Steroid−DNA Interactions Increasing Stability, Sequence-Selectivity, DNA/RNA Discrimination, and Hypochromicity of Oligonucleotide Duplexes

Colleen F. Bleczinski and Clemens Richert*

Contribution from the Department of Chemistry, Tufts University, Medford, Massachusetts 02155, USA

Received June 16, 1999

Abstract: Cholic acid and its deoxy derivatives were found to increase the melting point of oligonucleotide duplexes when coupled to their 5′-termini. For duplexes of mixed-sequence octamers, the melting point was 8−11 °C higher with the steroid appendage than without. For the self-complementary hexamer TGCGCA, a 21 °C melting point increase was measured in the presence of steroid appendage and −18 kcal/mol in $\Delta\Delta H°$. The affinity increases were accompanied by increased discrimination against mismatches at the two terminal base pairs and increased hypochromicity, as well as an improved DNA/RNA discrimination for the non selcomplementary sequence. Cross-peaks in the NOESY spectrum of $(\text{chl-T*GCGCA})_2$, where chl denotes the cholic acid residue and T* a 5′-amino-5′-deoxythymidine residue, point toward stacking interactions in which the methyl groups of the steroid face the nucleobases. Our results indicate that bile acids can form a specific complex with terminal T:A base pairs of double-stranded DNA.

Introduction

Non-covalent interactions of small molecules with nucleic acids have been described as falling into three classes: external electrostatic binding, groove binding, and intercalation.[1] Accordingly, most DNA-binding molecules possess cationic functional groups, size complementarity to one of the grooves, an aromatic ring system, or combinations of these. One class of typical aliphatic small molecules with important roles in selective molecular recognition in biology are the steroids. Cationic steroids, particularly steroidal diamines, have long been known to bind to nucleic acids.[2] Non-cationic steroids, however, are not commonly thought of as nucleic acid-binding molecules, even though early reports on the low-level binding of hormonal steroids to heat-denatured DNA exist.[3] In fact, conjugates of cholesterol and cholic acid with oligonucleotides have been designed with the goal of engaging the steroids in interactions with lipophilic structures, rather than the nucleic acids themselves.[4] For duplexes of cholesterol−DNA conjugates with a DNA target strand, Gryaznov and Lloyd found melting point increases of 5.5 °C for a 5′-phosphodiester linked construct and 2.0 °C for a construct bearing cholesterol at the 3′-terminus.[4f] Other cholesterol-conjugates have been reported to give no duplex-stabilizing effect.[5]

In the context of spectrometrically monitored selection experiments,[6] we became aware of an unexpectedly strong duplex-stabilizing effect of a cholate moiety covalently linked to the 5′-terminus of DNA.[7] Here we report results indicating that this appended steroid stacks sequence-selectively on exposed terminal T:A base pairs. As a result, oligodeoxyribonucleotides bearing this binding-motif for the terminal base pair behave like "high-fidelity hybridization probes", whose RNA/DNA and mismatch discrimination is more pronounced than that of their unmodified controls.

* To whom correspondence should be addressed.

(1) *Nucleic Acids in Chemistry and Biology, 2nd ed.*; Blackburn, G. M.; Gait, M. J., Eds.; Oxford University Press: Oxford, 1996.

(2) Selected references: (a) Mahler, H. R.; Goutarel, R.; Khuong-Huu, Q.; Ho, M. T. *Biochemistry* **1966**, *5*, 2177−2191. (b) Mahler, H. R.; Green, G.; Goutarel, R.; Khoung-Huu, Q. *Biochemistry* **1968**, *7*, 1568−1582. (c) Gabbay, E. J.; Glaser, R. *Biochemistry* **1971**, *10*, 1665−1674. (d) Sobell, H.; Tsai, C.-C.; Jain, S.; Gilbert, S. *J. Mol. Biol.* **1977**, *114*, 335−365. (e) Patel, D. J.; Canuel, L. L. *Proc. Natl. Acad. Sci.* **1979**, *76*, 24−28. (f) Patel, D. J. *Acc. Chem. Res.* **1979**, *12*, 118−125. (g) Gourevitch, M. I.; Puigdomenech, P. *Int. J. Biol. Macromol.* **1986**, *8*, 97−104. (h) Hsieh, H.-P.; Muller, J. G.; Burrows, C. J. *J. Am. Chem. Soc.* **1994**, *116*, 12077-1278. (i) Muller, J. G.; Ng, M. M. P.; Burrows, C. J. *J. Mol. Recogn.* **1996**, *9*, 143−148.

(3) (a) Ts'O, P. O. P.; Lu, P. *Proc. Natl. Acad. Sci. U.S.A.* **1964**, *51*, 17−24. (b) Cohen, P.; Kidson, C. *Proc. Natl. Acad. Sci. U.S.A.* **1969**, *63*, 458−464. (c) Cohen, P.; Chin, R.-C.; Kidson, C. *Biochemistry* **1969**, *8*, 3603−2609. (d) Kidson, C.; Thomas, A.; Cohen, P. *Biochemistry* **1970**, *9*, 1571−1576. (e) Chin, R.-C.; Kidson, C. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 2448−2452.

