of 17-mer TFO sequences containing nonconsecutive $^{\Psi}$C, and simulations for triplex containing TFO5-DNALNA$^{\Psi}$C and another with the same TFO sequence with LNA changed to DNA using classical MD (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: alessandra.villa@ki.se. Phone: +46-8-52481081 (A.V.).

ORCID

Jesper Wengel: 0000-0001-9835-1009

Alessandra Villa: 0000-0002-9573-0326

Present Address

[#](A.U.) Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, 48824, United States.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Departamento Administrativo de Ciencia, Tecnología e Innovación (COLCIENCIAS), Colombia (Ph.D. grant resolución 02007/24122010 to Y.V.P.-M.), Nanyang Technological University Research Scholarship (Y.D.H.), and the Swedish Research Council (VT 2015-04992, K2015-68X-11247-21-3) and EU Marie Skłodowska-Curie [ITN 71613, MMBio] for support. The authors are indebted to Søren Morgenthaler Echwald, Anapa Biotech A/S, Hørsholm, Denmark, for providing TINA-containing ONs.

■ ABBREVIATIONS

BQQ, benzoquinoquinoxaline; C, G, A, T, DNA bases (cytosine, guanine, adenine, thymine); EMSA, electrophoretic mobility shift assay; ERE, estrogen response element; human TFF, human trefoil factor; LNA, locked nucleic acid; MD, molecular dynamics; ^{me}C, 5-methyl-cytidine; $^{me}\underline{C}$, 5-methyl-cytidine LNA; NMR, nuclear magnetic resonance; ON, oligonucleotide; SASA, solvent-accessible surface area; T,

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

thymine LNA; TFO, triplex-forming oligonucleotide; TINA, p-twisted intercalating nucleic acid; $^{\Psi}$C, pseudoisocytidine

REFERENCES

(1) Zain, R.; Sun, J.-S. Do natural DNA triple-helical structures occur and function in vivo? *Cell. Mol. Life Sci.* **2003**, *60*, 862−870.

(2) Baran, N.; Lapidot, A.; Manor, H. Formation of DNA Triplexes Accounts for Arrests of DNA Synthesis at d(TC)n and d(GA)n Tracts. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 507−511.

(3) Daube, S. S.; von Hippel, P. H. Functional transcription elongation complexes from synthetic RNA−DNA bubble duplexes. *Science* **1992**, *258*, 1320−1324.

(4) Dayn, A.; Samadashwily, G. M.; Mirkin, S. M. Intramolecular DNA triplexes: Unusual sequence requirements and influence on DNA polymerization. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 11406−11410.

(5) Møllegaard, N. E.; Buchardt, O.; Egholm, M.; Nielsen, P. E. Peptide nucleic acid.DNA strand displacement loops as artificial transcription promoters. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 3892−3895.

(6) Hampel, K. J.; Lee, J. S. Two-dimensional pulsed-field gel electrophoresis of yeast chromosomes: Evidence for triplex-mediated DNA condensation. *Biochem. Cell Biol.* **1993**, *71*, 190−196.

(7) Veselkov, A. G.; Malkov, V. A.; Frank-Kamenetskll, M. D.; Dobrynin, V. N. Triplex model of chromosome ends. *Nature* **1993**, *364*, 496.

(8) Ito, T.; Smith, C. L.; Cantor, C. R. Sequence-specific DNA purification by triplex affinity capture. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 495−498.

(9) Ito, T.; Smith, C. L.; Cantor, C. R. Triplex affinity capture of a single copy clone from a yeast genomic library. *Nucleic Acids Res.* **1992**, *20*, 3524.

(10) Ito, T.; Smith, C. L.; Cantor, C. R. Affinity capture electrophoresis for sequence-specific DNA purification. *Genet. Anal.: Biomol. Eng.* **1992**, *9*, 96−99.

(11) Vary, C. P. Triple-helical capture assay for quantification of polymerase chain reaction products. *Clin. Chem.* **1992**, *38*, 687−694.

(12) Olivas, W. M.; Maher, L. J. Analysis of duplex DNA by triple helix formation: Application to detection of a p53 microdeletion. *BioTechniques* **1994**, *16*, 128−132.

(13) Havre, P. A.; Glazer, P. M. Targeted mutagenesis of simian virus 40 DNA mediated by a triple helix-forming oligonucleotide. *J. Virol.* **1993**, *67*, 7324−7331.

