Articles

Triplex Formation on DNA Targets: How To Choose the Oligonucleotide

Pierre Vekhoff,[‡,§] Alexandre Ceccaldi,[‡,§] David Polverari,[‡,§] Jean Pylouster,[‡,§] Claudio Pisano,[∥] and Paola B. Arimondo*[,‡,§]

UMR 5153 CNRS, Muséum National d'Histoire Naturelle USM0503, 43 rue Cuvier, 75231 Paris cedex 05, France, INSERM UR565, 43 rue Cuvier, 75231 Paris cedex 05, France, and Sigma Tau SpA, Pomezia, Italy

Received June 9, 2008; Revised Manuscript Received September 12, 2008

ABSTRACT: Triplex-forming oligonucleotides (TFOs) are sequence-specific DNA binders. TFOs provide a tool for controlling gene expression or, when attached to an appropriate chemical reagent, for directing DNA damage. Here, we report a set of rules for predicting the best out of five different triple-helical binding motifs (TM, UM, GA, GT, and GU, where M is 5-methyldeoxycytidine and U is deoxyuridine) by taking into consideration the sequence composition of the underlying duplex target. We tested 11 different triplex targets present in genes having an oncogenic role. The rules have predictive power and are very useful in the design of TFOs for antigene applications. Briefly, we retained motifs GU and TM, and when they do form a triplex, TFOs containing G and U are preferred over those containing T and M. In the case of the G-rich TFOs, triplex formation is principally dependent on the percentage of G and the length of the TFO. In the case of the pyrimidine motif, replacement of T with U is destabilizing; triplex formation is dependent on the percentage of T and destabilized by the presence of several contiguous M residues. An equation to choose between a GU and TM motif is given.

In the past decade, the identification of genes that play a key role in the progression and maintenance of specific diseases, such as oncogenes or tumor suppressors in cancers, calls for the use of agents able to act on specific DNA sequences. Thus, there is a need for sequence-specific double-stranded DNA binding agents. Three groups of molecules are known to recognize DNA in a sequence-specific manner: zinc fingers, minor groove binders, and triplex-forming oligonucleotides (TFOs).[1]

Zinc fingers are derived from natural ligands that bind in the DNA major groove, the zinc finger proteins (*1*). Zinc fingers can be synthesized by the transcription/translation machinery of the cell from introduced vectors to produce a protein able to target a DNA sequence. The binding activity of the zinc finger can be covalently linked to a trans effector (*2*), such as a transcription factor (*3*) or a catalytic domain of a protein [nucleases (*4*) or methylases (*5*–*8*)]. This strategy suffers from the difficulties in selecting the best zinc fingers with high specificity (not all 3 bp are recognized) and cellular toxicity (*9*).

A number of small molecules bind specifically to the minor groove of the DNA helix. Minor groove binders (MGBs) include carboxyamide ligands of *N*-methylpyrrole and *N*-methylimidazole that form hydrogen bonds and adapt to the curvature of the DNA helix (*10*). A major limitation of the MGBs is their chemical stability and the length of the target DNA sequence (8−10 bp). Many attempts to increase the length of the target DNA to at least 16 bp have been made.

Finally, triplex-forming oligonucleotides (TFOs) are major groove ligands that target unique DNA sequences by forming DNA triple helices thanks to specific hydrogen bonding interactions between the TFO and the oligopurine strand of the duplex (*11*). The use of TFO is limited to the presence of oligopyrimidine•oligopurine sequences and by the stability of the triplex. Nevertheless, triplex-forming oligonucleotides constitute an interesting DNA sequence-specific tool for many applications. They have been used, for example, to target cleaving or cross-linking agents, transcription factors, or nucleases to a chosen site on the DNA (*12*–*14*). Moreover, they are used not only as biological tools to induce modifications on DNA with the aim of controlling gene expression, such as by site-directed mutagenesis in mice (*15*), but also as biotechnological tools in various assays, for example, to test translocation of proteins on DNA (*16*) or topoisomerase activity (*17*). In this work, we present a simple formula for predicting theoretically the oligonucleotide that will form a triplex on a given DNA sequence.

Three classes of triple helices exist that differ in sequence composition and relative orientation of the backbone of the third strand versus that of the oligopurine strand of the duplex (*18*). In the pyrimidine motif, which we here call a TC

* To whom correspondence should be addressed. E-mail: arimondo@mnhn.fr. Phone: +33 1 40793859. Fax: +33 1 40793705.
‡ UMR 5153 CNRS USM0503 MNHN.
§ INSERM UR565.
∥ Sigma Tau SpA.

[1] Abbreviations: TFO, triplex-forming oligonucleotide; M, 5-methyldeoxycytidine; U, deoxyuridine; MGB, minor groove binder; PNA, peptide nucleic acid; LNA, locked nucleic acid; UCSC, University of California at Santa Cruz; NCBI, National Center for Biotechnology Information; HEPES, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid; nt, nucleotide; T_m, temperature of half-dissociation; TDS, thermal differential spectra; DF, dispersion factor.

10.1021/bi801087g CCC: $40.75
Published on Web 10/28/2008

triplex, the third strand is parallel to the oligopurine strand of the oligopurine•oligopyrimidine duplex, forming T•A*T and C•G*C^+ triplets. The pK_a of the imino group of cytosine, which must be protonated, is well below 7, making TC triplex formation pH-dependent (*19*). In the GA triplex, the third strand is oriented antiparallel to the purine strand of the duplex and forms reverse Hoogsteen C•G*G and T•A*A triplets. In the GT triplex, the third strand can be either antiparallel to the purine strand of the duplex in forming reverse Hoogsteen C•G*G and T•A*T triplets or parallel in forming Hoogsteen C•G*G and T•A*T triplets (*20*). The binding of the TFO to the target duplex generally results in an interaction thermodynamically weaker than the one observed between the two strands of the duplex itself (*21*). To stabilize the triple helix, chemical modifications have been incorporated at the bases (mainly C and T) (*18, 22*), in the backbone (PNA) (*23*), or in the sugar [e.g., morpholino (*24*) and LNA (*25*)].

We have analyzed how to choose a TFO directed against a defined target oligopyrimidine•oligopurine DNA duplex sequence using oligonucleotides that are minimally modified (TM, UM, GA, GT, and GU TFOs). We evaluated TFOs with deoxycytidines substituted with 5-methyldeoxycytidine (M), because it has been shown that this modification increases the triplex stability and thus decreases the pH dependence (*26, 27*). We also tested the substitution of thymidine with deoxyuridine that is stabilizing in the case of intramolecular GU triplexes (*28*). Since the discovery of TFOs, various chemical modifications of the backbone, bases, or sugars have been introduced to increase the stability of the triple-helical structure. However, in our study, we limited our choice to least modified oligonucleotides (only two modified bases were used), because it is of major interest to be able to use for eventual therapeutical applications of the antigene strategy low-cost TFOs and, thus, preferentially unmodified ones. As target sequences, we chose regions in 11 genes involved in the development and maintenance of tumors (Table 1). The interest in using these targets is that the corresponding TFO may be eventually used to down-regulate the expression of these genes.

EXPERIMENTAL PROCEDURES

Genome Information. To find the oligopyrimidine•oligopurine tracts (or "target sequences") in the 11 chosen genes, we developed a bioinformatics motif search program based on C language. The oligopyrimidine•oligopurine target sequence must match the following criteria: 15−30 bp long with 25−65% guanine (G). Sequence information of the human genome was taken from the UCSC database (version hg15; April 10, 2003). Annotations of the genes were obtained from the NCBI database. Promoter regions were assumed to be the 1000 bp region upstream of the beginning of the gene. Results were stocked in a MySQL database, and a web interface was used to facilitate the access to the database. All results were verified by BLAST on the NCBI database and PatternN at http://bioinfo.hku.hk/services/analyseq/cgi-bin/patternn_in.pl.

Oligonucleotides and Nomenclature. All the DNA oligonucleotides presented here were synthesized by Eurogentec (Belgium) and dissolved in 200 μL of doubly distilled water. Concentrations of all oligodeoxynucleotides were estimated by UV absorption using extinction coefficients at 260 nm (*29*). For each duplex, five motifs of TFO were tested: TM, UM, GA, GU, and GT, where deoxyuridine (U) replaces thymidine (T) and 5-methyldeoxycytidine (M) replaces cytidine (C). The nomenclature used refers to the two bases in the TFO; for example, a TM TFO has thymidines and 5-methyldeoxycytidines, and the name of the gene target. KitGA is, for example, the GA TFO designed against the target sequence present in the *KIT* gene. TFOs were designed in a parallel orientation for TM and UM TFOs or in an antiparallel orientation for GA, GT, and GU TFOs. The orientation of the triple helix is defined as the orientation of the oligopurine strand of the duplex.

Gel Retardation Assay. The oligopyrimidine strand of each duplex was 5′-end-labeled with [γ-^{32}P]ATP (France; Arlington Heights, IL) with T4 polynucleotide kinase (New England Biolabs, Beverly, MA). In the screening test, 10 μM TFOs were added to 20 nM radiolabeled duplex in 10 mM $MgCl_2$, 100 mM NaCl, 50 mM HEPES (pH 7.2), 10% sucrose, and 0.5 mg/mL tRNA and incubated for 24 h at room temperature. Electrophoresis was performed on a nondenaturing 12% (19:1) polyacrylamide gel containing 10 mM $MgCl_2$ and 50 mM HEPES (pH 7.2) at 25 °C.

To determine the concentration at which 50% of the triplex was formed (C_{50}), TFO concentrations ranging from 50 nM to 10 μM were added to 20 nM radiolabeled duplex under the conditions described above. Samples were incubated at 37 °C for 2 or 24 h. Electrophoresis was performed at 37 °C as described above. Experiments were run in at least triplicate.

Melting Experiments. All thermal denaturation studies were carried out using quartz cuvettes with an optical path length of 0.2−1 cm on an Uvikon 940 (or XL) spectrophotometer interfaced with a computer for data collection and analysis. The temperature control of the cell holder was achieved with a Neslab RTE-111 (or Haake C25P) circulating water bath. The temperature of the water bath was decreased from 91.5 to 1.5 °C and increased back to 91.5 °C at a rate of 0.2 °C/min using the thermoprogramer. Samples were kept for 2 h at the lowest temperature and 10 min at the highest temperature. The absorbances at 245, 260, 295, and 305 nm were recorded every 5 min, whereas absorbance spectra were recorded in the 220−420 nm range with a scan speed of 500 nm/min and a data interval of 1 nm. Thermal denaturation and renaturation experiments were carried out on a sample containing 1 μM duplex and 1.2 μM corresponding TFO. For the analysis of self-association, the TFOs were studied in a range between 1 and 10 μM. All samples were prepared in 20 mM sodium or lithium cacodylate (pH 7.2), 100 mM NaCl, KCl, or LiCl, respectively, and 0−10 mM $MgCl_2$.