(4) (a) Letsinger, R. L.; Zhang, G. G.; Sun, D. K.; Ikeuchi, T.; Sarin, P. S. *Proc. Natl Acad. Sci. U.S.A.* **1989**, *86*, 6553−6556. (b) Stein, C. A.; Pal, R.; Devico, A. L.; Hoke, G.; Mumbauer, S.; Kinstler, O.; Sarngadharan, M. G.; Letsinger, R. L. *Biochemistry* **1991**, *30*, 2439−2444. (c) Farooqui, F.; Sarin, P. S.; Sun, D.; Letsinger, R. L. *Bioconjugate Chem.* **1991**, *2*, 422−426. (d) De Smidt, P. C.; Le Doan, T.; de Falco, S.; van Berkel, T. J. C. *Nucleic Acids Res.* **1991**, *19*, 4695−4700. (e) Letsinger, R. L.; Chaturvedi, S. K.; Farooqui, F.; Salunkhe, M. *J. Am. Chem. Soc.* **1993**, *115*, 7535−7536. (f) Gryaznov, S. M.; Lloyd, D. H. *Nucleic Acids Res.* **1993**, *21*, 5909−5915. (g) Manoharan, M.; Johnson, L. K.; Bennett, C. F.; Vickers, T. A.; Ecker, D. J.; Cowsert, L. M.; Freier, S. M.; Cook, P. D. *Bioorg. Med. Chem. Lett.* **1994**, *4*, 1053−1060. (h) Zhang, G. R.; Farooqui, F.; Kinstler, O.; Letsinger, R. L. *Tetrahedron Lett.* **1996**, *37*, 6243−6246. (i) Manoharan, M.; Tivel, K. L.; Inamati, G.; Monia, B. P.; Dean, N.; Cook, P. D. *Nucleosides Nucleotides* **1997**, *16*, 1139−1140. (j) Manoharan, M.; Inamati, G.; Tivel, K. L.; Conklin, B.; Ross, B. S.; Cook, P. D. *Nucleosides Nucleotides* **1997**, *16*, 1141−1143. (k) Bijsterbosch, M. K.; Manoharan, M.; Tivel, K. L.; Rump, E. T.; Biessen, E. A. L.; DeVrueh, R. L. A., Cook, P. D.; van Berkel, T. J. C. *Nucleosides Nucleotides* **1997**, *16*, 1165−1168. (l) Letsinger, R. L.; Chaturvedi, S. *Bioconjugate Chem.* **1998**, *9*, 826−830.

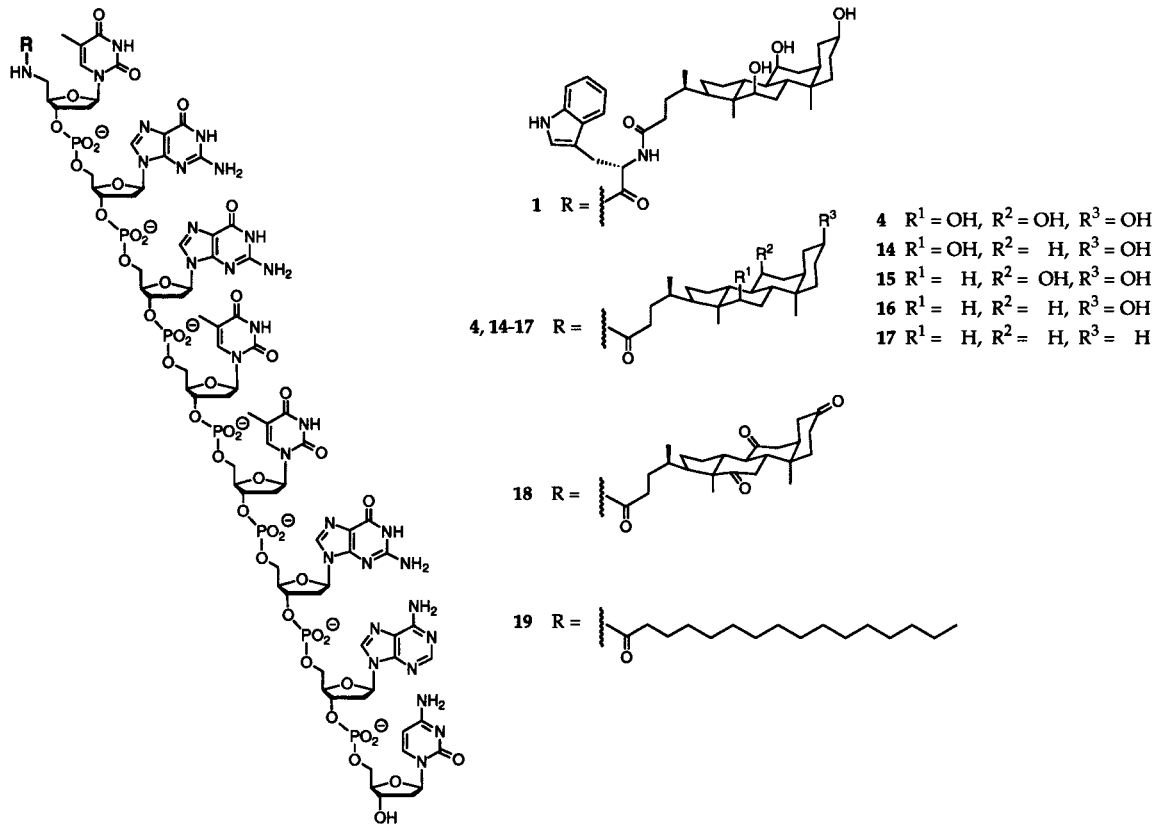
(5) (a) MacKellar, C.; Graham, D.; Will, D. W.; Burgess, S.; Brown, T. *Nucleic Acids Res.* **1992**, *20*, 3411−3417. (b) Zarytova, V. F.; Ivanova, E. M.; Chasovskikh, M. N. *Bioorg. Khim.* **1990**, *16*, 610−616.

(6) Berlin, K.; Jain, R. K.; Tetzlaff, C.; Steinbeck, C.; Richert, C. *Chem. Biol.* **1997**, *4*, 63−77.

(7) Bleczinski, C. F.; Richert, C. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 1737−1743.

