Nucleic Acids Research, 1993, Vol. 21, No. 19 **4511–4515**

Relative specificities in binding of Watson–Crick base pairs by third strand residues in a DNA pyrimidine triplex motif

J.A.Fossella, Y.J.Kim, H.Shih, E.G.Richards and J.R.Fresco*
Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

Received June 21, 1993; Revised and Accepted August 6, 1993

ABSTRACT

The specificity of binding of Watson–Crick base pairs by third strand nucleic acid residues via triple helix formation was investigated in a DNA pyrimidine triplex motif by thermal melting experiments. The host duplex was of the type $A_{10}-X-A_{10}$: $T_{10}-Y-T_{10}$, and the third strand $T_{10}-Z-T_{10}$, giving rise to 16 possible triplexes with Z:XY inserts, 4 duplexes with the Watson–Crick base pairs (XY) and 12 duplexes with mismatch pairs (XZ), all of whose stabilities were compared. Two Z:XY combinations confirm the primary binding of AT and GC target pairs in homopurine·homopyrimidine sequences by T and C residues, respectively. All other Z:XY combinations in the T:AT environment result in triplex destabilization. While some related observations have been reported, the present experiments differ importantly in that they were performed in a T:AT nearest neighbor environment and at physiological ionic strength and pH, all of which were previously untested. The conclusions now drawn also differ substantially from those in previous studies. Thus, by evaluating the depression in Tm due to base triplet mismatches strictly in terms of third strand residue affinity and specificity for the target base pair, it is shown that none of the triplet combinations that destabilize qualify for inclusion in the third strand binding code for the pyrimidine triplex motif. Hence, none of the mismatch triplets afford a general way of circumventing the requirement for homopurine·homopyrimidine targets when third strands are predominated by pyrimidines, as others have suggested. At the same time, the applicability of third strand binding is emphasized by the finding that triplexes are equally or much more sensitive to base triplet mismatches than are Watson–Crick duplexes to base pair mismatches.

INTRODUCTION

Work over three decades (reviewed in ref. 1) has shown that Watson–Crick duplexes, regardless of the type of backbone, can interact spontaneously with nucleic acid single strands to form triple helices. While homopurine·homopyrimidine duplexes are especially favorable targets for third strand binding (2), triplexes can be formed from other than homopolymer sequences: residues can be accommodated extrahelically in all three strands (3,4); triplexes can be formed between target duplexes and third strands in which both have a repeating sequence (2,5), in which the third strand contains both purine and pyrimidine residues (4), and even in which the target and third strand have an irregular sequence (6,7). These various types of triplex are illustrative of a high degree of specificity in the binding of base pairs and third strand residues, which has been summarized in the *third strand binding code* (2).

In view of a potential biological role for triplex formation, and the possibility of exploiting third strand binding to genomic target duplexes for experimental, therapeutic and diagnostic purposes, we have undertaken a more refined analysis of the specificity of such interactions. In this report, we provide a semi-quantitative description and analysis of the effects of all possible combinations of centrally located single third strand residues and target base pairs on the thermal stability of an otherwise regular triplex sequence in the pyrimidine triplex motif, i.e., a triplex with an all-pyrimidine third strand.

MATERIALS AND METHODS

Deoxyoligonucleotides

Synthesized and purified by the Midland Certified Reagent Co. (Midland, Texas), these 21-residue oligomers were examined before use by both anion exchange and reverse phase HPLC. The desired oligomers were found in each case to constitute >95% of the strands. The concentrations of the $A_{10}-X-A_{10}$ purine-rich and $T_{10}-Y-T_{10}$ pyrimidine-rich strands were calculated using the molar extinction coefficients for poly(dA) at 25°C, ϵ_{257} = 8600, and for poly(dT), ϵ_{265} = 8700.