RESULTS

Choice of the Oligonucleotides. We chose 11 genes implicated in cancer progression (Table 1). First, we determined the uninterrupted oligopyrimidine•oligopurine tracts (the target sequences) present in these genes using the triplex database described in Experimental Procedures. The following criteria were used: (i) a length of at least 18 nt for the target sequence, (ii) at least 40% guanine in the target sequence, (iii) no long stretches of guanines (more

Table 1: Choice of the TFOs, Their Target Sequences, and Their Localization

	Genes	Duplex containing the target sequence [a]	Localisation	Potential TFO sequences (5'-3')	TFO name
1	BCL2	5'GTAAAAAAAGAGGAGAAGAAAAAAAAGTAATA 3' 3'CATTTTTTTTCTCCTCTTCTTTTTTTTCATTAT 5'	18q21.3 exon non codant strand	TTTTMTMMTMTTMTTTTTTTM	bcl2TM
				UUUUMUMMUMUUMUUUUUUUM	bcl2UM
				GTTTTTTTTGTTGTGGTGTTTT	bcl2GT
				GUUUUUUUUGUUGUGGUGUUUU	bcl2GU
				GAAAAAAAAGAAGAGGAGAAAA	bcl2GA
2	BIRC1	5'GGGAGAGAGAAAAAGAGGAAGGGAGAGAGG 3' 3'CCCTCTCTCTTTTTCTCCTTCCCTCTCTCC 5'	5q13.1 exon non codant strand	MTMTMTMTTTTTMTMMTTMMMTMTM	birc1TM
				MUMUMUMUUUUUMUMMUUMMMUMUM	birc1UM
				TGTGTGGGTTGGTGTTTTTGTGTGT	birc1GT
				UGUGUGGGUUGGUGUUUUUGUGUGU	birc1GU
				AGAGAGGGAAGGAGAAAAAGAGAGA	birc1GA
3	BIRC2	5'GATGAAAAAAGAGAAGAGGAGAAGGAAA 3' 3'CTACTTTTTTCTCTTCTCCTCTTCCTTT 5'	11q22 exon codant strand	TTTTTTMTMTTMTMMTMTTM	birc2TM
				UUUUUUMUMUUMUMMUMUUM	birc2UM
				GTTGTGGTGTTGTGTTTTTTT	birc2GT
				GUUGUGGUGUUGUGUUUUUU	birc2GU
				GAAGAGGAGAAGAGAAAAAA	birc2GA
4	BIRC3	5'GAAATAAGGGAAGAGGAGAGAGAAAGAGCAACTG 3' 3'CTTTATTCCCTTCTCCTCTCTCTTTCTCGTTGAC 5'	11q22 exon codant strand	UUMMMUUMUMMUMUMUMUUUMUM	birc3UM
				TTMMMTTMTMMTMTMTMTTTMTM	birc3TM
				GTGTTTGTGTGTGGTGTTGGGTT	birc3GT
				GUGUUUGUGUGUGGUGUUGGGUU	birc3GU
				GAGAAAGAGAGAGGAGAAGGGAA	birc3GA
5	EGFR	5'GAGTCGAGAAGGGAAGGAGAGAGAAAAACTTTTA 3' 3'CTCACCGCTTCCCTTCCTCTCTCTTTTTGAAAAT 5'	7p11.2 exon non codant strand	TTTTTTMTMTTMTMMTMTTM	egfrTM
				UUUUUUMUMUUMUMMUMUUM	egfrUM
				GTTGTGGTGTTGTGTTTTTTT	egfrGT
				GUUGUGGUGUUGUGUUUUUU	egfrGU
				GAAGAGGAGAAGAGAAAAAA	egfrGA
6	KIT	5'GAGGAAAGAGAGAGAGGAGAAGAAGGGTGC 3' 3'CTCCTTTCTCTCTCTCCTCTTCTTCCCACG 5'	4q11-12 intron codant strand	MTTTMTMTMTMMTMTTMTTM	kitTM
				MUUUMUMUMUMMUMUUMUUM	kitUM
				GTTGTTGTGGTGTGTGTGTTTG	kitGT
				GUUGUUGUGGUGUGUGUUUG	kitGU
				GAAGAAGAGGAGAGAGAAAG	kitGA
7	NFKB1	5'ATTAGAAGAGAAGGAGAAGGAAAAATAGAAG 3' 3'TAATCTTCTCTTCCTCTTCCTTTTTATCTTC 5'	4q24 intron codant strand	MTTMTMTTMMTMTTMMTTTTT	nfκb1TM
				MUUMUMUUMMUMUUMMUUUUU	nfκb1UM
				TTTTTGGTTGTGGTTGTGTTG	nfκb1GT
				UUUUUGGUUGUGGUUGUGUUG	nfκb1GU
				AAAAAGGAAGAGGAAGAGAAG	nfκb1GA
8	NFKB1ß	5'GACTTGGAGAAGGAAGAAGAGGAGAGTGAGG 3' 3'CTGAACCTCTTCCTTCTTCTCCTCTCACTCC 5'	19q13 exon codant strand	MMTMTTMMTTMTTMTMMTMTM	nfκb1ßTM
				MMUMUUMMUUMUUMUMMUMUM	nfκb1ßUM
				GTGTGGTGTTGTTGGTTGTGG	nfκb1ßGT
				GUGUGGUGUUGUUGGUUGUGG	nfκb1ßGU
				GAGAGGAGAAGAAGGAAGAGG	nfκb1ßGA
9	MET	5'CTCACAGAAAAGAGAAAAAAGAGATCCAC 3' 3'GAGTGTCTTTTCTCTTTTTTCTCTAGGTG 5'	7q31 exon codant strand	TMTTTTMTMTTTTTTMTMT	metTM
				UMUUUUMUMUUUUUUMUMU	metUM
				TGTGTTTTTTTGTGTTTTGT	metGT
				UGUGUUUUUUGUGUUUUGU	metGU
				AGAGAAAAAAGAGAAAAGA	metGA
10	RET	5'GAACCCGGAGAGAAAAAAAGGGGAAAGGTAAG 3' 3'CTTGGGCCTCTCTTTTTTTCCCCTTTCCATTC 5'	10q11.2 intron non codant strand	TTTMMMMTTTTTTTTMTMTMM	retTM
				UUUMMMMUUUUUUUUMUMUMM	retUM
				TTTGGGGTTTTTTTTGTGTGG	retGT
				UUUGGGGUUUUUUUGUGUGG	retGU
				AAAGGGGAAAAAAAAGAGAGG	retGA
				MMTMTMTTTTTTTTGGGGTTT	retTMTG[c]
				MMUMUMTTTTTTTTGGGGTTT	retUMTG[c]
11	VEGF(1)[b]	5'GAAGAAGAGAAGGAAGAGGAGAGGGGGCCG 3' 3'CTTCTTCTCTTCCTTCTCCTCTCCCCCGGC 5'	6p12 exon codant strand	TTMTMTTMMTTMTMMTMTMM	vegf1TM
				UUMUMUUMMUUMUMMUMUMM	vegf1UM
				GGTGTGGTGTTGGTTGTGTT	vegf1GT
				GGUGUGGUGUUGGUUGUGUU	vegf1GU
				GGAGAGGAGAAGGAAGAGAA	vegf1GA
12	VEGF(2)[b]	5'GAGGAGGAAGAAGAGAAGGAAGAGGAGAGG 3' 3'CTCCTCCTTCTTCTCTTCCTTCTCCTCTCC 5'	6p12 exon codant strand	TMMTTMTTTMTMTTMMTTMTM	vegf2TM
				UMMUUMUUMUMUUMMUUMUM	vegf2UM
				GTGTTGGTTGTGTTGTTGGT	vegf2GT
				GUGUUGGUUGUGUUGUUGGU	vegf2GU
				GAGAAGGAAGAGAAGAAGGA	vegf2GA

[a] The target sequence is underlined. [b] VEGF1 and VEGF2 are two target sequences present in the VEGF gene. [c] We used additional TFOs for *RET*. RetTMTG and retUMTG are two TFOs that bind in a parallel orientation that contain a 3′ part of GT motif and a 5′ part of the TM and UM motifs, respectively.

than four) and no multiple repetitions of stretches of three or more guanines, and (iv) no stretches of more than seven adenines. Criterion (i) was chosen to maximize the likelihood that the target sequence was unique in the genome. Criteria