10.1021/ja9920415 CCC: $18.00
Published on Web 11/03/1999

Scheme 1



Results and Discussion

Oligonucleotide **1** (Scheme 1), whose 5′-terminus bears a cholic acid residue linked to a tryptophan residue, which in turn is linked to 5′-amino-5′-deoxy-DNA, was initially identified as a high affinity ligand to undecamer **2**, with a melting point 4 °C higher than that of the control duplex **3**/**2** (Table 1). The structure of both DNA hybrid **1** and target strand **2** were varied to probe the requirements for duplex stabilization. Removing the tryptophan residue between cholic acid and DNA in **1** yielded **4** with substantially higher duplex stability than the more lipophilic **1**. When offered RNA complementary strand **5**, the duplex stabilization shrank from +11 °C (DNA) to +5.7 °C (RNA), making **4** a hybridization probe with much better DNA/RNA discrimination than control octamer **3**. RNA selectivity was maintained over the entire ionic-strength range assayed (Figure 1), indicating that the effect could be employed under a variety of hybridization conditions and that the underlying interactions were not predominantly electrostatic, as found for peptide−DNA hybrids in the same test system.[8]

Uncharged hormonal steroids have been reported to bind to single-stranded nucleic acids.[3b] The three dangling residues of undecamer **2**, however, were found to be unimportant for duplex stabilization, since octamer **6** gave almost the same melting point increase for **4** versus **3** as **2** (10.4 vs 11.0 °C, Table 1). On the other hand, a terminal T:A base pair was required, as hepta- and hexamer complementary strands **7** and **8** showed little or no extra duplex stability when complexed with **4** rather than with **3**. More stringent proof of the need for a terminal T:A base pair for duplex stabilization came from experiments with complementary strands presenting a mismatched base to cholic acid-bearing oligonucleotides. Figure 2 shows melting curves of self-complementary hexamers with (**9**) and without (**10**) a 5′-cholic acid residue, together with the same curves for duplexes containing T:G mismatches at their termini (**11** and **12**). While the fully Watson Crick-paired duplex experienced a melting point increase of 20.9 °C from its two cholic acid appendages, for the duplex with terminal wobble base pairs,

Table 1. Melting Points of Oligonucleotide Duplexes with Different 5′-Residues

sequences[a,b]		T_m (°C)[c]
chl-W−T*GGTTGAC (**1**)	GTCAACCACGG (**2**)	39.5
chl-T*GGTTGAC (**4**)	GTCAACCACGG (**2**)	46.4
TGGTTGAC (**3**)	GTCAACCACGG (**2**)	35.4
chl-T*GGTTGAC (**4**)	r(GUCAACCACGG) (**5**)	37.4
TGGTTGAC (**3**)	r(GUCAACCACGG) (**5**)	31.7
chl-T*GGTTGAC (**4**)	GTCAACCA (**6**)	45.3
TGGTTGAC (**3**)	GTCAACCA (**6**)	34.9
chl-T*GGTTGAC (**4**)	GTCAACC (**7**)	33.4
TGGTTGAC (**3**)	GTCAACC (**7**)	31.2
chl-T*GGTTGAC (**4**)	GTCAAC (**8**)	19.5
TGGTTGAC (**3**)	GTCAAC (**8**)	21.6
dchl-T*GGTTGAC (**14**)	GTCAACCA (**6**)	44.4
cdchl-T*GGTTGAC (**15**)	GTCAACCA (**6**)	43.9
lchl-T*GGTTGAC (**16**)	GTCAACCA (**6**)	42.3
chol-T*GGTTGAC (**17**)	GTCAACCA (**6**)	43.0
dhchl-T*GGTTGAC (**18**)	GTCAACCA (**6**)	41.3
pal-T*GGTTGAC (**19**)	GTCAACCA (**6**)	38.3

[a] Sequences are given 5′ to 3′ terminus and are deoxyribonucleotides unless otherwise noted. Abbreviations for non-nucleic acid residues are: chl, cholic acid; dchl, deoxycholic acid; cdchl, chenodeoxycholic acid; lchl, lithocholic acid; chol, cholanic acid; dhchl, dehydrocholic acid; pal, palmitic acid. [b] T* denotes a 5′-amino-5′-deoxythymidine residue. [c] Experiments were performed at 1.2 ± 0.2 μM duplex concentration and 1 M salt concentration. For the duplexes of the bile acid hybrids with **6**, three to five melting points were averaged, giving standard deviations of ±0.98 °C (**4**), ±0.52 °C (**14**), ±0.28 °C (**15**), ±0.95 °C (**16**), ±0.34 °C (**17**), and ±0.72 °C (**18**). The remaining melting points are the average of two measurements differing in every case by less than 2 °C, but more typically by less than 1 °C. See Experimental Section for further details.

(8) Sarracino, D. A.; Steinberg, J. A.; Vergo, M. T.; Woodworth, G. F.; Tetzlaff, C. N.; Richert, C. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2511−2516.

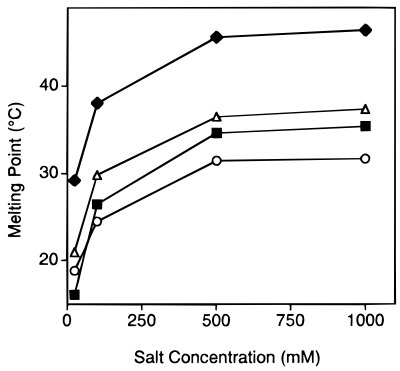


Figure 1. Melting points of duplexes of TGGTTGAC (**3**) with DNA complement (**2**, ■) and RNA complement (**5**, ○), compared to those of chl-T*GGTTGAC (**4**) with **2** (◆) and **5** (△) at different salt concentrations. See Table 1 for experimental conditions.