(14) Havre, P. A.; Gunther, E. J.; Gasparro, F. P.; Glazer, P. M. Targeted mutagenesis of DNA using triple helix-forming oligonucleotides linked to psoralen. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7879−7883.

(15) Frank-Kamenetskii, M. D.; Mirkin, S. M. Triplex DNA structures. *Annu. Rev. Biochem.* **1995**, *64*, 65−95.

(16) Soyfer, V. N.; Potaman, V. N. General Features of Triplex Structures. In *Triple-Helical Nucleic Acids*; Springer: New York, 1996; pp 100−150.

(17) Goñi, J. R.; de la Cruz, X.; Orozco, M. Triplex-forming oligonucleotide target sequences in the human genome. *Nucleic Acids Res.* **2004**, *32*, 354−360.

(18) Goñi, J. R.; Vaquerizas, J. M.; Dopazo, J.; Orozco, M. Exploring the reasons for the large density of triplex-forming oligonucleotide target sequences in the human regulatory regions. *BMC Genomics* **2006**, *7*, 63.

(19) Beal, P. A.; Dervan, P. B. Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation. *Science* **1991**, *251*, 1360.

(20) Helene, C.; Thuong, N. T.; Harel, A. Control of Gene Expression by Triple Helix-Forming Oligonucleotides. The Antigene Strategya. *Ann. N. Y. Acad. Sci.* **1992**, *660*, 27−36.

(21) Paugh, S. W.; Coss, D. R.; Bao, J.; Laudermilk, L. T.; Grace, C. R.; Ferreira, A. M.; Waddell, M. B.; Ridout, G.; Naeve, D.; Leuze, M.; LoCascio, P. F.; Panetta, J. C.; Wilkinson, M. R.; Pui, C.-H.; Naeve, C. W.; Uberbacher, E. C.; Bonten, E. J.; Evans, W. E. MicroRNAs Form Triplexes with Double Stranded DNA at Sequence-Specific Binding Sites; a Eukaryotic Mechanism via which microRNAs Could Directly Alter Gene Expression. *PLoS Comput. Biol.* **2016**, *12*, No. e1004744.

(22) Ono, A.; Ts'o, P. O. P.; Kan, L. S. Triplex formation of oligonucleotides containing 2′-O-methylpseudoisocytidine in substitution for 2′-deoxycytidine. *J. Am. Chem. Soc.* **1991**, *113*, 4032−4033.

(23) Kan, L.-S.; Lin, W.-C.; Yadav, R. D.; Shih, J. H.; Chao, I. NMR Studies of the Tautomerism in Pseudoisocytidine. *Nucleosides Nucleotides* **1999**, *18*, 1091−1093.

(24) Mayer, A.; Häberli, A.; Leumann, C. J. Synthesis and triplex forming properties of pyrrolidino pseudoisocytidine containing oligodeoxynucleotides. *Org. Biomol. Chem.* **2005**, *3*, 1653−1658.

(25) Shahid, K. A.; Majumdar, A.; Alam, R.; Liu, S.-T.; Kuan, J. Y.; Sui, X.; Cuenoud, B.; Glazer, P. M.; Miller, P. S.; Seidman, M. M. Targeted cross-linking of the human β-globin gene in living cells mediated by a triple helix forming oligonucleotide. *Biochemistry* **2006**, *45*, 1970−1978.

(26) Singh, V.; Fedeles, B. I.; Essigmann, J. M. Role of tautomerism in RNA biochemistry. *RNA* **2014**, *21*, 1−13.

(27) Khandogin, J.; Brooks, C. L. Constant pH molecular dynamics with proton tautomerism. *Biophys. J.* **2005**, *89*, 141−157.

(28) Esguerra, M.; Nilsson, L.; Villa, A. Triple helical DNA in a duplex context and base pair opening. *Nucleic Acids Res.* **2014**, *42*, 11329−11338.

(29) Aviñó, A.; Cubero, E.; González, C.; Eritja, R.; Orozco, M. Antiparallel triple helices. Structural characteristics and stabilization by 8-amino derivatives. *J. Am. Chem. Soc.* **2003**, *125*, 16127−16138.

(30) Semenyuk, A.; Darian, E.; Liu, J.; Majumdar, A.; Cuenoud, B.; Miller, P. S.; MacKerell, A. D.; Seidman, M. M. Targeting of an Interrupted Polypurine:Polypyrimidine Sequence in Mammalian Cells by a Triplex-Forming Oligonucleotide Containing a Novel Base Analogue. *Biochemistry* **2010**, *49*, 7867−7878.

