

Unnatural Nucleosides with Unusual Base Pairing Properties

Synthetic modified nucleosides designed to pair in unusual ways with the natural nucleic acid bases have many potential applications in nucleic acid biochemistry. These range from biochemical tools for probing nucleic acid structure or protein–nucleic acid interactions to tools for re-engineering DNA and ultimately proteins. Applications as components of nucleic acid–based diagnostic tools for clinical analysis have been envisioned. Furthermore, unnatural bases may be useful as components of antisense or antigene nucleic acid analogs in therapeutic applications. This unit serves to lay the foundation for future protocol units on unnatural base synthesis and application, with particular emphasis on unnatural base analogs that mimic natural bases in size, shape, and biochemical processing. A much more extensive compilation of unnatural nucleobases has recently been published by Luyten and Herdewijn (1998).

To design bases that mimic natural bases in function, it is useful to consider the factors that are essential for effective base pairing and stable duplex formation. In addition to structures that are configured to hydrogen bond and base stack within the spatial confines of duplex DNA and RNA, any surrogate base pair must also conform to specific dimensions and geometry if it is to function in roles that require recognition by nucleic acid–processing enzymes. To be isosteric with AT or GC base pairs, the C1' to C1' distance must be in the range of 10.8 to 11.0 Å, and λ_1 and λ_2 should be $\sim 50^\circ$ (Fig. 1.4.1; Saenger, 1984).

Just as it is important to know where to place hydrogen bond donor and acceptor sites, it is important to consider the availability of sur-

rounding free space when designing new nucleobases. This can generally be determined by inspecting the groove regions of nucleic acid duplex and triplex models. Where the foundation for the design is a natural nucleobase, substitution is allowed at C5 of C, T, or U. N4 of C and N6 of A are also possible attachment sites, but generally compromise base association. N2 of G is acceptable and, unlike all other sites, allows minor groove placement of an appendage. C8 of A and G have been used as sites for attachment of appendages, but substitution here may influence the conformation preference about the glycosidic bond. Replacement of the purine N7 with a carbon provides a site for attachment of appendages in a sterically tolerated position. There are many examples of appendages that, when added to the natural nucleic acid bases, enhance base-pair stability. These include substituents that increase the acidity of proton donor sites (Yu et al., 1993) or increase hydrophobicity and aromatic surface area for enhanced base stacking (Inoue et al., 1985), as well as appended cations for electrostatic interactions with the phosphodiester backbone (Ueno et al., 1998).

BASE PAIRS WITH ALTERNATIVE HYDROGEN BONDING SCHEMES

Purine-Pyrimidine-Like Base Pairs

Benner and co-workers have described a set of nucleobase analogs that resemble natural bases, but have reconfigured hydrogen bonding patterns (Piccirilli et al., 1990). Six orthogonal base pairs (**S.1** to **S.6**) are shown in Figure 1.4.2. Extensive studies on these nucleosides have revealed that even these subtle changes in structure can cause profound effects on thermodynamic (Voegel and Benner, 1994) and biochemical properties (Switzer et al., 1993; Horlacher et al., 1995). For example, the C-nucleoside pyrimidine mimics appear to base pair more weakly than equivalent base pairs composed of N-nucleosides. As far as biochemical properties, it appears that certain DNA polymerases require the pyrimidine O2 and purine N3 as recognition features for efficient template-mediated oligonucleotide synthesis (Horlacher et al., 1995).

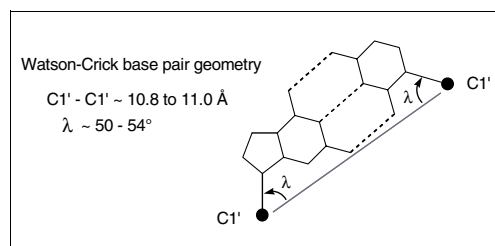


Figure 1.4.1 Base pair parameters. See *APPENDIX 1B* and Figure A.1B.4 for other base pairing schemes.

Contributed by Donald E. Bergstrom

Current Protocols in Nucleic Acid Chemistry (2001) 1.4.1-1.4.13

Copyright © 2001 by John Wiley & Sons, Inc.

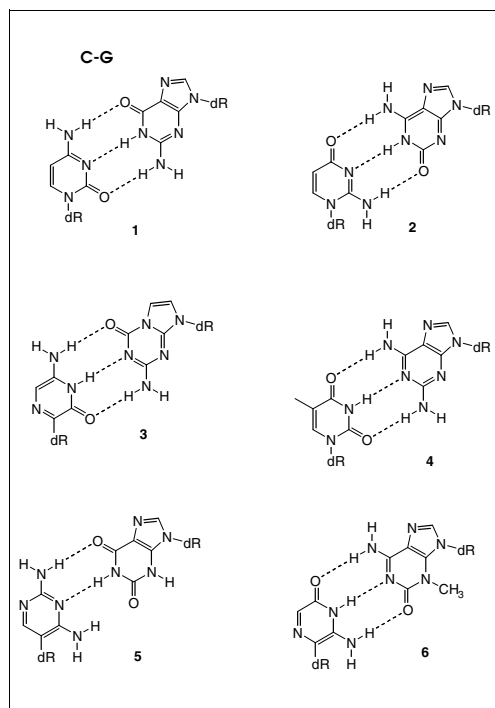


Figure 1.4.2 Structures of six Watson-Crick-type base pairs utilizing mutually exclusive hydrogen bonding schemes.

Self-Complementary Nucleobases

Another concept for new base pair development was proposed by Pochet and Marlière (1996). Based on the known ability of the mutagenic base 8-oxoguanine to base pair with A from a *syn* conformation, these researchers redesigned the base by removing the 6-oxo group. This yields the unnatural base 2-amino-8-oxopurine (**S.7**), which they postulate would pair with itself according to the arrangement shown in Figure 1.4.3. The results of biochemical studies with this unusual base have not yet been reported. One can imagine similar self-pairing potential for azole carboxamide nucleosides, as illustrated in Figure 1.4.3 for 1,2,4-triazole-3-carboxamide (**S.8**). However, T_m studies on a duplex containing 1,2,4-triazole-3-carboxamide showed that when placed oppo-

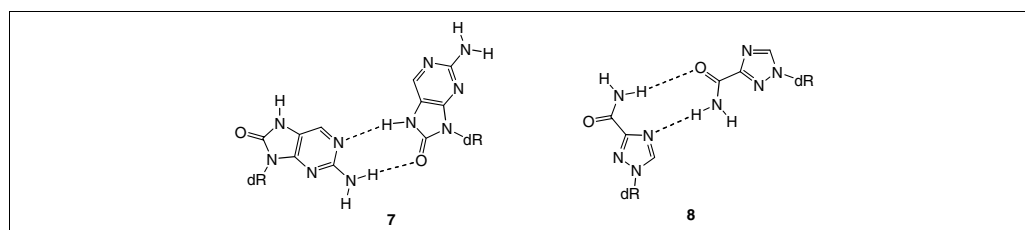


Figure 1.4.3 Self-pairing bases.

site itself, the duplex was destabilized more than when the triazole was placed opposite each of the natural bases. This result is in line with NMR studies on duplex DNA containing the triazole carboxamide opposite either G or T (Klewer et al., 2001). In both cases, the triazole prefers to adopt a conformation in which the amide group points out into the major groove rather than inward towards the opposing base pair. To achieve specific association through hydrogen bond interactions, one of the two triazole carboxamides in a self-pair would have to face with the amide projecting inward towards the opposing base. This example reflects parameters that one must consider when designing base pairs; the interior of the helix is a less hydrophilic environment and, without sufficient compensation, highly hydrophilic groups will prefer to assume positions that place them in a more hydrophilic environment.