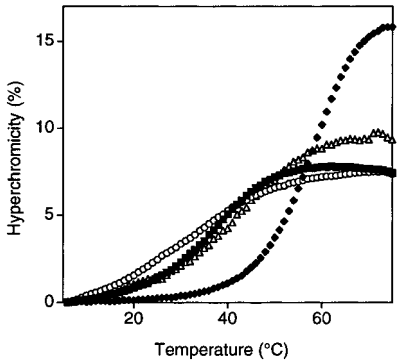


Figure 2. UV-melting curves of duplexes of TGCGCA (**10**, ■), chl-T*GCGCA (**9**, ◆), TGCGCG (**12**, ○), and chl-T*GCGCG (**11**, △) under the conditions given in Table 2.

Table 2. Melting Data for Dimerizing Oligonucleotides

sequences[a]	T_m (°C)[b]	hyperchromicity (%)
(chl-T*GCGCA)$_2$ (**9**)	57.4	16
(TGCGCA)$_2$ (**10**)	36.5	7.8
(chl-T*GCGCG)$_2$ (**11**)	43.4	9.8
(TGCGCG)$_2$ (**12**)	35	7.5
(chl-T*TTTTAAAAA)$_2$ (**13**)	44.6	39
(TTTTTAAAAA)$_2$ (**14**)	30.6	41

[a] The shorthand used is the same as that in Table 1. [b]Melting points are the average of two measurements at 7.9 ± 0.5 μM strand concentration (**9**−**12**) or 6.5 ± 0.5 μM strand concentration (**13**, **14**) and 1 M salt concentration.

this increase was only 8.4 °C (Table 2). Apparently, the cholic acid residues provide discrimination against mismatches at termini that are otherwise almost impossible to detect in melting experiments (ΔT_m = 1.5 °C between **10** and **12**). This was also found to be true for non-self-complementary sequences **3** and **6**, where the cholic acid induced enhanced discrimination against all three possible mismatches at the 3′-terminus in the target by elevating melting points of the fully matched sequence 4.0−5.5 °C more than those of the mismatched control (Table 3). Even at the penultimate position of the target, mismatch discrimination was increased, with $\Delta\Delta T_m$ values of 4.0−4.9 °C. A series of experiments with the same set of mismatched targets and strand concentrations up to 41 μM confirmed the "fidelity increases" with $\Delta\Delta T_m$'s of 2.7−5.8 °C (Supporting Information).

Not only did the cholic acid appendage induce a substantial increase in the fidelity of canonical Watson−Crick recognition by failing to stabilize the mismatches, it also induced an increase in the hyperchromicity accompanying duplex dissociation

Table 3. Melting Points of Oligonucleotide Duplexes with Complementary Strands Containing a Mismatched Nucleobase

oligonucleotide[a]	mismatched target	T_m (°C)[b]	discrimination (ΔT_m to fully matched target)
	Terminal Base Pair		
chl-T*GGTTGAC (**4**)	GTCAACC**T**	39.0	6.3
TGGTTGAC (**3**)	GTCAACC**T**	34.0	2.3
chl-T*GGTTGAC (**4**)	GTCAACC**G**	37.9	7.4
TGGTTGAC (**3**)	GTCAACC**G**	34.4	1.9
chl-T*GGTTGAC (**4**)	GTCAACC**C**	37.4	7.9
TGGTTGAC (**3**)	GTCAACC**C**	33.8	2.5
	Penultimate Base Pair		
chl-T*GGTTGAC (**4**)	GTCAAC**G**A	28.0	17.3
TGGTTGAC (**3**)	GTCAAC**G**A	23.9	12.4
chl-T*GGTTGAC (**4**)	GTCAAC**T**A	29.5	15.8
TGGTTGAC (**3**)	GTCAAC**T**A	25.1	11.2
chl-T*GGTTGAC (**4**)	GTCAAC**A**A	28.7	16.6
TGGTTGAC (**3**)	GTCAAC**A**A	23.7	12.6

[a] The shorthand used is the same as that in Table 1. [b]Melting points are the average of two measurements at 1.5 ± 0.2 μM strand concentration in 1 M salt solution, differing in every case by less than 2 °C, but more typically by less than 1 °C. See Experimental Part for further details.

(Figure 2), even though the fully saturated steroid moiety does not absorb at the wavelength monitored. This effect depends on duplex length and on the presence of a steroid at one or both termini, as expected, based on cooperativity considerations. In the case of hexamer **9**, hyperchromicity was doubled over that of control **10**, whereas the same effect was 27 and 17% for **4**:**6** and **3**:**6** and 23 versus 20% for duplexes of **4** and **3** with the longer target strand **2**. Only for a decamer duplex formed by the all A/T-duplexes (**13**, **14**) was this effect absent, but +14 °C in ΔT_m was still observed (Table 2).

