

DNA structural changes induced by intermolecular triple helix formation

Ibrahim Sayoh¹, David A. Rusling¹, Tom Brown² and Keith R. Fox^{1*}

¹School of *Biological Sciences, Life Sciences Building 85, University of Southampton, Southampton SO17 1BJ, UK*

²*Dept. of Chemistry University of Oxford, Oxford OX1 3TA, UK*

*To whom correspondence should be addressed

Email. k.r.fox@soton.ac.uk

Keywords: DNA triple helix, Triplex-forming oligonucleotide, DNase I, Footprinting, Fluorescence melting.

ABSTRACT

DNase I footprints of intermolecular DNA triplexes are often accompanied by enhanced cleavage at the 3'-end of the target site, at the triplex-duplex junction. We have systematically studied the sequence-dependence of this effect by examining oligonucleotide binding to sites flanked by each base in turn. For complexes with a terminal T.AT triplet, the greatest enhancement is seen with ApC, followed by ApG and ApT, with the weakest enhancement at ApA. Similar DNase I enhancements were observed for a triplex with a terminal C⁺.GC triplet, though with little difference between the different GpN sites. Enhanced reactivity to DEPC was observed at As that flank the triplex-duplex junction at AA(A or C) but not AA(G or T). Fluorescence melting experiments demonstrated that the flanking base affected the stability with a 4 °C difference in T_m between a flanking C and A. Sequences that produced the strongest enhancement correlated with those having the lower thermal stability. These results are interpreted in terms of oligonucleotide-induced changes in DNA structure and/or flexibility.

Introduction

Triplex-forming oligonucleotides (TFOs) bind sequence-selectively within the major groove of duplex DNA, forming specific hydrogen bonds to exposed groups on the base pairs (mainly to purines) ¹⁻⁴. Two main families of intermolecular triplexes have been described in which the third strand is either parallel or antiparallel to the purine-rich strand of the target. Parallel triplexes are characterised by T.AT and C⁺.GC triplets ^{4, 5}, while antiparallel complexes contain A.AT, G.GC and T.AT triplets ⁶. Parallel triplexes usually require conditions of low pH (<6.0), which is necessary for protonation of N3 of the third strand cytosine ^{7, 8}. They also require the presence of divalent cations (magnesium), which is especially necessary for stabilising the T.AT triplet ⁹⁻¹¹.

Several biophysical techniques have been used to probe the affinity, selectivity and structural effects of TFOs. Several high resolution NMR studies ¹²⁻¹⁸ and a limited number of X-ray crystallographic studies ¹⁹⁻²¹ have suggested that the underlying DNA duplex adopts a structure that is more like A-DNA, while retaining several B-like characteristics. UV and fluorescence melting studies have been used to determine the factors that affect triplex stability and to demonstrate selectivity for their intended target sequences^{22, 23}, while DNase I footprinting studies have shown the location of TFO target sites on long DNA fragments, and have been used to assess their selectivity and affinity^{24, 25}. DNase I has been the most commonly used footprinting probe, although the oligopurine TFO target sites are often relatively poor substrates for this enzyme.

Since DNase I cuts from the DNA minor groove, while the TFO is positioned in the major groove, TFO footprints cannot be caused by direct steric blockage of the enzyme. DNase I cleavage efficiency is known to be affected by local DNA structural variations, and A_n.T_n sequences are typically poor substrates for the enzyme on account of their narrow minor groove and rigid structure^{26, 27}. Crystal structures of the enzyme bound to short oligonucleotides showed that the DNA is bent away from the protein and this distortion may be an essential part of the catalytic mechanism, explaining why sequences, such as G_n.C_n are also poor substrates^{28, 29}. TFO-induced DNase I footprints could therefore be due to TFO-induced DNA structural changes or variations in the duplex flexibility.

Several studies have noted that there is often enhanced DNase I cleavage at the 3'-end of the TFO binding site at the triplex-duplex junction, and this is usually seen on the purine-rich strand ³⁰⁻³⁵. Enhancements in reactivity to diethylpyrocarbonate (DEPC) have also been noted at this location ³⁶. These enhancements are restricted to the triplex-duplex junction, and are only seen at a single bond at the 3'-end of the target, so cannot be explained by global changes in the ratio of the enzyme to free DNA ³⁷. Instead, these enhancements are thought to arise from TFO-induced changes in the local DNA structure that makes the phosphodiester bond at the duplex-triplex junction more susceptible to DNase I cleavage. These enhancements are often more pronounced than the footprints themselves and so are sometime evident in places where there is no clear DNase I footprint³⁸. In other instances, they have been used to estimate the location of the bound third strand³¹.

We have previously examined how the stability of the underlying duplex affects the apparent stability of the triplex ³⁹, but to date there have been no studies that systematically assess the sequence-dependence of these TFO-induced enhancements. We have often used the *tyrT* DNA fragment for these footprinting experiments, which we later modified to include a 17 base oligopurine tract *TyrT*(43-59) ⁴⁰. The 3'-end of this tract ends in an ApC step and in the present study we have changed this to ApX and GpX, where X = each base in turn and examined the interaction of these target with 12-mer and 11-mer TFOs using DNase I footprinting as well as the reaction with diethylpyrocarbonate. These studies have been augmented by fluorescence melting experiments with short oligonucleotides that are based on the same sequences.

Materials and Methods

Oligonucleotides

Three triplex-forming oligonucleotides were used for the footprinting experiments. 5'-CTCTTTTTTCTT (12-mer-T); 5'-CTCTTTTTTCTC (12-mer-C); and 5'-CTCTTTTTTCT (11-mer). These were provided by ATDBio and were synthesized using an Applied Biosystems ABI 394 automated DNA/RNA synthesizer using solid phase DNA phosphoramidite synthesis cycles. For the fluorescence melting experiments similar TFOs were prepared with 5'-dabcyl (Q) on the TFO and 5'-fluorescein on the duplex purine strand. The oligonucleotides were purified by gel filtration, dissolved in water and kept at -20 °C until required. Oligonucleotides for site-directed mutagenesis were also obtained from ATDBio and are listed in Table S1.

DNA sequences for footprinting

The base at the 3'-end of the oligopurine tract in *tyrT*(43-59) is an A, followed by C presenting an ApC step at the triplex-duplex boundary. The C at position 42 was changed to each base in turn by QuickChange PCR using the pairs of primers shown in Table S1. A set of four other derivatives was also prepared in which the A at position 43 was changed to G, followed by each of the other bases in turn, using the primers which are also shown in Table S1. The resulting plasmids were transformed into competent *E. coli* TG2 cells and plasmids were prepared using the QIAGEN miniprep kit. The sequences of the resulting plasmids were confirmed by MWG Eurofins. In this work the eight fragments are designated by the bases at positions 43 and 42 (*i.e.* AC, AG, AT, AA, GC, GG, GT and GA).

Plasmids containing the oligopurine target sites were digested with EcoRI and Aval and labelled at the 3'-end of the EcoRI site with α -³²P dATP using either reverse transcriptase or exo- Klenow fragment. The fragments of interest were separated from the remainder of the plasmid DNA on 6% polyacrylamide gels, eluted and dissolved in 10 mM Tris-HCl pH 7.4 containing 0.1 mM EDTA at a concentration of about 10 c.p.s./ μ L, as determined on a hand-held Geiger Counter.

DNase I footprinting

3 µL of TFO, diluted in 50 mM sodium acetate pH 5.0 containing 1 mM MgCl₂ were mixed with 1.5 µL of radiolabelled DNA and incubated at room temperature for 2 hours. DNase I cleavage was performed by adding 2 µL of DNase I (diluted to about 0.1 units/ml in 20 mM NaCl, 2 mM MgCl₂ and 2 mM MnCl₂) and digested for 2 minutes. The reactions were stopped by adding 5 µL of DNase I stop solution containing 80% formamide, 10 mM EDTA and 0.1% bromophenol blue. The samples were heated at 100 °C for 3 minutes, crash-cooled on ice before subjecting to polyacrylamide gel electrophoresis. A GA sequence marker was prepared by adding 1.5 µL of radioactively-labelled DNA to 20 µL of water and 5 µL of DNase I stop solution, and heating at 100 °C for 30 minutes with the cap open.

Reaction with diethylpyrocarbonate (DEPC)

3 µL of TFO, diluted in 50 mM sodium acetate pH 5.0 containing 1 mM MgCl₂ were mixed with 1.5 µL of radiolabelled DNA and incubated at room temperature for 2 hours, 3 µL of DEPC was then added and the reaction was left for 30 minutes with occasional mixing. The reaction was stopped by adding 2 µL of 3 M sodium acetate and the DEPC-modified products were precipitated, cleaved by adding 50 µL of 10% (v/v) piperidine, heated at 100 °C for 30 minutes and dried in a vacuum centrifuge. The pellets were washed with water, dried again and dissolved in 8 µL of DNase I stop solution.

Gel electrophoresis

The products of DNase I or DEPC digestion were separated on 8% denaturing polyacrylamide gels containing 8M urea. Gels were run at 1500 V for about 90 minutes, then fixed in 10% acetic acid and dried onto Whatman 3MM paper. Dried gels were exposed to phosphor screens overnight which were scanned with a Typhoon phosphorimager.