Solvent

Equimolar stock solutions of the single strands were prepared in 0.15 M NaCl/0.005 M $MgCl_2$/0.01 M cacodylate (Na^+)—pH 6.8. This solvent was selected for thermal melting experiments as it represents a reasonable approximation of the physiological ionic and pH environment, but for the basic peptides

* To whom correspondence should be addressed

or polyamines that may provide additional stabilization to a triplex structure.

Formation of duplex and triplex mixtures

Equimolar amounts of oligomers with particular X and Y residue inserts were mixed to form each of the four possible target duplexes with the Watson-Crick base pair inserts XY = AT, GC, TA, CG. These duplex mixtures were heated to 70°C and cooled over 24 hrs to 4°C to form duplexes with strands in complete register, as evidenced by the presence on PAGE of a single band with the mobility of a duplex, rather than higher molecular weight bands due to concatemers. Each duplex was combined at 4°C (to avoid pyrimidine strand exchange) with a stoichiometric amount of each of the four potential third strands with different Z inserts (Z=A, T, G or C) to generate 4 potential triplexes per family, each with a different third strand, i.e. 16 different triplexes. Each triplex mixture was incubated for 3–4 hrs at 4°C before it was melted.

Melting experiments

Absorbance-temperature profiles were obtained using an AVIV UV-VIS Model 14DS computer driven spectrophotometer equipped with a thermoelectric unit programmed to increase temperature stepwise every 2°C. The temperature was raised rapidly and the mixture allowed to equilibrate for 4.5 min., so that absorbance measurements were made on solutions at thermal equilibrium. Data collected at set wavelengths was corrected for baseline and computer plotted.

Since thermal stability of oligomer helices depends on oligomer concentration, thermal melting profiles were measured at a single Z+XY mixture concentration ($\sim 10^{-4}$ M in residues), and in the case of duplexes, at the concentration of XY in the Z+XY mixtures, in order to enable semiquantitative comparisons. By measuring melting profiles at several selective wavelengths (8), it was possible to observe for each mixture any triplex → duplex + single strand transition independently (283.5, 287nm) or along with the duplex → single strands transition (245,259nm), and to monitor any temperature-dependent baseline changes (320nm). There was no significant wavelength dependence for Tm values extracted from such profiles (Tm being determined within ±0.1°C—except when extrapolated, see below—as the temperature at which the differentiated profile δAbsorbance/δTemperature is at a maximum).

Experimental system

The ribo triplex poly(U:AU) and its deoxyribo analog poly(T:AT) are the most well studied (1). The pathways of their formation and thermal dissociation are well known, as are the hydrogen bonding schemes for the base triplets and the strand polarities. The deoxyribo form, 20 uniform T:AT triplets long, was selected as a host structure in which to insert a test base triplet combination Z:XY exactly in the middle, so that it would be flanked on either side by 10 T:AT triplets. Thus, the target consists of a strand with the sequence $A_{10}-X-A_{10}$ (the X strand) bound to a strand $T_{10}-Y-T_{10}$ (the Y strand). The bases X and Y were chosen so that XY is one of the four Watson–Crick pairs, AT, TA, GC, CG. The third strand is the Z strand, which has the sequence $T_{10}-Z-T_{10}$, where Z is one or other of the four bases. It is convenient to refer to such a triplex as Z:XY, which also defines the nature of the test triplet. The resultant triplexes, 21 base triplets long, contain target sequences of the size range to which nucleic acid single strands might be directed for third strand binding in biological systems. This length is short enough for the thermodynamic cost (or gain) of the test triplets to exert a readily discernible impact on the stability of the entire triplex.