The increase in sequence selectivity and DNA/RNA discrimination suggested that the cholic acid formed a tight, specific complex with the DNA. If so, the cholic acid should either bind with its concave or its convex side, i.e., with the α- or the β-face, to the terminal base pair. Since hydrogen bonds are known to be critical for selectivity in DNA−DNA and DNA−protein interactions, and the bile acids' hydrogen bond donor groups all protrude from the concave side, we hypothesized that the steroid bound with its α-face. The hydrogen bonds would presumably position the *cis*-connected A-ring of the steroid in the minor groove of the duplex. In the narrow minor groove of B-form DNA:DNA duplexes, a tight fit was imagined, while in the wider minor groove of A-form DNA:RNA duplexes,[9] the A-ring was thought to remain without tight interactions, explaining the observed DNA/RNA selectivity. To test this structural hypothesis, we probed the complex with oligonucleotides bearing analogues of cholic acid. Surprisingly, derivatives where one, two, or all three hydroxyl groups were missing on the cholic acid ring system (**14**−**17**, Scheme 1) gave melting points with **6** that were only 0.9−3.0 °C lower than that with parent compound **4** (Table 1). However, with the all-ketone cholic acid derivative dehydrocholic acid (**18**), where the ring is slightly distorted due to the changes in hybridization at carbons 3, 7, and 12, the melting point dropped by 4 °C. With a palmitic acid residue replacing the steroid (**19**), the duplex-stabilizing effect was down 7 °C. This suggested that (i) hydrogen bonds and unspecific hydrophobic forces were not dominating the duplex stabilization, and (ii) the concave side of the steroid ring system was not engaged in tight interactions with the DNA.

(9) A duplex of tryptophan-bearing **3** with **5** has been shown to be of the A-type (ref 7).

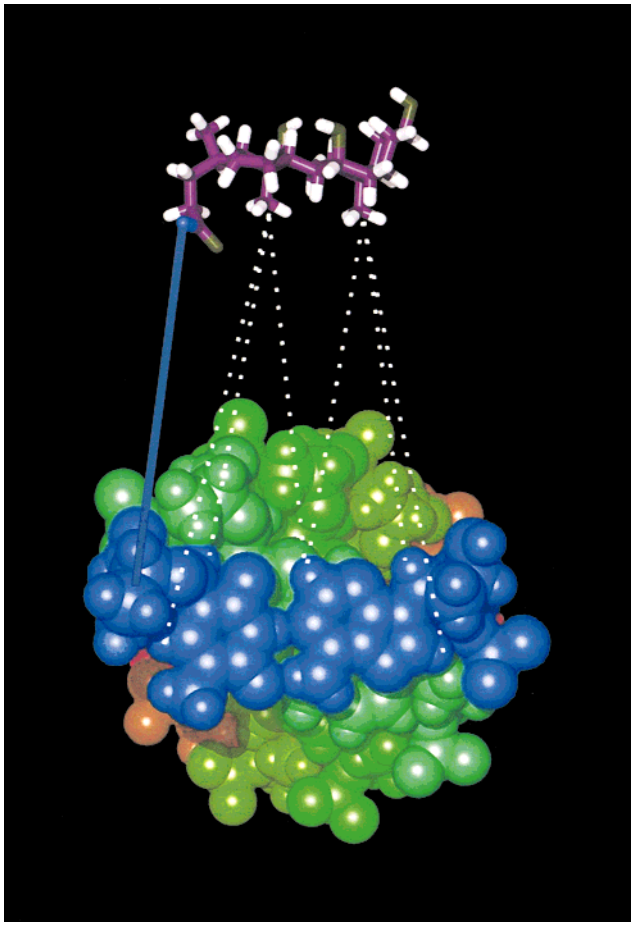


Figure 3. Schematic representation of key structural information obtained from cross-peaks in the NOESY spectrum of (chl-T*GCGCA)$_2$. Cholic acid and DNA are shown as separate molecules for clarity, with the covalent link indicated by a solid blue line. The DNA duplex conformation is that of a canonical B-form helix. The view is along the helix axis, and the terminal T:A base pair is shown in blue. Selected interproton distances of ≤6 Å between the C18 and C19 methyl groups of the cholic acid residue and the terminal nucleotides in (chl-T*GCGCA)$_2$ are indicated as broken white lines. See Supporting Information for NMR data. The figure was generated with VMD.[24]

An NMR study on the symmetrical duplex (chl-T*GCGCA)$_2$ (**9**) produced sufficient NOESY cross-peaks to locate the methyl groups at positions 18 and 19 of the steroid (i.e., the β-face) over the two terminal base pairs (Figure 3 and Supporting Information). Besides these methyl groups, several currently unassigned proton signals of the scaffold are shifted to higher field. Assuming that the unassigned protons are from the A-ring, the steroid ring system can be envisioned to pack against both the nucleobases and the protruding 2′-methylene group of the terminal adenosine, explaining the DNA/RNA discrimination observed. In a 3′-endo puckered ribonucleotide of the duplex with RNA, the 2′-CHOH group of the adenosine residue will be located further away from the two terminal nucleobases, making it more difficult for the bile acid to pack against the former and the latter simultaneously. The absence of cross-peaks between the methyl groups at positions 18 and 19 of the steroid and that of the thymine locates the cholic acid away from the major groove. Packing of a portion of the steroid backbone against the minor groove edge of the T:A base pair is therefore likely. A tight fit to this edge may explain the discrimination against mismatches, as T:A base pairs display a hydrophobic patch in the minor groove where C:G base pairs contain the polar 2-amino group of the guanosine.

Table 4. Thermodynamic Parameters for Duplex Dissociation, Derived from 1/Tm versus ln c_t Plots, for (chl-T*GCGCA)$_2$ (**9**) and Its Control Sequence T*GCGCA (**10**)[a]

duplex[b]	$\Delta H°$ (kcal/mol)	$\Delta S°$ (cal/mol K)	$\Delta G°$ (kcal/mol)
(TGCGCA)$_2$	46	124	7.9
(chl-T*GCGCA)$_2$	64	170	12

[a] Melting points were determined at 165 mM salt concentration (150 mM NaCl, 15 mM phosphate buffer, pH 7.0). [b]The shorthand used is the same as that in Table 1. See Supporting Information for plots and correlation coefficients.