Fluorescence melting

Fluorescence melting experiments were performed as previously described²³. For these experiments the TFOs were labelled with 5-dabcyl (Q); *i.e* 5'-**Q**-CTCTTTTTTCTT; 5'-**Q**-CTCTTTTTTCTC and 5'-**Q**-CTCTTTTTTCT. The target duplexes were labelled with fluorescein at the 5'-end of the purine strand *i.e*. 5'-F-GAGAAAAAAGAR**X**TGGTTG, where R = A or G and X = each base in turn; these

were annealed with the complementary oligonucleotides 5'-CAACCA~~X~~YTCTTTTTTCTC (Y = C or T and X = each base in turn). These sequences are listed in Table 1.

Melting profiles were determined using a Roche LightCycler in a total volume of 20 μ L in 50 mM sodium acetate pH 5.0, containing 200 mM NaCl in the presence of $MgCl_2$. The duplex concentration was 0.25 μ M for all experiments, with TFO concentrations of 9, 5, 3, 1 and 0.25 μ M. The mixtures were annealed by heating to 98 $^{\circ}$ C for 5 minutes and cooling to 35 $^{\circ}$ C at 0.1 $^{\circ}$ C/second; the reaction was held at 35 $^{\circ}$ C for 5 minutes before heating to 98 $^{\circ}$ C at 0.1 $^{\circ}$ C/second. The fluorescence profile was recorded for both the annealing and melting phases and no hysteresis was observed for any of these sequences. Melting temperatures (T_m) were estimated from the maxima in the first derivatives of the melting profiles using the LightCycler software.

Results

Terminal T.AT triplet

DNase I footprints for the interaction of the 12-mer-T oligonucleotide with four DNA fragments that contain the same triplex target site, but in which the 3'-A is followed by each base in turn, are shown in Figure 1. In each case clear concentration-dependent footprints are evident that extend to TFO concentrations below 1 μ M. In the original *tyrT*(43-59) sequence the flanking base is C, and results for this are shown in the left hand panel. As previously noted, the footprint extends above (5'-) the target site by about four nucleotides and is accompanied by enhanced cleavage at the 3'-(lower) triplex-duplex junction (indicated by the arrow). This enhancement shows a similar concentration dependence to the footprint itself, and has previously been interpreted as resulting from a triplex-induced conformational change at the triple-duplex junction. Similar concentration-dependent enhancements are evident with the other fragments with T, A and G flanking the 3'-A. Visual inspection of these gels suggests that the enhancement is most pronounced for AC, and weakest for AA. This might indicate that the dinucleotide ApC is most easily distorted by flanking triplex formation into a form that is most easily cleaved by DNase I.

Similar experiments for the interaction of the 11-mer-TFO with these sequences are shown in Figure 2A. The 11-mer TFO lacks the 3'-terminal nucleotide of 12-mer-T, and the base flanking the 3'-end of its target site is an A for all four fragments. The variable base is therefore one residue removed from the target site (*i.e.* AAN). DNase I footprints can be seen with all four fragments, though these require higher concentrations than with the 12-mer TFO, as a result of its shorter length, and are only clearest at the highest concentrations (3 μ M). Enhanced DNase I cleavage (indicated by the arrows) is again evident at each of the triplex-duplex junctions, and as expected, this is one base higher than with the 12-mer-T TFOs. These enhancements are generally weaker than those seen with the 12-mer and the intensity of the enhanced band is most pronounced with sequence AC and weakest with AA.

Similar experiments were performed with the 12-mer-C TFO, which differs from 12-mer-T by replacing the 3'-T with C, generating an 11-mer triplex of T.AT and C.GC triplets, that is followed a 3'-terminal C.AT triplet mismatch. Previous studies³¹ suggested that some triplexes, such as this one, with terminal mismatches produce

enhanced DNase I cleavage after both the canonical (T.AT) and non-canonical, mismatched (C.AT) triplets. The results of these experiments are shown in Figure 2B, and show similar concentration-dependent footprints to those seen with the 11-mer. In this instance no enhancements are evident with the sequences AA and AT. In contrast, two weakly enhanced bands are seen with sequences AC and AG. The upper of these is at the same position as seen with the 11-mer-TFO, corresponding to the triplex-duplex junctions after the canonical T.AT, while the lower corresponds to the location of the mismatched C.AT triplet. Once again, the sequence AC produces the strongest enhancement of DNase I cleavage.

Terminal C⁺.GC triplet

The results presented above describe triplexes that contain a 3'-terminal T.AT triplet (or a C.AT mismatch). In order to assess whether the identity of the terminal triplet affects these properties we changed the base at the 3'-end of the target oligopurine tract from A to G, generating four fragments in which this base (G) is flanked by each base in turn (fragments GA, GC, GT and GG). These form a 12-mer triplex with oligo 12-mer-C that is similar to that with 12-mer-T, but ends in a C⁺.GC triplet instead of T.AT. DNase I footprinting experiments with these four new variants of the *tyrT* sequence are shown in Figure 3. Concentration-dependent footprints are evident with all four fragments, which persist to concentrations of about 0.2 μ M. This is lower than the triplexes with the terminal T.AT triplet, as a result of the greater stability of C⁺.GC triplet. All these footprints are accompanied by enhanced cleavage at the 3'-end of the target site, at the triplex-duplex junction. In this instance the intensity of the enhanced bands is similar for all four flanking sequences.

We also investigated the interaction of the 11-mer TFO with these target sites that end in a 3'-G. This triplex is identical to that formed between this TFO and the targets ending in A, though the immediate 3'-flanking base is G instead of A (AGN instead of AAN). The results of these DNase I footprinting experiments are shown in Figure 4A and show concentration dependent footprints that persist to a concentration of about 1 μ M. This is about 10 times higher than the concentration that is required to produce DNase I footprints with 12-mer-C at this target sequence and is similar to that seen with the 11-mer-TFO and the AN sequences. These footprints are again accompanied by enhanced cleavage, which as expected, is located one band higher than with 12-mer-C. The intensities of these enhancements

are similar for all four sequences and are stronger than those with the AN targets, even though the underlying triplex is identical.

We also examined the interaction of the 12-mer-T TFO with the target sites that contain a 3'-guanine. This should generate a triplex with 11 canonical triplets (C⁺.GC and T.AT) ending with T.AT, followed by a mismatched T.GC triplet. DNase I footprints for this interaction are shown in Figure 4B. This produces DNase I footprints that are only apparent at the highest TFO concentrations (3 μ M). These footprints are accompanied by only weak enhancements, which are one base higher than those seen with 12-mer-C, at the same position as seen with the 11-mer, with no enhancements at the mismatched terminal T.GC triplet.

Reaction with diethylpyrocarbonate

Diethylpyrocarbonate mainly reacts at N7 of adenines. Its reaction is generally poor in duplex DNA, but it has been used to detect unusual or distorted DNA structures in which this base is more exposed ^{41, 42}. Enhanced reactivity has previously been demonstrated at a triplex-duplex junction ³⁶. Figure 5 shows the results of DEPC cleavage experiments with the four AN fragments, in the presence of the TFOs 12-mer-T, 11-mer and 12-mer-C. The control lanes of these footprints show some reaction with As within the oligopurine tract (especially in the run of six consecutive As) and at the 3'-end of the oligopurine tracts of AA and AT (but not AG and AC). DEPC cleavage at these sites is attenuated by interaction with the TFOs in a concentration dependent manner, as expected, as the TFO binds to N7 of A and prevents access to the probe. There is no evidence of any ligand-induced enhanced DEPC reactivity in the presence of 12-mer-T oligonucleotide (Figure 5A; first four panels), even with sequence AA, which contains an A immediately adjacent to the TFO binding site.

Similar experiments with the 11-mer TFO are shown in the middle four panels of this Figure (Figure 5B). In contrast to the results with 12-mer-T, enhanced reaction to DEPC is seen at the 3'-end of the TFO binding site in the sequence AA, and to a lesser extent in AC, though no enhancements are apparent with AT and AG. These enhancements are located at the A that is immediately 3'- to the 11-mer binding site (i.e. at AAC and AAA, with sequences AC and AA respectively).

Similar experiments showing the reaction with DEPC in the presence of 12-mer-C are shown in the final four panels of Figure 5 (Figure 5C). These cleavage patterns are

similar to those formed with the 11-mer TFO, and again show enhanced DEPC reactivity at the 3'-end of the triplex-duplex junction in the sequences AC and AA (at AAC and AAA respectively), and is again especially strong for AA.

Other probes

We also examined the effect of these TFOs on the reaction with permanganate (reacting with exposed Ts), micrococcal nuclease (cleaving at pA and pT) and hydroxyl radicals (generating an even ladder of cleavage products). None of the TFO-target combinations induced any enhanced reaction with any of these probes.