RESULTS

Thermal stability of triplexes with the 16 possible base triplet combinations

Figure 1 shows the UV melting profiles at 259nm (where the hyperchromic changes for the two transitions are similar though not identical) for the four families of Z+XY combinations. These four sets of profiles each contain very similar well-defined, upper transitions that are due to the melting of the target duplex. This assignment is supported by the fact that the members of each

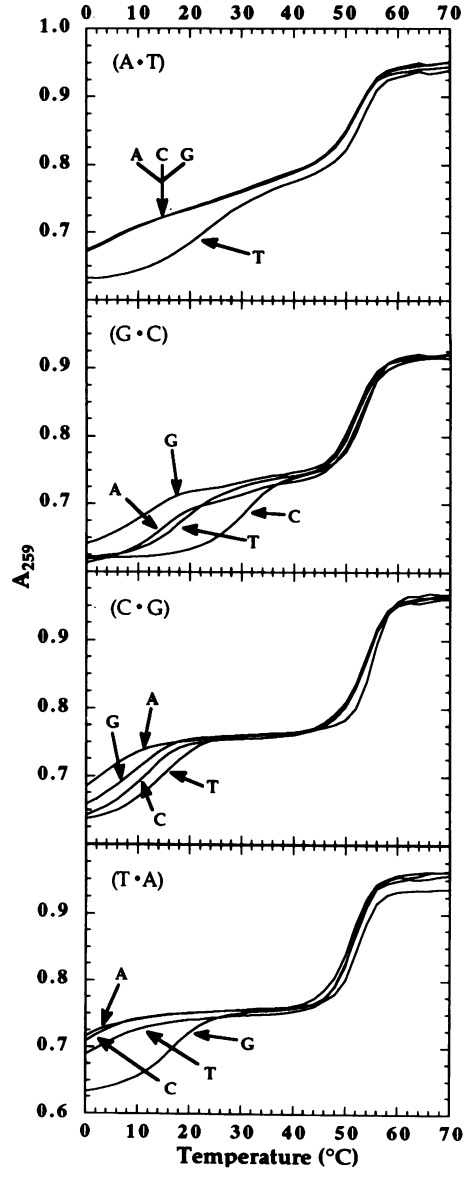

Figure 1. Equilibrium thermal melting profiles of triplexes with various test triplet inserts. The target doublet for each family of test triplets is given on the top left of each set of profiles and the third strand residues of the test triplet denote each profile.

family share essentially the same Tm value for that transition (Fig. 1), and by the close agreement of those upper transition Tm values with those for the corresponding duplexes XY at the concentration at which they are present in the Z+XY mixtures (Table 2, column III). It may also be seen from column III of Table 2 that the Tm values of the upper transitions in the Z+XY mixtures for which Z=Y are consistently ~1°C higher than those for the other members of the family. This is expected from the law of mass action since the Y strand is present at twice its concentration than in the other cases. Each profile also contains at least some indication of a lower transition from triplex → Watson–Crick duplex + single strand. The upper and lower transitions are separated in every case by a plateau region whose breadth depends on the differential stability of the triplex and its core duplex. Such biphasic melting is characteristic of polynucleotide triplexes at lower ionic strength (8).

Since the intermediate and high temperature plateaus of the melting profiles for each of the four families have essentially the same absorbance values, which is in keeping with all the Z+XY mixtures having the same oligomer concentrations, the absorbance values at the low temperature plateau for the lower transitions in the profiles in Fig. 1 must also be essentially the same. It is therefore possible, where necessary, to approximate the profiles below 0° with reasonable confidence.

Using the Tm values of all the lower transitions (Table 1), the relative destabilizing effects of the different Z residues in each family were evaluated. It can be seen from Table 1 that the nature of the Z base has a profound effect on the stability of the triplex. Thus, in the AT family, with Z=A, C, G, the host triplex is destabilized to the same extreme extent. By contrast, in the GC family, when Z=A, G, T the triplexes are only moderately destabilized relative to when the test triplet is C:GC. Moreover, it is apparent that the latter test triplet contributes more to the stability of the host triplex than does T:AT. In the CG and TA families the target base pair inserts are inverted so that the homopurine·homopyrimidine continuity of the target duplex is interrupted. In these families all four Z:CG test triplets and the G:TA test triplet moderately destabilize the host triplex relative to T:AT, while the other three Z:TA test triplets cause extreme destabilization.