Finally, a van't Hoff analysis of the melting transitions of the symmetrical duplexes of **9** and **10** at different concentrations[10] provided a thermodynamic signature of the stabilizing effect of the cholic acid residues. At near-physiological ionic strength (150 mM NaCl, 15 mM phosphate buffer), the free energy change of −4.1 kcal/mol brought about by the two steroid rings is due to a $\Delta\Delta H°$ of −18 kcal/mol, tempered by a relative increase in the entropic penalty for duplex formation by 46 entropy units (Table 4).

These results are noteworthy in several respects. First, based on salt titrations and structural considerations, the forces stabilizing the steroid-bearing duplexes are neither electrostatic attraction nor π-stacking interactions. Since the hydroxyl groups of the cholic acid are not critical for the stabilizing effect and the stabilization is enthalpic in nature, hydrogen bonding and hydrophobic effect also seem unlikely as governing forces. This leaves van der Waals forces between the alkyl components of the steroid and the nucleotides as the predominant stabilizing effect. Among these, those between the methyl groups at positions 18 and 19 of the cholic acid and the terminal A:T base pair are the most apparent in the structural picture obtained from the NMR data. Since placing methyl groups on aromatic rings is not commonly thought of when designing nucleic acid ligands, our results and those from nucleic acid[11] and non-nucleic acid model systems[12,13] may facilitate future drug design efforts.

Second, the observed DNA-binding mode of the covalently held steroid may be similar to that of other steroids with methyl groups on the β-face of their A/B and C/D ring junctions. Irehdiamine A, malouetine, various diaminoandrostanes, and several diamino bile acid derivatives induce hyperchromicity in double-stranded DNA and lower its melting point at high concentrations, indicating that they prefer to interact with exposed nucleobases.[2a−c,3g] Binding of dipyrandium, another steroidal diamine, to partly melted poly(dA:dT) shifts the highest field ^{1}H signals of the steroid to the same −0.1 ppm found for our C18/19 methyl groups in the stacked form of our complex,[2f] i.e., the methyl groups experience similar shielding from the nucleobases. Our results suggest direct stacking of the methyl groups on T:A base pairs,[14] rather than the more peripheral locations toward the groove suggested earlier.[15] Direct stacking of the methyl groups may also be considered for other 3,7-

(10) Marky, L. A.; Breslauer, K. J. *Biopolymers* **1987**, *26*, 1601−1620.
(11) Molinari, G.; Lata, G. F. *Arch. Biochem. Biophys.* **1962**, *96*, 486-490.
(12) Kim, E.-i.; Paliwal, S.; Wilcox, C. S. *J. Am. Chem. Soc.* **1998**, *120*, 11192−11193.
(13) Kobayashi, K.; Asakawa, Y.; Kato, Y.; Aoyama, Y. *J. Am. Chem. Soc.* **1992**, *114*, 10307−10313 and references therein.
(14) Similarly proposed for steroidal diamines in: Gourevitch, M.; Puidomenech, P.; Cave, A.; Etienne, G.; Mery, J.; Parello, J. *Biochimie* **1974**, *56*, 967−985 and ref 2 g.
(15) Sobell, H. M.; Tsai, C.-C.; Jain, S. C.; Gilbert, S. G. *J. Mol. Biol.* **1977**, *114*, 333−365.

substituted androstanes for which stacking of the α-face of rings C and D has been proposed.[16,17]

Third, the steroid−DNA hybrids may have practical applications. Covalently appended groups are frequently used to tune the properties of oligonucleotides.[18] So far, lipids have been coupled to oligonucleotides with a view toward improving cellular uptake or achieving duplex stabilization via lipid−lipid interactions.[4,19] The effects of the steroid substituents on target discrimination reported here should help to design more selective hybridization probes and antisense agents. In concert with selectivity-promoting modifications to the nucleobases and ribose rings,[20] high-fidelity hybridization probes may be constructed, in which the cholic acid moieties act as binding motifs for terminal T:A base pairs, deterring mismatched pairings and increasing hypochromicity of short duplexes. Independent of such applications, it is hoped that our results will stimulate further work on the exposed terminal base pairs of oligonucleotide duplexes as a binding site.

Experimental Section

Representative Protocol for Preparation of Bile Acid−DNA Hybrids. This representative procedure is for hybrid **16**, composed of lithocholic acid and DNA sequence 5′-T*GGTTGAC-3′, where T* denotes a 5′-amino-5′-deoxythymidine residue. Based on the MALDI spectra of crude products, the coupling of the bile acids to the amino-terminal DNA proceeded in >80% yield in all cases. With the current unoptimized workup and purification protocol, low recoveries are apparently due to adsorption of the lipophilically modified oligonucleotides to surfaces, explaining the comparatively low yields. Recoveries may be improved by silanizing or Teflon-coating all surfaces to which the hybrids are being exposed.