HYDROPHOBIC BASE PAIRS

There are now a significant number of examples of hydrophobic unnatural bases that pair with other hydrophobic bases or with themselves in duplex DNA with higher affinity than with any of the natural bases. 3-Nitropyrrole deoxyribonucleoside (**S.26**; Fig. 1.4.8), which was originally designed as a universal nucleobase, pairs with almost equal affinity to each of the natural bases, but a 12-mer duplex with nitropyrrole opposite itself is significantly more stable than the same duplex with nitropyrrole opposite each of the natural bases (Bergstrom et al., 1995; Zhang et al., 1998). The significance of the hydrophobic substituent (nitro) in mediating this effect is clear when one compares pyrrole-3-carboxamide, which in duplex DNA yields far more stable duplexes when paired opposite each of the natural bases than when paired opposite itself.

Because of these results, hydrophobic base pairs are considered to be attractive candidates for extension of the genetic alphabet. Two themes are possible: (1) the development of a

complementary hydrophobic pair, and (2) the creation of a single self-complementary hydrophobic base. The latter possibility is more attractive, because one need contend with the optimization of DNA replication with only one unnatural nucleoside.

A series of recent papers from a research effort at Scripps Research Institute led by Romesberg and Schultz has described considerable progress in the development of hydrophobic self-pairing bases (McMinn et al., 1999; Berger et al., 2000a; Ogawa et al., 2000; Wu et al., 2000). Three of these nucleoside analogs, 1- β -D-deoxyribose-7-azaindole (**S.10**; Ogawa et al., 2000) and two different 7-propynyl isocarboxtyril deoxyribonucleosides (**S.9**; R = H and CH₃; McMinn et al., 1999) are shown in Figure 1.4.4. These nucleoside analogs, as assessed by T_m measurements, self-pair significantly more effectively than they pair with any of the natural bases. Furthermore, both contain a lone pair of electrons that may be positioned optimally in the minor groove for interaction with DNA polymerase.

Kool and co-workers have designed a set of non-hydrogen-bonding base analogs for A and T (Fig. 1.4.4; **S.12** to **S.15**; Schweitzer and Kool, 1995; Guckian et al., 1998). In a 12-bp duplex, virtually all paired combinations of the hydrophobic bases gave higher T_m values than when each hydrophobic base was paired with any of the natural bases. Pyrene deoxyribonucleoside (**S.11**; Fig. 1.4.4) is noteworthy because of the high specificity with which it pairs opposite abasic sites (Matray and Kool, 1999).

The nonpolar nucleobase difluorotoluene (**S.14**), a thymine isostere lacking hydrogen-bonding functionality, can effectively substitute for thymine in both the template strand and as the incoming nucleoside triphosphate. These results suggest that shape recognition in the absence of hydrogen bonding is an important factor in base pair recognition (Guckian and Kool, 1997; Moran et al., 1997a,b; Guckian et al., 1998; Kool, 1998). Similarly, 4-methylbenzimidazole (**S.12**) is an effective surrogate for adenine when matched opposite the difluorotoluene nucleobase (Morales and Kool, 1999).

METAL-MEDIATED ASSOCIATION OF LIGAND NUCLEOBASE MIMICS

Another way of mediating specific association between nucleobase-like molecules contained on a deoxyribosephosphodiester backbone is to configure them to bind transition metals. The simplest configurations would involve metals that form either linear planar complexes (e.g., Hg) or square planar complexes (e.g., Cu). Two examples of nucleobase mimics that bind a metal in a square planar complex have been published. Meggers et al. (2000) synthesized oligonucleotides containing opposing pyridine and pyridine-2,6-dicarboxylate ligands. The duplex is significantly stabilized in the presence of copper (Fig. 1.4.5; **S.16**), but not by other metals.

Tanaka and Shionoya (1999) have described the synthesis of the diaminophenyl deoxyri-

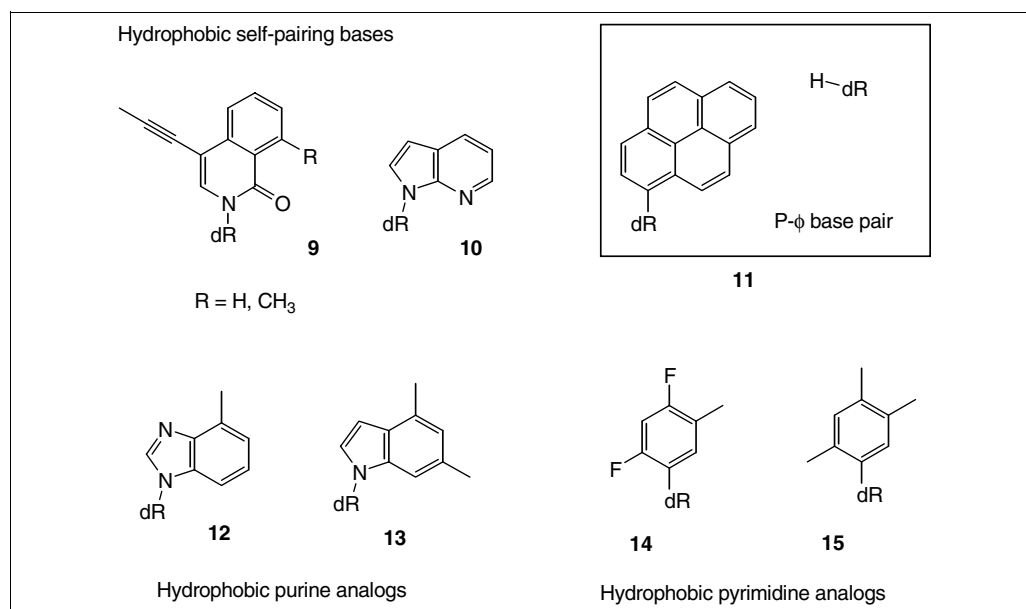


Figure 1.4.4 Hydrophobic nucleobase analogs.

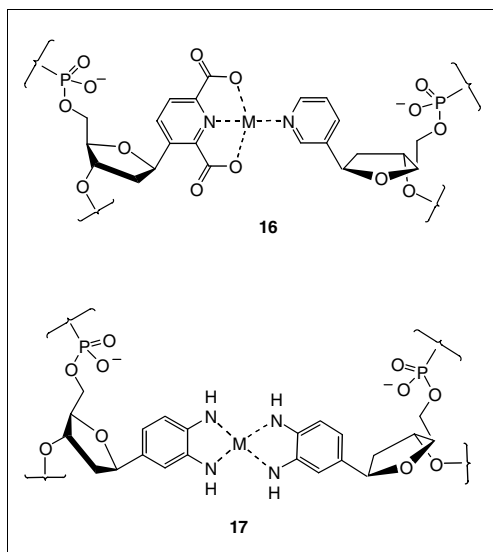


Figure 1.4.5 Metal-mediated duplex association through metal-chelating nucleoside analogs. M, metal.

bonucleoside **S.17**, which binds platinum in a square planar complex to yield a stable structure resembling a natural base pair. Incorporation of this modified base into DNA and metal-mediated assembly of duplexes has not yet been described. However, the concept is noteworthy and could be the prelude to the development of highly novel nucleic acid analogs with unusual applications in material science and nanotechnology.

DEGENERATE BASES

Chemical and radiological damage of nucleic acid bases frequently leads to modified bases that exhibit the ability to direct the incorporation of more than one natural base. Harnessing this ability for the development of mo-

lecular tools could be useful. One example is use of multicoding bases for the introduction of low-level mutations to generate random sequences in protein engineering.