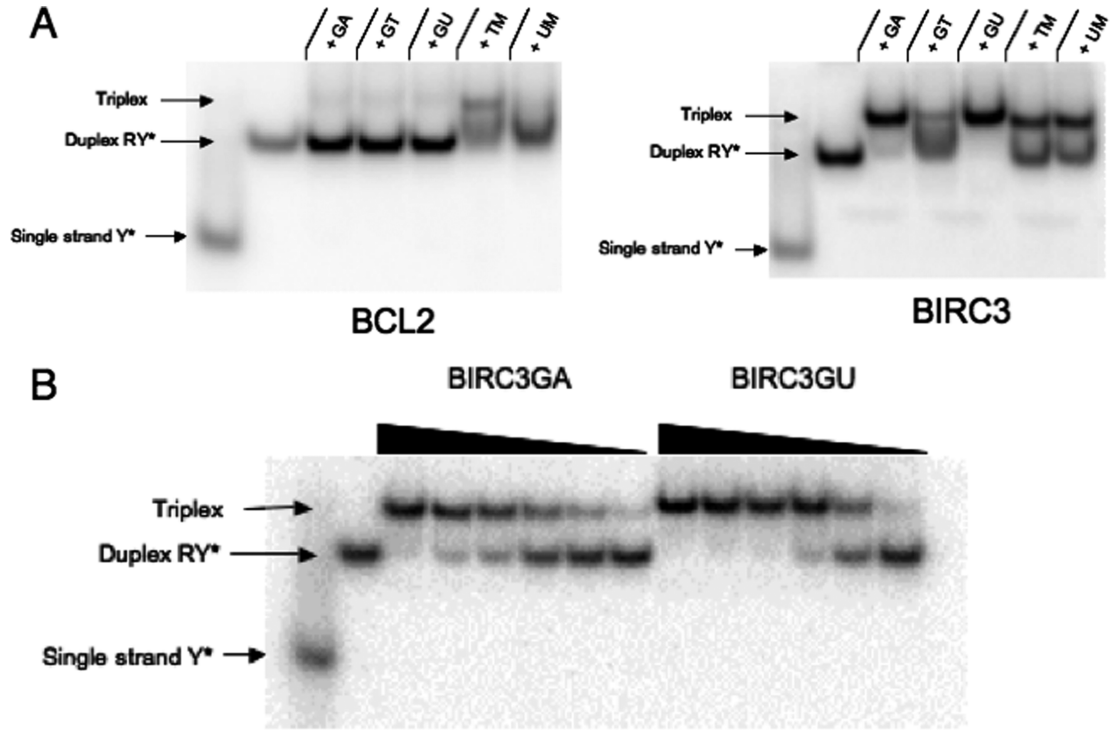


FIGURE 1: (A) Analysis of triplex formation by gel shift. Bcl2TFO or birc3TFO (10 μM) was added to 20 nM bcl2RY or birc3TFO duplex, respectively, radiolabeled on the oligopyrimidine strand (Y*) and analyzed after 24 h at room temperature on a 12% (19:1) polyacrylamide gel in 50 mM HEPES (pH 7.2) and 10 mM $MgCl_2$. (B) Concentration dependence of triplex formation. A range of concentrations, 10, 2, 1, 0.2, 0.1, and 0.05 μM, of birc3GA and birc3GU were added to 20 nM birc3RY duplex radiolabeled on the oligopyrimidine strand (Y*) and analyzed as described above after 24 h at 37 °C.

(ii), (iii), and (iv) were chosen to ensure stable triplex formation (*28*), while criterion (iii) was imposed to avoid as much as possible the formation of secondary structures such as quadruplexes of guanines that would compete with triplex formation (*30–33*). We ensured that sequences chosen were unique within the human genome by using BLASTN on the human genome sequence found on the NCBI website (http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=960_6). Each target sequence is unique in the human genome, except those for *MET* and *VEGF2* genes, which are present eleven and four times in the genome, respectively. The lengths of the TFOs are between 19 and 26 nt. Corresponding duplexes for gel retardation assays were designed using the gene sequence with the oligopyrimidine·oligopurine sequence positioned in the middle of the duplex sequence. Five TFOs were used for each target: TM and UM TFOs were designed in the parallel orientation, relative to the oligopurine strand of the target duplex, while GA, GU, and GT TFOs in the antiparallel orientation (Table 1).

Screening Test. We first determined the optimal TFO for each target, based on the assumption that the "optimal" TFO must form a stable triplex at pH 7.2 and 37 °C. The test consisted of assessing by gel shift experiments the ability of the TFO to form a triplex with the corresponding radiolabeled duplex after incubation for 24 h at 37 °C. Two examples are shown in Figure 1A. Data are summarized in Table 2. These data gave us a first indication about the rules that dictate triplex stability.

First, whatever the target sequence, a triple helix could be formed with at least one motif except for the *RET* target. Second, for pyrimidine motifs and the GA motif, all ranges of stability were observed: very stable (+++) and stable (++ and +) and no formation (−). In contrast, for the GU motif, a very stable triplex was formed (+++) or none at all (−): no intermediate stabilities were observed. The GT motif triplexes were always less stable than those formed by other motifs, in a manner independent of the orientation of the triplex, parallel or antiparallel (data not shown). In addition, whenever an antiparallel GT triplex was observed, the equivalent GU TFO formed a more stable triplex. Third, the two pyrimidine motif third strands (TM and UM) followed the same rules for triplex formation. In the case of the *BIRC1* and *BIRC3* targets, all five types of TFOs formed a triplex in the screening test, albeit with different affinities. Lastly, in the case of the *RET* target, we used additional TFOs, because no triplex was formed in the presence of the

Table 2: Triplex Formation after Incubation for 24 h at 37 °C and pH 7.2[a]

duplex	GA	GT	GU	TM	UM
BCL2	−	−	−	++	+
BIRC1	+++	+	+++	++	++
BIRC2	−	−	−	+	−
BIRC3	+++	+	+++	+	+
EGFR	+++	−	+++	−	−
KIT	+	−	+++	++	++
MET	−	−	−	+++	+++
NFKB1	+	−	−	++	++
NFKB1B	+++	+	+++	−	−
RET	+	−	−	−	−
VEGF1	++	+	+++	−	−
VEGF2	−	−	−	−	+

[a] The ability of the oligonucleotides to form a triplex with duplex target as measured by gel shift is indicated by a plus sign. The number of plus signs qualitatively indicates the capacity of the oligonucleotide to form a triple helix on its target: 100% of triplex formation is indicated by three plus signs and more than 50% triplex formation by two plus signs. A single plus sign indicates that less than 50% triplex was observed. No triplex formation is indicated by a minus sign.

Table 3: C_{50} Values Determined at pH 7.2 and 37 °C[a]

gene targeted	composition of the corresponding TFO	C_{50}(2 h) (μM)	C_{50}(24 h) (μM)
BCL2	TM	1.7 ± 0.2	1.2 ± 0.1
BIRC1	GA	0.3 ± 0.1	0.12 ± 0.02
	GU	0.45 ± 0.08	0.4 ± 0.1
	TM	2.6 ± 0.4	1.9 ± 0.2
	UM	>10	3.9 ± 0.6
BIRC3	GA	0.53 ± 0.06	0.24 ± 0.03
	GU	0.14 ± 0.02	0.13 ± 0.01
EGFR	GA	0.75 ± 0.03	0.76 ± 0.06
	GU	0.56 ± 0.14	0.63 ± 0.04
KIT	GU	0.73 ± 0.08	0.77 ± 0.07
	TM	1.0 ± 0.2	0.9 ± 0.2
	UM	1.0 ± 0.3	1.0 ± 0.3
MET	TM	0.60 ± 0.09	0.50 ± 0.02
	UM	3.1 ± 0.3	2.0 ± 0.2
NFKB	TM	>10	5.0 ± 0.6
	UM	>10	5.5 ± 0.6
NFKB1B	GA	0.41 ± 0.09	0.40 ± 0.08
	GU	0.14 ± 0.03	0.14 ± 0.02
VEGF1	GA	6.9 ± 0.5	7.2 ± 0.6
	GU	0.20 ± 0.02	0.20 ± 0.04

[a] Mean values of at least three experiments are reported with the standard error.

five TFO motifs. RetTMTG and retUMTG are two TFOs that bind in a parallel orientation that contain a 3′ part of the GT motif and a 5′ part of the TM and UM motifs, respectively. None of these designed TFOs formed a triplex on the *RET* target.

Triplex Stabilities. For each TFO oligonucleotide that formed a stable triplex (++ or +++), the affinity for the target was determined by evaluating triplex formation by gel shift analysis over a range of concentrations of TFO from 50 nM to 10 μM. We determined the concentration at which 50% of the triplex is formed (C_{50}) after incubation for 2 and 24 h at 37 °C (an example is shown in Figure 1B). Values are reported in Table 3.

Triplex formation was complete after incubation for 2 h at 37 °C for the purine motif triplexes (GA and GU), whereas the kinetics of triplex formation were slower for the pyrimidine third strands. This is in agreement with previous studies (*33*). Furthermore, the TM TFOs clearly form more stable triple helices than those with the UM motif. It has been suggested that methyl groups affect both the molecular polarizability of the base to which they are attached and the extent of hydrophobic interactions between neighboring atomic groups and solvent molecules (*34–36*). These two effects result in a lower affinity of binding of UM TFOs than of TM TFOs (*37*).

GA and GU TFOs bind with a similar stability on certain DNA targets (as in the case of EGFR, BIRC3, BIRC1, and NFKB1B) but with different stabilities on others (as for VEGF1 and KIT). No apparent rules explain these differences. It may be that self-association of the TFO interferes with triplex formation in certain cases since G-rich oligonucleotides are known to self-associate by forming GA DNA homoduplexes or G-quadruplexes. To investigate this hypothesis, the structures formed by the individual TFOs were analyzed by melting assays followed by UV absorbance spectroscopy under different salt conditions.

Characterization of TFO Secondary Structure by UV Absorbance Spectrophotometry. To study whether the oligonucleotides form a secondary structure that competes with triplex formation, the denaturation and reannealing profiles of the TFOs were followed by UV absorbance. The experiments were performed with different monovalent cations (Li^+, Na^+, or K^+), with various TFO concentrations (1, 2,

Table 4: Characterization of the Secondary Structures of the GA TFOs

Gene	Molecularity	Mg2+ (mM)	Tm (°C)	TDS
BIRC1	Inter	10	38	Homoduplex
		0	23	
BIRC3		10	36	
		0	<20	
KIT		10	28	
		0	<20	
EGFR		10	33	
		0	23	
NFKB1B		10	34	
		0	<20	
VEGF1		10	36	
		0	25	

Table 5: Characterization of the Secondary Structures of the GU TFOs

Gene	Monovalent cation	Tm (°C)	TDS in Na^+	Molecularity
BIRC1	Li^+	<20	N.C. [1]	N.D. [2]
	Na^+			
	K^+			
BIRC3	Na^+	<20	Quadruplex	
KIT	Li^+	<20	N.C. [1]	Intra
	Na^+	22		
	K^+	28		
EGFR	Li^+	<20	Quadruplex	
	Na^+	27		
	K^+	32		
NFKB1B	Li^+	<20		
	Na^+	28		
	K^+	36		
VEGF1	Na^+	34		

[1] Not conclusive. [2] Not possible to determine.

and 10 μM), and with different Mg^{2+} concentrations (0 or 10 mM). The nature of the monovalent cation is known to affect G-quadruplex formation (*38*). The TFO concentration dependence gives an indication of the molecularity of the structure (intra- or intermolecular), whereas Mg^{2+} is essential for GA homoduplex formation (*39*). By recording the thermal differential spectra (TDS) for each TFO alone, we determined the nature of the structures (*40*). The TDS is obtained by the subtraction of the UV absorbance spectrum of the unfolded state (spectrum recorded at a temperature well above the melting temperature, T_m) and the spectrum of the folded state (recorded well below its T_m). The T_m is defined as the temperature at which 50% of the structure is unfolded.