The DNA sequence was synthesized trityl-off on a 1 μmol scale on an ABI DNA synthesizer model 381A following the manufacturers recommendations (system software version 1.5). Phosphoramidites for dA^{Bz}, dC^{Bz}, and T were obtained from Perseptive Biosystems; dG^{dmf} was from ABI/Perkin-Elmer. The amino-terminal residue was introduced as previously reported.[21] A mixture of hydroxybenzotriazole (HOBT, 15.3 mg, 100 μmol), benzotriazole tetramethyluronium hexafluorophosphate (HBTU, 34.1 mg, 90 μmol), and lithocholic acid (37.7 mg, 100 μmol) was dried at 0.1 Torr for 1 h and dissolved in DMF (600 μL). The solution was treated with diisopropylethylamine (40 μL, 234 μmol), leading to a slight darkening. The reaction mixture was injected into a reaction chamber containing 5 mg of oligonucleotide-bearing controlled pore glass (∼0.2 μmol in DNA strands) and reacted for 85 min with flow-mixing every 5 min. The chamber was freed of the reaction mixture and the glass support washed with CH_3CN (2 × 3 mL), followed by drying at 0.1 Torr. The support was transferred to a polypropylene reaction vessel and treated with ammonium hydroxide (1 mL) for 16 h at room temperature. The supernatant was aspired, excess ammonia was removed with an air stream, and the pH was adjusted to 7.0 with glacial acetic acid. The solution was filtered (0.2 μm Whatman PVDF syringe-tip filter) and treated with ammonium acetate solution (200 μL, 1.0 M, pH 6.0). The resulting solution of the crude product was HPLC-purified using a C4 reversed phase column and a hyperbolical gradient of CH_3CN in triethylammonium acetate solution (0.1 M, pH 7.0). After freeze-drying, the product-containing fraction was re-lyophilized from deionized water (0.5 mL) to remove residual triethylammonium acetate. All oligonucleotides were HPLC-purified and characterized by MALDI-TOF mass spectrometry (see Supporting Information for spectra). Yields were determined from the integration of HPLC traces on a C18 (Alltech) or C4 (Vydac) column with a gradient of acetonitrile in 0.1 M triethylammonium acetate as described above. Calculated masses are average masses and *m*/*z* found are those of pseudomolecular ions ($[M - H]^-$), detected as the maximum of the unresolved isotope pattern.

lchl-T*GGTTGAC (16): yield 39%, HPLC: CH_3CN gradient 0% for 5 min, to 42% in 40 min, elution at 43.3 min. MALDI-TOF MS for $C_{103}H_{139}N_{30}O_{49}P_7$ $[M - H]^-$: calcd 2798.2, found 2796.5.

chl-W−T*GGTTGAC (1): yield 31%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 50% in 40 min, elution at 40.7 min. MALDI-TOF MS for $C_{114}H_{149}N_{32}O_{52}P_7$ $[M - H]^-$: calcd 3016.4, found 3012.6.

chl-T*GGTTGAC (4): yield 52%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 50% in 40 min, elution at 36.6 min. MALDI-TOF MS for $C_{103}H_{139}N_{30}O_{51}P_7$ $[M - H]^-$: calcd 2830.2, found 2830.1.

chl-T*GCGCA (9): yield 53%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 50% in 40 min, elution at 36.5 min. MALDI-TOF MS for $C_{82}H_{113}N_{24}O_{37}P_5$ $[M - H]^-$: calcd 2181.8, found 2181.1. See Supporting Information for assigned one- and two-dimensional ^1H NMR spectra.

chl-T*GCGCG (11): yield 71%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 39% in 40 min, elution at 43.8 min. MALDI-TOF MS for $C_{82}H_{113}N_{24}O_{38}P_5$ $[M-H]^-$: calcd 2197.8, found 2194.3.

chl-T*TTTTAAAAA (13): yield 21%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 50% in 40 min, elution at 36.2 min. MALDI-TOF MS for $C_{124}H_{165}N_{36}O_{61}P_9$ $[M - H]^-$: calcd 3414.6, found 3413.7.

dchl-T*GGTTGAC (14): yield 42%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 50% in 40 min, elution at 39.4 min. MALDI-TOF MS for $C_{103}H_{139}N_{30}O_{50}P_7$ $[M - H]^-$: calcd 2814.2, found 2816.3.

cdchl-T*GGTTGAC (15): yield 48%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 50% in 40 min, elution at 38.8 min. MALDI-TOF MS for $C_{103}H_{139}N_{30}O_{50}P_7$ $[M - H]^-$: calcd 2814.2, found 2815.1.

chol-T*GGTTGAC (17): yield 22%, HPLC on C4 phase: CH_3CN gradient 0% for 5 min, to 52% in 40 min, elution at 43.8 min. MALDI-TOF MS for $C_{103}H_{139}N_{30}O_{48}P_7$ $[M - H]^-$: calcd 2782.2, found 2779.2.

dhchl-T*GGTTGAC (18): yield 40%, HPLC on C4 phase: CH_3CN gradient 0% for 5 min, to 35% in 40 min, elution at 41.4 min. MALDI-TOF MS for $C_{103}H_{133}N_{30}O_{51}P_7$ $[M - H]^-$: calcd 2824.2, found 2821.2.

pal-T*GGTTGAC (19): yield 39%, HPLC on C18 phase: CH_3CN gradient 0% for 5 min, to 75% in 40 min, elution at 38.5 min. MALDI-TOF MS for $C_{95}H_{131}N_{30}O_{48}P_7$ $[M - H]^-$: calcd 2678.1, found 2678.7.

NMR Spectroscopy. Cholic acid−DNA hybrid **9** was synthesized (65 OD_{260} units), HPLC-purified, and product-containing fractions were dried and relyophilized five times from 5% ammonium hydroxide (500 μL) to remove residual triethylamine. NMR samples were prepared by dissolving the residue in H_2O/D_2O (9:1) or D_2O containing 150 mM NaCl and 10 mM phosphate buffer (KH_2PO_4/K_2HPO_4) at pH 7.0 (uncorrected for deuterium effect). Spectra were acquired with solutions 2 mM in **9** in microtubes (Shigemi Co., Tokyo, Japan). Acquisitions were performed on a Bruker DPX 600 spectrometer, and 500 and 750 MHz spectrometers custom designed at the Francis Bitter Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, MA. Experiments were performed at 10 °C with a WATERGATE gradient

(16) Waring, M. J.; Henley, S. M. *Nucleic Acids Res.* **1975**, *2*, 567−586.

(17) A third structural proposal can be found in: Hui, X.; Gresh, N.; Pullman, B. *Nucleic Acids Res.* **1989**, *17*, 4177−4187.

