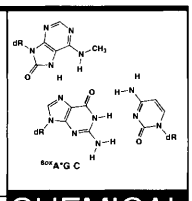
Towards a general triple helix mediated DNA recognition scheme



CHEMICAL SOCIETY REVIEWS

Svetlana O. Doronina and Jean-Paul Behr

Laboratiore de Chimie Génétique associé au CNRS, ULP—Faculté de Pharmacie, BP 24, F-67401 Illkirch Cedex, France

This review describes new perspectives offered by the synthesis of non-natural nucleosides to overcome current limitations and extend the triplex-mediated DNA recognition scheme to any sequence. Alternate strand purine binding, direct pyrimidine recognition, and binding to the whole base-pair are described. The review highlights structural requirements to the design of modified nucleosides, as well as perturbing events such as tautomeric ambiguity and intercalation for the extended heterocyclic bases.

1 Introduction

Nucleic acids triple helices were first described in 1957 by Felsenfeld and Rich,[1] just a few years after Watson and Crick discovered the double-helical nature of DNA. The present interest in triplexes follows the discovery in 1987 by the groups of Hélène and Dervan that short oligonucleotides can bind in a sequence specific manner to a duplex target in DNA under suitable conditions.[2,3] Since binding occurs in the major groove, triplex-forming oligonucleotides (TFOs) have the potential to interfere with regulatory proteins which bind to the same site, hence controlling gene expression. This exciting perspective has stimulated extensive work on triple helical complexes during the past decade (reviewed in refs. 4 and 5).

Much effort has been devoted to increasing the stability of triplexes under physiological conditions, as well as overcoming the major drawback of this approach, *i.e.* the requirement for oligopurine tracts on the target DNA. This limitation arises from the recognition mechanism which involves the purine's two remaining 'Hoogsteen' hydrogen-bonding sites in the major groove of the DNA duplex (Fig. 1); pyrimidines having only one hydrogen-bonding site vacant cannot be efficiently bound. This review will mainly focus on new perspectives offered by the synthesis of non-natural nucleosides to overcome current limitations and extend the triplex-mediated DNA recognition scheme to any sequence.

Other recent advances to stabilizing triple helices in a sequence-independent manner, such as conjugation of polyamines or triplex helix-specific intercalating agents, or the widely explored ribose-phosphate backbone modifications[5] [peptide nucleic acid (PNA), oligonucleotide $N^{3'}$–$P^{5'}$ phosphoramidates and deoxyribonucleic guanidine (DNG)] will not be discussed here.

Essentially two families of DNA triple helices have been characterized that differ in their third-strand sequence composition and relative orientation (Fig. 2). In the pyrimidine-*purine–pyrimidine (Py*Pu–Py) family a homopyrimidine (Py) third strand is bound (*) parallel to the purine strand of target duplex (Pu–Py) in the major groove of DNA through isomorphous T*A–T and G^+*G–C base triplets. Within the T*A–T triplex plane a thymine of the TFO interacts with a Watson–Crick A–T base pair by making two Hoogsteen hydrogen bonds with adenine. The requirement of a protonated cytidine for guanosine recognition results in pH-dependent binding with optimal stability much below physiological pH (5.6–6.0).

In the purine*purine–pyrimidine (Pu*Pu–Py) family, a purine-rich third strand is bound antiparallel to the purine strand of the target duplex. DNA recognition in this motif involves guanosine binding by reverse-Hoogsteen hydrogen bonding to guanosine of the G–C base pair (G*G–C base triplet), and either adenosine or thymidine binding to adenosine of the A–T base pair (A*A–T or T*A–T). Base triplets within this family are not isomorphous, *i.e.* location and orientation of the third-strand deoxyribose (dR) is sequence dependent. This leads to helix

J. P. Behr and S. O. Doronina

Svetlana O. Doronina was born in Tomsk, Russia. After receiving her Master's degree in Chemistry from Novosibirsk State University in 1985, she became a Junior Researcher in the laboratory of Organic Synthesis at the Novosibirsk Institute of Bioorganic Chemistry. In 1992 she came to France and since 1993 she started her PhD studies dealing with the DNA recognition by modified oligonucleotides under supervision of Dr J. P. Behr at the University Louis Pasteur of Strasbourg.

Jean-Paul Behr is Research director at the Centre National de la Recherche Scientifique. He was born in 1947 and graduated from the Chemical Engineering School in Strasbourg. He joined Jean-Marie Lehn in 1969 for PhD work in physical organic chemistry. After a postdoctoral stay in Sheffield to study molecular motion in biological membranes, he rejoined J. M. Lehn to contribute to the development of artificial receptors, catalysts and carriers. In 1989 he moved to the Faculty of Pharmacy to head the Laboratoire de Chimie Génétique. His major interests are now in the development of DNA-binding molecules relevant to biotechnology and gene therapy.

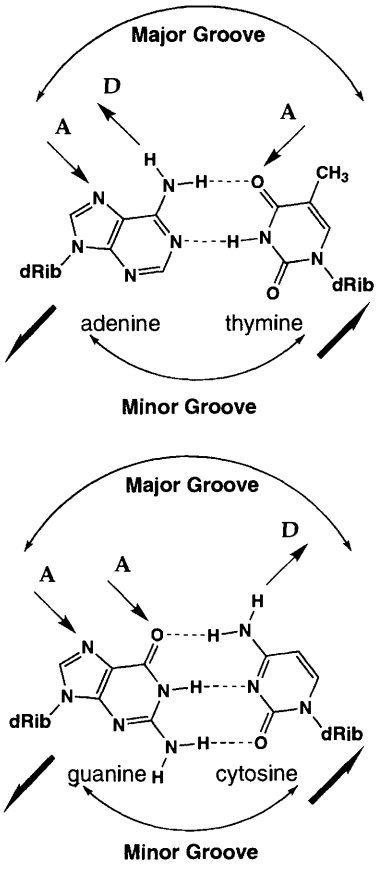


Fig. 1 Watson–Crick hydrogen bond formation (dotted lines) between complementary nucleic bases leaves several H-bonding sites [donor (D) or acceptor (A)] vacant in the major groove of the resulting duplex. Purines (adenine and guanine) have distinct bidentate 'Hoogsteen' sites vacant whereas pyrimidines (cytosine and thymine) have only a single remaining binding site in the major groove.

distortion for mixed sequences especially in the case of G*G–C/T*A–T sequences. These triplexes are not pH-dependent but are destabilized by physiological monovalent cation concentrations.

As shown above, DNA recognition by triple helix formation relies on hydrogen bonding interaction within the base triplets. However, the design of new binding motifs should take into account that third-strand binding depends on many factors besides the stability of isolated planar triplets. Indeed, stacking, van der Waals and dipolar interactions between the neighbouring heterocyclic bases are widely involved in helix stability. Structural isomorphism of base triplets too is an important concern to avoid ribose-phosphate backbone distortion: triplets are isomorphous if their N–dR bonds are superimposable (*e.g.* C^+*G–C and T*A–T in Fig. 2) and we shall see later that non-isomorphism leads to considerable triplex destabilization for random purine sequences. Cationic heterocycles may be favourable in isolated sites due to their strong H-bonding capacity and to attractive electrostatic interactions in a polyanionic context. Yet in contiguous sequences electrostatic repulsion between adjacent charges may result in overall destabilization as found when targeting G-stretches with protonated cytidines. Chimeric nucleosides are therefore designed to be mostly neutral.

Sequence-selective recognition not only means affinity for the target nucleic base, but also discrimination against the three other ones (which is much more stringent). With respect to the latter criterion, tautomerism of heterocyclic bases must seriously be considered. Whereas natural bases are > 99.99% in the Watson–Crick tautomeric form shown in Fig. 1, providing both replication fidelity and some possibility of evolution, most polyaza-heterocycles have several tautomeric structures, which may hydrogen bond to more than just one natural base. Ideally, the development of base-modified nucleoside analogues with unambiguous tautomerism would clarify this problem.