The TM and UM TFOs did not form any stable secondary structure above 10 °C that could compete with triplex formation (data not shown). In contrast, the six GU TFOs and the six GA TFOs studied appear to form secondary structures. Results are summarized in Tables 4 and 5 (and in Tables S1 and S2 of the Supporting Information). The T_m values of the secondary structures were dependent on the TFO concentration for GA TFOs, indicating the formation of intermolecular structures. In contrast, secondary structures of GU TFOs were not concentration-dependent, suggesting an intramolecular self-association.

The melting temperatures of GA TFOs were dependent on the TFO and Mg^{2+} concentrations (Figure 2A,B). The structures of GA TFOs were similar, whatever monovalent cation was used (Na^+ or K^+). The TDS of the GA TFOs were characterized, under all experimental conditions, by a negative peak at 290 nm and a major positive peak at 258 nm (Figure 2C) and were quasi-superimposable for all TFOs. The profile corresponds to the one of a GA homoduplex (*40*), a complex in which the GA TFO forms a bimolecular parallel duplex mainly by G·G and A·A base pairing. In conclusion, the Mg^{2+} dependence, the lack of an effect of the monovalent cation used, and the TDS strongly indicate that the secondary structures formed by the GA TFOs were GA homoduplexes.

The stabilities of the secondary structures were different for each TFO; the highest T_m, of 38 °C, was observed for birc1GA at 10 μM in 10 mM Mg^{2+}. We tested, for birc1GA, vegf1GA, and birc3GA that form the most stable secondary structures, whether the homoduplex could compete with triplex formation at 37 °C, and we observe no competition under our experimental conditions (data not shown).

The identification of the secondary structure formed by the GU TFOs was not straightforward (Table 5). The T_m values of the secondary structures do not depend on the TFO concentration (Figure 3A). For birc1GU and birc3GU, it was not possible to determine the molecularity, because the structures were not sufficiently stable. vegf1GU, nfkb1bGU, and birc3GU formed G-quadruplexes as identified by the TDS signature (Figure 3B): positive peaks at 243 and 273 nm (±2 nm) and a major negative peak at 295 nm (±1 nm) as observed for known quadruplexes (*40*).

The TDS of the other GU TFOs did not present a clear signature in the buffer used for the gel retardation assays, which contained Na^+. To determine which structure is formed, we then measured the TDS in K^+ or Li^+. K^+ strongly stabilized quadruplex formation, and Li^+ destabilized quadruplexes (Figure 3C). In the presence of K^+, the TDS for egfrGU and birc1GU were typical of a quadruplex of guanines (Figure 3D), whereas the TDS profile for kitGU,

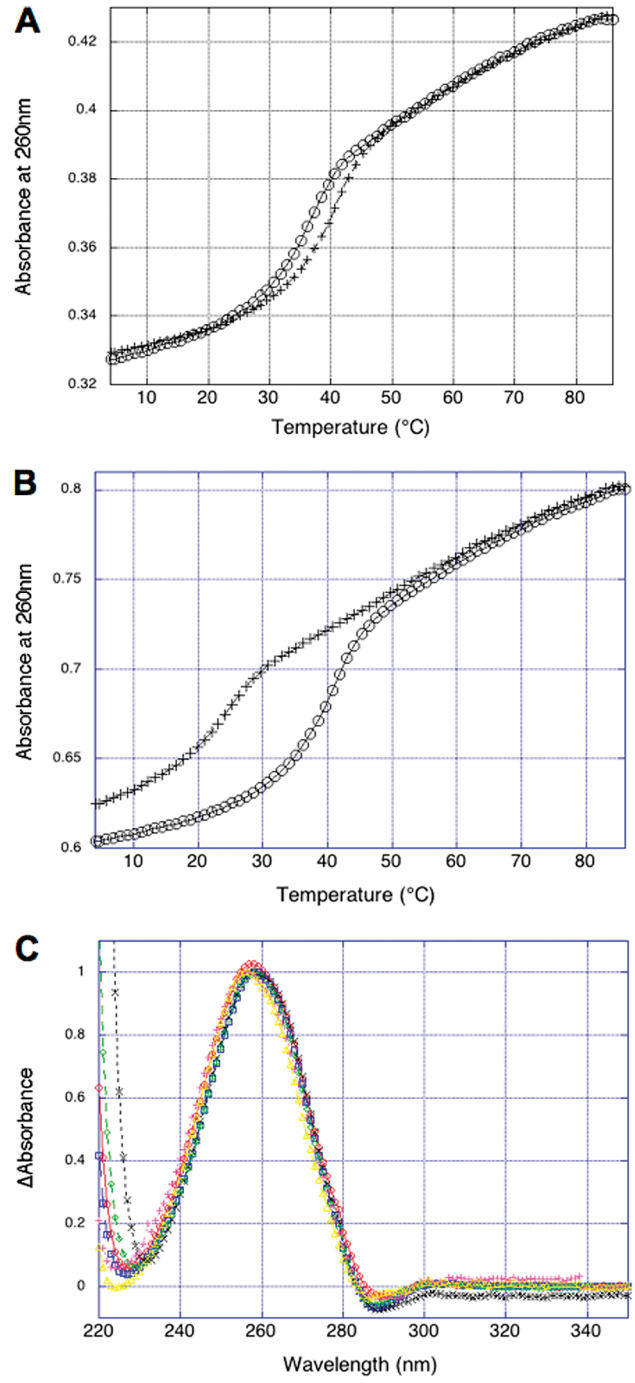


FIGURE 2: Self-association of GA TFOs. Thermal denaturation profiles followed at 260 nm by UV absorbance spectrophotometry measured in 10 mM Na^+ cacodylate (pH 7.2) and 100 mM NaCl of (A) TFO birc1GA (1 μM) with (○) or without (+) 10 mM $MgCl_2$ and (B) TFO birc1GA at 1 (○) or 10 μM (+) in 10 mM $MgCl_2$. (C) Thermal differential spectra of birc1GA (blue), vegfGA (red), kitGA (green), egfrGA (black), nfkb1bGA (purple), and birc3GA (yellow) at 10 μM in 10 mM Na^+ cacodylate (pH 7.2), 100 mM NaCl, and 10 mM $MgCl_2$.

even in the presence of K^+, was not conclusively that of a quadruplex (Table 5). A slight Mg^{2+} dependence was observed. It is likely that different structures coexist for the kit GU TFO. In the presence of Li^+, no secondary structure was observed above 10 °C for kitGU, egfrGU, birc1GU, or nfkb1bGU.

In general, the T_m values of the secondary structures were lower than those for GA TFOs, indicating that GU TFOs

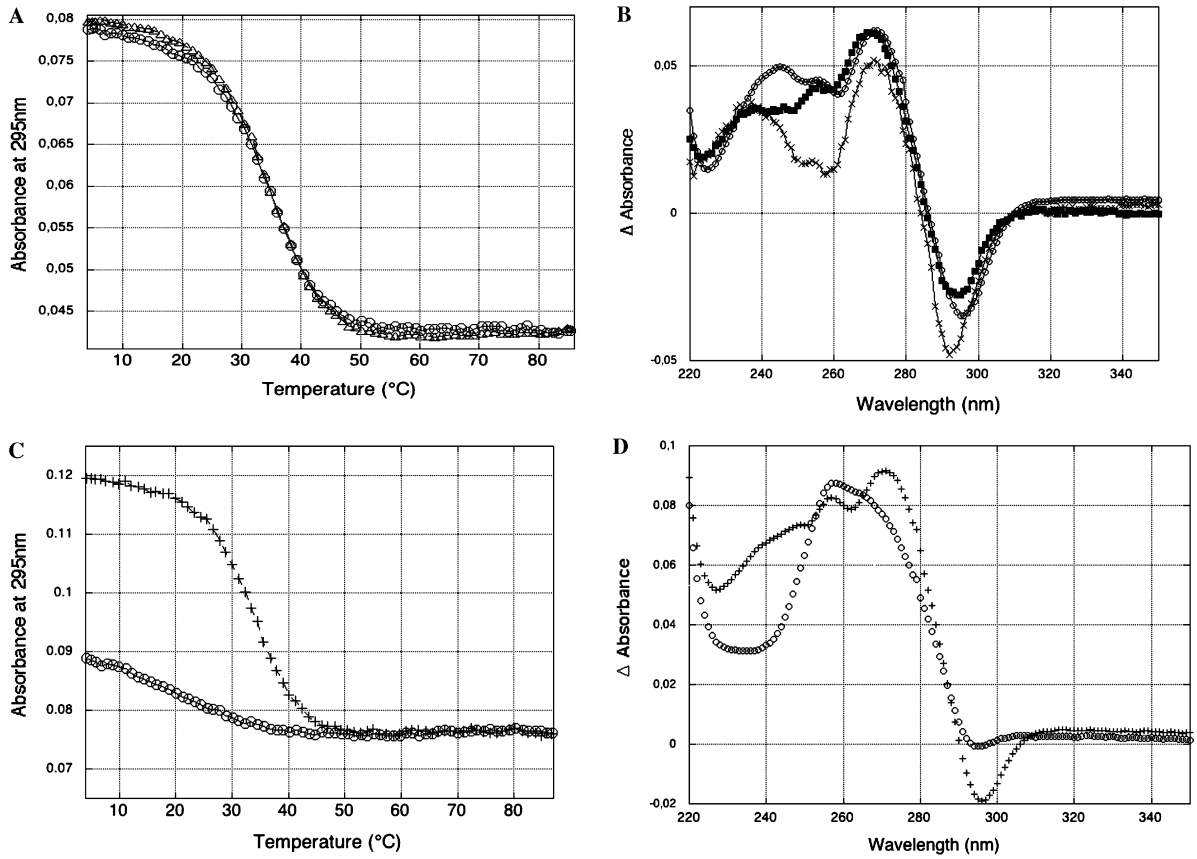


FIGURE 3: (A) Thermal denaturation profiles followed at 295 nm by UV absorbance spectrophotometry of TFO vegf1GU in 10 mM Na^+ cacodylate (pH 7.2), 10 mM $MgCl_2$ and 100 mM NaCl at 1 (○) or 10 μM of TFO (△). (B) Thermal differential spectra of nfkb1BGU (×), birc3GU (■), and vegf1GU (○) in 10 mM Na^+ cacodylate (pH 7.2), 10 mM $MgCl_2$ and 100 mM. (C) Thermal denaturation of vegf1GU followed at 295 nm by absorbance spectrophotometry in 10 mM Li^+ cacodylate (pH 7.2) and 100 mM LiCl (○) or in 10 mM K^+ cacodylate (pH 7.2) and 100 mM KCl (+). (D) Thermal differential spectra of egfrGU in 10 mM Li^+ cacodylate (pH 7.2) and 100 mM LiCl (○) or in 10 mM K^+ cacodylate (pH 7.2) and 100 mM KCl (+).