(31) Escudé, C.; Nguyen, C. H.; Kukreti, S.; Janin, Y.; Sun, J.-S.; Bisagni, E.; Garestier, T.; Hélène, C. Rational design of a triple helix-specific intercalating ligand. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 3591−3596.

(32) Zain, R.; Marchand, C.; Sun, J.-s.; Nguyen, C. H.; Bisagni, E.; Garestier, T.; Hélène, C. Design of a triple-helix-specific cleaving reagent. *Chem. Biol.* **1999**, *6*, 771−777.

(33) Højland, T.; Kumar, S.; Babu, B. R.; Umemoto, T.; Albaek, N.; Sharma, P. K.; Nielsen, P.; Wengel, J. LNA (locked nucleic acid) and analogs as triplex-forming oligonucleotides. *Org. Biomol. Chem.* **2007**, *5*, 2375−2379.

(34) Klinge, C. M. Estrogen receptor interaction with estrogen response elements. *Nucleic Acids Res.* **2001**, *29*, 2905−2919.

(35) Geny, S.; Moreno, P. M. D.; Krzywkowski, T.; Gissberg, O.; Andersen, N. K.; Isse, A. J.; El-Madani, A. M.; Lou, C.; Pabon, Y. V.; Anderson, B. A.; Zaghloul, E. M.; Zain, R.; Hrdlicka, P. J.; Jorgensen, P. T.; Nilsson, M.; Lundin, K. E.; Pedersen, E. B.; Wengel, J.; Smith, C. I. E. Next-generation bis-locked nucleic acids with stacking linker and 2′-glycylamino-LNA show enhanced DNA invasion into supercoiled duplexes. *Nucleic Acids Res.* **2016**, *44*, 2007−2019.

(36) Sharma, B. D.; McConnell, J. F. The crystal and molecular structure of isocytosine. *Acta Crystallogr.* **1965**, *19*, 797−806.

(37) Hirao, I.; Kimoto, M.; Yamakage, S.-i.; Ishikawa, M.; Kikuchi, J.; Yokoyama, S. A unique unnatural base pair between a C analogue, pseudoisocytosine, and an A analogue, 6-methoxypurine, in replication. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1391−1393.

(38) Kan, L.-S.; Lin, W.-C.; Yadav, R. D.; Shih, J. H.; Chao, I. NMR Studies of the Tautomerism in Pseudoisocytidine. *Nucleosides Nucleotides* **1999**, *18*, 1091−1093.

(39) Filichev, V. V.; Pedersen, E. B. Stable and selective formation of Hoogsteen-type triplexes and duplexes using twisted intercalating nucleic acids (TINA) prepared via postsynthetic Sonogashira solid-phase coupling reactions. *J. Am. Chem. Soc.* **2005**, *127*, 14849−14858.

(40) Filichev, V. V.; Gaber, H.; Olsen, T. R.; Jørgensen, P. T.; Jessen, C. H.; Pedersen, E. B. Twisted intercalating nucleic acids−intercalator influence on parallel triplex Stabilities. *Eur. J. Org. Chem.* **2006**, 3960−3968.