(18) Selected references: (a) Baudoin, O.; Marchand, C.; Teulade-Fichou, M. P.; Vigneron, J. P.; Sun, J. S.; Garestier, T.; Helene, C.; Lehn, J. M. *Chem. Eur. J.* **1998**, *4*, 1504−1508. (b) Rump, E. T.; de Vrueh, R. L. A.; Sliedregt, L. A. J. M.; Biessen, E. A. L.; van Berkel, T. J. C.; Bijsterbosch, M. K. *Bioconjugate Chem.* **1998**, *9*, 341−349. (c) Ali, O. M.; Franch, T.; Gerdes, K.; Pedersen, E. B. *Nucleic Acids Res.* **1998**, *26*, 4919−4924. (d) Saison-Behmoaras, T. E.; Duroux, I.; Thuong, N. T.; Asseline, U.; Helene, C. *Antisense Nucleic Acid Drug Dev.* **1997**, *7*, 361−368. (e) Klysik, J.; Kinsey, B. M.; Hua, P.; Glass, G. A.; Orson, F. M. *Bioconjugate Chem.* **1997**, *8*, 318−326. (f) Robles, J.; Maseda, M.; Beltran, M.; Concernau, M.; Pedroso, E.; Grandas, A. *Bioconjugate Chem.* **1997**, *8*, 785−788. See also: Helene, C.; Giovannangeli, C.; Guieysse-Peugeot, A. L.; Praseuth, D. *CIBA Symp. Ser.* **1997**, *209*, 94−102, and Kool, E. T. *Acc. Chem. Res.* **1998**, *31*, 502−510 for alternative approaches to improving the binding properties of oligonucleotides.

(19) Shea, R. G.; Marsters, J. C.; Bischofsberger, N. *Nucleic Acids Res.* **1990**, *18*, 3777−3783.

(20) Selected recent publications: (a) Lin, K.-Y.; Matteucci, M. D. *J. Am. Chem. Soc.* **1998**, *120*, 8531−8532. (b) Kuwahara, M.; Arimitsu, M.; Sisido, M. *J. Am. Chem. Soc.* **1999**, *121*, 256−257.

(21) Tetzlaff, C. N.; Schwope, I.; Bleczinski, C. F.; Steinberg, J. A.; Richert, C. *Tetrahedron Lett.* **1998**, *39*, 4215−4218.

pulse sequence[22] or with presaturation during the recycle delay to suppress the solvent signal. Spectra were collected with 2k data points in t_2, and 256 or 400 increments in t_1, and processed after zero-filling the latter to 1k. Peak assignment for the nucleic acid portion followed well-established protocols for B-DNA,[23] employing NOESY cross-peaks between resonances of H8 and H6 of the nucleobases and H1′ resonances of their own nucleotide and the H-1′ of their 5′-neighboring nucleotide, as shown in part in Figure S1 of the Supporting Information. Methyl groups of the steroid were identified by their integration value and, for CH_3 at position 21, the multiplicity of their signal. TOCSY and NOESY cross signals from CH_3-21 to H-17 allowed identification of CH_3-18 via NOESY cross-peaks.

Melting Curves and van't Hoff Analysis. UV melting curves were measured between 5 and 75 °C at 260 nm and a heating or cooling rate of 1 °C/min, as previously reported.[7] Melting points reported are the average of at least two measurements. For the study whose results are reported in Table 1, solutions were prepared in nuclease-free water with 0.1 mM EDTA and ammonium acetate buffer, pH 6.0. Results reported in Tables 2, 3, and S1 were obtained in deionized water with the desired salt concentration obtained through addition of ammonium acetate buffer. Under these conditions, duplexes **3** + **6** and **4** + **6**, on the comparison with whom the ΔT_m values in Table 3 are based, have melting points of 36.3 and 45.3 °C. Thermodynamic parameters were determined from ln c_t versus $1/T_m$ plots using $1/T_m = (R/\Delta H°)\ln c_t + \Delta S°/\Delta H°$, where T_m is the melting point, c_t is the total strand concentration, and R is the molar gas constant (8.3145 J mol^{-1} K^{-1}).[10]

Acknowledgment. This work was supported by the NIH, National Institute of General Medical Science, (Grant GM54783 to C.R.). The Center for Magnetic Resonance at the Francis Bitter Magnet Laboratory is supported by Grant No. RR00995, from the National Center for Research Resources at the NIH. The authors thank David Sarracino for helpful discussions and Professor David Walt for suggesting the initial mismatch study.

Supporting Information Available: NMR spectra of **9**, van't Hoff plots for **9** and **10**, melting points of duplexes with mismatched base pairs, an HPLC chromatogram and MALDI mass spectra of all hybrids (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA9920415

(22) Sklenar, V.; Piotto, M.; Leppik, R.; Saudek, V. *J. Magn. Reson. Ser. A* **1993**, *102*, 241−245.

(23) (a) Scheek, R. M.; Boelens, R.; Russo, N.; van Boom, J. H.; Kaptein, R. *Biochemistry* **1984**, *23*, 1371−1376. (b) Feigon, J.; Leupin, W.; Denny, W. A.; Kearns, D. R. *Biochemistry* **1983**, *22*, 5943−5951. (c) Patel, D. J.; Shapiro, L.; Hare, D. *J. Biol. Chem.* **1986**, *261*, 1223−1229. (d) Wijmenga, S. S.; Mooren, M. M. W.; Hilbers, C. W. In *NMR of Macromolecules, A Practical Approach*; Roberts, G. C. K., Ed.; Oxford University Press: Oxford, UK, 1993; pp 217−288.

(24) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.