Nucleic Acids Research, 1999, Vol. 27, No. 7 **1569–1577**

SURVEY AND SUMMARY

Towards mixed sequence recognition by triple helix formation

Darren M. Gowers[+] and Keith R. Fox*

Division of Biochemistry and Molecular Biology, School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK

Received November 9, 1998; Revised and Accepted January 28, 1999

ABSTRACT

The formation of intermolecular DNA triple helices offers the possibility of designing compounds with extensive sequence recognition properties which may be useful as antigene agents or tools in molecular biology. One major limitation of this approach is that these structures are generally restricted to homopurine·homopyrimidine target sites. This review describes the strategies that have been employed to overcome this drawback and outlines the potential for triplex formation at mixed sequence DNA targets.

INTRODUCTION

The formation of triple-stranded nucleic acids was discovered in 1957 from biophysical experiments on synthetic polynucleotides in which 2:1 mixtures of poly(U) and poly(A) were found to form a specific three-stranded structure (1,2). Interest in these structures increased dramatically in the late 1980s with the realisation that triplex-forming oligonucleotides could be used as DNA sequence reading agents (3,4), with potential uses as antigene agents and tools in molecular biology (5–10). Two main classes of triple helix have been characterised which differ according to the orientation and base composition of the third strand. Pyrimidine-rich third strands bind parallel to the duplex purine strand and include T·AT and C^+·GC triplets, whereas purine-rich oligonucleotides bind in an antiparallel orientation and include G·GC, A·AT and T·AT triplets. (In the following, the notation X·ZY denotes a triplet in which the third strand base X interacts with a ZY base pair forming hydrogen bonds to base Z.) The structures of these triplets are shown in Figure 1. In both motifs the third strand lies in the DNA major groove where it forms specific interactions with substituents on the duplex bases. The details of the structures and applications of these standard triplexes have been considered in several recent reviews (5–10). Parallel triplexes generally require conditions of low pH, necessary for protonation of the third strand cytosines, whereas the formation of antiparallel structures is pH independent. Both triplex motifs are strongly stabilised by the presence of divalent metal ions, which are thought to screen the charge interactions between the three negatively charged phosphodiester backbones (11,12).

Problems

Despite their exquisite sequence recognition properties there are several outstanding problems which limit the use of triplex-based technologies. Amongst these are the pH dependency of the C^+·GC triplet (requiring the synthesis of novel cytosine analogues), the low stability of triplexes compared with their duplex counterparts (requiring the addition of stabilising ligands or duplex-binding agents), poor cellular uptake and nuclease instability. A further problem, which is addressed in this review, is that all the triplets described above involve third strand recognition of only the purine base of the duplex base pairs. Recognition of pyrimidine residues is harder to achieve and usually restricts triplex formation to homopurine·homopyrimidine tracts. Overcoming this restriction, to include recognition of pyrimidine residues, will significantly increase the potential use of triplexes and much effort is currently directed at extending the triplex recognition code to include all possible base sequences. The studies outlined below describe some of the strategies which have been employed for recognising pyrimidine interruptions using either natural bases or synthetic base analogues.

PYRIMIDINE RECOGNITION

Natural bases

Several studies have attempted to define a general triplex recognition code by examining the stability of all possible base triplet combinations using natural DNA bases (13–18). Using this approach G·TA and T·CG consistently emerge as the best triplets for pyrimidine recognition. Each of these triplets, which are described in detail below, contain only one hydrogen bond to the third strand base and so are less stable than the triplets used for recognising purine residues. The T·CG triplet can be accommodated within both parallel and antiparallel triplexes, while G·TA is limited to parallel structures. As a result there are currently no natural bases for specifically recognising a TA inversion within an antiparallel complex, though C·TA appears to be the least

*To whom correspondence should be addressed. Tel: +44 1703 594374; Fax: +44 1703 594459; Email. krf1@soton.ac.uk

[+]Present address: Department of Biochemistry, University of Bristol, School of Medical Sciences, University Walk, Bristol BS8 1TD, UK

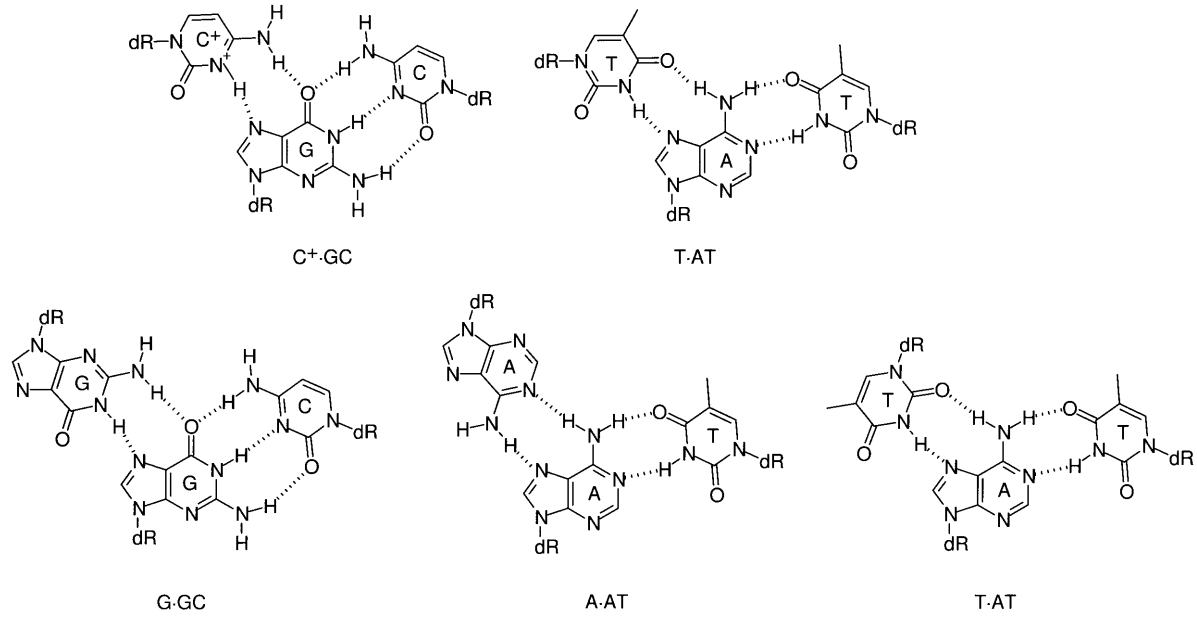

Figure 1. Chemical structures of parallel triplets T·AT and C$^+$·GC (top) and antiparallel triplets G·GC, A·AT and T·AT (bottom).