Bases that have the ability to pair with one natural base on template-directed primer extension but code for another base on template-directed replication by DNA polymerase have been referred to as convertides. Such bases are potentially useful as components of PCR/ligation-based detection assays (Day et al., 1999a,b). Two types of convertides have been described: (1) analogs that alter hydrogen bonding pattern and base pairing properties through tautomerism, and (2) analogs that alter their hydrogen bonding profile through changes in conformation. Brown and co-workers have described a purine analog that can base pair as either A (**S.20**) or G (**S.21**) and a pyrimidine analog that can base pair as C (**S.19**) or T (**S.18**; Fig. 1.4.6; Lin and Brown, 1989; Brown and Lin, 1991a; Negishi et al., 1997; Hill et al., 1998). The purine analog can adopt the hydrogen bonding configurations of adenine and guanine (Fig. 1.4.6) and the pyrimidine analog can adopt the hydrogen bonding configurations of thymine and cytosine through tautomerism.

Isosteric azole carboxamide nucleobases capable of mimicking different sets of natural bases have been described. These compounds can mimic more than one natural nucleobase by virtue of rotation about bonds α and χ (Fig. 1.4.7). Modeling studies with azole carboxamides paired with the four natural bases in a B-DNA duplex show that two important parameters of base-pair geometry are maintained: the C1' to C1' distances are in the range of 10.8 to 11.0 Å, and λ_1 and λ_2 are $\sim 50^\circ$. Models of

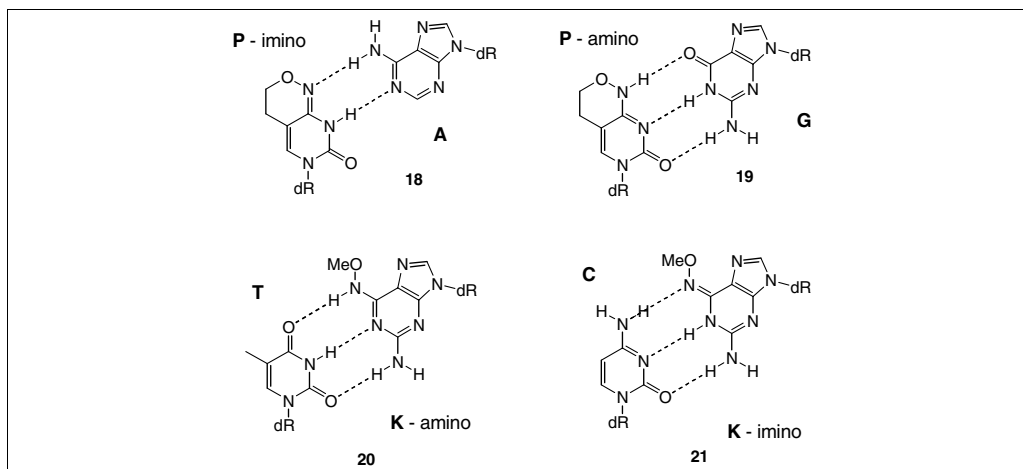


Figure 1.4.6 Tautomericly degenerate base analogs.

azole carboxamide base pairs with a natural base pair yield C1' to C1' distances that fall within 0.2 Å and λ_1 and λ_2 within 3° of the natural base pairs (Bergstrom et al., 1996; Johnson et al., 1997; Zhang et al., 1998). 1,2,4-Triazole-3-carboxamide deoxyribonucleoside illustrates the difficulty of predicting both structural and biochemical behavior of nucleoside analogs. Based on modeling studies, the author of this unit predicted that the triazole would base pair preferentially with G and T by the pairing motifs **S.22** and **S.23** (Fig. 1.4.7). Evidence in favor of this hypothesis included T_m data that showed that this analog gives significantly more stable duplexes when paired with G and T than when paired with C and A. However, an NMR study indicated that 1,2,4-triazole-3-carboxamide paired opposite T assumes a conformation in which the carboxamide group is not hydrogen bonded to T as shown in Figure 1.4.7 (**S.24**). On the other hand, 1,2,4-triazole-3-carboxamide directs the incorporation of C and T (but not G) when a template containing this base analog is replicated by *Taq* DNA polymerase. The geometrically acceptable base pairs in this case would be structures **S.25** and **S.23**.

UNIVERSAL NUCLEOSIDES

Universal Nucleoside Concept

The concept of a “wildcard” or “universal” nucleoside is generally understood to mean a nucleoside analog that can base pair equally

well with all of the natural bases in a nucleic acid duplex. Applications in which a nucleoside analog functions only passively as a partner for a natural base within a duplex represent only one aspect of the behavior necessary for a compound to be truly regarded as a universal nucleoside. Recognition of the triphosphate as a universal nucleotide substrate by DNA polymerase, or suitability of oligonucleotides containing the analog as substrates for enzymes such as restriction endonucleases or ligases, would also be useful functions. Although it is unlikely that a single nucleoside analog can be devised which would function as a universal nucleoside under all circumstances, it is likely that a variety of analogs will be found with universal nucleoside properties optimal for particular techniques.

Universal Base Specifications

The specifications for an ideal universal base are the following.

1. It should form equally stable base pairs with each of the natural nucleobases in all sequence contexts as determined by thermal melting experiments (T_m).
2. It should be accepted as template base by DNA polymerase (preferably *Taq* DNA polymerase) as assessed by PCR replication of universal base-containing template and steady-state kinetic experiments for single-nucleotide primer extension.

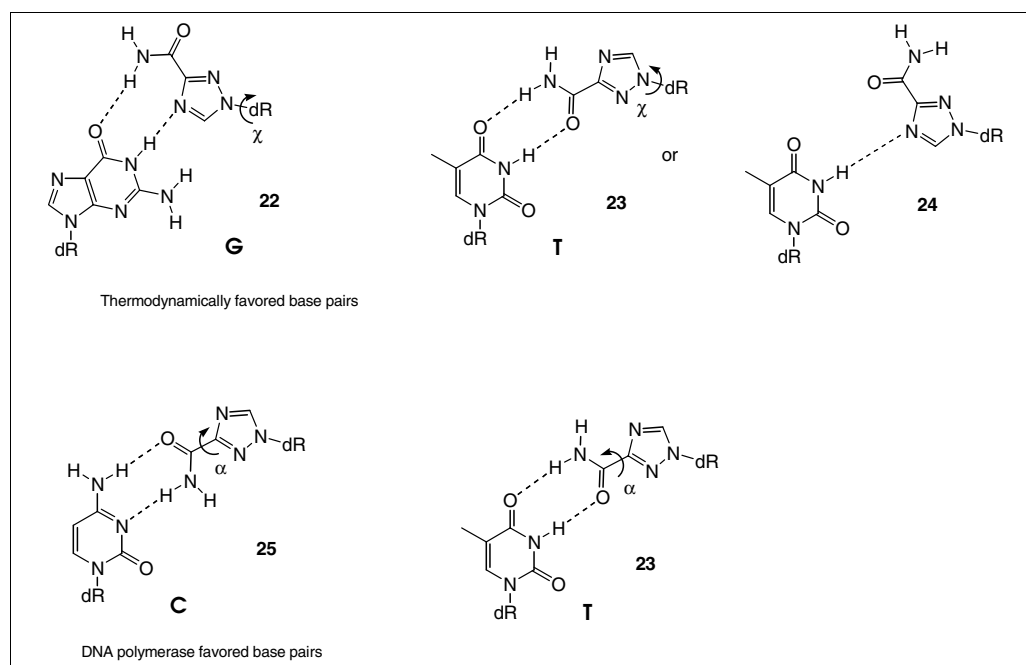


Figure 1.4.7 Conformationally degenerate base analogs.