Triplex stability

The experiments described above demonstrate that triplex formation can affect the susceptibility of flanking bases to some enzymes and chemical cleavage agents. However, these techniques are not sufficiently sensitive to detect any changes in triplex affinity. We therefore examined the stability of triplexes that are flanked by different base pairs by thermal melting studies using fluorescently labelled synthetic oligonucleotides. In these experiments the 12-mer third strand TFO was labelled at the 5'-end with dabcy1 while the purine strand of the target duplex was labelled at its 5'-end with fluorescein. The sequences of these oligonucleotides are shown in Table S2 and these were chosen to correspond to the 12-mer target site in the tyrT fragments. When the triplex is assembled, the fluorophore and quencher are in close proximity and the fluorescence is quenched. The triplex melts when the temperature is increased, separating the fluorophore and quencher, leading to a large increase in fluorescence. These experiments were performed in 50 mM sodium acetate pH 5.0 containing 1 mM MgCl₂ and 200 mM NaCl.

Fluorescence melting curves for the four target duplexes in the presence of 5 μ M of the 12-mer-T oligonucleotide are presented in Figure 6A and T_m values derived from these are presented in Table 2. These results show a clear difference in stability of these different complexes, even though they all contain the same triplex. There is a 4 °C difference in T_m between the highest (AG) and lowest (AC)

The results of similar experiments with the 11-mer TFO on these four target sequences are shown in Figure 6B and Table 2. As expected, these 11-mer triplexes melt at lower temperatures than those of the 12-mer-T triplexes, but again we find that the different complexes display significantly different melting temperatures.

These differences are less pronounced than for the 12-mer triplexes but the triplex with sequence AC melts about 2 °C lower than AG, even though these sequences only differ two base pairs distal to the triplex target site.

We also examined the stability of similar triplexes that contain a terminal 3'-C⁺.GC triplet instead of T.AT, using target duplexes that contain a GC base pair at the 3'-end of the polypurine tract, flanked by each base in turn. The melting curves of these four targets with TFO-12-mer-C are shown in Figure 6C and the T_m s are presented in Table 2. Sequence GT has the highest T_m (49.7 °C), though this is only slightly higher than GC (49.3 °C), GA (48.7 °C) and GG (48.6 °C). As expected, these T_m values are higher than those formed with the four targets with a 3'-terminal T.AT triplet as a result of the greater stability of the C⁺.GC triplet.

The results of similar experiments with the shorter 11-mer TFO on these four target sequences flanked by a GC base pair are also shown in Table 2. As expected, the T_m s of these 11-mer triplexes are between 8-10 °C lower than those of their 12-mer counterparts. This is a smaller reduction than between the 11mer and 12-mer-T for the targets ending with an AT base pair, reflecting the greater additional stability that is afforded by the C⁺.GC triplet compared to T.AT. As expected, there are only small differences between the T_m s of these 11-mer complexes, which only vary by a single base pair that is located two base pairs distal to the triplex.

Figure 6D shows the results of melting experiments with the AN target duplexes and the 12-mer-C TFO, generating a C.AT mismatch at the 3'-end of the triplex. The T_m values estimated from these data are shown in Table 2. Although these triplexes have slightly different melting temperatures, the differences are much less pronounced than with the 12-mer-T. Similar experiments were also performed with the 12-mer-T and the GN- targets sequences, generating a T.GC mismatch at the 3'-end of the triplex and the T_m s are shown in Table 2. It can be seen that there are no significant differences between these melting curves. Comparing the T_m of these triplexes with those generated by the fully matched 12-mer, it can be seen that the terminal mismatch decreases the triplex stability by about 10 °C. (Table 2). It appears that addition of a 3'-T.GC is destabilising compared to the 11-mer, while addition of a 3'-C.AT triplet increases triplex stability.

DISCUSSION

These results demonstrate that a common feature of intermolecular DNA triplex formation is the presence of enhanced DNase I cleavage at the triplex-duplex junction, at the 3'-end of the oligopurine strand of the target. For triplexes with a terminal T.AT triplet, the strongest enhancement is seen when this is followed by an ApC step (sequence AC), with the weakest observed at ApA. In principle, it would be best to compare cleavage of this band in the triplex, with that in the uncomplexed DNA, in order to assess the –fold enhancement in the presence of the triplex-forming oligonucleotide at each target site. However, this is often not possible to determine, as cleavage of this region is vanishingly low in the absence of the oligonucleotide. However, we have estimated the relative enhancements at these steps by comparing the intensity of the enhanced band with cleavage of other regions of the fragments that are not affected by triplex binding. These results are shown in Table 3 and confirm the strongest enhancement at ApC followed by ApG and ApT, with the weakest at ApA.

We assume that these enhancements reflect triplex-induced changes in the local DNA structure. How might these be reflected in altered DNase I activity? Crystal structures of DNase I bound to short oligonucleotides reveal that the enzyme functions by inserting an exposed loop into the DNA minor groove. Regions with a narrow groove (such as $A_n.T_n$ tracts) are therefore poor substrates for the enzyme. Ligand-induced changes in groove width have previously been used to explain enhancements adjacent to some small molecule binding sites ^{43, 44}. However, these changes are unlikely to account for these TFO-induced enhancements, as they are restricted to a single phosphodiester bond at the triplex-duplex junction. We can envisage two other possibilities to account for these changes. The BI to BII phosphate backbone configuration is known to depend on the local dinucleotide sequence, and DNase I favours the BII configuration ⁴⁵. It may be that triplex formation favours the BII configuration at the triple-duplex boundary, thereby enhancing enzyme cleavage. An alternative, though related, explanation is that DNase I is known to bend the DNA away from the minor groove at the cleavage site, and this bending may be an important part of the enzyme's catalytic mechanism ^{28, 29, 46}. Triplex formation may therefore alter the local deformability of the scissile phosphodiester bond. This is consistent with the observation that ApC is known to be

one of the most deformable dinucleotides and ApA is one of the most rigid ^{47, 48}. Such changes in DNA flexibility may not be restricted to the immediate sequence, but can be propagated into neighbouring sequences, and thereby account for the observation that the 11-mer produced greater enhancement with AC than the other three dinucleotides, even though the ApC step was one base removed from the enhancement.

These enhancements are not restricted to flanking ApN sites, but are also evident at GpN, at triplex-duplex junctions following a 3'-terminal C⁺.GC triplet. However, these enhancements are less sequence dependent and are similar magnitude for all GpN sites. It may be that the presence of a C⁺.GC triplet at the end of the triplex (instead of T.AT), which is known to give greater affinity, makes it more rigid, and less deformable.

The fluorescence melting experiments with oligonucleotides containing different flanking sequences showed subtle but significant changes in triplex stability, especially for the ApN series, with a 4 °C difference between the highest (ApG) and the lowest (ApC). It may be significant that the sequences with the lowest T_m generally correspond to the ones that produce the greatest enhancement in DNase I cleavage. It may be that, for the less stable complexes, some of the TFO binding energy is used to distort the DNA helix, or affect its dynamic flexibility, thereby reducing the inherent stability of the complex. In light of this, the targeting of oligopurine sites by TFOs might be improved by considering the flexibility of this flanking sequence.

The changes in reactivity to DEPC were surprising and showed no changes for interaction of the 12-mer T with any of the sequences, even at AA, with an A as the 3'-adjacent base. In contrast, the shorter (11-mer) sequence produced clear enhancements with the sequence AA, and also with AC. *i.e.* in the sequences AAA and AAC (in which the underlined base corresponds is the terminal triplet and the one in bold shows enhanced DEPC reactivity). This result is similar to that with the 12-mer C at these sequences (producing 11 canonical triplets and ending with a C.AT mismatch), for which DEPC enhancements were again observed with AA and AC. It therefore seems that DEPC enhancements are generated in sequences

AA(A/C) but not AA(G/T) (the interaction of the 12-mer T with AA ends with the sequence AAT).