2 Current limitations of homopurine triplexes

2.1 Py*Pu–Py triplexes

Protonation of cytidine is required in order to establish two hydrogen bonds with guanosine (Fig. 2). The pK_a of isolated cytidine is 4.3. However, even though its apparent pK_a increases when incorporated in oligonucleotides due to the polyanionic environment, optimum C^+*G–C triplet stability still requires acidic conditions. This rather limits the potential use of such oligomers *in vivo*, where intracellular pH is highly regulated at *ca.* 7.3. The first (and up to now one of the best) solution to provide Py*Pu–Py triplex stability at physiological pH was replacement of cytidine by 5-methylcytidine[6] (m^5C). This modification enhanced considerably the stability of triple helices at neutral pH and may serve as a reference for G–C base pair recognition for newly designed structures. The influence of cytidine methylation on triplex stability does not appear to result from the 0.2 units pK_a increase, but is mainly of entropic origin.[7] Recently numerous synthetic nucleosides have been developed that display the hydrogen bonding pattern of protonated cytidine, as we shall see now.

2.1.1 Pyrimidine-like analogues of cytidine

Pyrimidine-like pseudoisocytidine[8] (^{iso}C), pyrazine base[9] (pyDDA), 5-methyl-6-oxocytidine[10] ($m^{5ox}C$) lead to B*G–C triplets (B = ^{iso}C, pyDDA, $m^{5ox}C$) isomorphous to the canonical T*A–T triplet [Fig. 3(**A**)]. These heterocycles already have a hydrogen atom at the N^3 position which allows them to bind to guanine in a pH-independent fashion. Triplexes containing these modified bases are more stable at neutral pH than cytidine-containing triplexes but generally do not achieve the stability of 5-methylcytidine-containing triple helices. Thus, triple to double helix transition temperatures (T_m) of triplexes containing $m^{5ox}C$ are still significantly lower at neutral pH than those observed for the corresponding m^5C^+-containing triplexes. This likely reflects the fact that cations are better H-bond donors that show further stabilization on a polyanionic target.

To our knowledge, there are no data about the binding specificity of ^{iso}C and pyDDA. $m^{5ox}C$ has been shown to be very selective, as the only target base pair (besides G–C) that led to detectable triple helix formation was C–G, but with significantly less stability.

Both $m^{5ox}C$ and ^{iso}C can adopt several tautomeric structures. X-ray crystallography as well as duplex-DNA melting temperature studies suggest that the N^3-H tautomer is preferred for $m^{5ox}C$. Indeed $m^{5ox}C$-containing duplexes are less stable than their natural counterparts. This could be due to steric clash between N^3-H of $m^{5ox}C$ and N^1-H of G which prevents Watson–Crick base pairing.

^{iso}C has tautomeric structures in which the hydrogen atom can be bound to either N^1 or N^3 [Fig. 3(**B**)]. ^{iso}C-containing duplexes also show decreased stability, presumably because of steric clash [Fig. 3(**C**)]. These data lead to the supposition that the tautomeric ambiguity of these nucleoside analogues should not hamper their selectivity of binding.

$m^{5ox}C$ and ^{iso}C seem also to be promising candidates for targeting of G–C stretches. As mentioned previously, such sequences are particularly difficult targets in the Py*Pu–Py motif because of ionic repulsion between adjacent protonated residues when C^+ or m^5C^+-containing TFOs are used. Such unfavourable interactions should not be present in adjacent neutral $m^{5ox}C$ or ^{iso}C nucleotides. Indeed, a triple helix containing the $(^{iso}C)_6$ sequence was shown to melt at least 22 °C higher at pH 7.5 than the corresponding $(m^5C^+)_6$ triplex. Contiguous $m^{5ox}C$ did not lead to stable triplexes. Undesirable

steric effects brought about by the O^6 carbonyl atom or poor base stacking could be at the origin of the decreased stability.

To illustrate the impact of non-natural base tautomerism on specific nucleic base recognition the example of the pyrido[2,3-*d*]pyrimidine nucleoside[11] (F) is edifying. This heterocycle can exist in two tautomeric forms F1 and F2 [Fig. 4(**A**)], which can either bind to guanine or adenine in the Watson–Crick sense to form an antiparallel double helix, F1-G (3 H-bonds) being considerably more stable than F2-A (2 H-bonds) [Fig. 4(**B**)]. In Py*Pu–Py triple helixes F1 may bind to G–C or C–G base pairs through a single H-bond, or F2 to A–T through two H-bonds [Fig. 4(**C**)]. Binding experiments have demonstrated that F recognizes the A–T base pair when incorporated into TFOs, with an affinity similar to that of the canonical T*A–T triplet. This example shows that when tautomers are of comparable stabilities, complementary strand binding selects the tautomers giving the most stable among possible structures, *i.e.* F1-G (3 H-bonds) and F2*A–T (2 H-bonds).

2.1.2 Purine-like cytidine analogues

Some purine-like non-natural nucleotides (shown in Fig. 5) have been described for replacement of protonated cytidine: 3-methyl-5-amino-1*H*-pyrazolo[4,3-*d*]pyrimidine-7-one[12] (P1), N^6-methyl-8-oxoadenosine[13] (^{8ox}A) or N^7-guanosine[14] (^{7}G). From affinity cleaving analysis, the stability of base triplets containing N^7-guanosine decreases in the order ^{7}G*G–C $\gg$ ^{7}G*C–G $\gg$ ^{7}G*A–T $\approx$ ^{7}G*T–A with affinities comparable to those of m^{5}C. Nearly the same results were obtained for the pyrazole analogue P1. Its chemical isomer P2 [deoxyribose is linked to pyrazole N^2 for P2 and N^1 and P1, see Fig. 5(**C**)], did not show reasonable affinity to any Watson–Crick base pair; this is probably due to energetically unfavourable distortion of the third-strand backbone added to the fact that the methyl group of P2 disfavours the *anti* conformation.

The known tendency of 8-substituted adenosine analogues to be predominantly in the *syn* conformation led to the use of 8-oxoadenosine as a C^+ substitute [Fig. 5(**B**)]. T_m for a given triplex containing ^{8ox}A*G–C was 22 °C at pH 7.0 and 8.0, whereas the corresponding transition in the control triplex (containing canonical C$^+$*G–C) decreased from 28 °C at pH 7.0 to 17 °C at pH 8.0. ^{8ox}A is also able to form ^{8ox}A*C–G and ^{8ox}A*U–A triads, the latter one being stabilized by a hydrogen bond between the N^6 exocyclic amino group of 8-oxoadenosine and the O^4 of uridine. However, these mismatched triplets are significantly less stable than ^{8ox}A*G–C.