3. The corresponding deoxyribonucleoside triphosphate should be suitable as a substrate for DNA polymerase as assessed by steady-state kinetic experiments for single-nucleotide primer extension.

4. The base should be a substrate for DNA ligase as primer component as assessed by ligase chain reaction.

This list is by no means complete, but represents those specifications that would be of the greatest use for the manipulation of DNA in conventional molecular biology techniques.

The factors that contribute to base pair stability as assessed by thermal denaturation studies are not directly related to template preference by DNA polymerase. This has been illustrated by comparison of a set ofazole carboxamide nucleosides for which the order of thermodynamic stabilities differs substantially from their template coding properties (Hoops et al., 1997).

Nonpolar Nucleobase Analogs

The development of universal nucleic acid bases has progressed along two parallel lines. The first class of analogs includes molecules that cannot associate through hydrogen bonding, but because of size, shape, and hydrophobicity prefer to occupy the interior of a duplex (Fig. 1.4.8). The first analog of this class, phenyl deoxyribonucleoside (**S.28**), was reported in 1984 (Millican et al., 1984). It is highly destabilizing and pairs with significant

discrimination ($A > G = T > C$). By comparison, 3-nitropyrrole (**S.29**) pairs with little discrimination between each of the four natural bases (Bergstrom et al., 1995, 1997); however, it is significantly destabilizing relative to a natural base. This has been used to advantage in at least two separate applications. 3-Nitropyrrole has been used to increase the selectivity of hybridization-based detection of single-nucleotide polymorphisms (Guo et al., 1997), and as a tool to elevate the fidelity of thermostable *Thermus thermophilus* (*Tth*) DNA ligase for the ligation of oligonucleotide primers (Luo et al., 1996).

The introduction of a second aromatic ring (e.g., 5-nitroindole; **S.30**) led to substantial improvement in duplex stability without too great of a loss in pairing nondiscrimination (Loakes and Brown, 1994). Neither 3-nitropyrrole nor 5-nitroindole deoxyribonucleosides are effective substrates or template components for DNA polymerase (Loakes et al., 1995). They preferentially direct the incorporation of the less polar natural bases A and T (Hoops et al., 1997). More recently it has been shown that both nondiscriminatory base pairing and high stability can be achieved with nucleosides based on the quinolone heterocycle (**S.31**; Berger et al., 2000b). The emphasis of the latter study has been to develop analogs that are effective substrates for DNA polymerases.

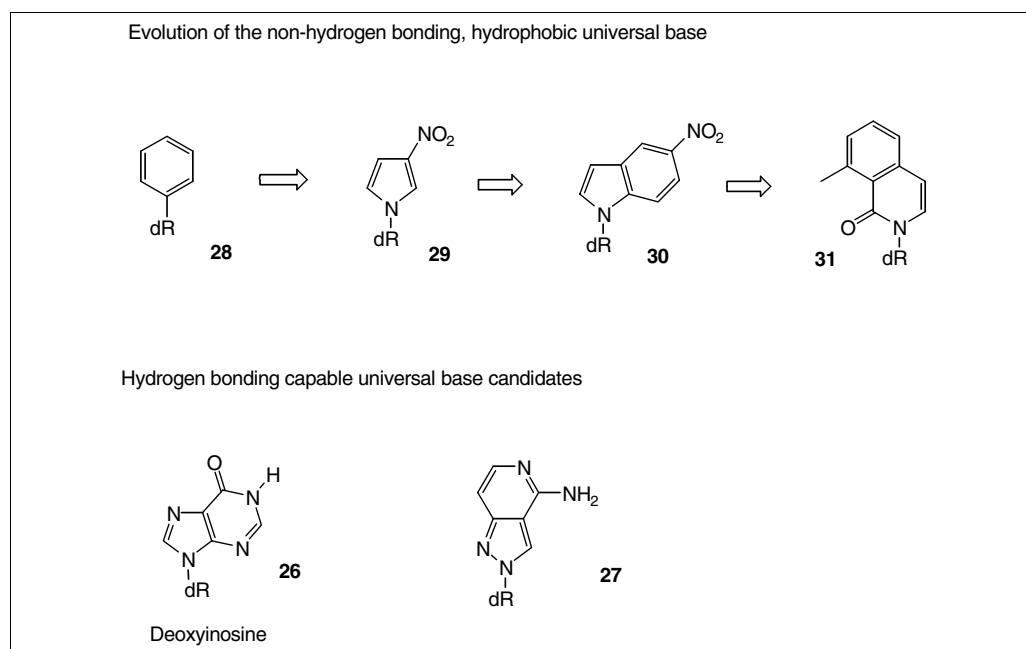


Figure 1.4.8 Universal base candidates.

Polar Hydrogen Bonding Nucleobase Analogs

The first studies of nucleoside analogs specifically designed to base pair with more than one of the four primary DNA bases appeared over a decade ago (Millican et al., 1984; Eritja et al., 1986; Seela and Kaiser, 1986; Habener et al., 1988; Lin and Brown, 1989; Brown and Lin, 1991b). The most extensively studied example is 2'-deoxyinosine (S.26; Fig. 1.4.8), which has been in use as a putative universal nucleoside in oligonucleotide probes and primers since 1985 (Ohtsuka et al., 1985). Structural studies on deoxyinosine-modified oligonucleotides show that dI can base pair to dC, dA (Corfield et al., 1987; Uesugi et al., 1987), T (Cruse et al., 1989; Carbonnaux et al., 1990), and dG (Oda et al., 1991). However, it is not a true universal nucleoside because the base pairs dI-dX (X = dA, dC, dG, T) differ in stability by as much as 2 to 3 kcal/mol (Martin and Castro, 1985; Kawase et al., 1989). More importantly, primers constructed with multiple sites of deoxyinosine substitution frequently give undecipherable results in sequencing experiments. More recently Seela and Debelak (2000) have developed a nucleoside analog, N^8 -(2'-deoxyribofuranoside) of 8-aza-7-deazaadenine (S.27; Fig. 1.4.8), which pairs with all four natural bases with significantly less discrimination than inosine, but with relatively high affinity.

TRIPLEX CONSTITUENTS

Duplex formation occurs through Watson-Crick pairing of purine and pyrimidine bases, which involves hydrogen bonding of NH3 and O4 of thymine with N1 and NH6 of adenine, and O2, N3, and NH4 of cytosine with NH2, NH1, and O6 of guanine. This leaves two sites on each of the purine bases (N7 and NH6 of adenine and N7 and O6 of guanine) free for hydrogen bonding. A third oligonucleotide strand can associate with a duplex through hydrogen bond formation to these sites within the major groove to form a triplex. Triplex formation with natural nucleotides generally assumes one of two themes: parallel association between a homopyrimidine strand and a homopurine-homopyrimidine duplex following the base association schemes $C^+ \bullet GC$ (S.32; Fig. 1.4.9) and $T \bullet AT$ (S.33), or antiparallel association involving the triplets $G \bullet GC$ (S.34), $A \bullet AT$ (S.35), and $T \bullet AT$. In each of these cases, base-base recognition from the third strand involves recognition of N7 and NH6 (adenine) or O6 (guanine). The reason that these base configurations are more stable than other pos-

sible arrangements of bases in a triplet configuration stems from geometry preferences, hydrogen bonding, and base stacking factors. Neither cytosine nor thymine can occupy the center strand, since both have only a single site (O6 of thymine and NH4 of cytosine) available for hydrogen bonding to a base in a third strand located in the major groove. On the other hand, each of the base-triplets illustrated in the figure has two hydrogen bonds between the central purine and the third-strand base. As a result, triplex formation is generally limited to polypurine strings within one stand of the duplex. This significantly limits the number of potential targets in an organism. $T \bullet TA$, $C \bullet CG$, $C \bullet TA$, $T \bullet CG$, $A \bullet TA$, $A \bullet CG$, $G \bullet TA$, and $G \bullet CG$, all of which could have at best one hydrogen bond between the third strand base and the central pyrimidine, are not stable. Furthermore, since C must be protonated at N3 in order to hydrogen bond to N7 of G, triplex formation through this motif is favored only at low pH.