The foregoing observations are summarized in Table 1 in terms of the Tm values for triplex dissociation. The host triplex, with its T:AT test triplet, has a Tm value of 22.6°C. Introducing C:GC as the test triplet raises Tm for the resultant triplex substantially, to 31°C. A linear relation between Tm and pH in 0.15 M Na^+/5 mM Mg^{2+} from pH 5.2 to 7.6, with a slope of −5.5°C/pH unit (data not shown), suggests that this test triplet forms with 2 H-bonds between a protonated third strand C residue and the target duplex, i.e., C^+:GC. The energetic cost of this protonation of C more than 2 units above its intrinsic pK may be compensated by the better stacking of C than T, and by the introduction of an isolated positive charge in the middle of the polyanionic helix. Moreover, third strand backbone regularity is not disturbed when C:GC replaces T:AT as the two triplets are isostructural.

Strand exchange

In three of the four combinations for each family of triplexes the two pyrimidine strands differ by only the Z residue. This creates the possibility of strand exchange between the Y- and Z-containing strands, which would result in a mixture with some XZ duplex and Y third strand (Y:XZ) in addition to the starting XY duplex and Z third strand (Z:XY). Such an occurrence would lead to erroneous interpretation of the observed Tm values, since the triplex transitions might display deceptively high Tm values and the duplex transitions deceptively low values. To test this possibility, all the Watson–Crick duplexes (XY) and mismatch duplexes (XZ) were formed at the concentration of XY or XZ in the Z+XY mixtures and their melting profiles determined. These profiles all showed common lower plateaus, well defined single cooperative transitions and matching upper absorbance values (data not shown). Table 2 lists the Tm values for the upper transitions in the original Z:XY triplex profiles (Column III), and for comparison the Tm values determined from the XY and XZ duplex profiles (Column V). A comparison of the Tm values for the XY and XZ duplexes possible in each family with those of the upper transitions clearly shows that the latter Tm values agree very closely with those for the corresponding XY Watson–Crick duplexes. In contrast, the Tm values for the mismatched XZ duplexes are significantly lower (by 5–13°C). Moreover, the first derivative profiles of the upper transitions are neither broader nor more asymmetric than those for the transitions of the corresponding XY duplexes (data not shown). Hence, these results indicate that the upper melting transitions observed for the Z+XY mixtures do not contain significant contributions from duplex combinations resulting from strand exchange (Y:XZ in addition to Z:XY). That such exchange is not detectable in the Z+XY mixtures is important because it confirms that the Tm values observed for the triplex transitions are those of the intended Z:XY combinations.

DISCUSSION

Differential sensitivity of the stability of duplexes to mismatches and of triplexes to third-strand mismatches

The availability of Tm values resulting from all possible test triplets (Table 1) and test doublets (Table 2) enables a comparison to be made of the sensitivity of the stability of triplexes to single mismatches in the third strand and of duplexes to mismatches. Figure 2 illustrates the differential sensitivity of duplexes and triplexes to such mismatches. It is evident that in comparison to the most stable triplex in the AT and GC families, triplexes with *third strands* containing single mismatches suffer much greater depression of Tm than do duplexes with single non-Watson–Crick test base pairs relative to the most stable fully Watson–Crick complementary duplex in each family. Thus, while the widths at the half-height of each first derivative profile are essentially the same in the two cases, ΔTm/mismatch for triplexes ranges between 12.7 and >20°C, whereas it varies only between 6 and 7.7°C for duplexes. It is apparent, therefore, that Watson–Crick duplexes are rather less sensitive to single mismatches than are triplexes. This finding is reasonable given

Table 1. T_m values for triplexes with various test triplets

Test Residue (Z) in Third Strand	Test Base-Pair (XY) in Watson-Crick Duplex: AT	TA	GC	CG
A	3.0	-5.0*	13.5	3.2
G	4.3	16.1	9.6	7.8
T	22.6	-0.3*	18.3	14.0
C	3.8	-3.0*	31.0	10.0

* T_m values <0°C were extrapolated. See text.