The common drawback of these compounds is the lack of structural isomorphism between B*G–C (B = P1, ^{7}G, ^{8ox}A) and T*A–T triplets, which prevents complex formation for P2. Although NMR analysis[15] does not reveal any major backbone distortion for P1 and ^{7}G, the three orders of magnitude decrease in the affinity of ^{7}G or P1-containing oligonucleotides for targeting an alternated $(GA)_5$ *vs.* a contiguous G_6 site[16]

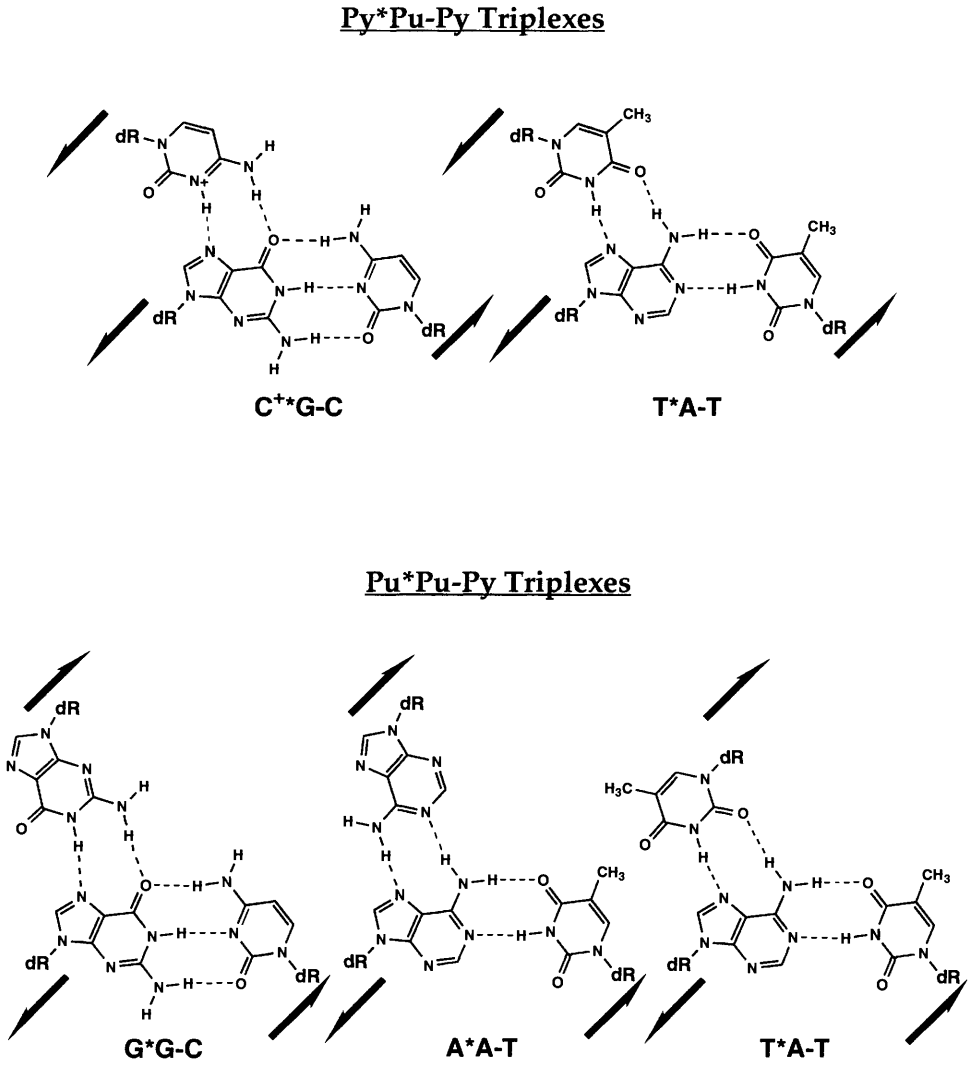


Fig. 2 The two families of homopurine-targeted triple helices. In X*Pu–Py, '–' and '*' represent Watson–Crick and Hoogsteen hydrogen bonds, respectively. Half-arrows indicate the relative deoxyribose-phosphate (dR) backbone orientation.

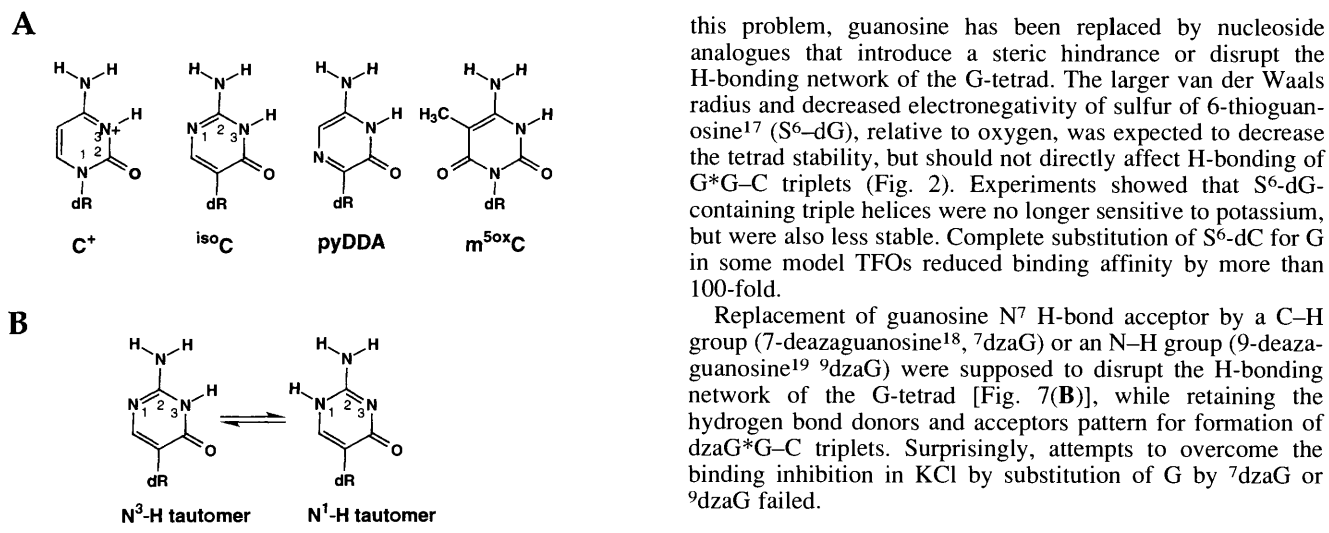


Fig. 3 Pyrimidine-like cytidine mimics. ^{iso}C, pyDDA and m^{5ox}C display the hydrogen bonding pattern of protonated cytidine (**A**). Possible N^3-H and N^1-H tautomeric forms of pseudoisocytidine (^{iso}C) (**B**). N^3-H (a) and N^1-H (b) tautomers of ^{iso}C in Watson–Crick type ^{iso}C–G base pair (**C**).

confirms the energy penalty for nonisomorphism of adjacent triplets.

2.2 Pu*Pu–Py triplex

The obvious advantage of Pu*Pu–Py triplexes is their pH independence: no protonation is required for the G*G–C, A*A–T (or T*A–T) base triplets, which are stabilized by two reverse-Hoogsteen-type hydrogen bonds between bases of the third strand and the purine strand of the Watson–Crick duplex (Fig. 2). T*A–T triplets are generally preferred to A*A–T base triplets because the latter, although well characterized for homopolymers, have stabilities critically dependent on the length of the homoA tract and on the presence of divalent cations. Reverse-Hoogsteen T*A–T triplets are able to bind within a G*G–C triple helical structure, but triplexes are far from isomorphism (Fig. 6); A*A–T and G*G–C triplets are not isomorphous either, but more so than T*A–T/G*G–C. The net result is that the stability of Pu*Pu–Py-type triple helices is usually determined by the content of G-residues.

2.2.1 Inhibition by monovalent cations

The development of G-rich TFOs as potential antigene drugs is hampered by the observation that triple helix formation is inhibited by monovalent cations, and especially by potassium which is the predominant intracellular cation. Inhibition is due to the involvement of the TFO into another process. At physiological K^+ levels, guanosine-rich oligonucleotides self-associate by stacking of K^+/guanine quartets [Fig. 7(**A**)] which are stabilized by a combination of mutual bidentate H-bonding and coordination of the four O^6 atoms of guanosine to K^+ located at the centre of the quadruplex. A number of such structures have been identified, including intra- and intermolecular associations between oligonucleotides. To overcome this problem, guanosine has been replaced by nucleoside analogues that introduce a steric hindrance or disrupt the H-bonding network of the G-tetrad. The larger van der Waals radius and decreased electronegativity of sulfur of 6-thioguanosine[17] (S^6–dG), relative to oxygen, was expected to decrease the tetrad stability, but should not directly affect H-bonding of G*G–C triplets (Fig. 2). Experiments showed that S^6-dG-containing triple helices were no longer sensitive to potassium, but were also less stable. Complete substitution of S^6-dC for G in some model TFOs reduced binding affinity by more than 100-fold.