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177

(41) Faucon, B.; Mergny, J.-L.; Hélène, C. Effect of third strand composition on the triple helix formation: Purine versus pyrimidine oligodeoxynucleotides. *Nucleic Acids Res.* **1996**, *24*, 3181−3188.
(42) Lee, J. S.; Woodsworth, M. L.; Latimer, L. J. P.; Morgan, A. R. Poly(pyrimidine) poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res.* **1984**, *12*, 6603−6614.
(43) Vekhoff, P.; Ceccaldi, A.; Polverari, D.; Pylouster, J.; Pisano, C.; Arimondo, P. B. Triplex formation on DNA targets: How to choose the oligonucleotide. *Biochemistry* **2008**, *47*, 12277−12289.
(44) Volker, J.; Klump, H. H. Electrostatic effects in DNA triple helixes. *Biochemistry* **1994**, *33*, 13502−13508.
(45) Sugimoto, N.; Wu, P.; Hara, H.; Kawamoto, Y. pH and Cation Effects on the Properties of Parallel Pyrimidine Motif DNA Triplexes. *Biochemistry* **2001**, *40*, 9396−9405.
(46) Chin, T.-M.; Lin, S.-B.; Lee, S.-Y.; Chang, M.-L.; Cheng, A. Y.-Y.; Chang, F.-C.; Pasternack, L.; Huang, D.-H.; Kan, L.-S. "Paper-clip" type triple helix formation by 5′-d-$(TC)_3Ta(CT)_3Cb(AG)_3$ (*a* and *b* = 0−4) as a function of loop size with and without the pseudoisocytosine base in the Hoogsteen strand. *Biochemistry* **2000**, *39*, 12457−12464.
(47) Egholm, M.; Christensen, L.; Deuholm, K. L.; Buchardt, O.; Coull, J.; Nielsen, P. E. Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA. *Nucleic Acids Res.* **1995**, *23*, 217−222.
(48) Hansen, M. E.; Bentin, T.; Nielsen, P. E. High-affinity triplex targeting of double stranded DNA using chemically modified peptide nucleic acid oligomers. *Nucleic Acids Res.* **2009**, *37*, 4498−4507.
(49) Leitner, D.; Schröder, W.; Weisz, K. Influence of sequence-dependent cytosine protonation and methylation on DNA triplex stability. *Biochemistry* **2000**, *39*, 5886−5892.
(50) Zain, R.; Polverari, D.; Nguyen, C.-H.; Blouquit, Y.; Bisagni, E.; Garestier, T.; Grierson, D. S.; Sun, J.-S. Optimization of Triple-Helix-Directed DNA Cleavage by Benzoquinoquinoxaline−Ethylenediaminetetraacetic Acid Conjugates. *ChemBioChem* **2003**, *4*, 856−862.
(51) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X.-J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C.-S.; Westhof, E.; Wolberger, C.; Berman, H. M. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* **2001**, *313*, 229−237.
(52) Wallace, J. A.; Shen, J. K. Predicting pK_a values with continuous constant pH molecular dynamics. *Methods Enzymol.* **2009**, *466*, 455−475.
(53) Goh, G. B.; Knight, J. L.; Brooks, C. L. Constant pH molecular dynamics simulations of nucleic acids in explicit solvent. *J. Chem. Theory Comput.* **2012**, *8*, 36−46.
(54) Bennett, C. H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.* **1976**, *22*, 245−268.
(55) Jørgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926.
(56) Brooks, B. R.; Brooks, C. L.; MacKerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(57) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.
(58) Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *J. Chem. Phys.* **1994**, *100*, 9050−9063.
(59) Xu, Y.; Nilsson, L.; MacKerrel, A. D. An Additive Charmm Force Field for Modified Nucleic Acids. *Biophys. J.* **2015**, *108*, 235a−236a.
(60) MacKerell, A. D. CHARMM Force Field Files. http://mackerell.umaryland.edu/charmm_ff.shtml#charmm (accessed Feb 6, 2017).
(61) Zheng, G.; Lu, X.-J.; Olson, W. K. Web 3DNA—A web server for the analysis, reconstruction, and visualization of three-dimensional nucleic-acid structures. *Nucleic Acids Res.* **2009**, *37*, W240−W246.
(62) Knight, J. L.; Brooks, C. L. λ-Dynamics Free Energy Simulation Methods. *J. Comput. Chem.* **2009**, *30*, 1692−1700.
(63) Knight, J. L.; Brooks, C. L. Applying efficient implicit nongeometric constraints in alchemical free energy simulations. *J. Comput. Chem.* **2011**, *32*, 3423−3432.
(64) Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. Accelerating molecular dynamic simulation on graphics processing units. *J. Comput. Chem.* **2009**, *30*, 864−872.
(65) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(66) Nilsson, L. Efficient Table Lookup Without Inverse Square Roots for Calculation of Pair Wise Atomic Interactions in Classical Simulations. *J. Comput. Chem.* **2009**, *30*, 1490−1498.
(67) Lavery, R.; Moakher, M.; Maddocks, J. H.; Petkeviciute, D.; Zakrzewska, K. Conformational analysis of nucleic acids revisited: Curves+. *Nucleic Acids Res.* **2009**, *37*, 5917−5929.
(68) Singh, S. K.; Koshkin, A. A.; Wengel, J.; Nielsen, P. LNA (locked nucleic acids): Synthesis and high-affinity nucleic acid recognition. *Chem. Commun.* **1998**, 455−456.
(69) Mergny, J.-L.; Lacroix, L. Kinetics and thermodynamics of i-DNA formation: Phosphodiester versus modified oligodeoxynucleotides. *Nucleic Acids Res.* **1998**, *26*, 4797−4803.

DOI: 10.1021/acsomega.7b00347
ACS Omega 2017, 2, 2165−2177