For these reasons, significant effort has been expended to (1) design third-strand unnatural bases that can bind opposite the pyrimidine bases, and (2) develop structural variations of cytosine that are protonated at neutral pH.

Nucleobase design for triplex formation has been extensively reviewed (Ganesh et al., 1996; Doronina and Behr, 1997; Gowers and Fox, 1999). Rather than reiterate the extensive studies that have been done on unnatural nucleosides as third-strand components, this review will only provide a few examples to illustrate the different types of approaches that attempt to solve the problem.

The triplets composed of $C^+ \bullet GC$ and $T \bullet AT$ are isomorphous. On the other hand, the $G \bullet GC$, $A \bullet AT$, and $T \bullet AT$ are not, which leads to significant dependence of triplex stability on the duplex sequence as well as on the relative number of GC and AT base pairs.

Since the third-strand association of C by G requires that C be protonated, contiguous protonated C's may lead to some destabilization through charge-charge repulsion. Consequently, nucleic acid chemists have sought nucleobase analogs that are protonated at this site at physiological pH. The approaches include modification of the cytosine to increase the pK_a of protonated N3, and replacement of cytosine with unnatural bases that have hydrogen bond donor-donor configuration required to pair with the acceptor G. Figure 1.4.10 (S.36 to S.38) illustrates a few of the reported analogs designed to mimic protonated C.

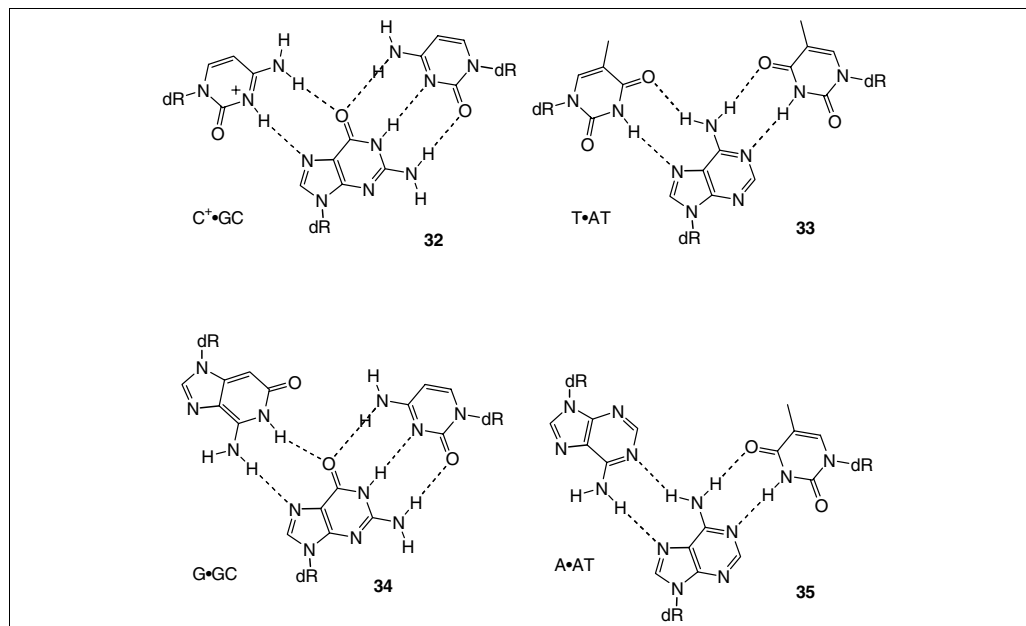


Figure 1.4.9 Base association on triplex formation.

The three structures shown in Figure 1.4.10 reflect three very different design strategies for mimicking protonated C. The first of these, **S.36**, contains the protonated-C mimic *N*⁶-methyl-8-oxo-deoxyadenosine (designated as M in the figure; Krawczyk et al., 1992). The 8-oxo group shifts the equilibrium about the glycosidic bond toward a conformation (shown in the figure) that positions the hydrogen-bond donor sites on M for interaction with N7 and O6 of G. The second triple, **S.37**, also contains a neutral protonated-C mimic, but in this case

the base is a pyrazopyrimidine (P1) with both hydrogen-bond donating groups contained in the pyrimidine ring (Koh and Dervan, 1992). The third triple, **S.38**, contains a protonated 2-aminopyridine (designated as P in the figure; Cassidy et al., 1997). This is a more accurate structural mimic of protonated C than M or P1. The greater basicity of the pyridine ($pK_a \sim 6$) means that the equilibrium will be shifted more towards the protonated form at physiological pH than in the case of C ($pK_a = 4.35$).

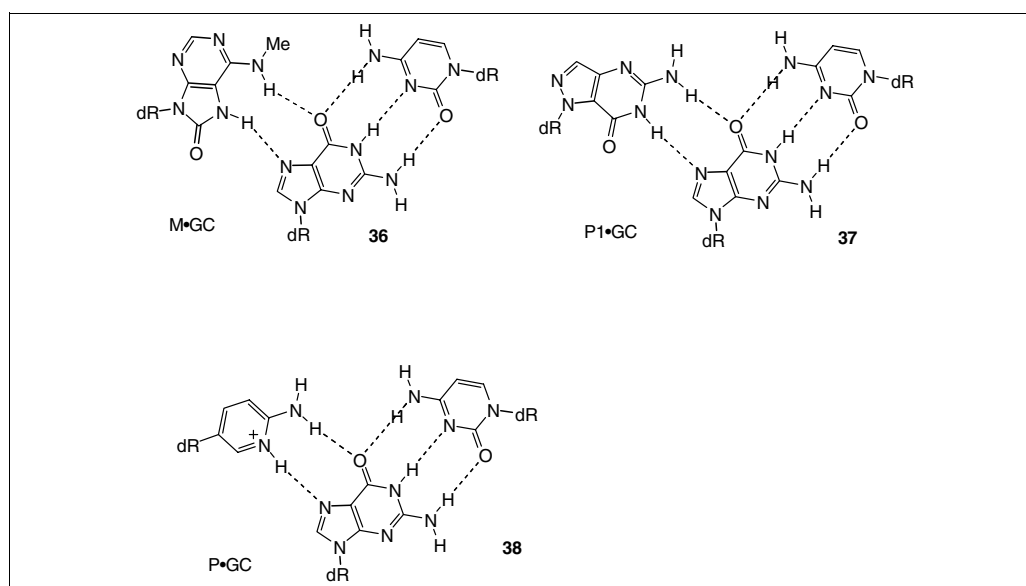


Figure 1.4.10 Protonated cytosine analogs for triplex association. See text for definitions of M, P, and P1.

The expansion of base pair recognition beyond C⁺•GC, T•AT, G•GC, A•AT, and T•AT is not the only issue. The T_m values for third-strand dissociation are typically far lower than for duplex DNA of equivalent length. Consequently, further modifications have been explored to enhance triplex stability. Commonly, this has been achieved by attaching intercalators to one end of the third strand. Significant effort has even been expended to develop triplex-specific intercalators. An alternative is to develop base analogs that provide additional stabilization elements. One example, recently reported by Fox and co-workers, is the use of the nucleoside analog 5-(3-aminopropargyl)-2'-deoxyuridine as a component of the third strand in place of thymidine (Bijapur et al., 1999). The increase in triplex stability, presumably due to association of the protonated amino group with the phosphodiester anion, was substantial.