Replacement of guanosine N^7 H-bond acceptor by a C–H group (7-deazaguanosine[18], ^{7}dzaG) or an N–H group (9-deazaguanosine[19] ^{9}dzaG) were supposed to disrupt the H-bonding network of the G-tetrad [Fig. 7(**B**)], while retaining the hydrogen bond donors and acceptors pattern for formation of dzaG*G–C triplets. Surprisingly, attempts to overcome the binding inhibition in KCl by substitution of G by ^{7}dzaG or ^{9}dzaG failed.

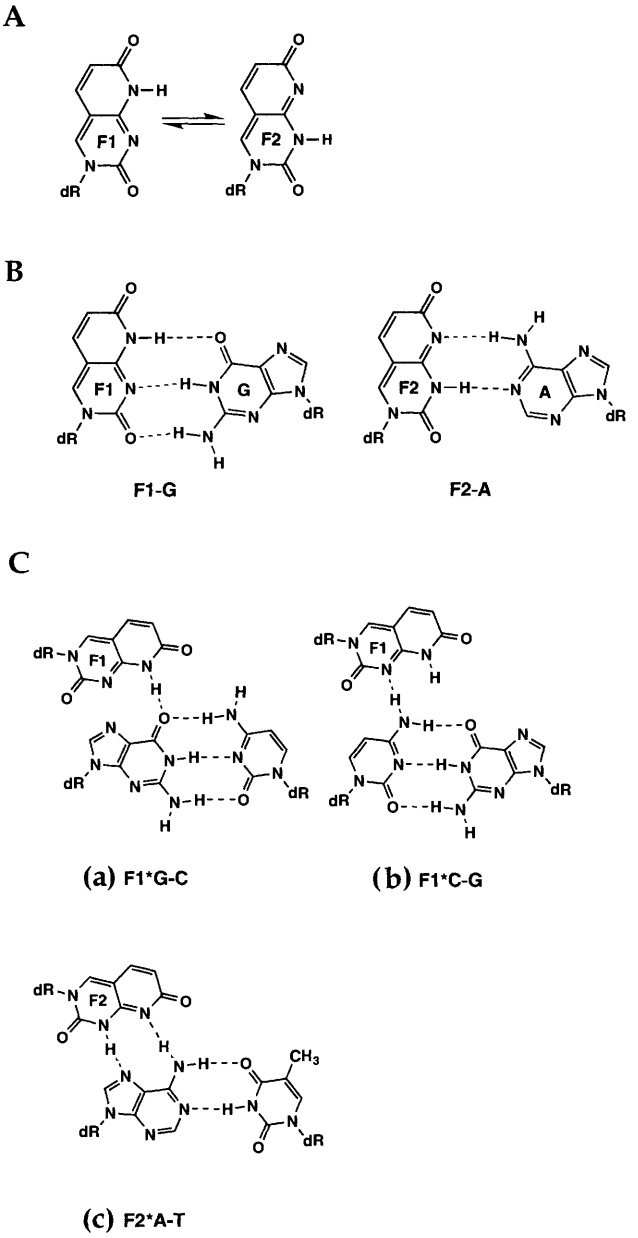


Fig. 4 The tautomers (**A**) of pyrido[2,3-*d*]pyrimidine nucleoside (F) bind either single or double stranded DNA. (**B**) Single-strand binding through F1–G and F2–A Watson–Crick type base pairs. (**C**) Triple helix stabilization with F1 (*a*, *b*) and F2 (*c*).

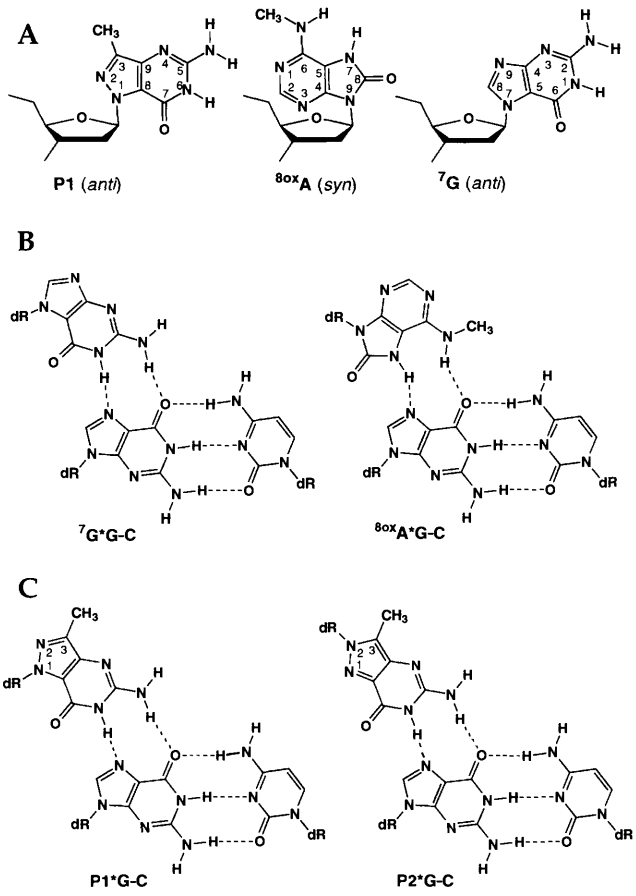


Fig. 5 Structures of purine-like cytidine mimics. (**A**) In contrast to the *anti* glycosidic bond conformation of most purine nucleosides, ^{8ox}A has predominantly the *syn* conformation. Nonisomorphous hydrogen bonding schemes of ^{7}G*G–C and ^{8ox}A*G–C (**B**), and of P1*G–C and P2*G–C (**C**) triplets within a Py*Pu–Py triplex.

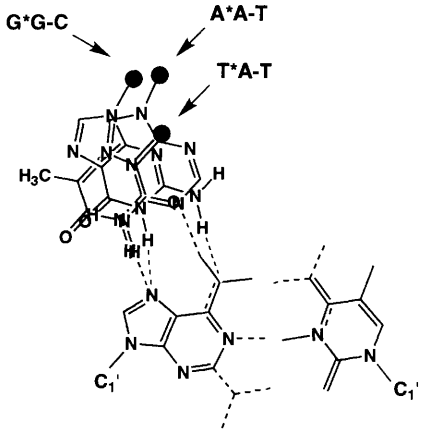


Fig. 6 Non-isomorphism of natural base triplets within a Pu*Pu–Py triple helix. (●) indicates the ribose C$^{1'}$ atom of the third strand.

2.2.2 Towards isomorphous base triplets

As mentioned above, another problem within this triplex family is the non-isomorphism of base triplets, especially for the T*A–T/G*G–C triplexes. This results in deformation of the deoxyribose-phosphodiester backbone at the transition from a bicyclic purine to a monocyclic pyrimidine, thus decreasing the stability of triple helices. TFOs containing only purines could have a better chance to be isomorphic. 7-Deazaxantosine[18] (dzaX) and 2-aminopurine[20] (amP) have been used instead of T (or A) for A–T base pair recognition (Fig. 8). dzaX seems to be very promising: in comparative studies under physiological K$^+$ concentration, the TFO containing G and dzaX showed a greater than 100-fold increase of affinity for the target sequence as compared to the G/T TFO.

amP*A–T triplets are isomorphous with G*G–C triplexes in the Hoogsteen orientation [Fig. 8(**C**)] which is different from the typical Pu*Pu–Py reverse-Hoogsteen family [Figs. 2, 8(**B**)]. The synthesis of modified TFOs containing amP has been described but no binding experiments have been reported so far.