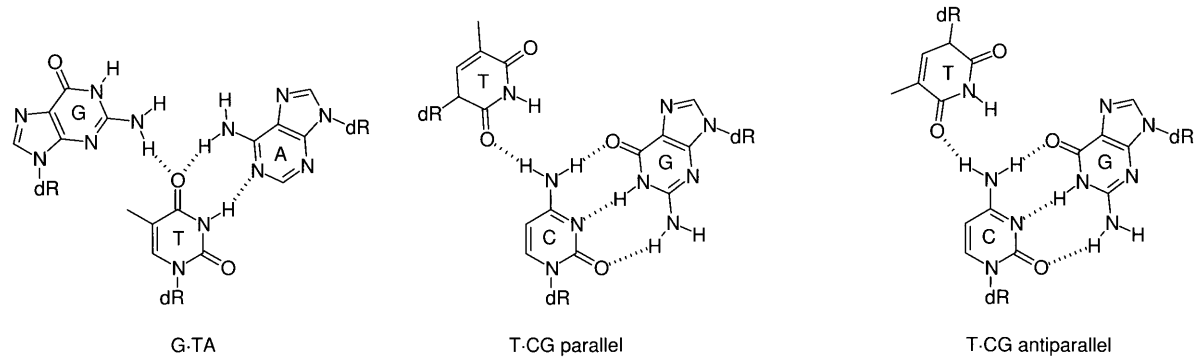

Figure 2. Structures of the G·TA triplet (left) and the T·CG triplet in both parallel and antiparallel orientations.

unfavourable combination (16). The chemical structures of G·TA and T·CG triplets are shown in Figure 2. Other mismatched triplets cause a large decrease in stability, which is greater at the centre than at the ends of the triplex (19–21). For antiparallel GT-containing oligonucleotides it has been suggested that point mutations at the 3′-end of the third strand have a greater effect on triplex formation than changes at the 5′-end (21).

G·TA triplet

The formation of the G·TA triplet was first proposed by Griffin and Dervan (13) from affinity cleavage experiments examining each of the triplet combinations across a TA inversion in a parallel DNA triplex. The properties of this triplet have since been examined in detail using both inter- and intramolecular triplexes.

Intermolecular studies. Inclusion of a single G·TA triplet within a longer T·AT flanking sequence was shown by footprinting studies to form a stable and specific complex (14); similar results were obtained by thermal denaturation studies (15). More recent work has examined the stability of triplexes containing multiple adjacent G·TA triplets and has suggested that each additional G·TA causes a 30-fold decrease in third strand affinity (22). Complexes containing up to three adjacent G·TA triplets at the centre of a parallel triplex can be formed but require the presence of a triplex-binding ligand (22). Formation of the G·TA triplet requires the 2-amino group of the third strand guanine; placing inosine opposite a TA base pair, generating an I·TA triplet, does not lead to stable triplex formation (13). The preferred glycosidic angle of the guanine base has been studied by using 8-bromoguanosine in place of G. This pseudobase, which adopts the *syn* arrangement, showed no triplex-forming ability, suggesting that the guanine adopts an *anti* configuration in the G·TA triplet (13). This property, which is common to parallel T·AT and C$^+$·GC triplets, has been confirmed by several NMR studies (23,24). Using deoxyuridine in place of thymine at a TA base pair (i.e. forming a G·UA base triplet) gives a triplex with significantly lower thermal stability than G·TA (25), suggesting that the 5-methyl

group of thymine is involved in stacking or hydrophobic interactions. The G·TA triplet is more stable when flanked by T·AT than C^+·GC (19), an effect which arises from the presence of an additional hydrogen bond to thymine in the adjacent T·AT triplet (see below).

Other studies have examined the possibility of targeting duplex regions of $(AT)_n$ with GT-containing oligonucleotides, generating complexes containing alternating G·TA and T·AT triplets (26). Although blocks of alternating T·AT and G·TA triplets alone are not stable, even in the presence of a triplex-binding ligand, these complexes can be induced to form by attaching this region to an adjacent block of consecutive T·AT triplets. In this way the duplex $A_{11}(AT)_6$·$(AT)_6T_{11}$ can be targeted with the third strand $T_{11}(TG)_6$, although this interaction still requires the presence of Mn^{2+} or a triplex-binding ligand. Further studies have shown that these complexes can be extended to longer $(AT)_n$ tracts (up to n = 11) (27) and can be stabilised with shorter T·AT tracts. The stability of these complexes increases with the length of the $(AT)_n$ tracts suggesting that the region of alternating G·TA and T·AT triplets makes a positive contribution to triplex stability. Similar complexes containing a few C^+·GC triplets in the anchoring tail are more stable and form in the presence of Mg^{2+}, without addition of a stabilising ligand (27). This is consistent with the observation that C^+·GC imparts a greater stability to triplex structures than T·AT (28–30) and raises the possibility that one means of facilitating triplex formation at pyrimidine inversions is to increase the stability of the surrounding canonical T·AT and C^+·GC triplets.

Intramolecular triplexes. NMR studies on short intramolecular (fold-back) triplexes, in which the three strands are linked together producing a single molecule consisting of seven or eight triplets, have revealed further features of the G·TA triplet (23,24,31–34). Of the two amino protons that could be used by guanine to bind the thymine O4 group, the structure using the proton closest to N3 is favoured, since this leads to optimized stacking interactions and a more regular phosphodiester backbone (34).

The TA base pair in the G·TA triplet shows little perturbation from a Watson–Crick conformation. However, the presence of G in an otherwise pyrimidine-rich strand produces some local changes in the structure of the third strand and the guanine is tilted out of the average plane of its TA target, avoiding steric clash with the thymine 5-methyl group. This produces a favourable stacking interaction between guanine and the thymine on its 5′-side (34,35). This interaction is thought to be a major determinant in the stability of the G·TA triplet (33). In this orientation an additional hydrogen bond can be formed between the unused amino proton of guanine and the duplex thymine of the 3′-adjacent T·AT triplet (23,32,34). As a result G·TA is more stable when flanked by T·AT than C^+·GC, as shown in affinity cleavage experiments (19). In order to facilitate these structural changes, the sugar pucker of the third strand guanine changes from C2′-*endo* to C3′-*endo*, positioning the sugar in a more favourable position within the backbone. The third strand also undergoes a slight overwinding at the TpG step and an underwinding at GpT, leading to a somewhat compressed and extended backbone conformation. This improves the base overlap at TpG and reduces the stacking at GpT (23). Model building studies have also suggested that the G·TA triplet causes a displacement in the position of the third strand deoxyribose group (36).

T·CG triplet

The stable interaction of thymine with a CG base pair in a parallel triplex was first proposed by Yoon *et al.* (37) who used electrophoretic mobility shift analysis to distinguish between triplex and duplex DNA forms. Since this triplet contains a single hydrogen bond between O2 of thymine and the exocyclic N4 of cytosine it is weaker than the canonical T·AT triplet. In most reports, T·CG appears to be weaker than G·TA (25,35). This hydrogen bonding pattern can also be adopted by a third strand cytosine, generating a C·CG triplet. Indeed, several reports have also suggested the C·CG triplet as a candidate for recognition of CG (38,39). The preference for a third strand pyrimidine for recognition of CG can also be attributed to its lower steric hinderance, since a purine in this position would clash with the exocyclic amino group of cytosine (20). The T·CG triplet can be formed in both parallel (40) and antiparallel (41,42) structures. As with G·TA, up to three consecutive T·CG triplets can be accommodated within a DNA triplex if the interaction is stabilised by a triplex-binding ligand (22). The third strand T can be replaced with U, generating a stable U·CG triplet. It is interesting to note that an antiparallel U·CG base triplet has been proposed to occur in the catalytic domain of group I introns (43).