Table 6: Features of the GA TFOs and Correlations with Triplex Structure Formation

TFOs	nt	G	G content	GA	AG	AA	GG	GGG	G at 5′	G at 3′	C_{50}(24 h) (μM)	T_m of competitive structure
						Stable Triplex						
birc1GA	26	12	0.46	8	9	5	3	1	0	1	0.12 ± 0.02	36
birc3GA	23	11	0.48	8	7	4	3	1	1	0	0.24 ± 0.03	30
nfkb1bGA	21	11	0.52	7	7	3	3	0	1	2	0.40 ± 0.08	32
egfrGA	23	10	0.43	6	7	7	3	1	0	1	0.76 ± 0.06	30
vegf1GA	20	10	0.50	7	6	3	3	0	2	0	7.2 ± 0.6	33
average	22.6	10.8	0.48	7.2	7.2	4.4	3.0	0.6	0.8	0.8		
ratio to G			1	15.0	15.0	9.2	6.3	1.3	1.7	1.7		
						No Triplex						
kitGA	22	10	0.45	8	8	4	1	0	1	1	≫10	
bcl2GA	22	6	0.27	5	4	11	1	0	1	0	≫10	
nfkbGA	21	8	0.38	5	6	7	2	0	0	1	≫10	
vegf(2)GA	20	9	0.45	7	6	4	2	0	1	0	≫10	
retGA	20	8	0.40	3	4	8	4	2	0	1	≫10	
birc2GA	20	7	0.35	6	5	7	1	0	1	0	≫10	
metGA	19	5	0.26	5	5	8	0	0	0	0	≫10	
average	20.6	7.6	0.37	5.6	5.4	7.0	1.6	0.3	0.6	0.4		
ratio to G			1	15.2	14.8	19.1	4.3	0.8	1.6	1.2		

are less prone than GA TFOs to self-associate. This represents an advantage of GU TFOs over GA TFOs.

Next we searched for simple rules to guide the choice of the TFO based on the sequence composition of the underlying duplex target.

Rules for the Choice of the GA TFOs. To elucidate the elements that are favorable to triplex formation, we have compared the following features: length of the TFO, G content, ApA, GpA, ApG, and GpG steps, stretches of three guanines, and the number of guanines at the 5′ end or 3′ end (Table 6). The TFOs were classified according to their ability to form a triplex (C_{50} at 24 h). The average values for the different motifs were normalized to the G content (number of guanines per length) of the oligonucleotide. The comparison of the normalized average for the oligonucleotides that form a triplex and those that do not shows that three features are important for triplex formation in the case of the GA TFOs. (1) A high G content is necessary for triplex formation; however, if the G content is too high, triplex formation is disfavored due to the competition between

Table 7: Features of the GU TFOs and Correlations with Triplex Structure Formation

TFOs	nt	G	G content	GU	UG	UU	GG	GGG	G at 5′	G at 3′	C_{50}(24 h) (μM)
					Stable Triplex						
birc3GU	23	11	0.48	8	7	4	3	1	1	0	0.13 ± 0.01
nfkb1bGU	21	11	0.52	7	7	3	3	0	1	2	0.14 ± 0.02
vegf1+GU	21	11	0.52	7	7	3	3	0	2	1	0.18 ± 0.05
birc3++GU	24	12	0.50	8	8	4	3	1	1	1	0.20 ± 0.06
vegf1GU	20	10	0.50	7	6	3	3	0	2	0	0.20 ± 0.04
birc3+GU	21	11	0.52	7	7	3	3	1	1	3	0.35 ± 0.08
birc1GU	26	12	0.46	8	9	5	3	1	0	1	0.39 ± 0.1
egfrGU	23	10	0.43	6	7	7	3	1	0	1	0.63 ± 0.04
LIMIT30GU12	30	12	0.40	9	10	8	2	0	0	1	0.69 ± 0.09
kitGU	22	10	0.45	8	8	4	1	0	1	1	0.77 ± 0.07
average	23	11	0.48	7.5	7.6	4.4	2.7	0.5	0.9	1.1	
ratio to G%			1	15.6	15.8	9.2	5.6	1.0	1.9	2.3	
					No Triplex						
LIMIT26GU11	26	11	0.42	8	9	6	2	0	0	1	≫10
LIMIT30GU11	30	11	0.37	9	10	9	1	0	0	1	≫10
LIMIT23GU9	23	9	0.39	6	7	7	2	0	0	1	≫10
bcl2GU	22	6	0.27	5	4	11	1	0	1	0	≫10
LIMIT21GU10d	21	10	0.48	6	7	4	3	0	0	1	≫10
LIMIT21GU10c	21	10	0.48	6	7	4	3	0	0	1	≫10
LIMIT21GU10b	21	10	0.48	7	6	5	3	0	1	1	≫10
LIMIT21GU10a	21	10	0.48	7	8	3	2	0	0	1	≫10
LIMIT21GU9	21	9	0.43	6	7	5	2	0	0	1	≫10
nfkbGU	21	8	0.38	5	6	7	2	0	0	1	≫10
birc2+GU	21	8	0.38	6	6	7	1	0	1	1	≫10
birc2aGU	21	8	0.38	7	7	6	0	0	1	1	≫10
vegf2GU	20	9	0.45	7	6	4	2	0	1	0	≫10
retGU	20	8	0.40	3	4	8	4	2	0	1	≫10
birc2GU	20	7	0.35	6	5	7	1	0	1	0	≫10
metGU	19	5	0.26	5	5	8	0	0	0	0	≫10
average	22	8	0.40	6.2	6.5	6.3	1.8	0.1	0.4	0.8	
ratio to G%			1	15.5	16.3	15.8	4.5	0.3	0.9	1.9	

triplex formation and self-association. (2) The average number of ApA motifs is 4.4 for the TFOs that form triplexes and 7.0 for those that do not (ratio for triplex formed to not formed is 0.5). Therefore, A stretches seem to be penalizing. (3) Finally, analysis of the bases at the 3′ and 5′ ends of the TFOs (ratios of 1.4 and 1 for formed to not formed, respectively) shows that a G at the 3′ end increases the extent of triplex formation, in agreement with previous observations (*39*).

In summary, GA TFOs form more stable secondary structures than the corresponding GU TFOs and form similar or less stable triplexes, with the exception of birc1 TFOs.

Rules for the Choice of the GU TFOs. The following parameters were considered: length of the TFO, content of G, number of stretches of three guanines (there were no stretches of four guanines with the exception of vegf2GU), and number of UpG, GpU, and GpG motifs (Table 7). The TFOs were sorted by triplex stability using C_{50} values at 24 h. First, the TFOs that formed the most stable triplex were a 23-mer having 48% G (birc3) and a 21-mer with 52% G (nfkb1b). Strikingly, there was a minimal barrier of G necessary to ensure triplex formation (40−50%). Above this percentage, TFOs formed stable triplexes, and below, they did not. The stabilizing effect of the G percentage (*%G*) is further confirmed by a Spearman correlation between C_{50} and *%G* of −0.704 (p value = 0.029 with $\alpha = 0.05$) (see the Supporting Information and Table S3), performed only on the TFOs that form stable triplexes. This correlation factor is further confirmed when all TFOs are included in the calculation. Accordingly, the percentage of G needed was dependent on the length of the TFO: longer TFOs that formed stable triplexes had lower G content than shorter TFOs.

Table 8: Sequences of Additional GU TFOs (5′ to 3′)[a]

LIMIT20GU9	GUUGUGGUGUUGUGU**G**U**G**UU
LIMIT21GU9	U**G**UUUGGUUGUGGUUGUGUUG
LIMIT21GU10a	U**GUG**UGGUUGUGGUUGUGUUG
LIMIT21GU10b	GGUUUGGUUGUGGUUGUGUUG
LIMIT21GU10c	U**GG**UUGGUUGUGGUUGUGUUG
LIMIT21GU10d	UU**GG**UGGUUGUGGUUGUGUUG
LIMIT23GU9	UUUUUGUGUGUGGUU**U**GGUUGUG
LIMIT26GU11	UGUGU**U**GGUUGGUGUUUUUGUGUGUG
LIMIT30GU11	UGUGUU**U**GUGUUGGUUGUUUUUGUGUUGUG
LIMIT30GU12	UGUGUU**GG**UGUUGGUUGUUUUUGUGUUGUG
vegf1+GU	GGUGUGGUGUUGGUUGUGUU**G**
birc3+GU	GUGUUUGUGUGUGGUGUUGGG
birc3++GU	GUGUUUGUGUGUGGUGUUGGGUU**G**
birc2+GU	GUUGUGGUGUUGUGUUUUUU**G**

[a] LIMIT20GU9 was designed from birc2GU with two U substitutions. LIMIT21GU9 and LIMIT21GU10 (a−d) were designed from nfkbGU. LIMIT23GU9 was designed from egfrGU via replacement of a G with a U. LIMIT26GU11 was designed from birc1GU. LIMIT30GU11 and LIMIT30GU12 were designed from birc1GU. In bold are the substituted bases.

To further characterize this barrier, we designed several additional GU TFOs containing 20, 21, 23, 26, and 30 nt (Table 8). To increase the percentage of G, we replaced one or more U residues with G in previously analyzed sequences, and vice versa, G with U, to decrease the percentage of G. The names of these TFOs derive from the length of the TFO and the number of guanines. LIMIT20GU9 is therefore a 20 nt TFO that contains nine guanines. The corresponding duplex targets were designed.

