Bioorganic Chemistry **30,** 62–80 (2002)
doi:10.1006/bioo.2001.1232, available online at http://www.idealibrary.com on IDEAL®

REVIEW

Phosphates, DNA, and the Search for Nonterrean Life: A Second Generation Model for Genetic Molecules[1]

Steven A. Benner[2] and Daniel Hutter

Departments of Chemistry and Anatomy and Cell Biology, University of Florida, P.O. Box 117200, Gainesville, Florida 32611-7200

Received August 24, 2001

Phosphate groups are found and used widely in biological chemistry. We have asked whether phosphate groups are likely to be important to the functioning of genetic molecules, including DNA and RNA. From observations made on synthetic analogs of DNA and RNA where the phosphates are replaced by nonanionic linking groups, we infer a set of rules that highlight the importance of the phosphodiester backbone for the proper functioning of DNA as a genetic molecule. The polyanionic backbone appears to give DNA the capability of replication following simple rules, and evolving. The polyanionic nature of the backbone appears to be critical to prevent the single strands from folding, permitting them to act as templates, guiding the interaction between two strands to form a duplex in a way that permits simple rules to guide the molecular recognition event, and buffering the sensitivity of its physicochemical properties to changes in sequence. We argue that the feature of a polyelectrolyte (polyanion or polycation) may be required for a "self-sustaining chemical system capable of Darwinian evolution." The polyelectrolyte structure therefore may be a universal signature of life, regardless of its genesis, and unique to living forms as well.

Key Words: phosphate-backbone; polyelectrolyte; sulfone; PNA; genetic molecule; nonterrean life.

INTRODUCTION

Ever since the emergence of the Watson–Crick model for DNA (*1*,*2*), chemists and biologists alike have underestimated the complexity of the molecular system that stands behind genetics and inheritance. This was almost certainly due, in part, to the simplicity, elegance, and utility of the model itself. The "first generation model" explicates genetic inheritance using the structure of DNA and just three very simple structural concepts: nucleobase stacking, size complementarity, and hydrogen bonding

[1] This paper is dedicated to Professor Frank H. Westheimer on the occasion of his 90th birthday.

[2] To whom correspondence and reprint requests should be addressed. Fax: (352) 392-7918. E-mail: benner@chem.ufl.edu.

0045-2068/02 $35.00

complementarity. Duplexes are stabilized by base stacking. The specificity of base pairing arises from size and hydrogen bonding complementarity between bases.

The well known Watson–Crick base pairing rules (A pairs with T, G pairs with C) follow from these premises. These rules permit molecular biologists to design molecular recognition systems without much background in organic chemistry. This power eludes more sophisticated chemists attempting to achieve molecular recognition in nonnucleic acid systems.

The simplicity, elegance, and utility of the first generation model has, however, been almost beguiling to scientists, who have often expressed surprise as complexity in nucleic acids, not captured by this model, has emerged from experimental work. When Rich and his coworkers announced the existence of Z-DNA (*3*), for example, the structural biology community was surprised, even though most polymers (indeed, most organic molecules) show a diversity of conformational states. Considerable research activity focuses on "bending DNA" (*4*), in part because it is surprising (under the model) that DNA might bend (even though other biopolymers do). The discovery of catalysis by RNA introns (*5*) and the RNA component of RNase P (*6*) was regarded as surprising, even though "catalysis" by RNA itself had been observed earlier as a normal property of functionalized polymers (*7*), and was suggested for elements of the translation machinery (*8,9*). The ability of RNA to provide general acid–base catalysis was viewed as noteworthy (*10*), even though this ability follows axiomatically from the chemical structure of RNA.

With DNA, however, the simple elegance of the first generation model created in the minds of chemists and biologists alike the expectation that derivatives and analogs of DNA would also behave in a rule-based fashion. Perhaps the most striking example of how the first generation model failed to guide nucleic acid chemistry comes from the "antisense" industry, which rose, flourished, and nearly vanished over a period of only a decade (*11–13*). The antisense concept was simple and entirely reasonable given the first generation model. Many diseases arise through the presence of unwanted DNA that is expressed as mRNA and encodes unwanted proteins. Many drugs work by binding the unwanted target (generally an unwanted protein). What could be simpler than targeting the unwanted mRNA that preceded the unwanted protein? Often, the sequence of the mRNA is known, especially for infectious diseases. The Watson–Crick model suggested simple design principles to create a molecule that binds to the mRNA (A pairs with T (or U), G pairs with C).