Parallel T·CG triplets. NMR studies with a 7mer parallel intramolecular triplex show that the T·CG triplet is stabilised by a single hydrogen bond between the O2 of thymine on the third strand to the free C4-amino proton on the duplex cytosine. Structurally, the T·CG triplet is similar to G·TA; the pairing alignments of T·C and G·T are identical in the two triplets. The base twist either side of the thymine residue of T·CG shows similar degrees of over- and underwinding as with G·TA. These perturbations may extend to the bases on either side of the third strand thymine. Surprisingly, small structural changes are evident in the purine duplex strand of a T·CG-containing triplex, although their basis is unclear. The sugar conformation also adopts a C3′-*endo* pucker to minimise backbone distortions (40).

Antiparallel T·CG triplets. Due to the rotatable nature of thymine, the T·CG interaction can also be formed within a purine motif antiparallel triplex (16,41,42,44). These studies generally incorporate a T·CG triplet at the centre of a G·GC- and A·AT- or T·AT-containing triplex. Various base analogues have been tested to determine the important functional groups in this interaction (45). In particular studies with oligonucleotides containing pyridin-2-one and pyridin-4-one suggest that the O^4 of thymine plays an important role in this interaction. One study with 5-halogenated-dU analogues showed that 5-fluoro-2′-deoxyuridine has a slightly higher affinity for CG inversions than T and was much better than the 5-iodo or 5-bromo derivatives (45). In contrast, a 19mer oligonucleotide directed against a target site in the hamster adenine phosphoribosyltransferase gene showed no difference in stability between 5-fluorouracil, thymine or imidazole for recognition of a CG inversion (46), though in this case it should be noted that the unmodified oligonucleotide bound with very high (sub-nanomolar) affinity. NMR studies have shown that the thymine carbonyl group used to bind the CG base pair is transposed from O2 in the parallel motif to the O4 group in antiparallel complexes (41,42). Such an observation is in keeping with a different ribose–phosphate arrangement adopted by the third strand in R·RY complexes. Despite their opposite strand orientations, the parallel and antiparallel forms of the T·CG triplet are nearly superimposable (41), underlining their structural

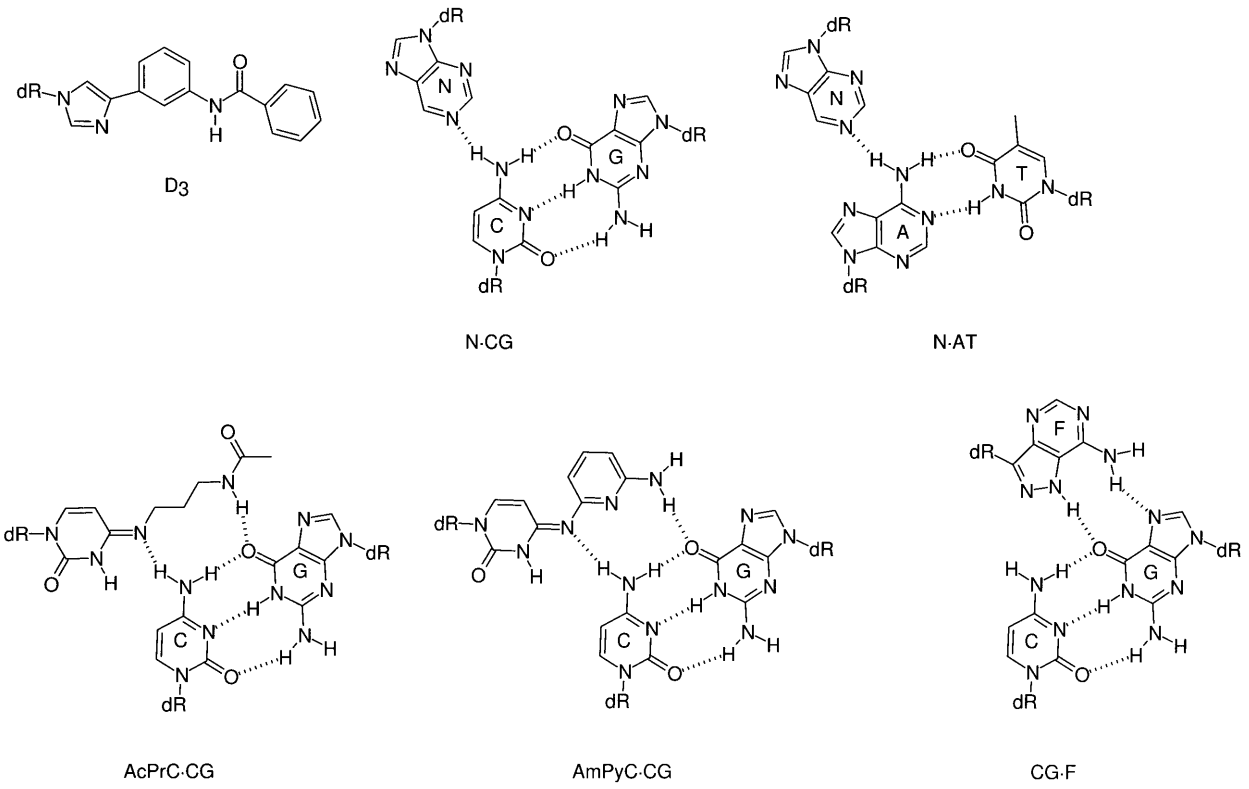

Figure 3. Structures of the synthetic base analogue 1-(2-deoxy-β-D-ribofuranosyl)-4-(3-benzamidophenyl)imidazole (D_3), deoxynebularine bound to CG (N·CG) and AT (N·AT), N^4-(3-acetamidopropyl)cytosine bound to CG (AcPrC·CG), N^4-(6-amino-2-pyridinyl)C bound to CG (AmPyC·CG) and deoxyformycin bound to CG (F·CG).

similarities. Both possess *anti* glycosidic bonds, C3′-*endo* sugar pucker and a slight 5′ tilt by the thymine residue to maintain the π-stacking interaction along the helix long axis. Molecular dynamics simulations suggest that the antiparallel T·CG triplet can be further stabilised by a water-mediated hydrogen bond between the thymine N3 hydrogen and the guanine carbonyl group at C6 (44).

Abasic sites

An alternative strategy for triplex formation at pyrimidine-containing sites is to avoid the offending pyrimidine residue by skipping a base in the third strand or opposing it with a non-selective residue. This has been attempted using oligonucleotides containing abasic sites such as 1,2-dideoxy-D-ribose (46,48). However, this residue produces triplexes with low stability, especially when multiple linkers are used in each oligonucleotide. This low stability is thought to be due to the loss of base stacking interactions in the third strand. The least destabilizing abasic linker described to date is propanediol, which has been shown to read over CG pairs in the Ha-*ras* promoter and prevent access of the Sp1 protein (49,50). It seems that triplex-forming oligonucleotides containing abasic sites are unlikely to be generally useful since not only are they less stable, due to the loss of essential stacking interactions, but there is a loss of stringency at this site.