References

- [1] Felsenfeld, G., Davies, D. R., and Rich, A. (1957) Formation of a three-stranded polynucleotide molecule, *J. Am. Chem. Soc* 79, 2023-2024.
- [2] Thuong, N. T., and Hélène, C. (1993) Sequence-specific recognition and modification of double-helical DNA by oligonucleotides, *Angewandte Chemie-International Edition in English* 32, 666-690.
- [3] Fox, K. R. (2000) Targeting DNA with triplexes, *Curr. Med. Chem.* 7, 17-37.
- [4] Moser, H. E., and Dervan, P. B. (1987) Sequence-specific cleavage of double helical DNA by triple helix formation, *Science* 238, 645-650.
- [5] Doan, T. L., Perrouault, L., Praseuth, D., Habhouh, N., Decout, J. L., Thuong, N. T., Lhomme, J., and Hélène, C. (1987) Sequence-specific recognition, photo-cross-linking and cleavage of the DNA double helix by an oligo-[alpha]-thymidylate covalently linked to an azidoproflavine derivative, *Nucleic Acids Res.* 15, 7749-7760.
- [6] Beal, P. A., and Dervan, P. B. (1991) Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation, *Science* 251, 1360-1363.
- [7] Sugimoto, N., Wu, P., Hara, H., and Kawamoto, Y. (2001) pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes, *Biochemistry* 40, 9396-9405.
- [8] Collier, D. A., and Wells, R. D. (1990) Effect of length, supercoiling, and pH on intramolecular triplex formation. Multiple conformers at pur•pyr mirror repeats, *J. Biol. Chem* 265, 10652-10658.
- [9] James, P. L., Brown, T., and Fox, K. R. (2003) Thermodynamic and kinetic stability of intermolecular triple helices containing different proportions of C⁺.GC and T.AT triplets, *Nucleic Acids Res.* 31, 5598-5606.
- [10] Floris, R., Scaggiante, B., Manzini, G., Quadrifoglio, F., and Xodo, L. E. (1999) Effect of cations on purine purine pyrimidine triple helix formation in mixed-valence salt solutions, *Eur. J. Biochem.* 260, 801-809.
- [11] Malkov, V. A., Voloshin, O. N., Soyfer, V. N., and Frank-Kamenetskii, M. D. (1993) Cation and sequence effects on stability of intermolecular pyrimidine-purine-purine triplex, *Nucleic Acids Res.* 21, 585-591.
- [12] Asensio, J. L., Lane, A. N., Dhesi, J., Bergqvist, S., and Brown, T. (1998) The contribution of cytosine protonation to the stability of parallel DNA triple helices, *J. Mol. Biol* 275, 811-822.
- [13] Asensio, J. L., Brown, T., and Lane, A. N. (1999) Solution conformation of a parallel DNA triple helix with 5' and 3' triplex-duplex junctions, *Structure* 7, 1-11.
- [14] De los Santos, C., Rosen, M., and Patel, D. (1989) NMR studies of DNA (R⁺)_n·(Y⁻)_n·(Y⁺)_n triple helices in solution: Imino and amino proton markers of T·A·T and C·G·C⁺ base-triple formation, *Biochemistry* 28, 7282-7289.
- [15] Radhakrishnan, I., Gao, X. L., Delossantos, C., Live, D., and Patel, D. J. (1991) NMR Structural Studies of Intramolecular (Y⁺)N·(R⁺)N(Y⁻)N DNA Triplexes in Solution - Imino and Amino Proton and Nitrogen Markers of G.TA Base Triple Formation, *J* 30, 9022-9030.
- [16] Radhakrishnan, I., and Patel, D. J. (1993) Solution structure of an intramolecular purine•purine•pyrimidine DNA triplex, *J. Am. Chem. Soc* 115, 1615-1617.
- [17] Radhakrishnan, I., and Patel, D. J. (1994) Solution Structure and Hydration Patterns of a Pyrimidine.Purine.Pyrimidine DNA Triplex Containing a Novel T.CG Base-Triplet, *J. Mol. Biol.* 241, 600-619.
- [18] Radhakrishnan, I., and Patel, D. J. (1994) DNA triplexes: Solution structures, hydration sites, energetics, interactions, and function, *Biochemistry* 33, 11405-11416.
- [19] Rhee, S., Han, Z. J., Liu, K. L., Miles, H. T., and Davies, D. R. (1999) Structure of a triple helical DNA with a triplex-duplex junction, *Biochemistry* 38, 16810-16815.
- [20] Abdallah, H. O., Ohayon, Y. P., Chandrasekaran, A. R., Sha, R. J., Fox, K. R., Brown, T., Rusling, D. A., Mao, C., and Seeman, N. C. (2016) Stabilisation of self-assembled DNA crystals by triplex-directed photo-cross-linking, *Chem. Commun.* 52, 8014-8017.

- [21] VanMeervelt, L., Vlieghe, D., Dautant, A., Gallois, B., Precigoux, G., and Kennard, O. (1995) High-resolution structure of a DNA helix forming (C.G)*G base triplets *Nature* 374, 742-744.
- [22] Mergny, J. L., and Lacroix, L. (2003) Analysis of thermal melting curves, *Oligonucleotides* 13, 515-537.
- [23] Darby, R. A. J., Sollogoub, M., McKeen, C., Brown, L., Risitano, A., Brown, N., Barton, C., Brown, T., and Fox, K. R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler, *Nucleic Acids Res.* 30, e39.
- [24] Hampshire, A. J., Rusling, D. A., Broughton-Head, V. J., and Fox, K. R. (2007) Footprinting: A method for determining the sequence selectivity, affinity and kinetics of DNA-binding ligands, *Methods* 42, 128-140.
- [25] Rusling, D. A., Powers, V. E. C., Ranasinghe, R. T., Wang, Y., Osborne, S. D., Brown, T., and Fox, K. R. (2005) Four base recognition by triplex-forming oligonucleotides at physiological pH, *Nucleic Acids Res.* 33, 3025-3032.
- [26] Drew, H. R., and Travers, A. A. (1984) DNA structural variations in the Escherichia-coli *tyrT*-promoter, *Cell* 37, 491-502.
- [27] Drew, H. R. (1984) Structural specificities of 5 commonly used DNA nucleases, *J. Mol. Biol.* 176, 535-557.
- [28] Suck, D., Lahm, A., and Oefner, C. (1988) Structure refined to 2Å of a nicked DNA octanucleotide complex with DNase-I, *Nature* 332, 464-468.
- [29] Lahm, A., and Suck, D. (1991) DNase I-induced DNA conformation. 2 Å Structure of a DNase I-octamer complex, *J. Mol. Biol.* 222, 645-667.
- [30] Bijapur, J., Keppler, M. D., Bergqvist, S., Brown, T., and Fox, K. R. (1999) 5-(1-propargylamino)-2'-deoxyuridine (U-P): a novel thymidine analogue for generating DNA triplexes with increased stability, *Nucleic Acids Res.* 27, 1802-1809.
- [31] Cardew, A. S., Brown, T., and Fox, K. R. (2012) Secondary binding sites for heavily modified triplex forming oligonucleotides, *Nucleic Acids Res.* 40, 3753-3762.
- [32] Cassidy, S. A., Strekowski, L., Wilson, W. D., and Fox, K. R. (1994) Effect of a triplex-binding ligand on parallel and antiparallel DNA triple helices using short unmodified and acridine-linked oligonucleotides, *Biochemistry* 33, 15338-15347.
- [33] Fox, K. R. (1994) Formation of DNA Triple Helices Incorporating Blocks of G.GC and T.AT Triplets Using Short Acridine-Linked Oligonucleotides, *Nucleic Acids Res.* 22, 2016-2021.
- [34] Sollogoub, M., Darby, R. A. J., Cuenoud, B., Brown, T., and Fox, K. R. (2002) Stable DNA triple helix formation using oligonucleotides containing 2'-aminoethoxy,5-propargylamino-U, *BIOCHEMISTRY-USA* 41, 7224-7231.
- [35] Stonehouse, T. J., and Fox, K. R. (1994) DNase-I footprinting of triple-helix formation at polypurine tracts by acridine-linked oligopyrimidines - Stringency, structural-changes and interaction with minor-groove binding ligands, *Biochimica et Biophysica Acta-Gene Structure and Expression* 1218, 322-330.
- [36] Collier, D. A., Mergny, J. L., Thuong, N. T., and Helene, C. (1991) Site-Specific Intercalation at the Triplex-Duplex Junction Induces A Conformational Change Which Is Detectable by Hypersensitivity to Diethylpyrocarbonate, *Nucleic Acids Res.* 19, 4219-4224.
- [37] Goodisman, J., and Dabrowiak, J. C. (1992) Structural changes and enhancements in DNase I footprinting experiments, *Biochemistry* 31, 1058-1064.
- [38] Chandler, S. P., and Fox, K. R. (1993) Triple-Helix Formation at A8Xa8.T8Yt8, *FEBS Lett.* 332, 189-192.
- [39] Rusling, D. A., Rachwal, P. A., Brown, T., and Fox, K. R. (2009) The stability of triplex DNA is affected by the stability of the underlying duplex, *Biophys. Chem.* 145, 105-110.
- [40] Brown, P. M., Madden, C. A., and Fox, K. R. (1998) Triple-helix formation at different positions on nucleosomal DNA, *Biochemistry* 37, 16139-16151.

- [41] Jeppesen, C., and Nielsen, P. E. (1988) Detection of intercalation-induced changes in DNA-structure by reaction with diethyl pyrocarbonate or potassium-permanganate - evidence against the induction of Hoogsteen base-pairing by echinomycin, *FEBS Lett.* 231, 172-176.
- [42] Kahl, B. F., and Paule, M. R. (2009) The Use of Diethyl Pyrocarbonate and Potassium Permanganate as Probes for Strand Separation and Structural Distortions in DNA, In *DNA-Protein Interactions: Principles and Protocols, Third Edition* (Moss, T., and LeBlanc, B., Eds.), pp 73-85.
- [43] Fox, K. R., and Waring, M. J. (1984) DNA structural variations produced by actinomycin and distamycin as revealed by DNAase-I footprinting, *Nucleic Acids Res.* 12, 9271-9285.
- [44] Low, C. M. L., Drew, H. R., and Waring, M. J. (1984) Sequence-specific binding of echinomycin to DNA - evidence for conformational-changes affecting flanking sequences, *Nucleic Acids Res.* 12, 4865-4879.
- [45] Heddi, B., Abi-Ghanem, J., Lavigne, M., and Hartmann, B. (2010) Sequence-Dependent DNA Flexibility Mediates DNase I Cleavage, *J. Mol. Biol.* 395, 123-133.
- [46] Suck, D., and Oefner, C. (1986) Structure of DNase-I at 2.0 Å resolution suggests a mechanism for binding to and cutting DNA, *Nature* 321, 620-625.
- [47] Travers, A. A. (2004) The structural basis of DNA flexibility, *Philosophical Transactions of the Royal Society of London Series a-Mathematical Physical and Engineering Sciences* 362, 1423-1438.
- [48] ElHassan, M. A., and Calladine, C. R. (1996) Propeller-twisting of base-pairs and the conformational mobility of dinucleotide steps in DNA, *J. Mol. Biol.* 259, 95-103.