The only challenge to implementing the antisense strategy seemed to be the backbone. Nucleic acids have a repeating phosphodiester linkage in their backbone, which join adjacent nucleosides in the chain. The phosphodiester linkages are the targets of nucleases that degrade DNA. Likewise, and perhaps more severely, the repeating phosphodiesters make DNA a polyanion. As reviewed by Westheimer more than a decade ago in his now-famous paper on "Why Nature Chose Phosphates" (*14*), anions (and anionic charges carried by phosphates in particular) prevent molecules from crossing cell membranes (*15*).

These two features, charge and sensitivity to enzymes, of DNA made it unlikely that one could ever use DNA directly as a drug. Most of it would be degraded, and the part that was not would not make it across cell membranes to reach its target.

Fortunately, the first generation model for DNA structure did not propose any

particular role for the polyanionic backbone in the molecular recognition event, although they clearly understood the importance of placing the phosphates in the duplex structure in a way that permitted them to interact with water. The phosphate-sugar backbone otherwise served simply as a scaffold. Perhaps other scaffolds would do? Indeed, Westheimer remarked that perhaps amides might serve as well.

For this reason, we (and others) reasoned, why not make analogs of DNA where the backbone phosphate was replaced? In particular, the replacement would be non-ionic, and stable to enzymatic degradation. The expectation, based on the first generation model, was that virtually any backbone-modified DNA analog would retain the molecular recognition properties of DNA itself. A sample of analogs of DNA where the phosphodiesters are replaced by a neutral linker, following this logic, is shown in Fig. 1 (*16*).

Upon closer consideration from a physical organic perspective, the structure for DNA is not as inevitable as it might seem from the first generation model, however. Indeed, from a physical organic perspective, it does not correspond closely to how a chemist might set out to design two molecules that will bind with great selectivity to each other. Consider just three features of the DNA structure, and contrast them with how chemists might design molecular recognition systems:

(a) DNA strands are floppy. Rigid molecules would seem better for molecular recognition.

(b) DNA uses hydrogen bonding to achieve specificity, according to the model. In water, hydrogen bonds are abundant. It seems curious in water to use hydrogen bonding as the key to molecular recognition.

(c) Receptor and ligands are both anions. In the design of a molecular recognition

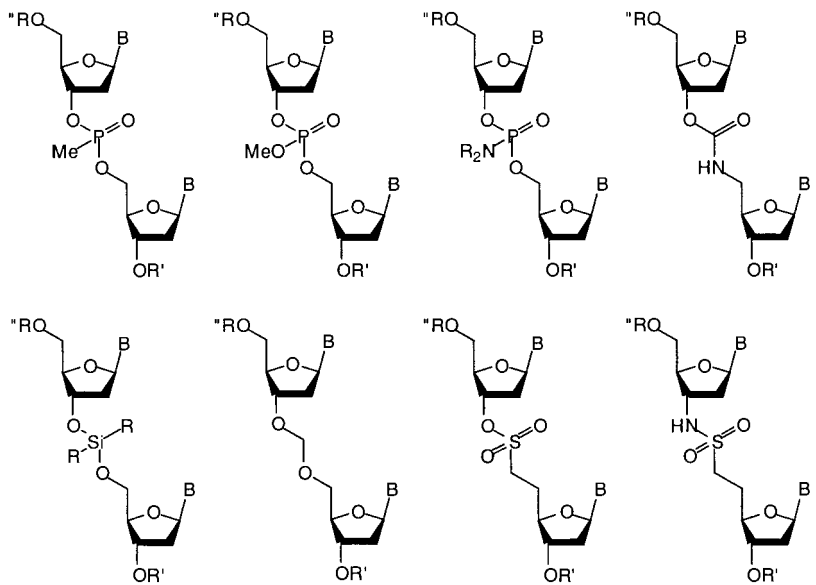

FIG. 1. A selection of analogs of DNA where the phosphate is replaced by an uncharged analog. *R* is alkyl or aryl; *R′* and *R″* are nucleic acid or hydrogen.

system, charge complementarity might be sought. Few chemists would design a polyanion to bind to another polyanion.

The paradigm of physical organic chemistry, which involves systematic synthesis of analogs of DNA and examining their properties, could be used to test the Watson–Crick model. After all, the Watson–Crick model had solved the problem. The solution was elegant. What was there to test?