Table 2. Comparison of T_m values for upper transitions of Z+XY mixtures with those for all possible XY and XZ mixtures

Family, XY	Third Strand, Z	Tm of Upper Transition for Triplex Profiles	Possible Duplex Combinations	Tm of Possible Watson-Crick and Mismatch Duplexes
I	II	III	IV	V
AT	T*	53.3	AT	51.3
	G	52.0	AG	44.3
	A	52.3	AA	43.6
	C	52.0	AC	43.6
CG	T	54.0	CT	45.3
	G*	55.3	CG	54.0
	A	54.0	CA	43.0
	C	54.0	CC	41.0
TA	T	51.6	TT	46.6
	G	51.3	TG	47.0
	A*	52.3	TA	51.3
	C	51.3	TC	43.6
GC	T	53.3	GT	47.0
	G	52.3	GG	46.0
	A	52.6	GA	46.0
	C*	53.3	GC	53.0

*The member of each family for which the pyrimidine strands are identical.

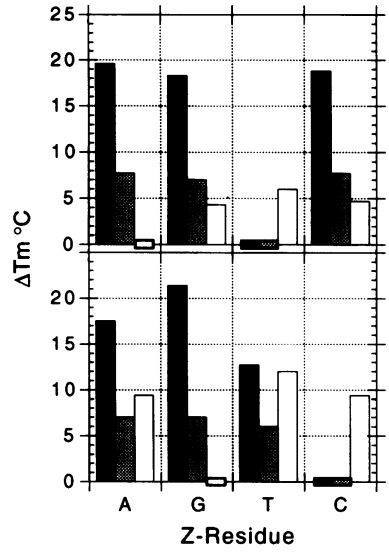

Figure 2. Differential sensitivity of triplexes and duplexes to single mismatches. The depression of Tm, ΔTm, is shown as a consequence of replacing the third strand test residue, Z, in T:AT or C^+:GC, with a mismatching one. ΔTm is also shown for the two comparable base pair mismatches. (Top plate—■ T:AT—Z:AT; ⊠ AT—AZ; □ AT−ZT. Bottom plate—■ C:GC−Z:GC; ⊠ GC−GZ; □ GC−ZC). It can be seen that ΔTm for the base triplet mismatches are generally much larger than for the base pair mismatches.

that the interaction of third strands and target duplexes is weaker than the interaction of complementary strands. This is likely due to the difference in charge density of the two types of helices that are contained within cylinders of essentially the same diameter. However, the finding is not relevant as regards the targeting of a particular third strand to matching (i.e., intended) vs. mismatching (i.e., unintended) target duplex sequences (see below).

Criteria and evaluation of sensitivity of third strand binding to mismatches due to the target Watson−Crick pair

Given that third strands are intended to bind uniquely to particular targets in a genomic sea of targets differing by at least one base pair, it is appropriate to consider to which target base pair(s) and with what affinity each type of third strand residue is most likely to bind in the motif being studied. Hence, our goal is to evaluate the degree of specificity with which target base pairs can be bound by third strand residues via triple helix formation. With the oligomers employed, third strand binding is determined in the first instance by the interaction of the 20 third strand T residues which flank the central test Z residue (10 on each side) with the matching AT pairs of the target. In addition, there is a positive or negative contribution from the test triplet itself. The essential issue, then, is to interpret the degree of reduction (or gain) in third strand binding due to replacement of the central T:AT by the test triplet. In addressing these considerations, we take into account the *affinity* of the third strand residue for the target pair, i.e., strength of binding, and the *specificity* of binding, i.e., uniqueness of binding. Both these criteria are important in attempting to decide whether or not particular third strand residues give rise to particularly functional triplets. Towards this end, we resort here to deductions based upon a comparison of Tm values.