Dabcyl-labelled TFOs

5'-Q-CTCTTTTTTCTT-3' (12-mer-T)

5'-Q-CTCTTTTTTCTC-3' (12-mer-C)

5'-Q-CTCTTTTTTCT-3' (11-mer)

Target duplexes used to form 12-mer triplexes with a 3'-T.AT triplet

AC: 5'-CAACCAGTTCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAACTTGGTTG-3'

AG: 5'-CAACCACTCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAAGTTGGTTG-3'

AA: 5'-CAACCATTCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAAATGGTTG-3'

AT: 5'-CAACCAATCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAATTTGGTTG-3'

Target duplexes used to form 12-mer triplexes with a 3'-C⁺.GC triplet

GC: 5'-CAACCAGCTCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAGCTTGGTTG-3'

GG: 5'-CAACCACCTCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAGGTTGGTTG-3'

GA: 5'-CAACCATCTCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAGATTGGTTG-3'

GT: 5'-CAACCAACTCTTTTTTCTC-3' 5'-F-GAGAAAAAAGAGTTTGGTTG-3'

Table 1. Oligonucleotides used in fluorescence melting experiments. Q = Dabcyl, F = fluorescein. The pyrimidine-containing TFOs were labelled with 5'-dabcyl while the target duplexes were labelled with fluorescein at the 5'-end of the purine strands. The variant nucleotides at the 3'-end of the target are underlined.

Sequence	12-mer-T	11-mer	12-mer-C
AC	40.7±0.4	38.2±0.5	42.8±0.2
AG	45.0±0.3	40.2±0.2	41.8±0.2
AA	43.7±0.3	40.4±0.2	42.1±0.4
AT	42.3±0.2	39.8±0.3	40.7±0.2
GC	36.4±0.2	39.7±0.4	49.3±0.1
GG	37.6±0.3	40.4±0.3	48.6±0.1
GA	38.5±0.2	40.9±0.2	48.7±0.1
GT	38.4±0.4	41.0±0.1	49.7±0.1

Table 2: T_m values ($^{\circ}\text{C}$) for the interaction of the three TFOs with different target sites. Reactions were performed in 50 mM sodium acetate pH 5.0 containing 200 mM NaCl. The target duplex concentration was 0.2 μM with 5 μM TFO.

Target sequence	Relative enhancement (arbitrary units)
AC	1.00 \pm 0.01
AG	0.54 \pm 0.04
AA	0.38 \pm 0.01
AT	0.62 \pm 0.05

Table 3: Relative intensities of the enhanced bands at the highest oligonucleotide concentration (3 μM), relative to bands in the rest of the fragment that are not affected by the oligonucleotide.

Legends to Figures

Figure 1: DNase I footprinting of the 12-mer-T oligonucleotide with fragments AX. The experiments were performed in 50 mM sodium acetate pH 5.0, containing 1 mM MgCl₂. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. Con indicates the control in the absence of added oligonucleotide and tracts labelled GA correspond to markers for purines. The solid bars indicate the location of the target site for the 12-mer, while the arrows indicate the location of the enhancements at the 3'-end of the target site. The DNA was labelled at the 3'-end, so the gel runs 5'-3' from top to bottom.

Figure 2: DNase I footprinting of oligonucleotides 11-mer (A) and 12-mer-C (B) with fragments AX. The experiments were performed in 50 mM sodium acetate pH 5.0, containing 1 mM MgCl₂. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. Con indicates the control in the absence of added oligonucleotide and tracts labelled GA correspond to markers for purines. The solid bars indicate the location of the 11-mer (A) and 12-mer (B) target sites. Note that 12-mer-C generates a mismatched C.AT triplet at the lower (3'-end) of the target site. The arrows indicate the location of the enhancements at the 3'-end of the target site. The DNA was labelled at the 3'-end, so the gel runs 5'-3' from top to bottom.

Figure 3: DNase I footprinting of the 12-mer-C oligonucleotide with fragments GX. The experiments were performed in 50 mM sodium acetate pH 5.0, containing 1 mM MgCl₂. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. Con indicates the control in the absence of added oligonucleotide and tracts labelled GA correspond to markers for purines. The solid bars indicate the location of the target site for the oligonucleotide, while the arrows indicate the location of the enhancements at the 3'-end of the target site. The DNA was labelled at the 3'-end, so the gel runs 5'-3' from top to bottom.

Figure 4: DNase I footprinting of oligonucleotides 11-mer (A) and 12-mer-T (B) with fragments GX. The experiments were performed in 50 mM sodium acetate pH 5.0, containing 1 mM MgCl₂. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. Con indicates the control in the absence of added oligonucleotide and tracts labelled GA correspond to markers for purines. The solid bars indicate the location of the 11-mer (A) and 12-mer (B) target sites. Note that 12-mer-T generates a mismatched T.GC triplet at the lower (3'-end) of the target site. The arrows indicate the location of the enhancements at the 3'-end of the target site. The DNA was labelled at the 3'-end, so the gel runs 5'-3' from top to bottom.

Figure 5: DEPC footprinting of oligonucleotides 12-mer-T (A); 11-mer (B) and 12-mer-C with fragments AX. The experiments were performed in 50 mM sodium acetate pH 5.0, containing 1 mM MgCl₂. Oligonucleotide concentrations (μM) are shown at the top of each gel lane. Con indicates the control in the absence of added oligonucleotide and tracts labelled GA correspond to markers for purines. The solid bars indicate the location of the target sites for 12-mer-T (A), 11-mer (B) and 12-mer-C (C). Note that 12-mer-C generates a mismatched C.AT triplet at the lower (3'-end) of the target site. The asterisks indicate the location of the enhancements at the 3'-end of the target site. The DNA was labelled at the 3'-end, so the gel runs 5'-3' from top to bottom.

Figure 6: Fluorescence melting curves for the interaction of 5 μM dabcyI-labelled 12-mer-T, 11-mer and 12-mer-C triplex-forming oligonucleotides with the target sites AX and GX. Experiments were performed in 50 mM sodium acetate pH 5.0, containing 200 mM NaCl and 1 mM MgCl₂. The duplex purine strand was labelled at the 5'-end with fluorescein. The target duplex concentration was 0.25 μM. A) interaction of 12-mer-T with AX, B) interaction of 11-mer with AX; C) interaction of 12-mer C with GX, D) interaction of 12-mer-C with AX.

Plasmid	QuickChange Primer pairs
AG	5' -GAAGAGAAAAAAGA <u>AG</u> TGGTTGCGTAATTTTC-3' 5' -GAAAATTACGCAACCACTTCTTTTTTCTCTTC-3'
AT	5' -GAAGAGAAAAAAGA <u>AT</u> TGGTTGCGTAATTTTC-3' 5' -GAAAATTACGCAACCAATTCTTTTTTCTCTTC-3'
AA	5' -GAAGAGAAAAAAGA <u>AA</u> TGGTTGCGTAATTTTC-3' 5' -GAAAATTACGCAACCATTTCTTTTTTCTCTTC-3'
GC	5' -GAAGAGAAAAAAGAG <u>CT</u> TGGTTGCGTAATTTTC-3' 5' -GAAAATTACGCAACCAGCTCTTTTTTCTCTTC-3'
GG	5' -GAAGAGAAAAAAGAG <u>GT</u> TGGTTGCGTAATTTTC-3' 5' -GAAAATTACGCAACCACCTCTTTTTTCTCTTC-3'
GT	5' -GAAGAGAAAAAAGAG <u>TT</u> TGGTTGCGTAATTTTC-3' 5' -GAAAATTACGCAACCAACTCTTTTTTCTCTTC-3'
GA	5' -GAAGAGAAAAAAGAG <u>AT</u> TGGTTGCGTAATTTTC-3' 5' -GAAAATTACGCAACCATCTCTTTTTTCTCTTC-3'

Table S1

Sequences of oligonucleotides used in QuickChange PCR to generate the eight triplex targets AN and GN, where N is each base in turn

Fig. 1

5' - ...AAGTGTAGGAA GAGAAAAAGA **AN** TGGTTG ... - 3'
 3' - ...TTCACAATCCTT CTCTTTTTTCTT NACCAAC ... - 5'