3 Towards a general triple helix-mediated DNA recognition scheme

As discussed previously, triple helix formation is strictly limited to homopurine targets, as a consequence of DNA recognition

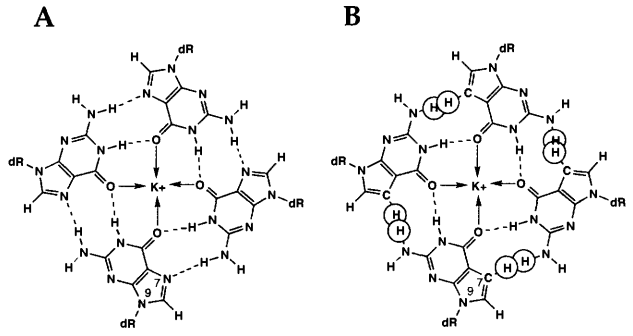


Fig. 7 (**A**) G-rich oligonucleotides form G-tetrads in the presence of potassium cations. (**B**) Putative disruption of the tetrad H-bonding network for ^{9}dzaG.

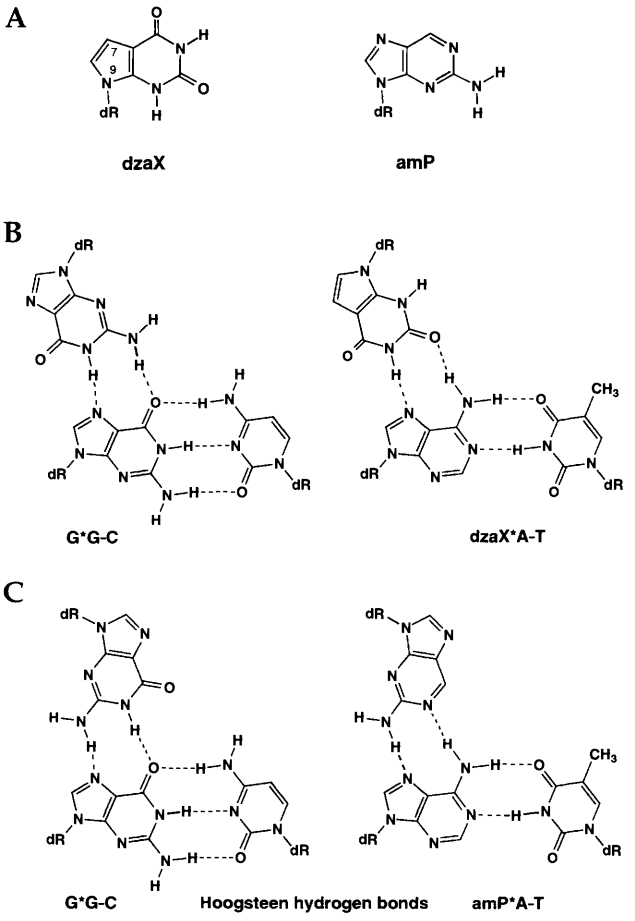


Fig. 8 7-Deazaxantosine (dzaX) and 2-aminopurine (amP) (**A**) could form isomorphous base triplets with A–T in a G*G–C context, by reverse-Hoogsteen bonding for dzaX (**B**), or Hoogsteen bonding for amP (**C**)

via Hoogsteen-type hydrogen bonds. As further stages towards a general sequence recognition scheme, two types of sequences have been tackled: alternated oligopurine–oligopyrimidine tracts and quasi-homopurine sequences with a single pyrimidine base.

3.1 'Switched' triple helices

When the DNA target is made of adjacent oligopurine-oligopyrimidine domains, a continuous oligopurine tract still runs along the helix yet switches strand at each domain junction. Short canonical triple helices of low stability could be formed with each homopurine sequence. However, if all short TFOs are linked *via* appropriate spacer groups, triplex formation is very much enhanced due to cooperative binding. This approach is called alternate-strand triple helix formation, or 'switched' triple helix. Indeed, short TFOs when joined together form a single third-strand oligomer that zigzags along the major groove, switching from one oligopurine strand to the next one on the other strand. Combining the two canonical triple helix families with opposite strand polarities (Fig. 2) leads to many examples of 'switched' triple helices.[21] Non-natural α-anomeric oligonucleotides,which generally bind in the reverse orientation as compared to β-oligonucleotides, are also attractive blocks to enrich the set of 'switched' TFOs.[22]

An alternate-strand recognition scheme exclusively using Py*Pu–Py triplexes is illustrated in Fig. 9(**A**). The alternated purine (R)–pyrimidine (Y) target DNA sequence is $R_7Y_7R_7$. The third strand is made of three short pyrimidine (Y) TFOs. On both ends, TFOs run in the same direction along the major groove, whereas the central one must have the opposite polarity to stay parallel to its R_7 target which is located on the opposite DNA strand.

The design of alternate third strands is a complex task requiring linkers adapted to each type of junction crossing the major groove. Molecular modelling is widely used to find optimally rigid linker groups with reduced entropy of binding. Several linkers have been described, such as hexaethylene glycol, propane-1,3-diol and xylose derivatives. Recently, a direct base-to-base linkage has been described,[21] that allows shorter linkers to be used for different types of junctions. Binding tests with a $C^{3'}$–$C^{3'}$U-linked oligonucleotide [Fig. 9(**B**)] as TFO (Py*Pu–Py motif) have shown triple helix formation, but the junction remains to be optimized to increase the complex stability. An as yet unanswered question is the effective participation of the bases close to the junctions in the stability and selectivity of the binding process.

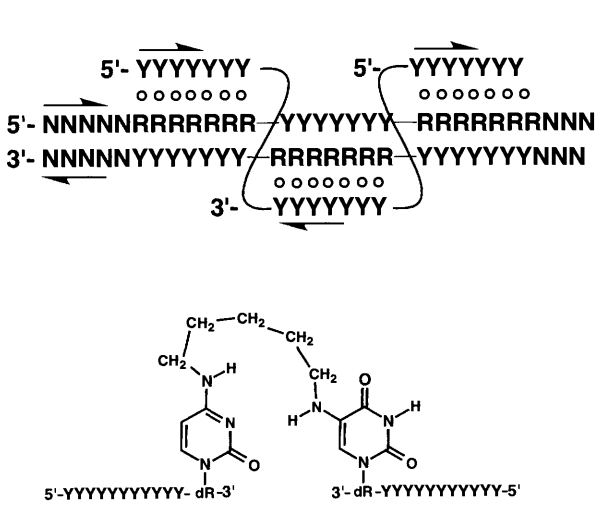


Fig. 9 (**A**) Alternate strand recognition in a Py*Pu–Py context. (**B**) Structure of a short strand-switching pentamethylene $C^{3'}$–$^{3'}$U-linker.

3.2 Direct pyrimidine base recognition

Several solutions have been described to bypass interruption of the target purine stretch by a single pyrimidine. Non-specific stabilization by incorporation of an intercalator at an internal site of the TFO is a simple and efficient solution.[23] Alternatively, pyrazole-, imidazole-, 1,2,4-triazole- and tetrazole-substituted nucleotides have been used in front of a pyrimidine base in an otherwise Pu*Pu–Py triple helix.[24] Azole incorporation enhanced TFO binding when compared to that of a natural base. These five-membered heterocycles were not designed for recognition of any particular base but rather to maintain some stacking interaction with neighbouring bases. Furthermore their rather small size reduces steric hindrance, especially when facing T which has a protruding methyl group. The pyrazole derivative seemed to be of particular interest: despite showing only weak preference for T–A over G–C, it discriminated strongly against A–T and C–G base pairs.

A fatal limitation to specific recognition of pyrimidine bases in Watson–Crick base pairs is that they possess only a single remaining hydrogen bond-forming site in the major groove (Fig. 1), that results in an intrinsically lower stability. Furthermore, thymine binding is obscured by its bulky methyl group. Considerable efforts to solve these problems have been undertaken, yet only limited success achieved.