Perhaps for this reason, prior to 1985, only a few chemists were applying their synthetic skills to DNA. For example, as early as 1970, Pitha *et al.* had attached vinyl groups to uracil and created polymers (*17*). This predated both synthesis and direct sequencing of DNA. In the 1970s, Eckstein had prepared phosphorothioate analogs of DNA (*18*), although with the goal more to develop an understanding of the mechanism of enzymatic reactions involving phosphorus as the electrophile than to test the Watson–Crick model itself.

Starting in Zurich in 1985, we began a research program to analyze the structure of DNA in this light. Strong programs were also developed elsewhere, notably within corporate laboratories at Central Research at Ciba-Geigy, Gilead, and at Isis (*12,16*). Each program was designed to modify systematically the structure of the nucleobases, the sugars, and the phosphates. Each has advanced our knowledge of molecular recognition in DNA. Several have had technological impact as well. For example, our work with the heterocyclic nucleobases has led to diagnostic products that are detecting very low levels of DNA in biological samples, DNA containing a single nucleotide change in a patient, and real time quantitation of mRNA (*19*). In the form of an "account," let us review some of what we have learned from these efforts, especially about the role of phosphates in the molecular recognition event.

A SMALL STEP FROM THE NATURAL PHOSPHODIESTER BACKBONE

Our first effort involved replacing the phosphodiester linkers in DNA by linkers that would be close in structure to the phosphodiesters, but not have their negative charges. Ideally, we wished to "add" a proton to each phosphorus nucleus (in a sense) to "make" the sulfate diester, a unit that would be isoelectronic to the phosphate diester. Simple chemical considerations ruled out the sulfate diester linkage directly, however. The primary 5′-carbon in the sulfate diester would almost certainly be too electrophilic to survive in water. We felt that a sulfonate would be too reactive as well, although heroic efforts by Widlanski and his coworkers (*20*) provided DNA analogs incorporating this structure; their reactivity was used to probe the active site of a polymerase (*21*). The sulfonamide was also considered as a structure that would lack the undesirable reactivity of a sulfonate (*22*).

We chose to address the reactivity problem by making the dimethylenesulfone linker (Fig. 2). A dimethylenesulfone is isoelectronic to a phosphodiester. It is also nonionic, stable to alkaline degradation, and not stereogenic (that is, it does not create diastereoisomers). We were also attracted by the fact that sulfones are intrinsically soluble in water and that Hanahan had shown that dimethylsulfoxide increased the frequency with which transforming DNA entered bacterial cells (*23*).

The preparation of these compounds required solutions to problems encountered generally in large scale organic synthesis. We needed to make four building blocks, in sufficient quantities to permit each to be the starting point of a multistep synthesis

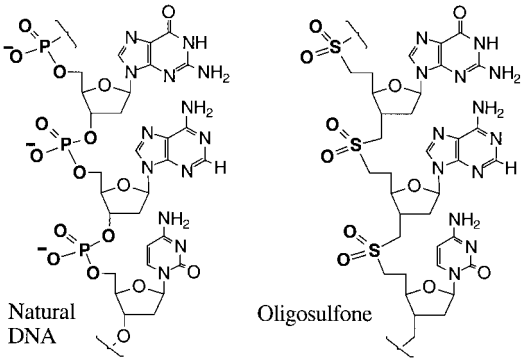

FIG. 2. A dimethylenesulfone-linked analog of DNA.

of an oligomer. All of the oligomers were prepared in solution (not on solid phase) so that they could be fully characterized. Many of the oligomers proved to be challenging analytical targets (see below). The details of synthetic efforts are given elsewhere (*24*,*25*).

The synthetic effort did not go unrewarded. One of the first sulfone analogs of RNA to be made was $G_{SO_2}C$, the sulfone analog of the RNA dinucleotide G_{PO_2}-C, which had been crystallized and studied by Rich and his coworkers (*26*,*27*). Both molecules were self-complementary in the Watson–Crick sense and might be expected to form a duplex. Indeed, the $G_{SO_2}C$ structure (Fig. 3) did so (*28*). The duplex it formed was remarkably similar to the duplex formed by the natural RNA dimer G_{PO_2}-C. In both molecules, the crystallographic unit cell contained two molecules in the form of a right-handed double helix with antiparallel orientation of the strands. In both molecules, the duplexes were joined via canonical Watson–Crick base pairs. In both, the two strands were related via a crystallographic twofold rotation axis. Conformations around the glycosyl bonds (*anti*), ribose puckers (C3′-*endo*-type), and all backbone torsion angles (Tables 1 and 2; Fig. 3) of the $G_{SO_2}C$ analog fell into the same ranges as found in natural RNA duplexes. Further, the overall dimensions of the two duplexes (e.g., relative S(P)⋯S(P) and C1′⋯C1′ distances, Tables 1 and 2) were the same to within 0.3 Å (Fig. 4).