As shown in Table 7, the 20-mer TFO vegf1GU that contains 10 guanines out of 20 nt (50% G) forms a triplex, whereas birc2GU with seven guanines (35% G) and retGU with eight guanines (40% G) do not; a GU TFO with nine guanines (LIMIT20GU9 with 45% G) also does not form a

Table 9: Features of the TM TFOs Classified According to Their Ability To Form a Triplex

TFO	nt	*M*	T content	MT	TM	TT	MM	TTT	MMM	C_{50}(24 h) (μM)
					Stable Triplex					
metTM	19	8	0.74	6	5	7	2	3	0	0.50 ± 0.02
kitTM	22	12	0.55	9	8	5	3	3	1	0.9 ± 0.2
bcl2TM	22	6	0.73	4	5	11	1	8	0	1.2 ± 0.1
birc1TM	26	10	0.54	8	8	4	1	1	0	1.9 ± 0.2
nfkbTM	21	5	0.62	5	5	8	0	6	0	5.0 ± 0.6
average	21	8	0.63	6.4	6.2	7.0	1.4	4.2	0.2	
ratio to %T			1	10.1	9.8	11.1	2.2	6.6	0.3	
					Unstable Triplex					
birc3TM	23	11	0.52	7	8	4	3	1	1	−10
birc2TM	20	7	0.65	5	6	7	1	4	0	−10
average	22	9	0.59	6.0	7.0	5.5	2.0	2.5	0.5	
ratio to %T			1	10.2	11.9	9.4	3.4	4.3	0.9	
					No Triplex					
nfkb1bTM	21	11	0.48	7	7	3	3	0	0	≫10
vegf2TM	20	9	0.55	6	7	4	2	0	0	≫10
egfrTM	23	10	0.57	7	6	6	3	3	1	≫10
vegf1TM	20	10	0.50	6	7	3	3	0	0	≫10
retTM	20	8	0.60	4	3	8	4	6	2	≫10
average	21	10	0.54	6.0	6.0	4.8	3.0	1.8	0.6	
ratio to %T			1	11.1	11.1	8.9	5.6	3.3	1.1	

triplex. Thus, for a 20-mer TFO, at least ten guanines (or 50% G) are needed.

In the case of 21-mers, a TFO containing 52% G (nfkb1bGU), i.e., containing eleven guanines, formed a triplex, whereas TFOs with eight guanines (nfkbGU, 38% G) and nine guanines (LIMIT21GU9, 43% G) did not. The barrier is ~48% G since LIMIT21GU10d formed a triplex, whereas LIMIT21GU10a, -b, and -c do not (all containing ten guanines).

In the case of 23-mers, a 39% G-containing TFO (nine guanines) did not form a triplex (LIMIT23GU9). For this length, the percentage of G must be >44% (egfrGU, ten guanines) for triplex formation.

A 26 nt TFO containing twelve guanines (birc1GU, 46% G) formed a triplex, whereas a TFO with eleven guanines (LIMIT26GU11, 42% G) did not. For this length, the number of guanines must be >12, i.e., 46% G.

A 30-mer with 37% G did not form a triplex (limit30GU11); 40% G content is needed to form a triplex with this length (limit30GU12). In summary, the G percentage is clearly critical for triplex formation: between 40% for 30-mer TFOs and 50% for 20-mer TFOs.

Among the six motifs that were analyzed (Table 7), GpG was the only motif that played an important role in triplex formation. The higher the number of GpG motifs in the TFO, the more stable the triplex. Indeed, the Spearman correlation between C_{50} and the number of GpG motifs is −0.703 [p value = 0.029 with $\alpha = 0.05$ (Table S3)]. This correlation factor is further confirmed when all TFOs are included in the calculation.

Another important parameter is the distance between the GpG pairs. The closer these dinucleotide motifs are to each other, the higher the affinity of the TFO for the duplex target. To further investigate the influence of the dispersion of the G content on the ability of the GU TFO to form a stable triplex, we defined a dispersion factor of G [DF(G)] as the number of nucleotides between stretches of 1 to n guanines divided by the number of intervals present in the sequence and normalized to the percentage of G. The Spearman correlation between C_{50} and the G dispersion factor is 0.689 (p value = 0.032 with $\alpha = 0.05$) for GU TFOs that form a triplex and 0.726 (p value < 0.0001 with $\alpha = 0.05$) for all GU TFOs. This means that for a given percentage of G, the less dispersed the stretches of G in the sequence are, the more stable the GU triplex.

A G at the 3′ end of the oligonucleotide is known to stabilize GA triplexes (*39*). We evaluated the stabilities of several TFOs with 3′ guanines in comparison to TFOs without. Vegf1+GU was designed from vegf1GU by adding a G at the 3′ end (Table 8). Importantly, all added bases were chosen so that they still form a triplet on the duplex target. We evaluated two sequences similar to birc3GU. We added one G for birc3++GU, and we removed 2 uridines to create birc3+GU. There was no clear effect of a G at the 3′ end of the GU TFO: it can slightly stabilize as in the case of vegf1+GU (where one G was added) or slightly destabilize as in the case of birc3++GU (where one G was also added). When 2 uridines were deleted at the 3′ end of birc3GU (birc3+GU) to leave a shorter TFO with a G at the 3′ end, the TFO was 3-fold less stable than the triplex formed by the parent. In the case of GA triplexes, the equivalent situation, where the deletion of two adenines leaves a shorter TFO with a G at the 3′ end, greatly stabilizes the triple-helical structure (*39*). On the other hand, Table 7 indicates that the presence of a G at the 5′ end of the GU TFOs is stabilizing. Accordingly, the Spearman correlation between the C_{50} and the number of guanines at the 5′ end is −0.461 for all TFOs (p value = 0.019 with $\alpha = 0.05$ [Table S3 of the Supporting Information)], which is consistent with a 5′ end stabilizing effect. This was an unexpected result and underscores another difference between GU and GA triplexes.

Rules for the Choice of TM TFOs. Of the eleven DNA sequences (Table 2), not all can be targeted by G-containing TFOs; therefore, we also analyzed the TM TFOs. The UM TFOs were excluded since they seem to follow the same rules as the TM TFOs but form less stable triplexes. In Table 9, the length of the TFO, the T content, and the number of the four possible motifs (MpM, MpT, TpM, and TpT) are

reported for each TFO sorted by their triplex stability (determined by C_{50} values at 24 h).

There was not a significant correlation between the percentage of T and the length of the TFO for triplex stability. Nevertheless, motifs containing a higher percentage of T were present in more stable triplexes. In the most stable triplexes, the content of MpT and TpM motifs (normalized to the amount of T) is quite constant, while the number of TpT motifs decreases (from 11.1 to 8.9) as stability decreases and the MpM motif is found most often in least stable triplexes (from 2.2 to 5.6). This is in agreement with literature reports that A•T*T triplets are much more stabilizing at pH >7.0 than G•C*C$^+$ triplets (*41*). Here, the A•T*T triplet is also much more stabilizing than the G•C*M + triplet. Thus, T-rich motifs stabilize triplex formation, whereas M-rich motifs, especially MpMpM, are very destabilizing.

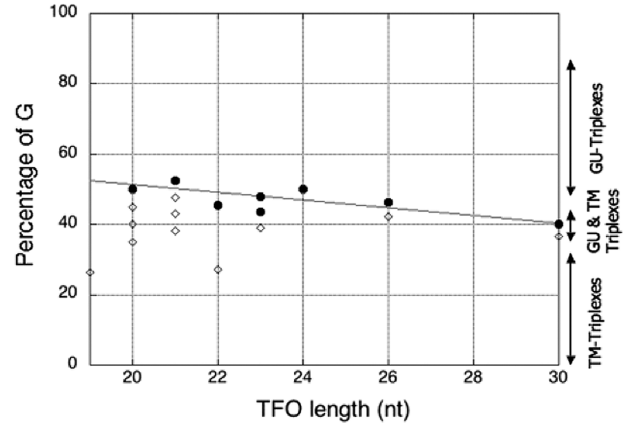


FIGURE 4: Representation of the percentage of guanine as a function of the length of the TFO in nucleotides. The circles correspond to data for the GU TFOs that form a stable triplex and the diamonds to the ones that do not. The straight line represents the limit between the GU (above) and TM triplexes (below).

DISCUSSION

We have analyzed the stability of five different triplex-forming motifs targeting oligopyrimidine•oligopurine sequences found in 11 genes implicated in cancer. We compared TFOs of motifs TM, UM, GA, GT, and GU. TFOs were designed in a parallel orientation, relative to the oligopurine strand of the target duplex, for TM and UM TFOs or in an antiparallel orientation for GA, GT, and GU TFOs. In an initial gel shift analysis, GT and UM TFOs formed the least stable triplexes; thus, to target a gene, GA, GU, and TM TFO motifs should be preferentially used.

Importantly, all our experiments were performed under near-physiological conditions (pH 7.2, 37 °C, 100 mM NaCl, and 10 mM $MgCl_2$). Incubation conditions are known to affect thermodynamic parameters of triplex formation, so the rules we have developed apply to these conditions.

As expected from the literature, there is a barrier of G percentage (40−50%) below which TFOs do not form stable triplexes. G-Rich TFOs can be useful because the kinetics of triplex association is fast (Table 3), C_{50} values are lower than those for pyrimidine TFOs, and the latter do not always form triplexes. G-Rich oligonucleotides can form, however, competitive secondary structures. The choice, in fact, of G-rich oligonucleotides must be a careful compromise between triplex formation and self-association (often G-quadruplex formation). For this reason, we did not evaluate sequences containing G stretches of more than four residues. It has been shown, for example, that above 70% G in the TFO, triplex formation was inhibited by K^+ cations because these ions stabilize G-quadruplex formation (*42*).

In a preliminary screen for triplex formation, the use of GA TFOs is justified for cost reasons; however, GU TFOs should be preferred in triplex applications because of various advantages. The replacement of a U with an A in a G-rich third strand modifies the nature of the competing secondary structure and thus its stability: GU TFOs are less prone to self-association. Interestingly, in the case of the GA TFOs, we observe the same trend of stability for homoduplex and triplex: birc1GA > birc3GA > nfkb1bGA > egfrGA > kitGA. An exception is vegf1GA that forms a stable secondary structure but a weak triplex (discussed below). Various differences in stability are observed between GA and GU triplexes. A significant difference between GA and GU TFOs is that GU TFOs form either very stable triplexes or none at all, whereas GA TFOs form triplexes with a range of stabilities (Table 2). Furthermore, on one hand, the analysis we have conducted confirmed that GA triplexes are stabilized by a G at the 3′ end of the TFO. On the other hand, in the case of the GU triplexes, it showed, for the first time, that a 3′-terminal G did not have a stabilizing effect but a 5′ G did. This is well-illustrated by the triplexes formed on the *VEGF(1)* target. The GU TFO forms a 35-fold stronger triplex than the GA TFO, where the presence of two adenines at the 3′ end are very destabilizing and two guanines at the 5′ end have no effect; the situation is reversed on the GU triplex, where two uridines at the 3′ end have no effect and the presence of two guanines at the 5′ end are stabilizing. Finally, GU TFOs form the most stable triplexes, with one exception (*BIRC1*). Altogether, these observations underline the differences between GA and GU triplexes and the advantage of the latter.