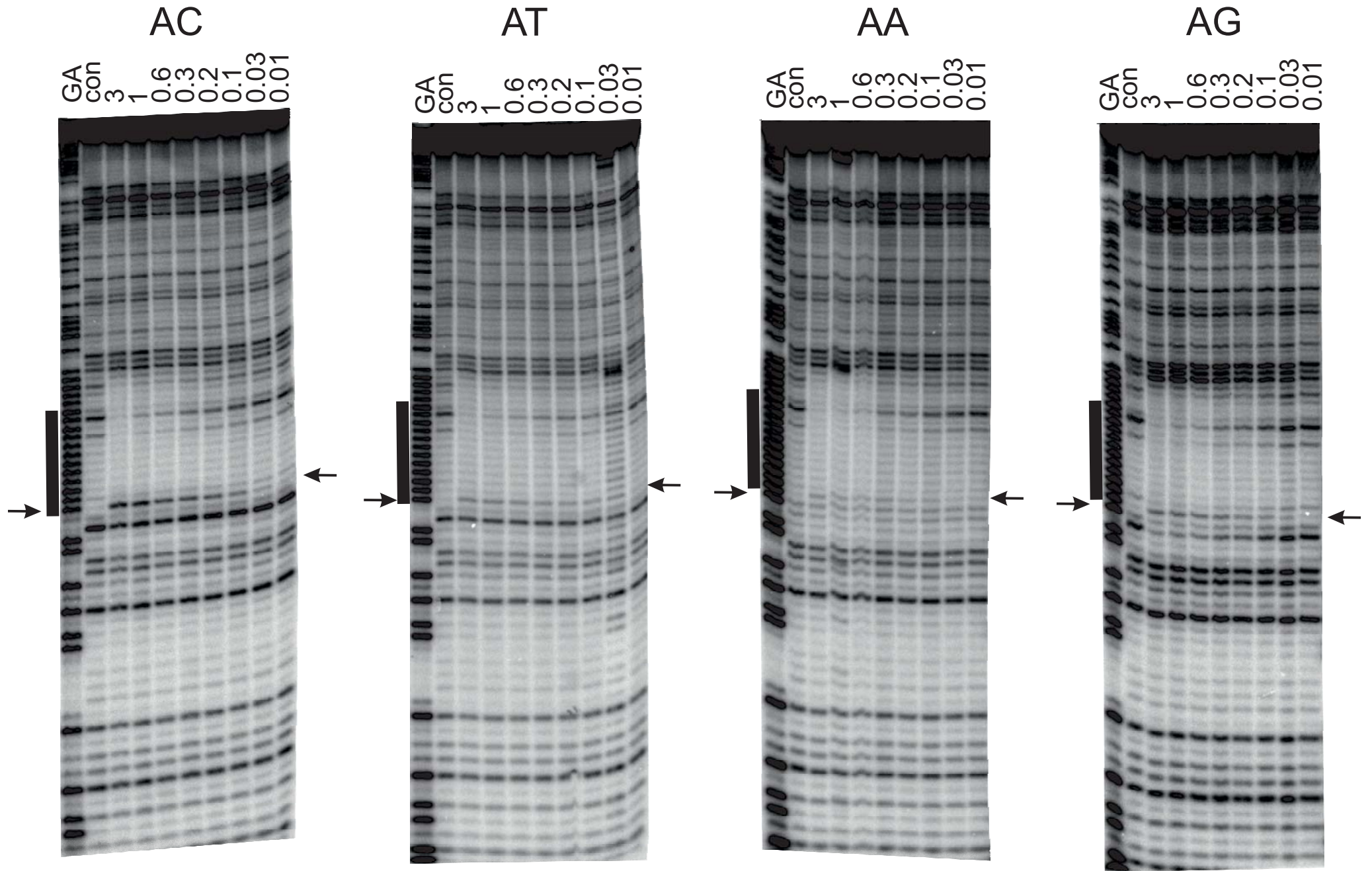


Fig. 2

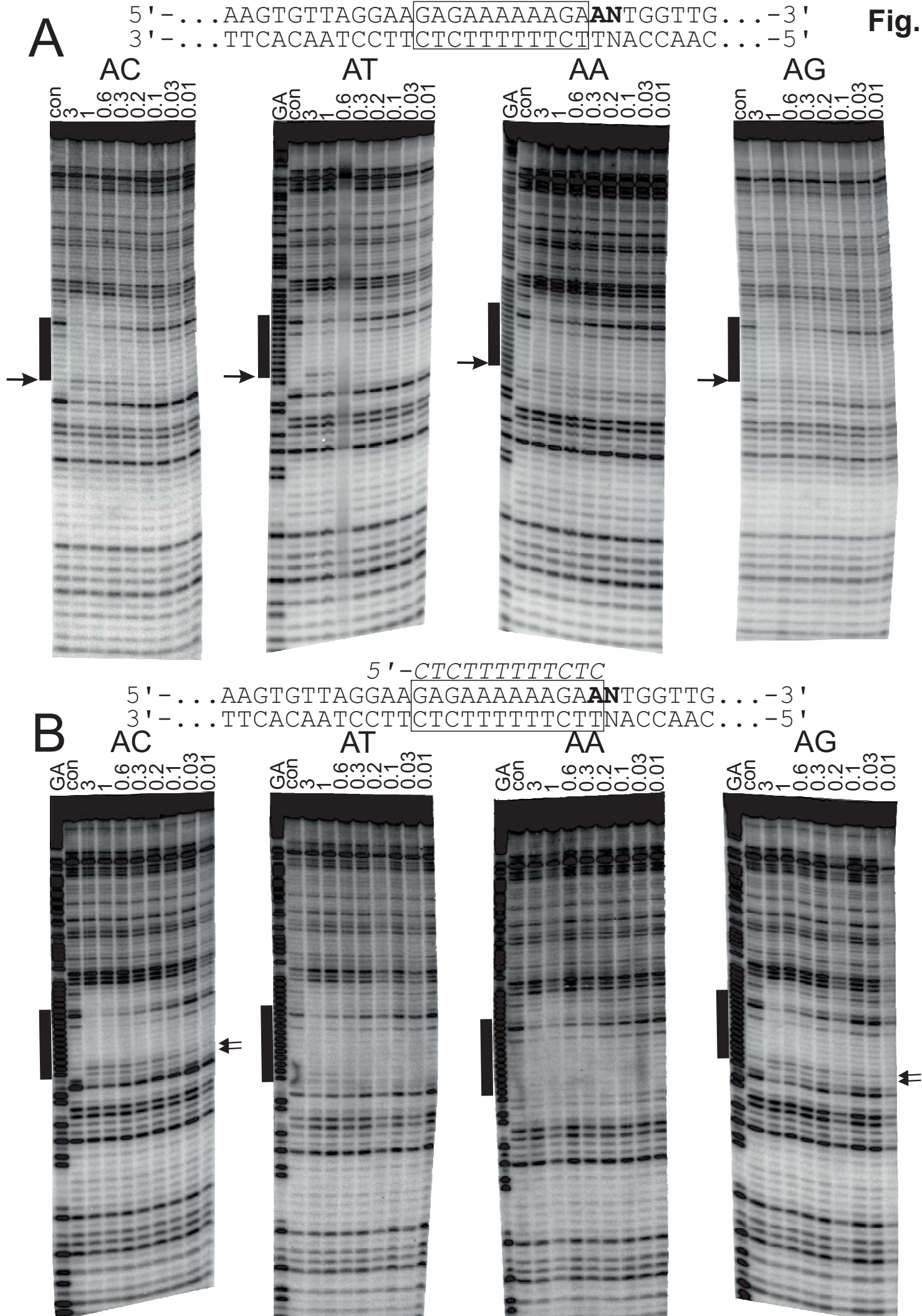


Fig. 3

5' - *CTCTTTTTTCTC*
 5' - ...AAGTGTTAGGAA GAGAAAAAAGAG **GN** TGGTTG ... - 3'
 3' - ...TTCACAATCCTT CTCTTTTTTCTC NACCAAC ... - 5'