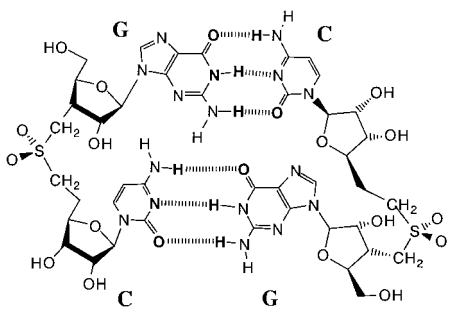

FIG. 3. Schematic of the duplex formed by $G_{SO_2}C$ in the crystal.

TABLE 1

Comparison of Backbone[a] and Glycosyl (χ) Torsion Angles as well as Pseudorotation Phase Angles (P) (in deg) for [$G_{SO_2}C$] (Top) and [$G_{PO_2^-}C$] (Bottom)

	α	β	γ	δ	ε	ζ	χ	P
G(1)			176	93	−140	−66	−178	12
C(2)	−58	−179	44	86			−170	29
G(1)			49	82	−150	−66	−157	7
C(2)	−75	−176	51	75			−152	17

[a] The angular notations for backbone torsions are O5′(C6′)$\underline{\beta}$C5′$\underline{\gamma}$C4′$\underline{\delta}$C3′$\underline{\varepsilon}$O3′ (C3″)$\underline{\zeta}$P(S)$\underline{\alpha}$O5′ (C6′).

Some differences were evident between $G_{SO_2}C$ and $G_{PO_2^-}C$ upon close examination, of course. In the $[G_{SO_2}C]_2$ duplex, for example, the glycosidic torsion angles were larger by 21° and 18° for the G and C residues than in the $[G_{PO_2^-}C]_2$ duplex (Table 1). The torsion angles α, δ, and ε were different by at least 10° in the sulfone duplex. The helical parameters of the duplex were modestly different in the sulfone duplex (Table 2). The rise in the natural duplex was slightly smaller than in the sulfone. The bases were stacked parallel in the sulfone with a spacing of 3.40 Å, instead of the 3.68 Å spacing with a roll of 8° observed in the stacking of bases in the natural RNA duplex. The $[G_{SO_2}C]_2$ duplex was unwound to give ca. 17 residues per helical turn, rather than the 11 residues in standard A-type RNA.

These differences in the overall structures of the RNA and sulfone duplexes can all be explained, however, without invoking the loss of charge. For example, the sulfone places the C6′ methylene group at the site occupied by O5′ in the natural duplex. This methylene group is larger than the oxygen, and may interact sterically with the C6 C-H group of cytosine. This interaction appeared to result in an increased C6′–C5′–C4′ angle, and a slight change in the cytosine ribose conformation (compare P values of residues C(2) of the two duplexes in Table 1). Torsion δ for the C residues appeared to be adjusted to reflect this change. Likewise, the unwinding in the $[G_{SO_2}C]_2$ double helix appeared to result predominantly from the fact that the S–C6′–C5′ bond angle is 111°, instead of the 120° found in RNA for the corresponding P–O5′–C5′ angle.

Thus, the absence of the negative charge in the sulfone-linked RNA analog appeared

TABLE 2

Comparison of Helical Parameters[a] for $[G_{SO_2}C]_2$ and $[G_{PO_2^-}C]_2$ (Angles in Degrees and Distances in Å)

	Rise	Twist	Inclination	Slide	Roll	S⋯S P⋯P	C1′⋯C1′	Buckle	Prop. twist
$G_{SO_2}C$	2.92	20.8	9.4	−3.2	0.7	18.0	10.6	6.0	0.8
GpC	2.59	34.7	28.0	−1.3	7.9	17.7	10.7	5.8	2.7

[a] Atoms selected to determine the helical operator were C1′, N1(9), C2′, O4′.