Novel base analogues for pyrimidine recognition

Stable recognition of pyrimidine inversions by base analogues presents two major problems. First, pyrimidine bases only present one hydrogen bonding site within the major groove, in contrast to purines which for selective recognition can involve two hydrogen bonds. As a result triplets for recognition of TA and CG are generally much weaker than T·AT, C^+·GC, A·AT or G·GC. Secondly, any compound used to recognise TA must avoid steric clash with the 5-methyl group of thymine. Nonetheless, a few synthetic nucleoside analogues have been prepared which selectively interact with pyrimidines in a homopurine triplex site, though there has been greater success in recognising CG than TA. Figures 3–5 summarise the chemical structures of some compounds which have been designed to bind at TA or CG inversion sites.

One of the first examples of a pyrimidine-selective third strand base used a functionalised benzimidazole nucleus (Fig. 3, top left), 1-(2-deoxy-β-D-ribofuranosyl)-4-(3-benzamidophenyl)imidazole (D_3) (51). This base generated stable triplets at both TA and CG Watson–Crick base pairs, with much lower stability when placed opposite GC and AT; the rank order of triplet stability was D_3·TA = D_3·CG > D_3·AT > D_3·GC. Although this synthetic base analogue was originally thought to discriminate between YR and RY pairs by shape-selective recognition of YR, NMR studies have subsequently shown that it acts by intercalating adjacent to TA or CG base pairs at the duplex YpR step, instead of directly interacting within the major groove (52,53).

The CG base pair has been recognised within an antiparallel triplex using 2′-deoxynebularine (N) (54; Fig. 3). An oligonucleotide containing two N residues formed a specific complex on plasmid DNA at a 15 bp site containing two CG base pairs within an oligopurine tract. Within the sequence context studied the N·CG triplet stabilised the triple helix by 1 kcal mol^{-1} compared with an

A·CG mismatch. However, this base analogue also forms a stable triplet with adenine and the rank order of stability is N·CG = N·AT >> N·GC = N·TA. Figure 3 (top centre and right) shows possible hydrogen bonding interactions for the N·CG and N·TA triplets.

Arguably the best base analogues described to date for stable recognition of CG within a parallel triplex are based on N^4-substituted cytosine derivatives (55–58). The first of these to be described was N^4-(3-acetamidopropyl)cytosine (55; Fig. 3, bottom left). The T_m of complexes containing X·CG triplets was greatest for X = N^4-(3-acetamidopropyl)C, although N^4-(3-amino-propyl)C formed a significant though less stable complex. In contrast, N^4-(3-carboxypropyl)C and N^4-(3-butyl)C did not form stable triplets at CG. These results, together with model building studies, suggest that the 3-amino hydrogen of N^4-(3-aminopropyl)C or the amide hydrogen of N^4-(3-acetamidopropyl)C form hydrogen bonds to the O6 carbonyl of guanine in the CG base pair. No stable triplexes were observed on placing N^4-(3-acetamidopropyl)C opposite TA or AT, though a triplet of lower stability was observed at GC. The observation that this base, possessing a flexible side chain, could support triplex formation, opened the possibility that further analogues with more rigid side chains might form more stable triplets. A further analogue with such a rigid side chain is N^4-(6-amino-2-pyridinyl)C (Fig. 3, bottom centre; 56,57). Molecular modelling studies with this base analogue suggested that the pyridine ring is capable of spanning the major groove, placing the 6-amino group close to the O6 or N7 of G. This group is essential for complex formation since (2-pyridinyl)C does not form a stable triplet at CG. Formation of the imino tautomer may also generate a second hydrogen bond with the N^4-amino group of the CG base pair. It is also likely that, as well as providing a rigid platform for positioning the hydrogen bonding groups, the side groups participate in stacking interactions with the neighbouring base pairs. It should also be noted that this analogue appears to form two distinct triplex structures, one of which may involve intercalation of the base analogue in a similar manner to that observed with D_3 (57). This analogue also forms a stable triplet with AT base pairs, which is more stable than the canonical T·AT triplet. This interaction may involve an additional hydrogen bond to N7 of A so that the third strand base is attached by three hydrogen bonds.

Simple azole-2′-deoxyribonucleosides have been proposed as non-specific base analogues for binding to pyrimidine interruptions within antiparallel triplexes (58,59). The analogues were designed in anticipation that the azole ring would be small enough to avoid steric clash with the inverted base pair, while retaining stacking interactions within the third strand. A family of azoles, including pyrazole, imidazole, 1,2,4-triazole and 1,2,3,4-tetrazole, were tested against sites containing three pyrimidine inversions and were shown to bind at least 100- to 1000-fold more tightly than oligonucleotides containing natural bases. These analogues also bound to GC pairs, though with lower affinity than guanine, but showed no interaction with AT pairs. Oligonucleotides containing imidazole and tetrazole groups bound to both TA and CG. The triazole analogue showed weaker binding to CG, while pyrazole showed no interaction with CG, binding only to TA and GC (58). Other groups have also used azole-containing oligonucleotides for both parallel and antiparallel triplex formation at mixed sequence target sites, though their binding is not strong (60).

An alternative strategy for recognising CG inversions is to design base analogues which reach across the major groove, forming hydrogen bonds with the opposing guanine. This is not possible with natural bases because of the large (3–5 Å) distortion required to achieve hydrogen bonding on the other side of the major groove. In this regard 2′-deoxyformycin A (Fig. 3, bottom right) has been shown to stabilise antiparallel triplexes at CG inversions by 10-fold, compared with oligonucleotides containing natural bases (61). In this analogue the position of the hydrogen bond donors is altered relative to those in guanine so that it is able to form two hydrogen bonds with guanine on the opposing DNA strand. Doronina and Behr (62) have suggested that the α- and β-anomers of 4-guanidinocytidine might be capable of binding CG and GC pairs, respectively (Fig. 4, top). In these untested analogues it is proposed that the phosphodiester backbone might be positioned in the centre of the major groove, in contrast to most triplexes in which it is closer to the purine strand. If the third strand is equidistant from the two duplex strands then guanines on opposing strands could be recognised by different anomers.

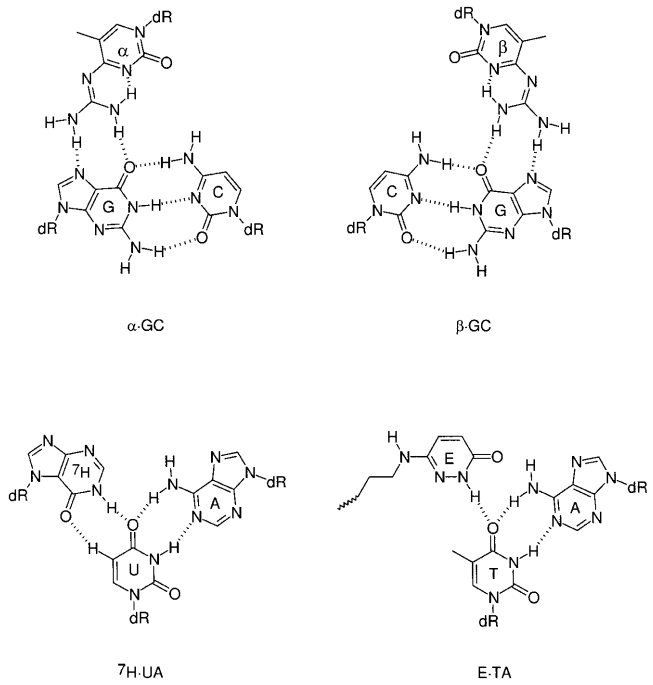

Figure 4. α- and β-anomers of 4-guanidinocytidine bound to GC and CG base pairs. In this analogue the phosphodiester backbone is positioned in the centre of the DNA major groove; the base analogue can therefore interact with guanine residues on either side of the major groove depending on its anomeric configuration. Hypoxanthine nucleoside analogue ^{7}H bound to a UA base pair (^{7}H·UA). Note that this involves an unconventional C-H…O hydrogen bond. Proposed structure of the 3-oxo-2,3,-dihydropyridazine·TA PNA triplet E·TA.