From Table 1 it is apparent that for each third strand test residue there is one test target base pair that is bound most strongly: T:AT, G:TA, C^+:GC, and A:GC. However, each of these triplets is not of comparable utility as their third strand residues vary in affinity for their target, and perhaps more importantly, they are not comparably specific for their target pair.

Thus, in the case of the T residue, the Tm value is highest for T:AT, and the Tm difference (ΔTm) is substantial between T:AT and T:CG or T:GC and is extreme between T:AT and T:TA. Therefore, T shows a high level of affinity for AT and favorable specificity relative to other target pairs. The C residue shows an even higher level of affinity for GC. (Note, however, that it is isolated in the nearest neighbor environment studied. The affinity of C for GC would undoubtedly drop if two or more third strand C residues were nearest neighbors, making all but the first more difficult to protonate (cf. pH dependence of poly(C^+:GC) triplex formation in ref. 2).) Moreover, ΔTm values between C:GC and C:AT, C:CG and C:TA are so large that these triplets are not likely to contribute to infidelity in third strand binding for targets of discriminating length. While the third strand G residue binds to TA most strongly, G:TA is nevertheless weak, indeed weaker than T:GC. Hence, the good specificity of G for TA in comparison to GC, CG and AT target pairs is of marginal significance. Finally, the third strand A residue shows very low affinity for even its most tightly bound target pair, GC.

We are thus left with two triplets, T:AT and C:GC with pyrimidine third strand residues, whereby AT and GC pairs can be bound with great affinity. Whereas C shows great specificity for GC, T shows a minor tendency to also bind to GC and an even lesser one to bind to CG. Of the purine third strand residues, G shows some selective potential for binding to TA, with an

affinity intermediate between those of T for GC and for CG, while A shows no significant affinity for any target pair.

Third strand binding code words in the pyrimidine triplex motif

Our rank order of stabilities for the 16 possible triplet combinations (evident from Table 1) is not in full agreement with that observed in two recent investigations. One of these (9) used a similar spectroscopic assay method, but different pyrimidine nearest neighbor environments, acid rather than neutral pH, and 1 M Na^+ instead of 5 mM Mg^{2+}. The second study (10) utilized very acid pH and 2-D gel electrophoretic assays of superhelical density to detect intramolecular triplex formation and stability. Despite these experimental differences, these two studies, as well as others that evaluated only a few base triplet combinations (11–13), report as found here, the highest levels of binding of AT base pairs by T third strand residues and GC pairs by C residues, confirming earlier results obtained with homopolynucleotide systems (cf. 1). As for the remaining 14 test triplets, the differences in rank order no doubt reflect the different nearest neighbor, pH and ionic conditions studied.

Where the other studies especially differ from the present work is in the interpretation of the significance of various test triplet stabilities. In particular, it is our view that the present data do not justify extending the third strand binding code for the pyrimidine triplex motif to TA and CG in addition to AT and GC target pairs (11,12). This is so because TA and CG target pairs cannot be bound under physiological conditions with a combination of sufficient affinity and specificity by G and T residues, respectively. As noted, while the level of specificity of G for TA pairs is high, its level of affinity is modest at best. Also, the fact that T binds to CG modestly is of no value, since T binds to GC to an even greater extent and AT to a very much greater extent.