A challenge for nucleic acid chemists has been to design and develop modified nucleobases that can recognize CG and TA base pairs through association from the pyrimidine side of the major groove. Two examples of modified nucleosides designed to bind the CG base pair are illustrated in Figure 1.4.11. In **S.39**, only one hydrogen bond to the cytosine is present (Prévoit-Halter and Leumann, 1999). In contrast, **S.40** was designed to extend across the base pair and hydrogen bond to G as well as C (Huang et al., 1996). Rothman and Richards have proposed a number of structures from modeling experiments for TA base pair recognition (Rothman and Richards, 1996). The proposed structures, which contain alkyne/alkene spacers to a five-membered ring heterocycle, are predicted to form hydrogen bonds to O4 of T and N6 of A. They should be interesting candidates for future investigation.

MODIFYING NATURAL BASES TO TUNE PAIRING AFFINITY

There are applications that would benefit from either decreasing base pair stability or increasing base pair stability while retaining base pair specificity. For natural sequences, CG base pairs contribute more to duplex stability than AT base pairs. This creates a problem in DNA-array-based strategies where it is advantageous for all sequences of the same length to have melting temperatures within a narrow temperature range. One would expect that to accomplish this without complicating sequence-specific effects would be difficult. A number of examples have been published that report progress in this direction. Nguyen et al. (1998) have reported that N^4 -ethylcytosine destabilizes a CG base pair to the extent that it resembles an AT base pair in stability. Alternatively, it is possible to increase the stability of AT base pairs by appending certain substituents at C5 of T in place of methyl. This is exemplified by the nucleoside analog 5-propynyl-2'-deoxyuridine, which has found application in antisense oligonucleotides (Wagner et al., 1993).

One would think that it would be possible to increase the affinity of association between A and T by adding an amino group to C2 of A to give 2,6-diaminopurine (2,6-DAP), which should hydrogen bond to T with three hydrogen bonds. However, in practice, this is not the case. The effect of the 2,6-DAP-T base pair on duplex DNA stability is dependent on sequence, and in some instances is not as stabilizing as an AT base pair. However, Matray et al. (2000) have discovered that 2,6-DAP is consistently stabilizing when it is incorporated into oligonucleotides containing the N3'→P5' phosphoramidate linkage. This effect may be related to the adoption of A-type helices by this backbone modification. This example illustrates the difficulty of designing unnatural bases for manipulating DNA properties. Too many param-

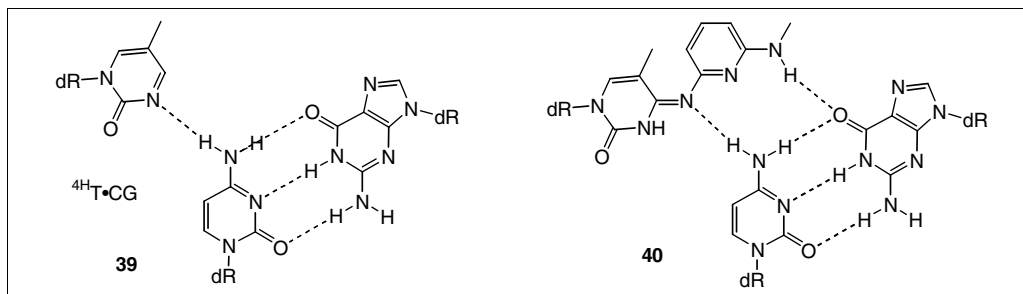


Figure 1.4.11 Nucleobase analogs for triplex recognition of CG base pairs.

ters contribute to base pair stability to enable unambiguous prediction of the effects on duplex properties of even minor structural changes in the heterocycle.

Yet another way that modified bases may be used to tune hybridization was reported by Kutyavin et al. (1996). 2-Thiothymine base pairs with adenine, but not with 2,6-DAP. This allowed Kutyavin and co-workers to develop a strategy for invasion of double-stranded DNA with formation of a stable three-arm junction. The complementary invading oligonucleotides contain 2-thiothymine and 2,6-DAP and do not form a stable duplex with each other, but they do hybridize effectively with the complementary natural sequences.

BASE PAIR RECOGNITION

The unifying characteristic of the compounds discussed in the previous sections is that they are designed to function as nucleic acid components in place of natural nucleotides. However, the scope and potential uses for new molecules designed for nucleobase recognition extend well beyond those described above for unnatural nucleosides. Work on the recognition of base sequences through azole oligomers that bind to the nucleobases through the minor groove has led to the development of highly specific inhibitors of duplex DNA. Design of analogs that read single base pairs provides the opportunity for development of highly sensitive mismatch detection, which could in turn lead to tools for discovery of single-nucleotide polymorphisms. One example is the bis-naphthyridine intercalator (**S.42**) that binds specifically to GG base pairs (Fig. 1.4.12; Nakatani et al., 2001). It should be possible to design other analogs that selectively bind other mismatches. A second example illustrated in the figure is the hexylureido isindolin-1-one derivative (**S.41**), which can associate with both bases of a CG base pair through hydrogen bonding in the major groove.

The molecular details of recognition of multiple base pairs in a sequence through specific association of the bases with DNA-binding molecules such as netropsin are well recognized. Extensive research by Dervan and co-workers has cumulated in the design of polyamides capable of binding long duplex segments with high specificity through hydrogen bonding interactions in the minor groove (Dervan and Burli, 1999; Gottesfeld et al., 2000).

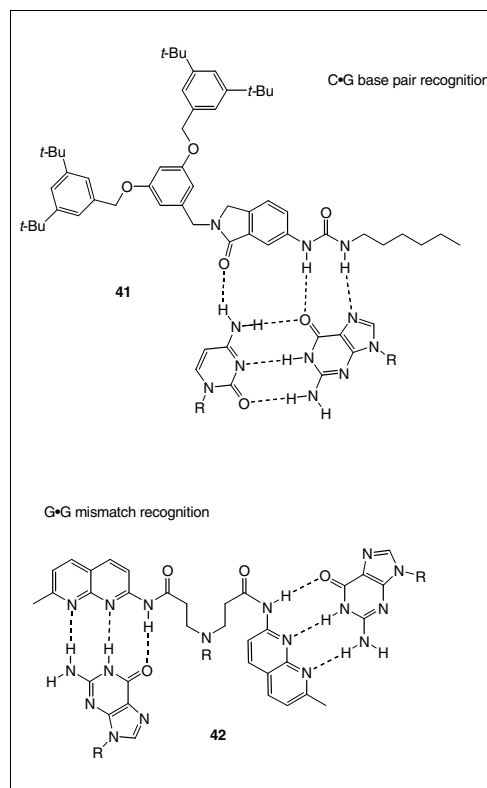


Figure 1.4.12 Molecules that recognize and bind combinations of natural bases.

CONCLUSION

The objective of this review is to highlight a rapidly developing area of nucleic acid chemistry: unnatural base design. The design of new unnatural base pairs, universal bases, triplex components, and other bases with unusual base pairing specificities will continue to provide an attractive arena for molecular designers. It is hoped that the unit will provide some guidance and inspiration for synthetic chemists seeking problems in nucleic acid chemistry. Many potential uses for unnatural bases have been identified, and in most instances totally successful solutions have not yet been established.