A further insight into the possibility of pyrimidine recognition is seen in studies with a hypoxanthine nucleoside analogue ^{7}H. Although this base is only capable of forming one conventional hydrogen bond, it forms a stable triplet at GC base pairs (63,64). This was empirically attributed to the formation of two C-H…O hydrogen bonds flanking the conventional hydrogen bond. A similar interaction was subsequently shown with uridine, forming a ^{7}H·UA triplet, supposedly incorporating a C-H…O bond with the C5 hydrogen of uridine (64; Fig. 4, bottom left). In support of this proposal no triplex was formed at a TA base pair, since the 5-methyl group of thymine obstructs the recognition process, preventing the formation of this unusual hydrogen bond. A

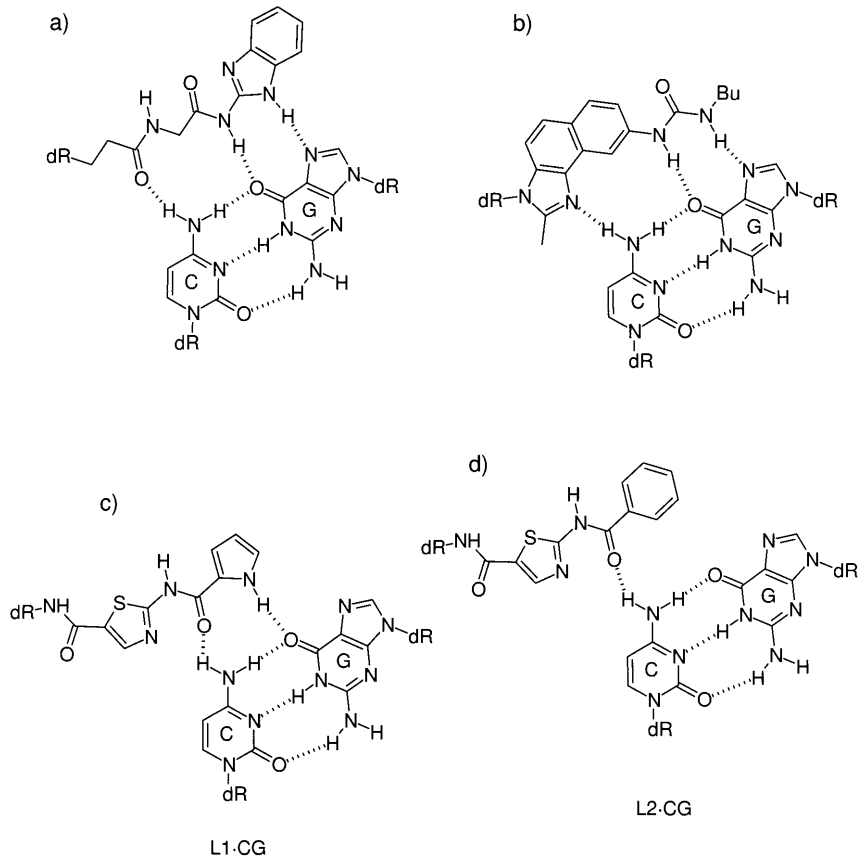

Figure 5. Recognition of CG by benzimidazole-glycyl (**a**), 2-methyl-8-(*N′-n*-butylureido)naphth[1,2-d]imidazole) (**b**), L1 (**c**) and L2 (**d**).

similar triplet has also been proposed for T·UA, although again this is not possible with thymine in the duplex strand. Although the ^{7}H base recognises both GC and UA pairs its use in DNA triplexes is limited as uracil is not a DNA base. It is therefore not clear how the unconventional C-H…O could be incorporated into the design of bases for specific recognition of pyrimidines.

From molecular modelling studies Mohan *et al.* (65) have suggested that the xanthine nucleus might provide a useful scaffold for designing triplex-forming bases capable of interacting with pyrimidine inversions. Similarly, other unnatural nucleosides have been proposed for recognition of TA (66,67) and CG base pairs (68) on the basis of molecular modelling studies with both A- and B-type triplex configurations. However, as yet none of these sugestions have been confirmed by experimental studies.

One of the few examples of synthetic base analogues which have been successfully employed for recognizing a TA inversion is 3-oxo-2,3,dihydropyridazine (E) (Fig. 4, bottom right; 69). This analogue, which was attached to the Hoogsteen strand of a bis-PNA, was designed with a longer linker so as to avoid steric clash with the 5-methyl group of thymine and a hydrogen bond donor positioned to bind the 4-oxo group of thymine. Although the E·TA triplet is less stable than either C^{+}·GC or T·AT it produces complexes with higher T_m values than positioning a G opposite the TA pair.

Base pair recognition

All the natural bases and most of the synthetic base analogues used in triplex formation function by making specific contacts with one base (usually purine) of each base pair. As a result recognition of pyrimidines is necessarily weaker and less stringent since there are fewer available hydrogen bonding positions than on the purine bases. An alternative strategy, which has so far received less attention, is to design base analogues which form hydrogen bond contacts with substituents on both bases of the Watson–Crick base pair. One example of this strategy is the N^4-substituted cytosine analogues, described above (55–57) for recognition of CG. Further examples for simultaneous recognition of both bases in CG pairs are benzimidazole-glycyl (70; Fig. 5a) and 2-methyl-8-(*N′-n*-butylureido)naphth[1,2-*d*]-imidazole) (Fig. 5b; 71), although to date successful interaction has only been demonstrated in chloroform. Lehmann *et al.* (72) have also designed novel base analogues L1 and L2 for interaction with CG pairs. These analogues should form two and one hydrogen bonds, respectively, with CG pairs (Fig. 5c and d). Although oligonucleotides containing these analogues bound across CG inversions, L1 and L2 bound with similar affinities, suggesting that the predicted hydrogen bonding patterns were not the major factor responsible for the interaction. In addition they

bound with similar affinities to sites containing TA inversions, displaying a similar pattern of selectivity to D_3. It therefore seems likely that these analogues act by intercalation at CG and TA steps.

Most of the modified heterocycles described above are poorly selective for TA or CG base pairs and bind with low affinity or stability when incorporated into oligonucleotides. Other attempts at designing analogues for recognition of pyrimidine inversions have also been unsuccessful. Indeed, a salutary lesson in the design of artificial bases is provided by Guzzo-Pernell *et al.* (73), who designed a novel series of modified cytosine analogues for TA and CG recognition, which were subsequently found not to bind any of the four base pairs when incorporated into triplex-forming oligonucleotides. The design of novel base analogues for specific recognition of CG and TA still lacks a definitive solution.