To evaluate these triplets further, we have also compared the sensitivity of the T:AT, G:TA and T:CG test triplets in the T:AT environment to increasing levels of stringency afforded by decreasing Na^+ concentration (data not shown). $\delta Tm/\delta log[Na^+]$ for the 3 → 2 transition is 58 ±1°C for T:AT, 61 ±1°C for G:TA, and 62 ±1°C for T:CG, with $\delta Tm/\delta log[Na^+]$ for the 2 → 1 transitions (16 ±1°C). Given the constancy of the dependence for the duplexes, the variations in the dependence between the triplex with T:AT and those with G:TA and T:CG are viewed as significant. Since $\delta Tm/\delta log[Na^+]$ for the 3 → 2 transitions is dominated by the 20 flanking T:AT triplets, it would appear that G:TA and T:CG are themselves much more sensitive than T:AT to increasing stringency. The fact that G:TA and T:CG (or T:GC) are not detected in homopolynucleotide binding experiments (under conditions where poly(G) self structure is not possible (2)), suggests also that when these potential triplets are nearest neighbors, third strand binding affinity is severely reduced. Moreover, it has been observed that when G:TA is flanked by C^+:GC on either or both sides, marked destabilization of third strand binding results (14). For these several reasons, it does not seem likely that these triplets represent a *general* way to circumvent the currently restrictive requirement of third strand binding for homopurine·homopyrimidine targets, as has been suggested (12). This is important because it removes a source of degeneracy of the binding code that would result in loss of specificity for target sequences.

Hence, we conclude that the specificities discussed above for T and C third strand residues bode well for the use of third strands directed to target duplexes as specific inhibitors of gene expression *in vivo*. So does the degree of destabilization due to single third strand mismatches (non-T:AT and non-C:GC triplets), which is equal to or greater than that for single mismatches in double stranded (anti-sense) complexes (Table 2). We note, in this connection, that whether or not the third strand binding mechanism is exploited for the modulation of gene expression in contemporary organisms, this mechanism has the necessary specificity to have found utility in a pre-protein RNA world.

ACKNOWLEDGEMENTS

This work was supported by NIH Grant GM 42936. We thank K.Yoon and A.Weis for helpful discussions.

REFERENCES

1. Broitman, S.L. and Fresco, J.R. (1993) *Progress in Nucleic Acid Res. and Mol. Biol.* **45**, 000–000.
2. Letai, A.G., Palladino, M.A., Fromm, E., Rizzo, V and Fresco, J.R. (1988) *Biochemistry* **27**, 9108–9112.
3. Fresco, J.R. and Alberts, B.M. (1960) *Proc. Natl. Acad. Sci. USA* **46**,311–321.
4. Lomant, A.J. and Fresco, J.R. (1975) *Progress in Nucleic Acid Res. and Mol. Biol.* **15**, 185–218.
5. Morgan, A.R. and Wells, R.D. (1968) *J. Mol. Biol.* **37**, 63–80.
6. Moser, H.E. and Dervan, P.B. (1987) *Science* **238**, 645–650.
7. Fresco, J.R. (1989) Abstracts, Conference on Oligodeoxynucleotides as Antisense Inhibitors of Gene Expression: Therapeutic Implications.
8. Blake, R.D., Massoulie, J., and Fresco,J.R. (1967) *J. Mol. Biol.* **30**, 291–308.
9. Mergny,J.-L., Sun, J.-S., Rougé, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J., and Héléne, C. (1991) *Biochemistry.* **30**, 9791–9798.
10. Belotserkovskii, B.P., Veselkov, A.G., Filippov, S.A., Dobrynin, V.N., Mirkin, S.M., and Frank-Kamenetskii, M.D. (1991) *Nucleic Acids Res.* **18**, 6621–6624.
11. Yoon, K., Hobbs, C.A., Koch, J., Sardaro, M., Kutny, R. and Weis, A.L. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 3840–3844.
12. Griffin, L.C. and Dervan, P.B. (1989) *Science* **245**,.967–971.
13. Roberts, R.W. and Crothers, D.M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9397–9401.
14. Kiessling, L.L., Griffin, L.C. and Dervan, P.B. (1992) *Biochemistry* **31**, 2829–2834.
15. Dolinnaya, N.G. and Fresco, J.R. (1992) *Proc. Natl. Acad. USA* **89**, 9242–9246.

This is paper no.19 in the series entitled 'Polynucleotides'. Paper no. 18 is ref. 15.