LITERATURE CITED

- Berger, M., Ogawa, A.K., McMinn, D.L., Wu, Y., Schultz, P.G., and Romesberg, F.E. 2000a. Stable and selective hybridization of oligonucleotides with unnatural hydrophobic bases. *Angew. Chem. Int. Ed. Engl.* 39:2940-2942.
- Berger, M., Wu, Y., Ogawa, A.K., McMinn, D.L., Schultz, P.G., and Romesberg, F.E. 2000b. Universal bases for hybridization, replication and chain termination. *Nucl. Acids Res.* 28:2911-2914.

- Bergstrom, D.E., Zhang, P., Toma, P.H., Andrews, C.A., and Nichols, R. 1995. Synthesis, structure, and deoxyribonucleic acid sequencing with a universal nucleoside: 1-(2'-Deoxy- β -D-ribofuranosyl)-3-nitropyrole. *J. Am. Chem. Soc.* 117:1201-1209.
- Bergstrom, D.E., Zhang, P., and Johnson, W.T. 1996. Design and synthesis of heterocyclic carboxamides as natural nucleic acid mimics. *Nucleosides Nucleotides* 15:59-68.
- Bergstrom, D.E., Zhang, P., and Johnson, W.T. 1997. Comparison of the base pairing properties of a series of nitroazole nucleobase analogs in the oligodeoxyribonucleotide sequence 5'-d(CGCXAATTYGCG)-3'. *Nucl. Acids Res.* 25:1935-1942.
- Bijapur, J., Keppler, M.D., Bergqvist, S., Brown, T., and Fox, K.R. 1999. 5-(1-Propargylamino)-2'-deoxyuridine (Up): A novel thymidine analogue for generating DNA triplexes with increased stability. *Nucl. Acids Res.* 27:1802-1809.
- Brown, D.M. and Lin, P.K.T. 1991a. The structure and application of oligodeoxyribonucleotides containing modified, degenerate bases. *Nucl. Acids Symp. Ser.* 24:209-212.
- Brown, D.M. and Lin, P.K.T. 1991b. Synthesis and duplex stability of oligonucleotides containing adenine-guanine analogues. *Carbohydr. Res.* 216:129-139.
- Carbonnaux, C., Fazakerley, G.V., and Sowers, L.C. 1990. An NMR structural study of deaminated base pairs in DNA. *Nucl. Acids Res.* 18:4075-4081.
- Cassidy, S.A., Slickers, P., Trent, J.O., Capaldi, D.C., Roselt, P.D., Reese, C.B., Neidle, S., and Fox, K.R. 1997. Recognition of GC base pairs by triplex-forming oligonucleotides containing nucleosides derived from 2-aminopyridine. *Nucl. Acids Res.* 25:4891-4898.
- Corfield, P.W.R., Hunter, W.N., Brown, T., Robinson, P., and Kennard, O. 1987. Inosine-adenine base pairs in a B-DNA duplex. *Nucl. Acids Res.* 15:7935-7949.
- Cruse, W.B.T., Aymani, J., Kennard, O., Brown, T., Jack, A.G.C., and Leonard, G.A. 1989. Refined crystal structures of an octanucleotide duplex with I.T. mismatch base pairs. *Nucl. Acids Res.* 17:55-72.
- Day, J.P., Bergstrom, D., Hammer, R.P., and Barany, F. 1999a. Nucleotide analogs facilitate base conversion with 3'-mismatch primers. *Nucl. Acids Res.* 27:1810-1818.
- Day, J.P., Hammer, R.P., Bergstrom, D., and Barany, F. 1999b. Nucleotide analogs and new buffers improve a generalized method to enrich for low abundance mutations. *Nucl. Acids Res.* 27:1819-1827.
- Dervan, P.B. and Burli, R.W. 1999. Sequence-specific DNA recognition by polyamides. *Curr. Opin. Chem. Biol.* 3:688-693.
- Doronina, S.O. and Behr, J.-P. 1997. Towards a general triple helix mediated DNA recognition scheme. *Chem. Soc. Rev.* 26:63-71.
- Eritja, R., Horowitz, D.M., Walker, P.A., Ziehler-Martin, J.P., Boosalis, M.S., Goodman, M.F., Itakura, M., and Kaplan, B.E. 1986. Synthesis and properties of oligonucleotides containing 2'-deoxynebularine and 2'-deoxyxanthosine. *Nucl. Acids Res.* 14:8135-8153.
- Ganesh, K.N., Kumar, V.A., and Barawkar, D.A. 1996. Synthetic control of DNA triplex structure through chemical modifications. In *Supramolecular Control of Structure and Bonding* (A.D. Hamilton, ed.) pp. 263-327. John Wiley & Sons, New York.
- Gottesfeld, J.M., Turner, J.M., and Dervan, P.B. 2000. Chemical approaches to control of gene expression. *Gene Expr.* 9:77-91.
- Gowers, D.M. and Fox, K.R. 1999. Towards mixed sequence recognition by triple helix formation. *Nucl. Acids Res.* 27:1569-1577.
- Guckian, K.M. and Kool, E.T. 1997. Highly precise shape mimicry by a difluorotoluene deoxynucleoside, a replication-competent substitute for thymidine. *Angew. Chem. Int. Ed. Engl.* 36:2825-2828.
- Guckian, K.M., Morales, J.C., and Kool, E.T. 1998. Structure and base pairing properties of a replicable nonpolar isostere for deoxyadenosine. *J. Org. Chem.* 63:9652-9656.
- Guo, Z., Liu, Q., and Smith, L.M. 1997. Enhanced discrimination of single nucleotide polymorphisms by artificial mismatch hybridization. *Nat. Biotechnol.* 15:331-335.
- Habener, J.F., Vo, C.D., Le, D.B., Gryan, G.P., Ercolani, L., and Wang, A.H.J. 1988. 5-Fluorodeoxyuridine as an alternative to the synthesis of mixed hybridization probes for the detection of specific gene sequences. *Proc. Natl. Acad. Sci. U.S.A.* 85:1735-1739.
- Hill, F., Loakes, D., and Brown, D.M. 1998. Polymerase recognition of synthetic oligodeoxyribonucleotides incorporating degenerate pyrimidine and purine bases. *Proc. Natl. Acad. Sci. U.S.A.* 95:4258-4263.
- Hoops, G.C., Zhang, P., Johnson, W.T., Paul, N., Bergstrom, D.E., and Davisson, V.J. 1997. Template directed incorporation of nucleotide mixtures usingazole-nucleobase analogs. *Nucl. Acids Res.* 25:4866-4871.
- Horlacher, J., Hottiger, M., Podust, V.N., Hubscher, U., and Benner, S.A. 1995. Recognition by viral and cellular DNA polymerases of nucleosides bearing bases with nonstandard hydrogen bonding patterns. *Proc. Natl. Acad. Sci. U.S.A.* 92:6329-6333.
- Huang, C.-Y., Bi, G., and Miller, P.S. 1996. Triplex formation by oligonucleotides containing novel deoxycytidine derivatives. *Nucl. Acids Res.* 24:2606-2613.
- Inoue, H., Imura, A., and Ohtsuka, E. 1985. Synthesis and hybridization of dodecadeoxyribonucleotides containing a fluorescent pyridopyrimidine deoxynucleoside. *Nucl. Acids Res.* 13:7119-7128.