Stabilisation of mismatches

One method which has been successfully employed for stabilising triplexes is to develop ligands which bind selectively to triplex and not to duplex DNA. Several such ligands have been described including BePI (74,75), naphthylquinoline derivatives (76,77), coralyne (78) and disubstituted amidoanthraquinones (79). As well as stabilising triplexes formed at oligopurine tracts these ligands can promote the formation of weaker triplexes, such as those containing base mismatches. In general these ligands do not affect the stringency of triplex formation and the relative binding strengths of different oligonucleotide substitutions are not affected (77). However, by increasing the strength of binding, by up to 1000-fold, complexes can be formed at sequences for which there are no clear rules. The naphthylquinoline derivatives have been shown to promote the formation of triplexes at sites containing up to three consecutive base pair inversions using T·CG and G·TA triplets (22). Studies with BePI have shown that the least destabilising triplets are the same in both the presence and absence of the ligand and that the third strand base is less important in the presence of the ligand (80). However, addition of the ligand does provide some discrimination between different inverted base pairs. In particular, 3-nitropyrrole discriminates CG from GC, TA and AT pairs in the presence, but not the absence of BePI (80).

A different approach to targeting sites containing pyrimidines has used an internal acridine group adjacent to the base facing the inverted purine·pyrimidine base pair (81,82). In this way, the loss of triplex stability at the inversion is partly overcome by the additional binding free energy of the intercalator, in a similar fashion to that achieved by attaching an intercalator to the 5′-end of the third strand. In the absence of a base inversion, inclusion of an internal acridine has little or no effect on triplex stability, possibly because it is a duplex- rather than a triplex-specific intercalator. For targets containing CG or TA inversions the acridine moiety increases the stability of triplexes with either natural or synthetic bases opposing the pyrimidine base. Recognition of the TA base pair is strongest using either acridine or propanediol with an acridine on its 3′-side. Recognition of CG is greatest with either cytosine with acridine on its 3′-side or guanine with acridine on its 5′-side (81).

Alternate strand recognition

A simple strategy for extending recognition beyond simple homopurine tracts, without the requirement for separate recognition

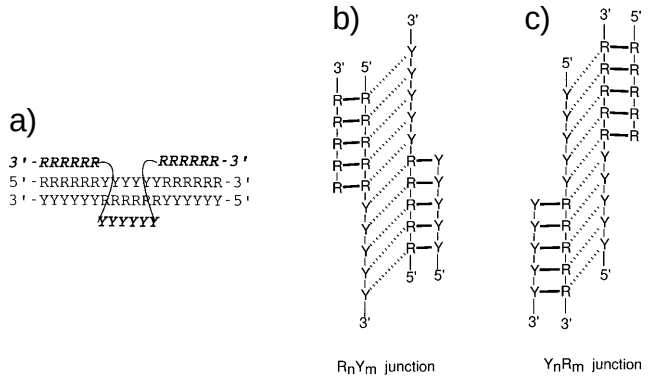

Figure 6. (**a**) Schematic representation of an alternate strand DNA triplex. The duplex target $R_6Y_6R_6.Y_6R_6Y_6$ is bound by the third strand $R_6Y_6R_6$ forming two blocks of antiparallel R·RY triplets flanking a block of parallel Y·RY triplets. The third strand is shown in bold italic. (**b**) Alternate strand recognition across an R_nY_m junction. (**c**) Alternate strand recognition across a Y_nR_n junction. In these representations the helix is viewed along the DNA major groove, with the base pairs running at an angle. Base pairs are joined by a dotted line, while interaction with the third strand is indicated by the solid lines. It can be seen that for recognition of R_nY_m the two blocks of triplets overlap at the RY junction, while for recognition of Y_nR_m there is a gap between the two triplex blocks.

of pyrimidines, involves binding of blocks of purines on opposite DNA strands (83). Recognition is achieved by a single third strand, portions of which are targeted against single oligopurine tracts. By linking these portions together there is a cooperative increase in binding strength and an increase in specificity. If only one triplex motif is employed (parallel or antiparallel) then joining the third strands of adjacent triplexes requires either a 3′–3′ or 5′–5′ linkage, changing third strand orientation at the junction. Linkers that have been used include 1,3 propanediol (84), 1,2 dideoxyribose (83) and a xylose moiety (85). Base-to-base linkage has also been reported (86,87). A modification of this strategy, which avoids the need for unusual backbone linkers, targets adjacent oligopurine blocks on alternate DNA strands by combining both parallel and antiparallel triplexes in one structure (Fig. 6a; 88). The third strand consists of both CT blocks (forming parallel structures) and GT or GA blocks (forming antiparallel triplexes). Since the two motifs have opposite orientations the polarity of the third strand is maintained as the oligonucleotide switches from one strand to the other. In these mixed motif structures, crossing of the major groove by the third strand is not a simple process. Because the phosphodiester backbone is tilted relative to the helix axis, recognition across an R_nY_m junction results in an overlap of two bases in the third strand, whereas for recognition of Y_nR_m an additional linker is required to bridge the gap (Fig. 6b and c; 88). As a result alternate strand recognition at R_nY_m junctions is generally easier than at Y_nR_m (88–91). This approach has been used to target the conserved long terminal repeat sequences of HIV DNA (92), for which a single 11 base oligonucleotide was able to bind across the strands forming a stable alternate strand complex.

Another strategy which extends the range of sequences which can be targeted by triplex formation uses a flexible linker to join two oligonucleotides which form complexes at separate sites (93). Two 12 base oligopurine duplex tracts, separated by one

helical turn, were simultaneously bound by hybrid oligonucleotides containing a linker of 20–25 rotatable bonds.

CONCLUSIONS

Although there has been significant progress towards resolving many of the practical issues of triplex formation, a general solution to the problem of mixed sequence recognition is still elusive. G·TA and T·CG are the most stable triplets using natural bases at pyrimidine inversions and these may have some limited uses, which may be improved by increasing the strength of neighbouring base triplets. There is, however, no method for recognising TA within an antiparallel triplex. Since pyrimidine bases offer only one potential hydrogen bonding site within the DNA major groove, there is limited scope for designing specific base analogues directed only at pyrimidine bases. The 5-methyl group of thymine presents a further obstacle for recognition of TA inversions. Attempts at recognising substituents on both bases of the Watson–Crick pair have also been disappointing and often involve intercalation of the proposed analogues at the YpR step, rather than base-specific recognition. It is also clear that stable triplex formation will require optimal stacking of the third strand bases. In this regard it should be noted that although T·AT and C^+·GC are isohelical within a parallel triplex (94), the position of the phosphodiester backbone varies for other base triplets, leading to backbone distortions at each triplet step. Attempts to target the purine base on the opposing strand may fail as a result of poor third strand stacking as the novel base reaches across the major groove. Successful recognition of mixed sequence DNA by triplex formation therefore remains an elusive goal, which requires the design, synthesis and evaluation of more novel base derivatives.