Several non-canonical natural base triplets of intermediate stability, such as G*T–A in a Py*Pu–Py context and T*C–G (Pu*Pu–Py) were discovered following a systematic study.[25] Both base triplets involve a single hydrogen bond between the third-strand base (G or T) and the target pyrimidine inversion (T or C) (Fig. 10).

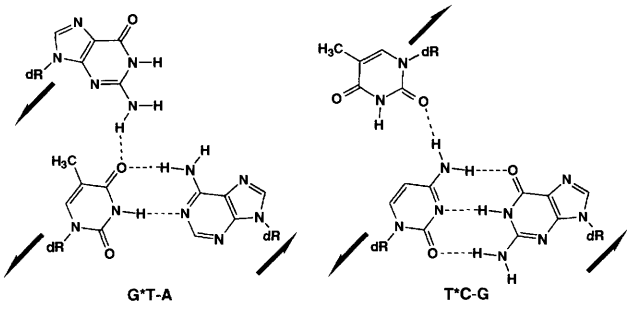


Fig. 10 Non-canonical G*T–A and T*C–G base triplets in a Py*Pu–Py and a Pu*Pu–Py context, respectively. In contrast to the canonical purine recognition scheme, only a single hydrogen bond can be drawn between third strand and target bases.

Some nucleoside analogues like nebularine[26] (N) recognize C–G in a Pu*Pu–Py triple helix. Once again a single hydrogen bond is established, and the N*C–G interaction is weaker than G*G–C, A*A–T or T*A–T. Moreover, nebularine has comparable affinities for C and A since it provides a hydrogen bonding acceptor site to the exocyclic amino group (Fig. 11).

Thus solutions found so far have limited destabilization rather than improving pyrimidine recognition.

3.3 Other-strand recognition

As stated above, proper pyrimidine recognition with energies comparable to that of purines requires more than one hydrogen bond. Since this is simply not feasible with pyrimidines, an attractive solution would be to bind the facing purine of the opposite strand instead. Along these lines, formycin A (^{for}A) was suggested to form two hydrogen bonds with the guanosine of the other strand at C–G inversions in a Pu*Pu–Py triplex,[27] as shown in Fig. 12(**A**). To do so however, a 3–5 Å translation as well as a rotation around the ribose-phosphate backbone are required, which would result in severe backbone distortion. Incorporation of three ^{for}A*C–G instead of G*C–G resulted in a ten-fold increase in binding affinity,[27] but this may also coincidentally be due to reduced steric hindrance of formycin relative to guanine.

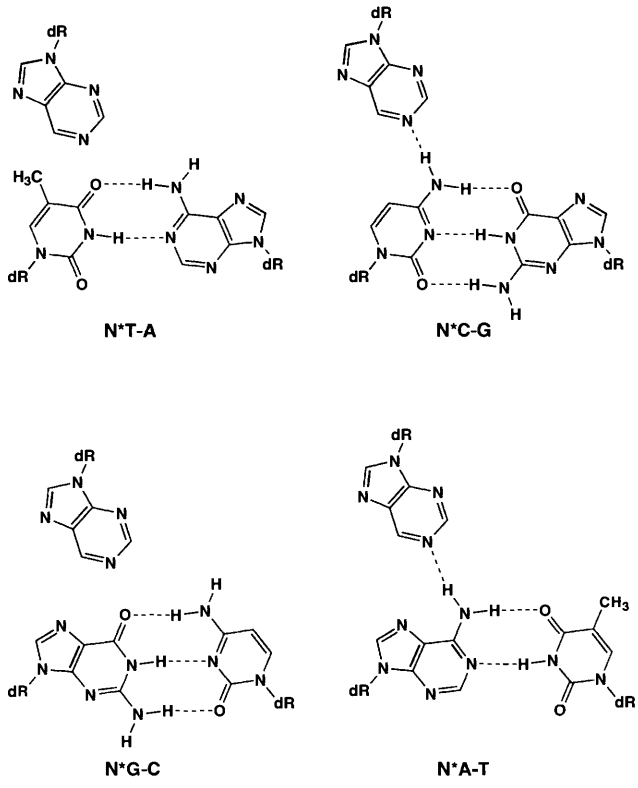


Fig. 11 Putative interactions between nebularine (N) and the four Watson–Crick base pairs in an antiparallel Pu*Pu–Py triple helix. Nebularine provides a hydrogen bonding acceptor site to the exocyclic amino groups of C and A.

Indeed, when comparing ^{for}A*C–G and G*C–G triplets within the Pu*Pu–Py backbone geometry [Figs. 12(**B**) and (**C**)], formycin escapes the G(N^1-H and N^2-H)/C(N^4-H) clash and may even take advantage of an attractive ^{for}A(N^1)/C(N^4-H) interaction.

The future success of this approach will rely on the possibility of crossing the DNA duplex major groove with more extended structures and without backbone distortion.

3.4 Base-pair recognition

The most promising way to reconcile affinity and selectivity is to use extended heterocyclic systems that can simultaneously utilize all major groove hydrogen-bonding sites of a Watson–Crick base-pair for molecular recognition. The first example of such a strategy was the use of 4-(3-benzamidophenyl)imidazole[28] (D3) expected to hydrogen bond simultaneously C(N^4-H) and G(O^6) of C–G in a Py*Pu–Py triple helix (Fig. 13). Other base pairs were supposed to lead to steric clashes with adenine, cytosine and the methyl group of thymine (Fig. 13). Unfortunately, experimentation showed the following order of stabilities: D3*T–A ≈ D3*C–G > D3*A–T > D3*G–C. The rationale to this was found later, when it was shown that D3 interacts by sequence-specific intercalation, rather than by hydrogen bonding.[29]

When N^4-(3-acetamidopropyl)cytidine (^{aap}C) was incorporated in front of C–G inversions, a considerable stabilizing effect was observed relative to third strands with N^4-butylcytidine or N^4-(3-carboxypropyl)cytidine whose side chains lack hydrogen bond donating groups. The length of the flexible acetamidopropyl arm of ^{aap}C is sufficient to span the duplex major groove and allows the terminal amide to form an additional hydrogen bond with O^6 of guanine in the target C–G base pair. To improve binding, a more rigid 6-amino-2-pyridinyl sub-

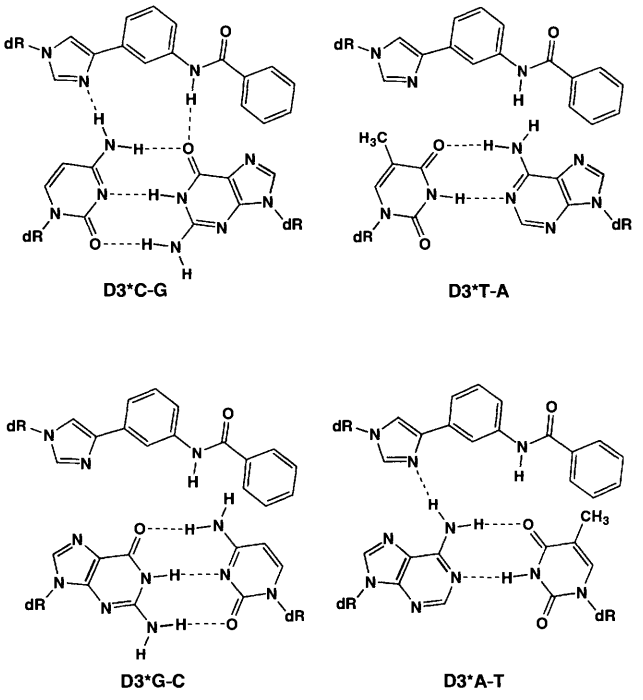


Fig. 13 Best hydrogen bonding fit between the synthetic base analogue D3 and the four base pairs in a Py*Pu–Py triple helix