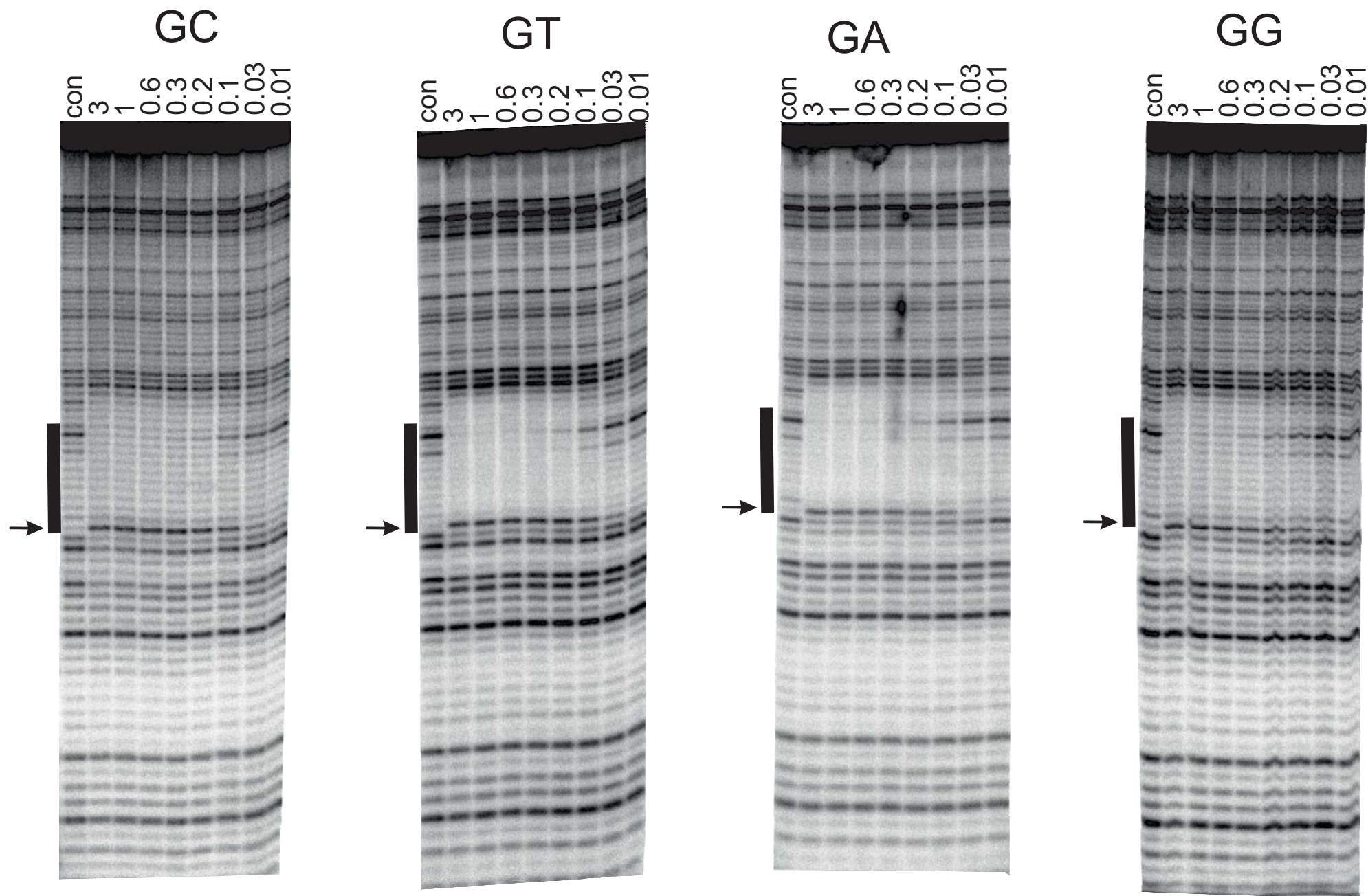


Fig. 4

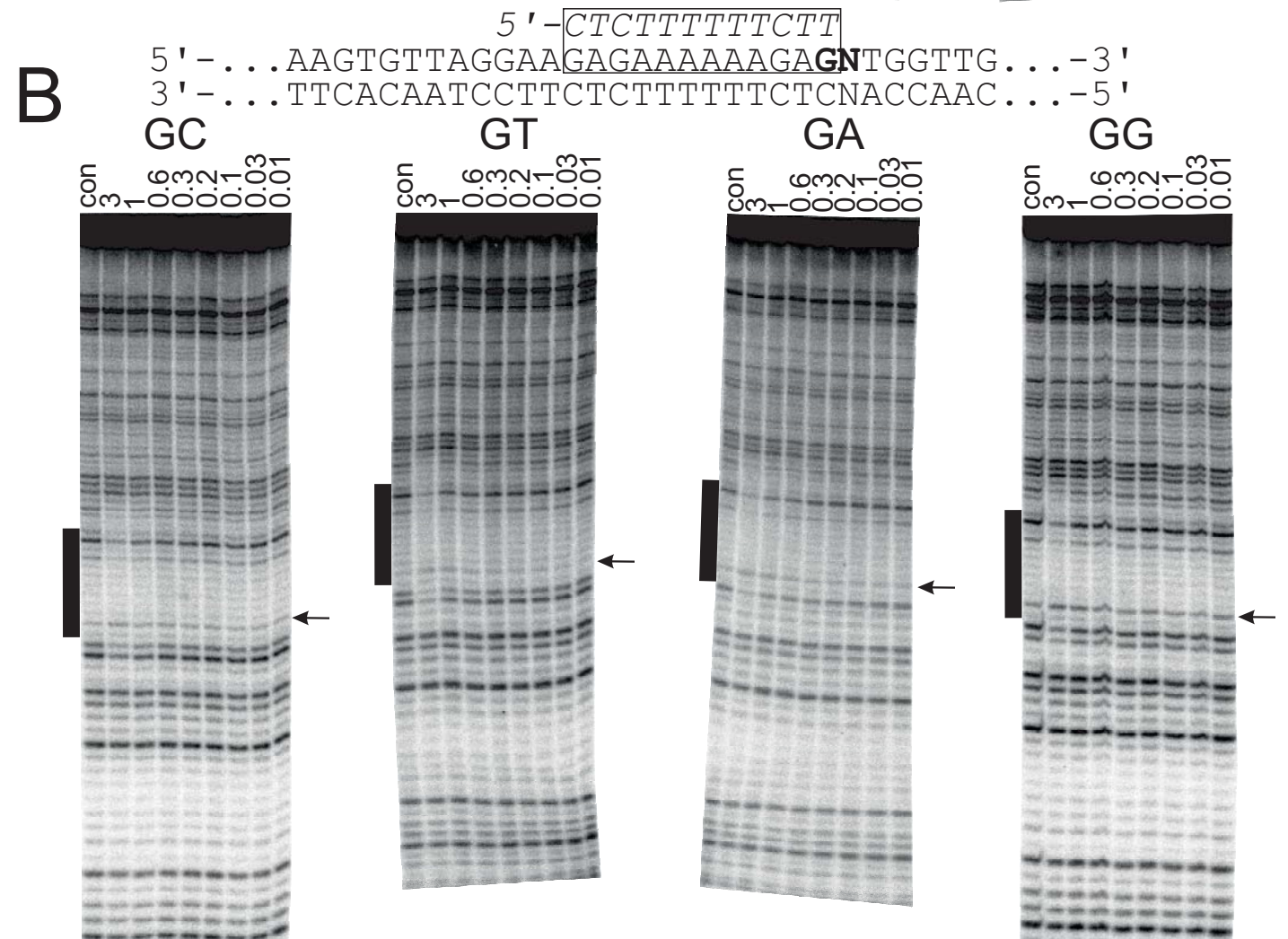
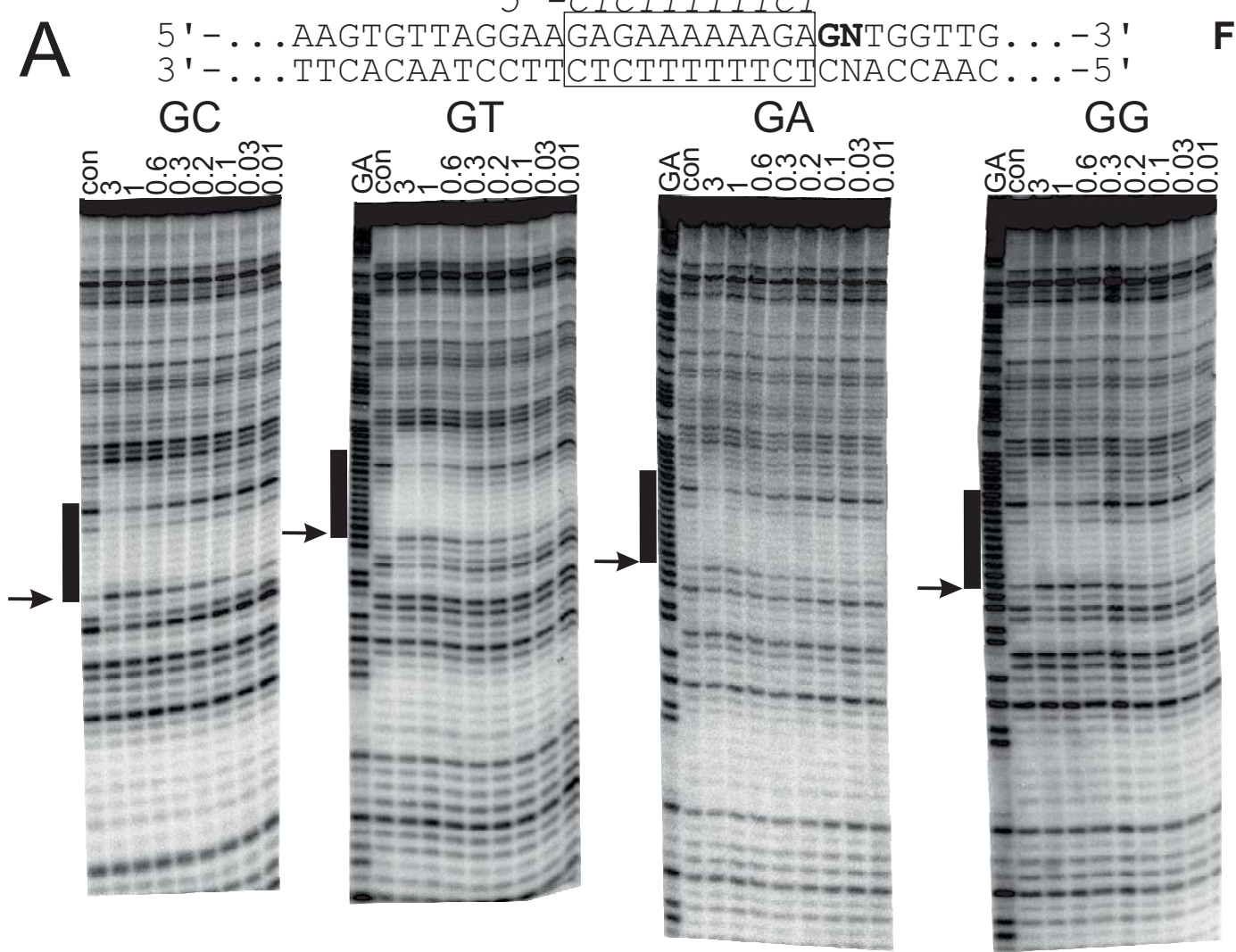