ACKNOWLEDGEMENT

Work in the author's laboratory is funded by the Cancer Research Campaign.

REFERENCES

1 Felsenfeld,G. and Rich,A. (1957) *Biochim. Biophys. Acta*, **26**, 457–468.
2 Felsenfeld,G., Davis,D.R. and Rich,A. (1957) *J. Am. Chem. Soc*., **79**, 2023–2024.
3 Le Doan,T., Perrouault,L., Praseuth,D., Habhoub,N., Decout,J.L., Thuong,N.T., Lhomme,J. and Hélène,C. (1987) *Nucleic Acids Res*., **15**, 7749–7761.
4 Moser,H.E. and Dervan,P.B. (1987) *Science*, **238**, 645–650.
5 Hélène,C. (1991) *Anticancer Drug Des*., **6**, 569–584.
6 Soyfer,V.N. and Potaman,V.N. (1996) *Triple Helical Nucleic Acids*. Springer-Verlag, New York, Berlin, Heidelberg.
7 Vasquez,K.M. and Wilson,J.H. (1998) *Trends Biochem. Sci*., **23**, 4–9.
8 Chan,P.P and Glazer,P.M. (1997) *J. Mol. Med*., **75**, 267–282.
9 Neidle,S (1997) *Anticancer Drug Des*., **12**, 433–442.
10 Thuong,N.T. and Hélène,C. (1993) *Angew. Chem*., **32**, 666–690.
11 Malkov,V.A., Voloshin,O.N., Soyfer,V.N. and Frank-Kamenetskii,M.D. (1993) *Nucleic Acids Res*., **21**, 585–591.
12 Potaman,V.N. and Soyfer,V.N. (1994) *J. Biomol. Struct. Dyn*., **11**, 1035–1040.
13 Griffin,L.C. and Dervan,P.B. (1989) *Science*, **245**, 967–971.
14 Chandler,S.P. and Fox,K.R. (1993) *FEBS Lett*., **332**, 189–192.
15 Fossella,A., Kim,Y.J., Shih,H., Richards,E.G. and Fresco,J.R. (1993) *Nucleic Acids Res*., **21**, 4511–4515.
16 Chandler,S.P. and Fox,K.R. (1996) *Biochemistry*, **35**, 15038–15048.
17 Greenberg,W.A. and Dervan,P.B. (1995) *J. Am. Chem. Soc*., **117**, 5016–5022.
18 Sun,J.-S., Mergny,J.-L., Lavery,R., Montenay-Garestier,T. and Hélène,C. (1991) *J. Biomol. Struct. Dyn*., **9**, 411–424.
19 Kiessling,L.L., Griffin,L.C. and Dervan,P.B. (1992) *Biochemistry*, **31**, 2829–2834.
20 Mergny,J.L., Sun,J.S., Rougée,M., Montenay-Garestier,T., Barcelo,F., Chomilier,J. and Hélène,C. (1991) *Biochemistry*, **30**, 9791–9798.
21 Chen,A.-J. and Van Dyke,M.W. (1994) *Nucleic Acids Res*., **22**, 4742–4747.
22 Gowers,D.M. and Fox,K.R. (1997) *Nucleic Acids Res*., **25**, 3787–3794.
23 Radhakrishnan,I. and Patel,D.J. (1994) *Structure*, **2**, 17–32.
24 Radhakrishnan,I., Gao,X., De los Santos,C., Live,D. and Patel,D.J. (1991) *Biochemistry*, **30**, 9022–9030.
25 Miller,P.S. and Cushman,C.D. (1993) *Biochemistry*, **32**, 2999–3004.
26 Chandler,S.P. and Fox,K.R. (1995) *FEBS Lett*., **360**, 21–25.
27 Gowers,D.M. and Fox,K.R. (1998) *Nucleic Acids Res*., **26**, 3626–3633.
28 Volker,J. and Klump,H.H. (1994) *Biochemistry*, **33**, 13502–13508.
29 Keppler,M.D. and Fox,K.R. (1997) *Nucleic Acids Res*., **25**, 4464–4469.
30 Asensio,J.L., Lane,A.N., Dhesi,J., Bergqvist,S. and Brown,T. (1998) *J. Mol. Biol*., **275**, 811–822.
31 Radhakrishnan,I., Patel,D.J. and Gao,X. (1992) *Biochemistry*, **31**, 2514–2523.
32 Radhakrishnan,I., Patel,D.J., Veal,J.M. and Gao,X. (1992) *J. Am. Chem. Soc*., **114**, 6913–6915.
33 Wang,G., Malek,S. and Feigon,J. (1992) *Biochemistry*, **31**, 4838–4846.
34 Radhakrishnan,I. and Patel,D.J. (1994) *Biochemistry*, **33**, 11405–11416.
35 Radhakrishnan,I., Patel,D.J., Priestley,E.S., Nash,H.M. and Dervan,P.B. (1993) *Biochemistry*, **32**, 11228–11234.
36 Best,G.C. and Dervan,P.B. (1995) *J. Am. Chem. Soc*., **117**, 1187–1193.
37 Yoon,K., Hobbs,C.A., Koch,J., Sardaro,M., Kutny,R. and Weiss,A.L. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 3840–3844.
38 Mergny,J.L., Collier,D., Rougée,M., Montenay-Garestier,T. and Hélène,C. (1991) *Nucleic Acids Res*., **19**, 1521–1526.
39 Belotserkovskii,B.P., Veselkov,A.G., Filippov,S.A., Dobrynin,V.N., Mirkin,S.M. and Frank-Kamenetskii,M.D. (1990) *Nucleic Acids Res*., **18**, 6621–6624.
40 Radhakrishnan,I. and Patel,D.J. (1994) *J. Mol. Biol*., **241**, 600–619.
41 Ji,J., Hogan,M.E. and Gao,X.L. (1996) *Structure*, **4**, 425–435.
42 Dittrich,K., Gu,J., Tinder,R., Hogan,M. and Gao,X. (1994) *Biochemistry*, **33**, 4111–4120.
43 Michel,F., Ellington,A.D., Couture,S. and Szostak,J.W. (1990) *Nature*, **347**, 578–580.
44 Weerasinghe,S., Smith,P. and Pettitt,B.M. (1995) *Biochemistry*, **34**, 16269–16278.
45 Durland,R.H., Rao,T.S., Revankar,G.R., Tinsley,J.H., Myrick,M.A., Seth,D.M., Rayford,J., Singh,P. and Jayaraman,K. (1994) *Nucleic Acids Res*., **22**, 3233–3240.
46 Vasquez,K.M., Wensel,T.G., Hogan,M.E. and Wilson,J.H. (1995) *Biochemistry*, **34**, 7243–7251.
47 Horne,D.A. and Dervan,P.B. (1991) *Nucleic Acids Res*., **19**, 4963–4965.
48 Kandimalla,E.R., Manning,A.N., Venkataraman,G., Sasisekharan,V. and Agrawal,S. (1995) *Nucleic Acids Res*., **23**, 4510–4517.
49 Mayfield,C., Ebbinghaus,S., Gee,J., Jones,D., Rodu,B., Squibb,M. and Miller,D. (1994) *J. Biol. Chem*., **269**, 18232–18238.
50 Mayfield,C. and Miller,D. (1994) *Nucleic Acids Res*., **22**, 1909–1916.
51 Griffin,L.C., Kiessling,L.L, Beal,P.A., Gillespie,P. and Dervan,P.B. (1992) *J. Am. Chem. Soc*., **114**, 7976–7982.
52 Koshlap,K.M., Gillespie,P., Dervan,P.B. and Feigon,J. (1993) *J. Am. Chem. Soc*., **115**, 7908–7909.
53 Wang,E., Koshlap,K.M., Gillespie,P., Dervan,P.B and Feigon,J. (1996) *J. Mol. Biol*., **257**, 1052–1069.
54 Stilz,H.U. and Dervan,P.B. (1993) *Biochemistry*, **21**, 2177–2185.
55 Huang,C.-Y., Cushman,C.D. and Miller,P.S. (1993) *J. Org. Chem*., **58**, 5048–5049.
56 Huang,C.-Y. and Miller,P.S. (1993) *J. Am. Chem. Soc*., **15**, 10456–10457.
57 Huang,C.-Y., Bi,G. and Miller,P.S. (1996) *Nucleic Acids Res*., **24**, 2606–2613.
58 Durland,R.H., Rao,T.S., Bodeepudi,V., Seth,D.M., Jayaraman,K. and Revankar,G.R. (1995) *Nucleic Acids Res*., **23**, 647–653.
59 Jayaraman,K., Durland,R.H., Rao,T.S., Revankar,G.R., Bodepudi,V., Chaudhary,N. and Guy-Caffey,J. (1995) *Nucleosides Nucleotides*, **14**, 951–955.
60 Gee,J.E., Revankar,K., Rao,T.S. and Hogan,M.E. (1995) *Biochemistry*, **34**, 2042–2048.
61 Rao,T.S., Hogan,M.E. and Revankar,G.R. (1994) *Nucleosides Nucleotides*, **13**, 95–107.
62 Doronina,S.O. and Behr,J.-P. (1997) *Tetrahedron Lett*., **39**, 547–550.
63 Marfut,J., Parel,S.P. and Leumann,C. (1997) *Nucleic Acids Res*., **25**, 1875–1882.
64 Marfut,J. and Leumann,C. (1998) *Angew. Chem*., **37**, 175–177.
65 Mohan,V., Cheng,Y.K., Marlow,G.E. and Pettitt,B.M. (1993) *Biopolymers*, **33**, 1317–1325.
66 Rothman,J.H. and Richards,W.G. (1996) *Mol. Simulat*., **18**, 13–42.