In several cases, GU TFOs did not form a stable triplex on the target sequence. Therefore, it is important to consider pyrimidine third strands. We considered only TM TFOs, because UM TFOs formed less stable triplexes in our initial analyses. The next step was thus to find a simple equation that allowed us to choose between a GU and a TM TFO on the basis of the sequence of the duplex target. When we graph (Figure 4) the percentage of G in the TFOs as a function of TFO length (in nucleotides), we can draw a straight line through the points that correspond to stable triplexes (circles), described by the equation

$$G_N = 73 - 1.1L \quad (R = 0.786) \qquad (1)$$

where G_N corresponds to the percentage of G necessary to form a triplex and L corresponds to the TFO length. This represents the limit between the formation of TM and GU triplexes under our experimental conditions (notably at pH 7.2 and 37 °C). Practically, if the percentage of G (*%G*) in the oligopurine sequence of the target is higher than the G_N calculated by eq 1, a GU TFO will form a stable triplex; below that, a TM TFO should be chosen, as for values around the G_N value both motifs can be used. This rule is valid only in the range of TFO lengths that we have tested (i.e., between 19 and 30 nt) and, as stated above, under the experimental conditions that were tested (pH 7.2, 100 mM NaCl, and 10 mM $MgCl_2$) and for targets containing stretches of G of fewer

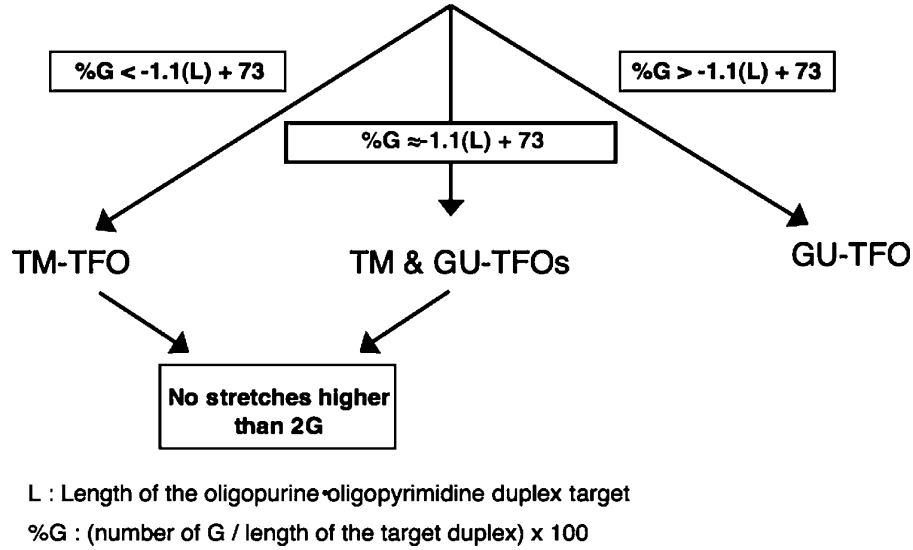
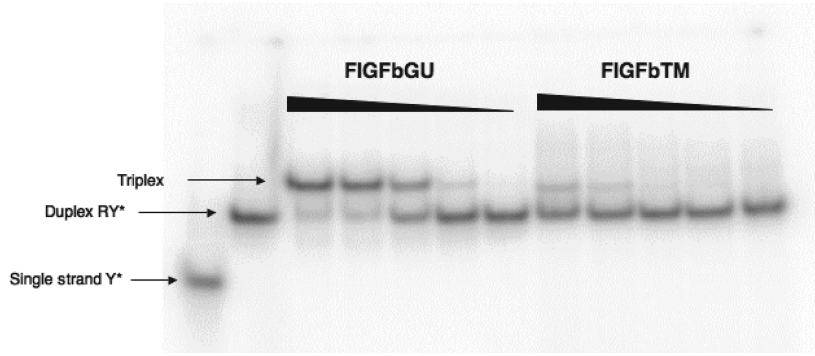


FIGURE 5: Synthetic representation of the steps for the design of the appropriate TFO for a given target sequence.

A)

Gene	Duplex containing the target sequence	Localisation	TFO sequences (5'-3')	TFO
FIGF	5′ AAGTGGAGAAGG AGGAGAGAGAGAGAGAGAG TGTG 3′ 3′ TTCACCTCTTCC TCCTCTCTCTCTCTCTC ACAC 5′	Xp22.31	TMMTMTMTMTMTMTMTMTM	figfbTM
			GUGUGUGUGUGUGUGGU	figfbGU

B)



C)

TFOs	C50 (μM) (2h)	C50 (μM) (24h)
figfbTM	>10	8.5 ±2.0
figfbGU	0.2 ±0.1	0.2 ±0.1

FIGURE 6: FIGF target. (A) Sequences of the duplex target and of the GU and TM TFOs. The triplex site is underlined. (B) Analysis of triplex formation by gel shift. A range of concentrations (1, 0.5, 0.2, 0.1, and 0.05 μM figfGU and 10, 5, 2, 1, and 0.5 μM figfTM) were added to 20 nM figfRY duplex radiolabeled on the oligopyrimidine strand (Y*) and analyzed after 2 h at 37 °C on a 12% (19:1) polyacrylamide gel in 50 mM HEPES (pH 7.2) and 10 mM $MgCl_2$. (C) C_{50} values calculated after 2 and 24 h for both TFOs.

than four. Nevertheless, this equation may be applicable to longer TFOs, as according to the equation a 73 bp target sequence containing only A can be targeted by a T_{73} TFO. Indeed, AT·T triplexes have been observed with poly-T polymers (*43*). Finally, these rules can also be used to help design TFOs that contain chemical modifications that do not alter greatly the structure of the triple helix.

The rule derived from eq 1 is useful but not sufficient. After making the first selection on the basis of this guideline, one should then consider the following features (Figure 5). For TM triplexes, the presence of contiguous M residues is destabilizing; this is probably due to the presence of adjacent positive charges on the cytidines that have a repulsive effect (*36*). In the case of GU TFOs, the presence of G stretches of more than four may eventually stabilize the formation of competitive G-quadruplex structures that are stabilized in the presence of K^+. These secondary rules help explain the behaviors of TFOs such as birc2TM, egfrTM, egfrGU, and retTM that do not simply follow the requirement given by eq 1. In the case of the *RET* target, G_N is 52 and *%G* is 40; therefore, a TM TFO should be used, but retTM does not form a stable triplex. This sequence has a stretch of four M residues in a row that are extremely destabilizing (i.e., four positive charges in a row) and also two adjacent M residues. Thus, it is not surprising that the triplex formed by the TFO is not stable. For *EGFR*, G_N is 50 and *%G* is 44; therefore, either motif could be used. However, in the case of egfrTM, an MpMpM and an MpM stretch are present that destabilize the triplex. For egfrGU, the presence of a GpGpG motif followed by a GpG motif allows formation of a G-quadruplex and the triplex is formed with only intermediate stability (C_{50} = 0.6 μM). Finally, there are some exceptions. G_N is 52 and *%G* 35 for the *BIRC2* target, but birc2TM does not form a very stable triplex under our experimental conditions. However, it is the only motif of the five tested for which we observe formation of a triplex (+, Table 2).

Importantly, the procedure given in Figure 5 is a guide. It does not ensure that a sequence designed using these rules will necessarily form a stable triplex, but these predictions hold in most cases. Chemical modifications can be used to further stabilize the triplex. We should also note that these predictions were tested for intermolecular triplex formation and may be less reliable for intramolecular ones.

To verify the validity of the rule, we used it to evaluate sequences known from the literature to form stable triplexes under similar experimental conditions (37 °C and 7.0 ≤ pH ≤ 7.5). For the sake of simplicity, only the oligopurine strand of the target sequence will be cited. Weisz et al. designed TFOs against a 7 bp target sequence, $^{5'}GAGAGA_2{}^{3'}$. According to the formula, $G_N = 73 - 1.1 \times 7 \approx 65$, while *%G* is 43. This percentage is lower than the percentage of G needed; thus, since no G-stretches are present, we predict that a TM motif should be used. In fact, the authors report that this type of TFO formed a stable triplex, whereas the other motifs did not. For the $^{5'}AG_2AG_2AG_2AG_2{}^{3'}$ sequence, $G_N = 73 - 1.1 \times 12 \approx 60$, while *%G* is ~67. Therefore, G-rich oligonucleotides should form a stable triplex. We have indeed used a GA TFO to target this sequence (*39*). In the same study, we target the sequence $^{5'}A_2GAG_2AG_2AG_2AG_2{}^{3'}$, for which $G_N = 73 - 1.1 \times 15 \approx 56$ and *%G* is ~60, well above the G_N. Indeed, GA and GT TFOs form very stable triplexes on this sequence, too. Another interesting example is the case of a 17 bp duplex target of the oligopurine sequence $^{5'}AG_2AGAGAGA_2GA_2G_3{}^{3'}$ (*33*). It gives a G_N of ≈54 and a *%G* of ~53. Here, both G-rich TFOs and TM TFOs may be used. However, we would predict that G_2 and G_3 tracks are destabilizing for TM TFOs and that only the G-rich TFO will form a triplex. This is exactly what the authors observed. Finally, we used the rules depicted in Figure 5 to design a TFO directed against the c-fos-induced growth factor (vascular endothelial growth factor D) FIGF, which plays an important role in tumorigenesis. We chose a 19 bp oligopurine·oligopyrimidine duplex target [located on chromosome X: 1532873669 (strand +) (Figure 6A)]. The calculated G_N is 52, and *%G* is 53; therefore, both a GU TFO and a TM TFO may form a triplex. However, the TM TFO is probably not very stable because the target sequence contains a pair of guanines. The corresponding GU and TM TFOs indeed form a triplex with C_{50} values at 24 h of 0.2 and 8.5 μM, respectively (Figure 6), confirming the validity of our simple rule.

The use of these twelve TFOs in cells to downregulate oncogenic expression is currently under evaluation.

CONCLUSION

The analysis of stabilities of triplexes formed by five different TFO motifs (GA, GT, GU, TM, and UM) on each of eleven different target sequences has permitted us to define some simple rules for predicting triplex stability. These rules are limited to the experimental conditions used in this study (pH 7.2, 100 mM NaCl, and 10 mM $MgCl_2$), to TFOs of 12−30 nt, and to sequences containing stretches of fewer than four guanines.

The most stable triplexes are formed by GU or TM TFOs. For G-rich TFOs, GU TFOs form in general the most stable triplexes with the fastest kinetics. TM TFOs bind with the highest stability in the pyrimidine motif. Their weaknesses are the requirement for the protonation of cytosines and the problem induced by adjacent cytosines. T-Stretches are, on the other hand, stabilizing.