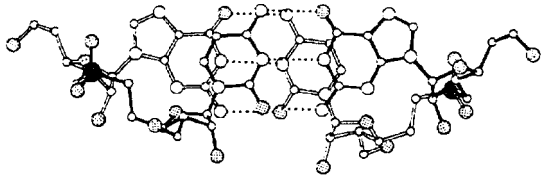

FIG. 4. The crystal structure of $G_{SO_2}C$ (*28*).

to have remarkably little impact on the overall duplex structure. Only in solvation did the change in charge seem to have an impact. Thus, the sulfone unit cell carried 8 water molecules and 1 methanol per strand; the natural duplex carried 9 water molecules. The major and minor groove base functional groups were extensively hydrated in both the sulfone analog and the natural RNA, but the solvation of the sulfone group itself was different. Instead of forming contacts with a sodium ion and three water molecules as seen with the phosphate, the sulfone group formed contacts with one water molecule, a cytosine NH_2 group from a neighboring duplex, and the CH_3 group of a methanol molecule.

The ability of the $G_{SO_2}C$ sulfone dimer to form duplexes suggested that a negative charge is not essential for Watson–Crick duplex formation. But the next round of experiments, with ($A_{SO_2}U$), suggested that the tendency to form Watson–Crick pairing might be weaker in a sulfone analog than in the corresponding phosphodiester. Here, the self-complementary $A_{SO_2}U$ duplex would be joined by only four hydrogen bonds, instead of the six that join the $G_{SO_2}C$ sulfone duplex. The natural $A_{PO_2^-}U$ dimer did form a duplex (*29*). In contrast, $A_{SO_2}U$ crystallized to give a single stranded structure (*30*), with Watson–Crick interactions not dominating the structure. The two structures available for the methylphosphonate linked deoxynucleotide dimers d($G_{POCH_3}C$) and d($A_{POCH_3}T$) found strikingly parallel results (*31,32*). The first formed a duplex joined by six hydrogen bonds involving two Watson–Crick pairs. The second formed a single strand, not a duplex joined by four hydrogen bonds involving two Watson–Crick pairs.

THE BREAKDOWN OF WATSON–CRICK BASE PAIRING IN LONGER UNCHARGED ANALOGS

These results encouraged us to prepare longer oligosulfones. Here, we encountered stronger evidence that the loss of the phosphodiester charge might have consequences. The first evidence came from tetrameric oligosulfone analogs of RNA. For example, the tetramer R-$T_{SO_2}T_{SO_2}T_{SO_2}$C-SR′ (where R and R′ were various protecting groups) formed dimers in tetrahydrofuran. The dimer was disrupted by adding small amounts of methanol, suggesting that hydrogen bonding might be involved. This dimer could not, however, be joined by canonical Watson–Crick pairing, as the system contained no A to pair with T and no G to pair with C.

Still longer sulfone-linked oligonucleotide analogs departed further from expectations formed by analogy to DNA. Particularly remarkable was the sulfone $A_{SO_2}U_{SO_2}G_{SO_2}G_{SO_2}U_{SO_2}C_{SO_2}A_{SO_2}U$. The molecule displayed a remarkable thermal denaturation curve, melting at ca. 80°C. Upon melting, a large hyperchromicity was observed (>200%; 25% is typical for melting of a DNA duplex). The sequence

was not, in the Watson–Crick sense, self-complementary. Further, the denaturation appeared to be unimolecular. Thus, we concluded that this oligosulfone folded to a rather stable conformation, indeed one of the most stable single stranded "RNA" structures known.

The results of these and many other experiments suggested that each oligosulfone has its own unique properties and reactivity. Different sulfones differing (in some cases) by only one nucleobase displayed different levels of solubility, aggregation behavior, folding, and chemical reactivity. Thus, the properties of tetranucleoside sulfone analogs and octanucleoside sulfone analogs was influenced dramatically by adding a single charge to one end of the molecule, or replacing a single nucleobase (*33*).

Further, Watson–Crick pairing seemed to be lost. For example, Huang found evidence for base pairing between d($U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_2}U_{SO_3^-}$) and its DNA complement (*33*). Eschgfaeller found, however, no evidence for the binding of an octamer differing only at one position, and replacing T for U: $T_{SO_2}T_{SO_2}T_{SO_2}C_{SO_2}T_{SO_2}T_{SO_2}T_{SO_2}T_{SO_3^-}$ to its complementary DNA sequence. On the other hand, the second molecule displayed good bioavailability in a mouse model (*34*,*35*).