Fig. 5

5'-CTCTTTTTTCT 11-mer
 5'-CTCTTTTTTCTC 12-mer-C
 5'-CTCTTTTTTCTT 12-mer-T
 5'-...AGGAAGAGAAAAAAGAAANTGGTTG...-3'
 3'-...TCCTTCTCTTTTTTCTTNACCAAC...-5'

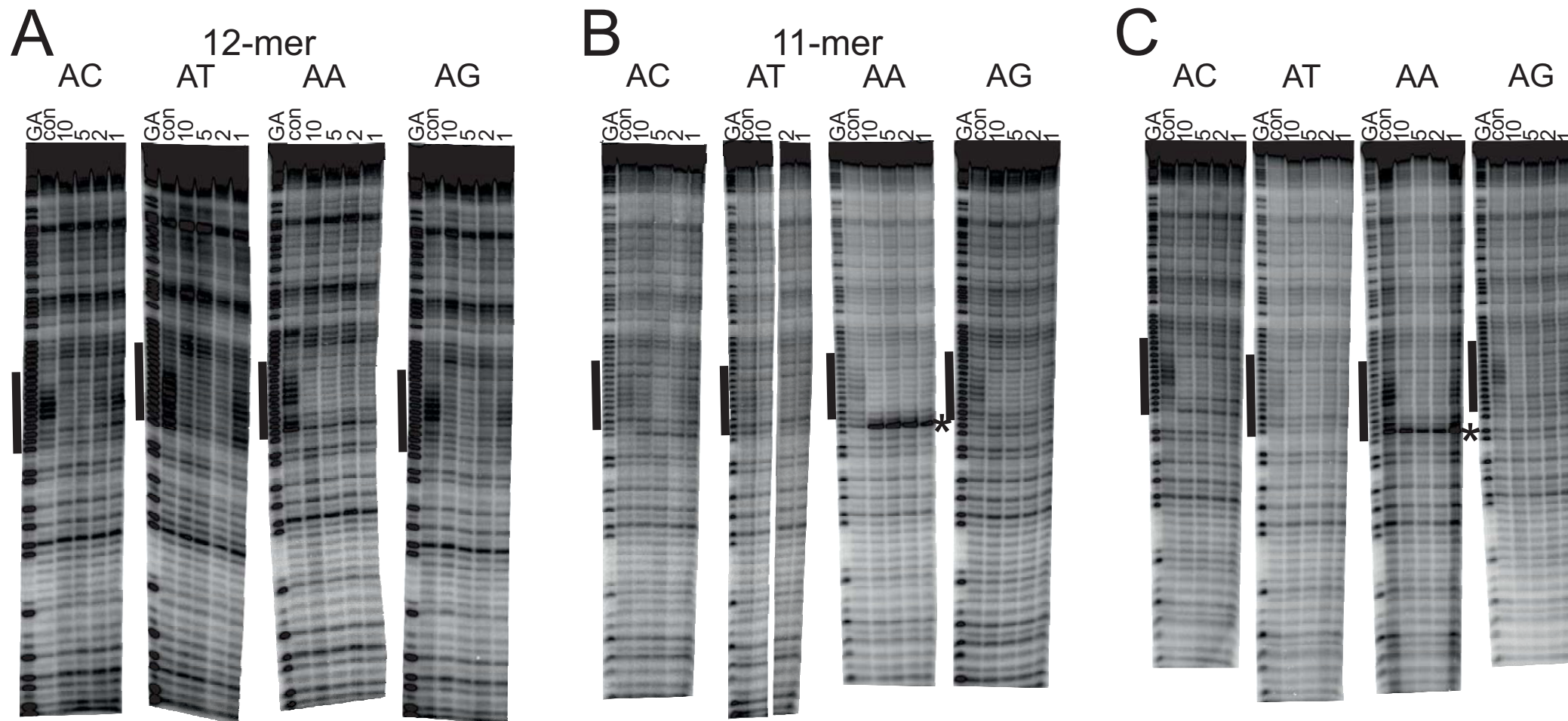
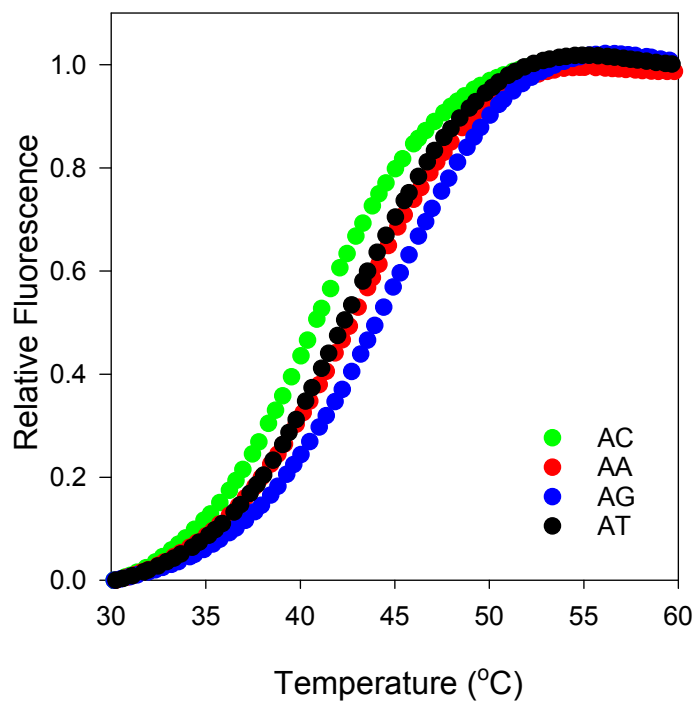
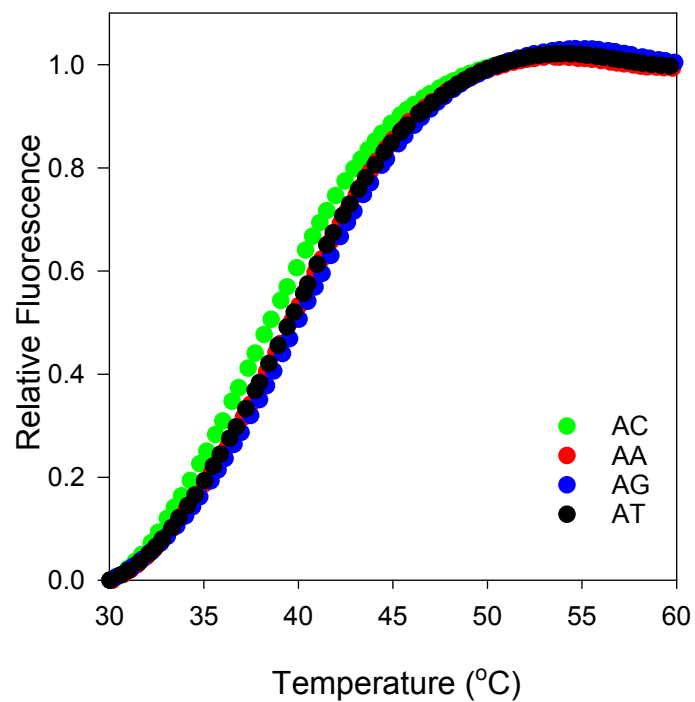


Fig. 6

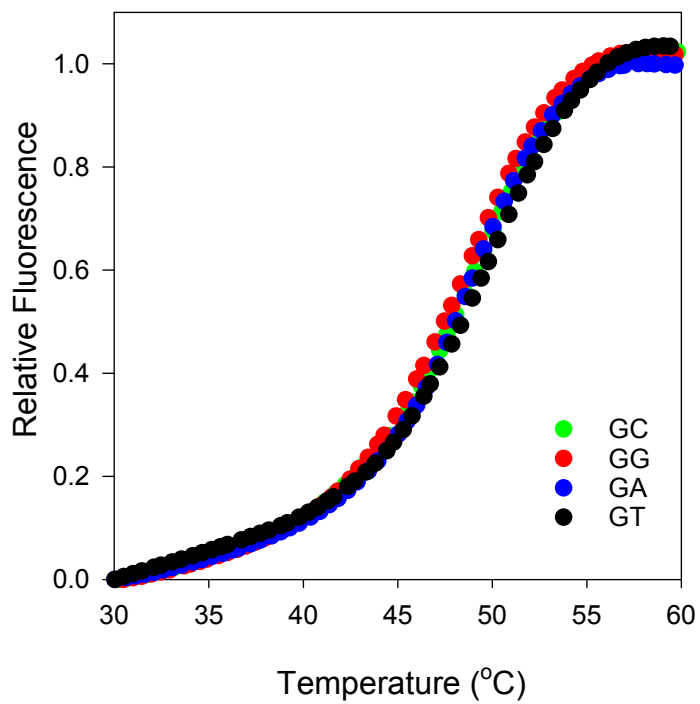
12mer-T



11mer



12mer-C matched



12mer-C- mismatch