67 Rothman,J.H. and Richards,W.G. (1996) *Biopolymers*, **39**, 795–812.
68 Rothman,J.H. and Richards,W.G. (1996) *J. Mol. Model.*, **2**, 456–466.
69 Eldrup,A.B., Dahl,O. and Nielsen,P.E. (1997) *J. Am. Chem. Soc.*, **119**, 1116–1117.
70 Sasaki,S., Nakashima,S., Nagatsugi,F., Tanaka,Y., Hisatome,M and Maeda,M. (1995) *Tetrahedron Lett.*, **52**, 9521–9524.
71 Zimmerman,S.C. and Schmitt,P. (1995) *J. Am. Chem. Soc.*, **117**, 10769–10770.
72 Lehmann,T.E., Greenberg,W.A., Liberles,D.A., Wada,C.K. and Dervan,P.B. (1997) *Helv. Chim. Acta*, **80**, 2002–2022.
73 Guzzo-Pernell,N., Tregear,G.W., Haralambidis,J. and Lawlor,J.M. (1998) *Nucleosides Nucleotides*, **17**, 1191–1207.
74 Mergny,J.L., Duval-Valentin,G., Nguyen,C.H., Perrouault,L., Faucon,B., Rougée,M., Montenay-Garestier,T., Nisagni,E. and Hélène,C. (1992) *Science*, **256**, 1681–1684.
75 Escudé,C., Nguyen,C.H., Mergny,J.-L., Sun,J.-S., Bisagni,E., Garestier,T. and Hélène,C. (1995) *J. Am. Chem. Soc.*, **117**, 10212–10219.
76 Wilson,W.D., Tanious,F.A., Mizan,S., Yao,S., Kiselyov,A.S., Zon,G. and Strekowski,L. (1993) *Biochemistry*, **32**, 10614–10621.
77 Chandler,S.P., Strekowski,L., Wilson,W.D. and Fox,K.R. (1995) *Biochemistry*, **34**, 7234–7242.
78 Lee,J.S., Latimer,L.J.P. and Hampel,K.J. (1993) *Biochemistry*, **32**, 5591–5597.
79 Fox,K.R., Polucci,P., Jenkins,T.C. and Neidle,S. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 7887–7891.
80 Kukreti,S., Sun,J.S., Loakes,D., Brown,D.M., Nguyen,C.H., Bisagni,E., Garestier,T. and Hélène,C. (1998) *Nucleic Acids Res.*, **26**, 2179–2183.
81 Zhou,B., Puga,E., Sun,J., Barestier,T. and Hélène,C. (1995) *J. Am. Chem. Soc.*, **117**, 10425–10428.
82 Kukreti,S., Sun,J.S., Garestier,T. and Hélène,C. (1997) *Nucleic Acids Res.*, **25**, 4264–4270.
83 Horne,D.A. and Dervan,P.B. (1990) *J. Am. Chem. Soc.*, **112**, 2435–2437.
84 Ono,A., Chen,C.-N. and Kan,L. (1991) *Biochemistry*, **30**, 9914–9921.
85 Froehler,B.C., Terhorst,T., Shaw,J.-P. and MaCurdy,S.N. (1992) *Biochemistry*, **31**, 1603–1609.
86 Zhou,B.W., Marchand,C., Asseline,U., Thuong,N.T. Sun,J.-S., Garestier,T. and Hélène,C. (1995) *Bioconjugate Chem.*, **6**, 516–523.
87 Asseline,U and Thuong,N.T. (1994) *Tetrahedron Lett.*, **35**, 5221–5224.
88 Beal,P.A. and Dervan,P.B. (1992) *Nucleic Acids Res.*, **20**, 2773–2776.
89 Jayasena,S. and Johnston,B.H. (1992) *Biochemistry*, **31**, 320–327.
90 Jayasena,S.D. and Johnston,B.H. (1992) *Nucleic Acids Res.*, **20**, 5279–5288.
91 Washbrook,E. and Fox,K.R. (1994) *Biochem. J.*, **301**, 569–575.
92 Bouziane,M., Cherny,D.I., Mouscadet,J.F. and Auclair,C. (1996) *J. Biol. Chem.*, **271**, 10359–10364.
93 Kessler,D.J., Pettitt,B.M., Cheng,Y.-K., Smith,S.R., Jayaraman,K., Vu,H.M. and Hogan,M.E. (1993) *Nucleic Acids Res.*, **21**, 4810–4815.
94 Giovannangeli,C., Rougée,M., Garestier,T., Thuong,N.T. and Hélène,C. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8631–8635.