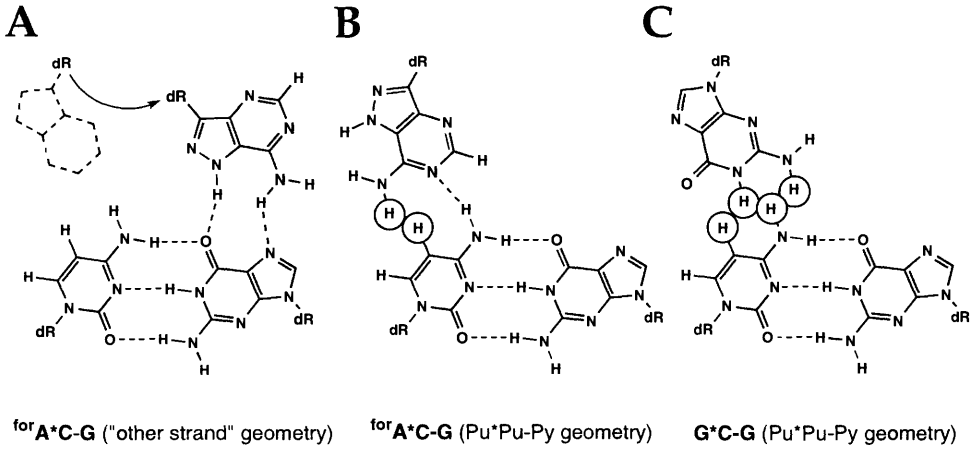


Fig. 12 (**A**) For cytosine recognition, ^{for}A could hydrogen bond to the guanine of the opposite DNA strand at the expense of a *ca.* 5 Å translation (arrow) through the major groove. The dotted line structure indicates the position of guanine in a G*G–C triplex. ^{for}A (**B**) and G (**C**) facing the same C–G inversion.

stituent was introduced at the N^4 position of cytidine (^{apy}C), thus providing two hydrogen bonding sites for interaction with C–G and three for A–T[30] [Fig. 14(**A**)]. Both triads involve an unusual imino tautomeric form of ^{apy}C [Fig. 14(**B**)] which is supported by ^1H NMR and UV spectroscopy of ^{apy}C.

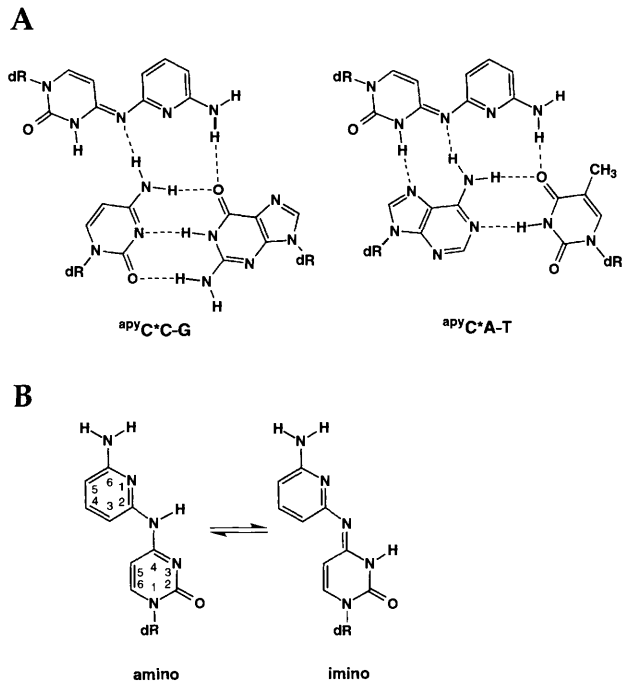


Fig. 14 (**A**) Possible hydrogen bonding schemes of the imino tautomer of ^{apy}C with C–G and A–T base pairs in a Pu*Pu–Py triple helix. (**B**) The amino/imino tautomeric equilibrium of ^{apy}C

As shown in Fig. 14(**A**), the pyridine ring spans the major groove so as to place the 6-amino group of ^{apy}C within hydrogen bonding distance of guanine O^6 or thymine O^4. This interaction is vital and removal of the amino group results in a much decreased affinity or even in no triplex formation at all. Additional stabilization could come from stacking of the pyridine ring with neighbouring bases. Experimentally, stability of the ^{apy}C*G–C base triplet was found to be comparable to that of the canonical C+*G–C; unfortunately A–T, and to a lesser extent T–A and G–C base pairs, also recognized. Here again, design [Fig. 14(**A**)] and experimental facts did not coincide. The ^{apy}C-containing oligonucleotide showed two distinct interconvertible binding modes: hydrogen bonding, as expected, and intercalation as already established for 4-(3-benzamidophenyl)imidazole (D3, see above).

Recently two novel structures for cytidine–guanosine recognition have been designed (Fig. 15) and tested in chloroform. Benzaminoimidazole-glycyl[31] (BIG) and 2-methyl-8-(*N'*-butylureido)naphth[1,2-*d*]imidazole[32] (UNI) were shown by NMR to bind to a C–G base pair by hydrogen bonding to both bases. Such experiments provide preliminary evidence about com-

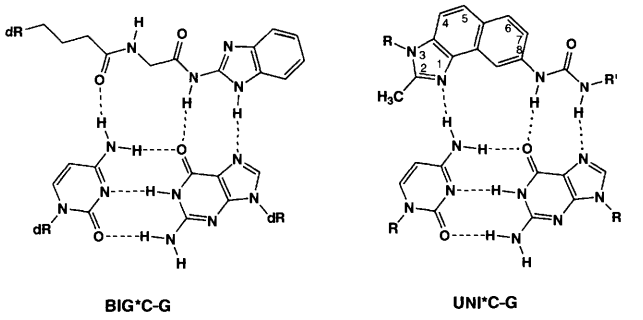


Fig. 15 Extended BIG and UNI nucleic base analogues for specific C–G base pair recognition

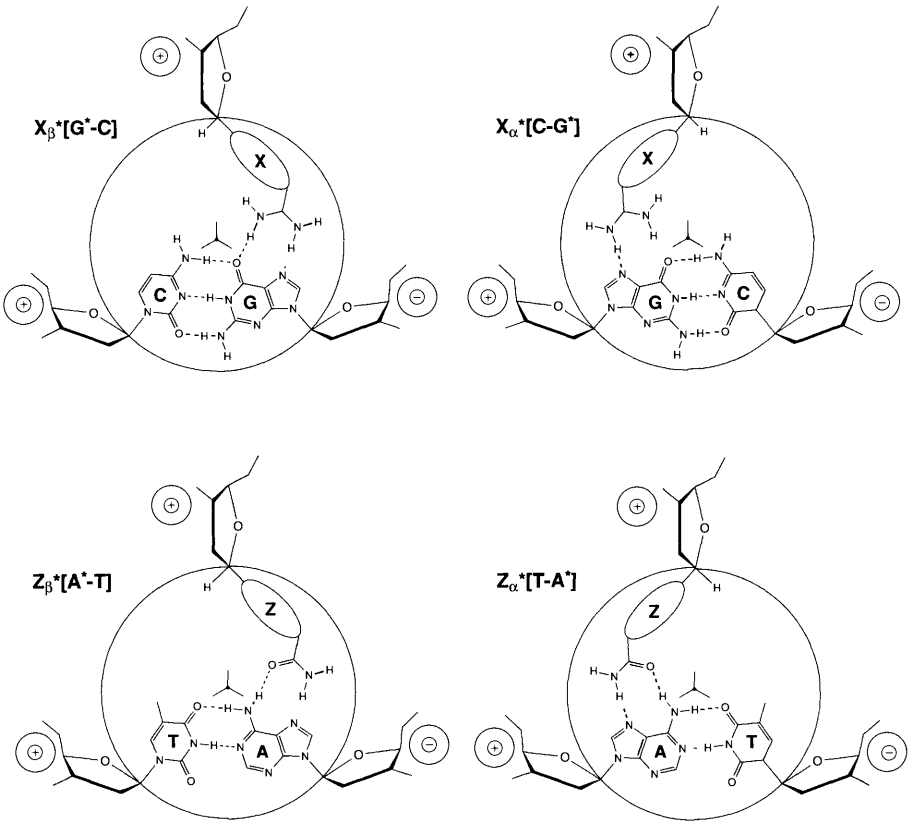


Fig. 16 A third-strand oligonucleotide centered in the major groove would be able to switch from one DNA strand to the other through the β/α anomerism.

plementarity with a given base pair. However they do not take into account base stacking interactions nor isomorphism which are also essential features for triple helix stability.