How then do we choose between a TM motif and a GU motif? We determined the percentage of G necessary (G_N) for GU triplex formation (Figure 5). If the percentage of G (*%G*) in the sequence of the target is above this G_N value, then a GU TFO is preferred; if it is below, a TM TFO is better. In the case where $\%G \approx G_N$, either motif can be used. In the case of TM triplexes, it is important to consider that MpM dinucleotides, and longer stretches of M, are destabilizing.

In summary, we present a simple formula for predicting theoretically the oligonucleotide that will form a triplex on a given DNA sequence. It will be helpful in the design of TFOs for antigene applications: as biotechnological tools or to probe the mechanisms of proteins acting on DNA or as sequence-specific DNA binders for the control of gene expression.

ACKNOWLEDGMENT

We are grateful to Julien Gros, Laurent Lacroix, Jean-Louis Mergny, and Loïc Ponger for fruitful discussions.

SUPPORTING INFORMATION AVAILABLE

Complete characterization of the secondary structures of GA TFOs and GU TFOs (Tables S1 and S2, respectively) and Spearman correlation calculations for GU TFOs (Table S3) with associated scatter plots with confidence ellipses. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Desjarlais, J. R., and Berg, J. M. (1993) Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. U.S.A. 90*, 2256–2260.
2. Mandell, J. G., and Barbas, C. F. (2006) Zinc Finger Tools: Custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res. 34*, w516–w523.
3. Beerli, R. R., and Barbas, C. F. (2002) Engineering polydactyl zinc-finger transcription factors. *Nat. Biotechnol. 20*, 135–141.
4. Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., and Holmes, M. C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature 435*, 646–651.
5. Minczuk, M., Papworth, M. A., Kolasinska, P., Murphy, M. P., and Klug, A. (2006) Sequence-specific modification of mitochondrial DNA using a chimeric zinc finger methylase. *Proc. Natl. Acad. Sci. U.S.A. 103*, 19689–19694.
6. Li, F., Papworth, M., Minczuk, M., Rohde, C., Zhang, Y., Ragozin, S., and Jeltsch, A. (2007) Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. *Nucleic Acids Res. 35*, 100–112.
7. Smith, A. E., and Ford, K. G. (2007) Specific targeting of cytosine methylation to DNA sequences in vivo. *Nucleic Acids Res. 35*, 740–754.
8. Nomura, W., and Barbas, C. F., III (2007) In vivo site-specific DNA methylation with a designed sequence-enabled DNA methylase. *J. Am. Chem. Soc. 129*, 8676–8677.
9. Wu, J., Kandavelou, K., and Chandrasegaran, S. (2007) Custom-designed zinc finger nucleases: What is next? *Cell. Mol. Life Sci. 64*, 2933–2944.
10. Dervan, P. B., Doss, R. M., and Marques, M. A. (2005) Programmable DNA binding oligomers for control of transcription. *Curr. Med. Chem.: Anti-Cancer Agents 5*, 373–387.
11. Rogers, F. A., Lloyd, J. A., and Glazer, P. M. (2005) Triplex-forming oligonucleotides as potential tools for modulation of gene expression. *Curr. Med. Chem.: Anti-Cancer Agents 5*, 319–326.
12. Arimondo, P. B., Thomas, C. J., Oussedik, K., Baldeyrou, B., Mahieu, C., Halby, L., Guianvarc'h, D., Lansiaux, A., Hecht, S. M., Bailly, C., and Giovannangeli, C. (2006) Exploring the cellular activity of camptothecin-triple-helix-forming oligonucleotide conjugates. *Mol. Cell. Biol. 26*, 324–333.
13. Eisenschmidt, K., Lanio, T., Simoncsits, A., Jeltsch, A., Pingoud, V., Wende, W., and Pingoud, A. (2005) Developing a programmed

restriction endonuclease for highly specific DNA cleavage. *Nucleic Acids Res. 33*, 7039–7047.
14. Kuznetsova, S., Ait-Si-Ali, S., Nagibneva, I., Troalen, F., Le Villain, J. P., Harel-Bellan, A., and Svinarchuk, F. (1999) Gene activation by triplex-forming oligonucleotide coupled to the activating domain of protein VP16. *Nucleic Acids Res. 27*, 3995–4000.
15. Vasquez, K. M., Narayanan, L., and Glazer, P. M. (2000) Specific mutations induced by triplex-forming oligonucleotides in mice. *Science 290*, 530–533.
16. Levy, O., Ptacin, J. L., Pease, P. J., Gore, J., Eisen, M. B., Bustamante, C., and Cozzarelli, N. R. (2005) Identification of oligonucleotide sequences that direct the movement of the *Escherichia coli* FtsK translocase. *Proc. Natl. Acad. Sci. U.S.A. 102*, 17618–17623.
17. Maxwell, A., Burton, N. P., and O'Hagan, N. (2006) High-throughput assays for DNA gyrase and other topoisomerases. *Nucleic Acids Res. 34*, e104.
18. Buchini, S., and Leumann, C. J. (2003) Recent improvements in antigene technology. *Curr. Opin. Chem. Biol. 7*, 717–726.
19. Manzini, G., Xodo, L. E., Gasparotto, D., Quadrifoglio, F., van der Marel, G. A., and van Boom, J. H. (1990) Triple helix formation by oligopurine-oligopyrimidine DNA fragments. Electrophoretic and thermodynamic behavior. *J. Mol. Biol. 213*, 833–843.
20. Sun, J. S., and Hélène, C. (1993) Oligonucleotide-directed triple-helix formation. *Curr. Opin. Struct. Biol. 3*, 345–356.
21. Shafer, R. H. (1998) Stability and structure of model DNA triplexes and quadruplexes and their interactions with small ligands. *Prog. Nucleic Acid Res. Mol. Biol. 59*, 55–94.
22. Seidman, M. M., and Glazer, P. M. (2003) The potential for gene repair via triple helix formation. *J. Clin. Invest. 112*, 487–494.
23. Nielsen, P. E. (2007) Peptide nucleic acids and the origin of life. *Chem. Biodiversity 4*, 1996–2002.
24. Lacroix, L., Arimondo, P. B., Takasugi, M., Hélène, C., and Mergny, J. L. (2000) Pyrimidine morpholino oligonucleotides form a stable triple helix in the absence of magnesium ions. *Biochem. Biophys. Res. Commun. 270*, 363–369.
25. Grunweller, A., and Hartmann, R. K. (2007) Locked nucleic acid oligonucleotides: The next generation of antisense agents? *BioDrugs 21*, 235–243.
26. Lee, J. S., Woodsworth, M. L., Latimer, L. J., and Morgan, A. R. (1984) Poly(pyrimidine) ·poly(purine) synthetic DNAs containing 5-methylcytosine form stable triplexes at neutral pH. *Nucleic Acids Res. 12*, 6603–6614.
27. Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., and van Boom, J. H. (1991) Effect of 5-methylcytosine on the stability of triple-stranded DNA: A thermodynamic study. *Nucleic Acids Res. 19*, 5625–5631.
28. Mills, M., Arimondo, P. B., Lacroix, L., Garestier, T., Klump, H., and Mergny, J. L. (2002) Chemical modification of the third strand: Differential effects on purine and pyrimidine triple helix formation. *Biochemistry 41*, 357–366.
29. Cantor, C. R., Warshaw, M. M., and Shapiro, H. (1970) Oligonucleotides interactions. III. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers 9*, 1059–1077.
30. Olivas, W. M., and Maher, L. J., III (1995) Competitive triplex/quadruplex equilibria involving guanine-rich oligonucleotides. *Biochemistry 34*, 278–284.
31. Noonberg, S. B., Francois, J. C., Garestier, T., and Hélène, C. (1995) Effect of competing self-structure on triplex formation with purine-rich oligodeoxynucleotides containing GA repeats. *Nucleic Acids Res. 23*, 1956–1963.
32. Arimondo, P. B., Garestier, T., Hélène, C., and Sun, J. S. (2001) Detection of competing DNA structures by thermal gradient gel electrophoresis: From self-association to triple helix formation by (G,A)-containing oligonucleotides. *Nucleic Acids Res. 29*, e15.
33. Faucon, B., Mergny, J. L., and Hélène, C. (1996) Effect of third strand composition on the triple helix formation: Purine versus pyrimidine oligodeoxynucleotides. *Nucleic Acids Res. 24*, 3181–3188.
34. Sowers, L. C., Shaw, B. R., and Sedwick, W. D. (1987) Base stacking and molecular polarizability: Effect of a methyl group in the 5-position of pyrimidines. *Biochem. Biophys. Res. Commun. 148*, 790–794.
35. Gamper, H. B., Jr., Kutyavin, I. V., Rhinehart, R. L., Lokhov, S. G., Reed, M. W., and Meyer, R. B. (1997) Modulation of Cm/T, G/A, and G/T triplex stability by conjugate groups in the presence and absence of KCl. *Biochemistry 36*, 14816–14826.
36. Leitner, D., Schroder, W., and Weisz, K. (2000) Influence of sequence-dependent cytosine protonation and methylation on DNA triplex stability. *Biochemistry 39*, 5886–5892.
37. Soto, A. M., Rentzeperis, D., Shikiya, R., Alonso, M., and Marky, L. A. (2006) DNA intramolecular triplexes containing dT → dU substitutions: Unfolding energetics and ligand binding. *Biochemistry 45*, 3051–9505.
38. Mergny, J. L., Phan, A. T., and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett. 435*, 74–78.
39. Arimondo, P. B., Barcelo, F., Sun, J. S., Maurizot, J. C., Garestier, T., and Hélène, C. (1998) Triple helix formation by (G,A)-containing oligonucleotides: Asymmetric sequence effect. *Biochemistry 37*, 16627–16635.
40. Mergny, J. L., Li, J., Lacroix, L., Amrane, S., and Chaires, J. B. (2005) Thermal difference spectra: A specific signature for nucleic acid structures. *Nucleic Acids Res. 33*, e138.
41. Roberts, R. W., and Crothers, D. M. (1996) Kinetic discrimination in the folding of intramolecular triple helices. *J. Mol. Biol. 260*, 135–146.
42. Reither, S., and Jeltsch, A. (2002) Specificity of DNA triple helix formation analyzed by a FRET assay. *BMC Biochem. 3*, 27.
43. Arnott, S., and Selsing, E. (1974) Structures for the polynucleotide complexes poly(dA) with poly(dT) and poly(dT) with poly(dA) with poly(dT). *J. Mol. Biol. 88*, 509–521.

BI801087G