These are not, of course, behaviors expected from DNA itself. The physical properties of two DNA molecules of the same length are usually nearly identical even when their sequences are quite different. Different DNA sequences of the same length generally move at (nearly) the same position in an electrophoresis experiment. Pairing is not, in general, dramatically altered by changing a single nucleobase. Virtually all DNA sequences are soluble in water, and insoluble in ethanol. Indeed, the constancy in the physical behavior of DNA, despite large change in sequence, is the key to the laboratory manipulation of DNA.

Replacing the phosphodiester groups by sulfone linkers generated a polymer that folded, had different properties, and whose properties changed dramatically upon small changes in sequence. Biological chemists are, of course, familiar with biopolymers that fold, have different properties, and whose properties can change dramatically by substitution of a single building block. These properties are all characteristic of proteins. Thus, we might view the results obtained with nonionic nucleotide analogs as saying that by replacing the anionic phosphodiester groups by uncharged sulfone groups, we have transformed DNA into a kind of "protein."

PHOSPHATES: THE KEY TO MOLECULAR RECOGNITION IN NUCLEIC ACIDS?

These results suggest three hypotheses that propose roles for the phosphodiester linkages, and how they might be important to the molecular recognition that characterizes DNA. They suggest three specific roles for the phosphodiester anions, in addition to the (perhaps obvious) role of making the polymer soluble in water. They all stress the universality of a polyelectrolyte as part of any molecule that needs to perform a genetic role.

Phosphates Force Interactions Between Strands as Far From the Backbone as Possible

As natural as the Watson–Crick hydrogen bonding pattern is, it is not necessarily the preferred one. Upon cocrystallization of equimolar amounts of adenine and thymine,

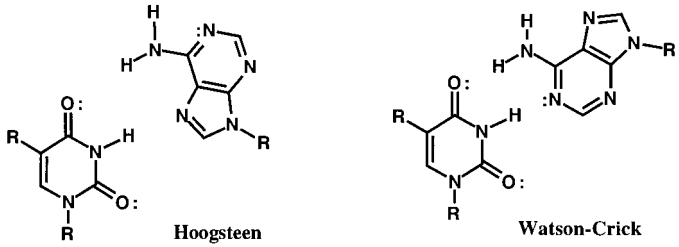

FIG. 5. A representation of the Hoogsteen (left) and Watson–Crick (right) base pairs.

Hoogsteen was surprised to discover that the adenine-thymine mixture paired through hydrogen bonding to N-7 of the adenine, not N-1 (as is seen in the standard DNA duplex structure) (Fig. 5).

From first principles, one expects a richly functionalized molecule to interact with other molecules through any of its functional groups. Proteins interacting with proteins offer many examples suitable for detailed study. What is also expected, however, is that two DNA strands, because of their polyanionic nature, cannot interact with each other from any point on their structure. In particular, the complex between two DNA strands is expected to place the polyanionic backbones as far from each other as possible. Conversely, the two strands should contact each other as far from the phosphate backbone as possible. This is, of course, the Watson–Crick "edge" of the heterocycles (Fig. 6).

The polyanionic nature of the backbone, under this hypothesis, is key to the molecular recognition between two DNA strands. It constrains the interaction of two strands to the regions of the molecule where those interactions are desired, and where those interactions can be described by simple rules.

Consistent with this hypothesis, Steinbeck and Richert (*36*) examined the solution structure of the dimethylenesulfone-linked $U_{SO_2}C$ dinucleotide analog using two-dimensional NMR and restrained molecular dynamics. In $CDCl_3$, the sulfone formed

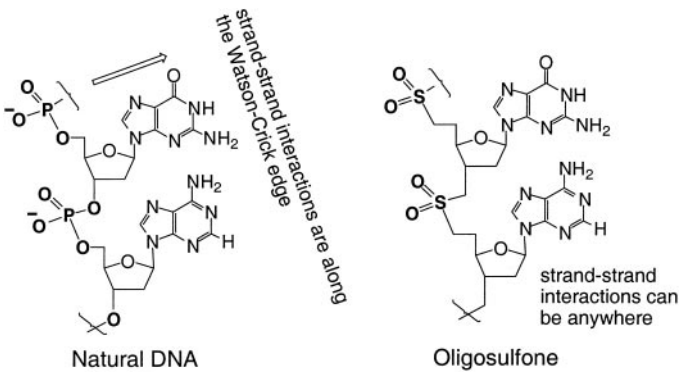

FIG. 6. The repulsion of the backbone charges forces strand–strand interactions in the polyanionic DNA to occur as far from the backbone as possible, along the Watson–Crick "edge" of the nucleobases.