4 Conclusion

A great deal of intellectual effort and organic synthesis has been invested in molecular recognition of double stranded DNA, with the aim of developing antiviral drugs and controlling endogeneous gene expression. Indeed, drugs based on the antigene strategy would have many advantages over those based on the antisense (anti-messenger RNA) strategy. Yet, up to now, the antigene strategy has essentially benefitted from pharmacological advances obtained with antisense oligonucleotides, such as nuclease resistance and intracellular delivery[33] due to their common generic structures. Increasing the stability of triple helices in physiological media is just one of the problems related to gene targeting (the adverse corollary being the slow kinetics of triplex formation[34]). In this review we deliberately focused on ways of extending triplex formation to non-homopurine sequences, as although polyPu–polyPy stretches are not uncommon in the mammalian genome, the consequences of making random sequences amenable to targeting are obvious. This remains a challenge for organic chemists.[35]

The challenge has been taken up as illustrated by the numerous examples described in this review. They highlight, besides unpredictable synthetic difficulties, predictable requirements for modifying nucleotides, in the area of complementarity, triplex isomorphism and base stacking. They also illustrate effects such as intercalation as opposed to hydrogen bonding for extended molecules, and tautomeric ambiguity of heterocyclic bases. Undoubtedly, integration of this knowledge will help chemists to play a pivotal role in the success of drug development based on sequence-selective DNA recognition.

Added in proof: Besides this review, our real contribution to a general triple helix mediated DNA recognition scheme has been rather small. Our approach was based on the natural complementarity of primary amides with adenine and of guanidine with guanine. The major difficulty of pyrimidine recognition was circumvented by targeting the opposite purine bases with the same molecular groups, but of the unnatural α-anomer nucleoside. Thus molecular recognition of the four nucleic bases (A, T, G, C) is achieved with only two unnatural bases (X, Z), degeneracy being removed by the anomeric (α, β) pluralism. For this to be possible, and for the four base triplets to be isomorphous, several geometrical constraints apply which severely limit the number of chemically reasonable structures (Fig. 16). Organic chemistry along these lines is in progress.[36]

5 References

1 G Felsenfeld, D. Davies and A Rich, *J Am Chem Soc*, 1957, **79**, 2023
2 H. E Moser and P. B. Dervan, *Science*, 1987, **238**, 645.
3 T. Le Doan, L. Perrouault, D. Praseuth, N. Habhoub, J. L. Decout, N. T Thuong, J Lhomme and C Hélène, *Nucleic Acids Res*, 1987, **15**, 7749
4 Y.-K Cheng and B. M. Pettitt, *Prog Biophys Mol Biol*, 1992, **58**, 225.
5 J.-S. Sun, T. Garestier and C Hélène, *Curr Opin Struct Biol*, 1996, **6**, 327
6 J. S. Lee, M. L. Woodsworth, L. J P Latimer and A. R. Morgan, *Nucleic Acids Res*, 1984, **12**, 6603
7 L E. Xodo, G. Manzini, F Quadrifoglio, G. A. van der Marel and J H. van Boom, *Nucleic Acids Res*, 1991, **19**, 5625
8 A. Ono, P O. P. Ts'o and L -S Kan, *J Am Chem Soc*, 1991, **113**, 4032.
9 U. von Krosigk and S. A. Benner, *J Am Chem Soc*, 1995, **117**, 5361
10 G. Xiang, R Bogacki and L. W McLaughlin, *Nucleic Acids Res*, 1996, **24**, 1963.
11 A. B. Staubli and P. B. Dervan, *Nucleic Acids Res*, 1994, **22**, 2637
12 J S. Koh and P. B. Dervan, *J Am Chem Soc*, 1992, **114**, 1470
13 P S. Miller, P Bhan, C. D Cushman and T L Trapane, *Biochemistry*, 1992, **31**, 6788
14 J. Hunziker, E. S. Priestley, H. Brunar and P B Dervan, *J Am Chem Soc*, 1995, **117**, 3233
15 I Radhakrishnan and D. J. Patel, *Biochemistry*, 1993, **32**, 11 228
16 E S Priestley and P B. Dervan, *J Am Chem Soc*, 1995, **117**, 4761.
17 T. S. Rao, R. H Durland, D. M Seth, M. A. Myrick, V Bodepudi and G. R Revankar, *Biochemistry*, 1995, **34**, 765
18 J F Milligan, S H. Krawczyk, S Wadwani and M D Matteucci, *Nucleic Acids Res*, 1993, **21**, 327
19 T. S. Rao, A F. Lewis, R. S. Durland and G. R. Revankar, *Tetrahedron Lett*, 1993, **34**, 6709.
20 V Roig, R Kurfurst and N T Thuong, *Tetrahedron Lett*, 1993, **34**, 1601
21 B -W Zhou, C. Marchand, U. Asseline, N T Thuong, J.-S Sun, T. Garestier and C. Hélène, *Bioconj Chem*, 1995, **6**, 516
22 S B Noonberg, J.-C. François, D. Praseuth, A.-L Guieysse-Peugeot, J. Lacoste, T Garestier and C. Hélène, *Nucleic Acids Res*, 1995, **23**, 4042.
23 B -W Zhou, E. Puga, J -S. Sun, T. Garestier and C Hélène, *J Am Chem Soc*, 1995, **117**, 10 425.
24 R H Durland, T. S. Rao, V Bodepudi, D. M Seth, K Jayaraman and G. R Revankar, *Nucleic Acids Res*, 1995, **23**, 647.
25 W A Greenberg and P B. Dervan, *J Am Chem Soc*, 1995, **117**, 5016.
26 H U. Stilz and P B Dervan, *Biochemistry*, 1993, **32**, 2177
27 T. S Rao, M E. Hogan and G. R. Revankar, *Nucleosides Nucleotides*, 1994, **13**, 95.
28 L. C. Griffin, L. L. Kiessling, P. A. Beal, P. Gillespie and P B. Dervan, *J Am Chem Soc*, 1992, **114**, 7976
29 K M. Koshlap, P Gillespie, P. B. Dervan and J. Feigon, *J Am Chem Soc*, 1993, **115**, 7908
30 C.-Y Huang, G. Bi and P. S. Miller, *Nucleic Acids Res*, 1996, **24**, 2606
31 S. Sasaki, S. Nakmashima, F. Nagatsugi, Y. Tanaka, M. Hisatome and M. Maeda, *Tetrahedron Lett*, 1995, **36**, 9521.
32 S C. Zimmerman and P. Schmitt, *J Am Chem Soc*, 1995, **117**, 10769
33 O Boussif, F. Lezoualc'h, M A. Zanta, M. D. Mergny, D Scherman, B. Demeneix and J. P. Behr, *Proc Natl Acad Sci USA*, 1995, **92**, 7297
34 S Baillet and J. P Behr, *Tetrahedron Lett*, 1995, **36**, 8981
35 N. Schmid and J. P. Behr, *Tetrahedron Lett*, 1995, **36**, 1447
36 S O Doronina, S Blanalt-Feidt and J P. Behr, to be published.

Received, 25th October 1996
Accepted, 22nd November 1996