- Johnson, W.T., Zhang, P., and Bergstrom, D.E. 1997. The synthesis and stability of oligodeoxyribonucleotides containing the deoxyadenosine mimic 1-(2'-deoxy-β-D-ribofuranosyl)imidazole-4-carboxamide. *Nucl. Acids Res.* 25:559-567.
- Kawase, Y., Iwai, S., and Ohtsuka, E. 1989. Synthesis and thermal stability of dodecadeoxyribonucleotides containing deoxyinosine pairing with four major bases. *Chem. Pharmacol. Bull.* 37:599-601.
- Klewer, D., Zhang, P., Bergstrom, D.E., Davisson, V.J., and Liwang, A.C. 2001. Conformations of nucleoside analog 1-(2'-deoxy-β-D-ribofuranosyl)-1,2,4-triazole-3-carboxamide in DNA duplexes with different sequence contexts. *Biochemistry* 40:1518-1527.
- Koh, J.S. and Dervan, P.B. 1992. Design of a non-natural deoxyribonucleoside for recognition of GC base pairs by oligonucleotide-directed triple-helix formation. *J. Am. Chem. Soc.* 114:1470-1478.
- Kool, E.T. 1998. Replication of non-hydrogen bonded bases by DNA polymerases: A mechanism for steric matching. *Biopolymers* 48:3-17.
- Krawczyk, S.H., Milligan, J.F., Wadwani, S., Moulds, C., Froehler, B.C., and Matteucci, M.D. 1992. Oligonucleotide-mediated triple helix formation using an N3-protonated deoxycytidine analog exhibiting pH-independent binding within the physiological range. *Proc. Natl. Acad. Sci. U.S.A.* 89:3761-3764.
- Kutyavin, I.V., Lukhtanov, E.A., Gorn, V.V., Meyers, R.B. Jr., and Gamper, H.B. Jr. 1996. Oligonucleotides containing 2-aminoadenine and 2-thiothymine act as selectively binding complementary agents. *Biochemistry* 35:11170-11176.
- Lin, T.K.T. and Brown, D.M. 1989. Synthesis and duplex stability of oligonucleotides containing cytosine-thymine analogues. *Nucl. Acids Res.* 17:10373-10383.
- Loakes, D. and Brown, D.M. 1994. 5-Nitroindole as an universal base analogue. *Nucl. Acids Res.* 22:4039-4043.
- Loakes, D., Brown, D.M., Linde, S., and Hill, F. 1995. 3-Nitropyrrole and 5-nitroindole as universal bases in primers for DNA sequencing and PCR. *Nucl. Acids Res.* 23:2361-2366.
- Luo, J., Bergstrom, D.E., and Barany, F. 1996. Improving the fidelity of *Thermus thermophilus* DNA ligase. *Nucl. Acids Res.* 24:3071-3078.
- Luyten, I. and Herdewijn, P. 1998. Hybridization properties of base-modified oligonucleotides within the double and triple helix motif. *Eur. J. Med. Chem.* 33:515-576.
- Martin, F.H. and Castro, M.M. 1985. Base pairing involving deoxyinosine: Implications for probe design. *Nucl. Acids Res.* 13:8927-8938.
- Matray, T.J. and Kool, E.T. 1999. A specific partner for abasic damage in DNA. *Nature* 399:704-708.
- Matray, T., Gamsey, S., Pongracz, K., and Gryaznov, S. 2000. A remarkable stabilization of complexes formed by 2,6-diaminopurine oligonucleotide N3'→P5' phosphoramidates. *Nucleosides Nucleotides Nucleic Acids* 19:1553-1567.
- McMinn, D.L., Ogawa, A.K., Wu, Y., Liu, J., Schultz, P.G., and Romesberg, F.E. 1999. Efforts towards expansion of the genetic alphabet: DNA polymerase recognition of a highly stable, self pairing hydrophobic base. *J. Am. Chem. Soc.* 121:11585-11586.
- Meggers, E., Holland, P.L., Tolman, W.B., Romesberg, F.E., and Schultz, P.G. 2000. A novel copper-mediated DNA base pair. *J. Am. Chem. Soc.* 122:10714-10715.
- Millican, T.A., Mock, G.A., Chauncey, M.A., Patel, T.P., Eaton, M.A.W., Gunning, J., Cutbush, S.D., Neidle, S., and Mann, J. 1984. Synthesis and biophysical studies of short oligodeoxynucleotides with novel modifications: A possible approach to the problem of mixed base oligodeoxynucleotide synthesis. *Nucl. Acids Res.* 12:7435-7453.
- Morales, J.C. and Kool, E.T. 1999. Minor groove interactions between polymerase and DNA: More essential to replication than Watson-Crick hydrogen bonds? *J. Am. Chem. Soc.* 121:2323-2324.
- Moran, S., Ren, R.X.-F., and Kool, E.T. 1997a. A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. U.S.A.* 94:10506-10511.
- Moran, S., Ren, R.X.-F., Rumney, S.I., and Kool, E.T. 1997b. Difluorotoluene, a nonpolar isostere for thymine, codes specifically and efficiently for adenine in DNA replication. *J. Am. Chem. Soc.* 119:2056-2057.
- Nakatani, K., Sando, S., and Saito, I. 2001. Scanning of guanine-guanine mismatches in DNA by synthetic ligands using surface plasmon resonance. *Nat. Biotechnol.* 19:51-55.
- Negishi, K., Williams, D.M., Inoue, Y., Moriyama, K., Brown, D.M., and Hayatsu, H. 1997. The mechanism of mutation induction by a hydrogen bond ambivalent, bicyclic N-4-oxy-2'-deoxycytidine in *Escherichia coli*. *Nucl. Acids Res.* 25:1548-1552.
- Nguyen, N.K., Bonfils, E., Auffray, P., Costaglioli, P., Schmitt, P., Asseline, U., Durand, M., Maurizot, J.C., Dupret, D., and Thuong, N.T. 1998. The stability of duplexes involving AT and/or G(4Et)C base pairs is not dependent on their AT/G(4Et)C ratio content. Implication for DNA sequencing by hybridization. *Nucl. Acids Res.* 26:4249-4258.
- Oda, Y., Uesugi, S., Ikehara, M., Kawase, Y., and Ohtsuka, E. 1991. NMR studies for identification of dI:dG mismatch base-pairing structure in DNA. *Nucl. Acids Res.* 19:5263-5267.

- Ogawa, A.K., Wu, Y., McMinn, D.L., Liu, J., Schultz, P.G., and Romesberg, F.E. 2000. Efforts toward the expansion of the genetic alphabet: Information storage and replication with unnatural hydrophobic base pairs. *J. Am. Chem. Soc.* 122:3274-3287.
- Ohtsuka, E., Matsuki, S., Ikehara, M., Takahashi, Y., and Matsubara, K. 1985. An alternative approach to deoxyoligonucleotides as hybridization probes by insertion of deoxyinosine at ambiguous codon positions. *J. Biol. Chem.* 260:2605-2608.
- Piccirilli, J.A., Krauch, T., Moroney, S.E., and Benner, S.A. 1990. Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet. *Nature* 343:33-37.
- Pochet, S. and Marlière, P. 1996. Construction of a self-complementary nucleoside from deoxyguanosine. *C. R. Acad. Sci. (Paris)* 319:1-7.
- Prévoit-Halter, I. and Leumann, C.J. 1999. Selective recognition of a C-G base pair in the parallel DNA triple-helical binding motif. *Bioorg. Med. Chem. Lett.* 9:2657-2660.
- Rothman, J.H. and Richards, W.G. 1996. Novel Hoogsteen-like bases for configurational recognition of the T-A base pair by DNA triplex formation. *Biopolymers* 39:795-812.
- Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer-Verlag, New York.
- Schweitzer, B.A. and Kool, E. 1995. Hydrophobic, non-hydrogen-bonding bases and base pairs DNA. *J. Am. Chem. Soc.* 117:1864-1872.
- Seela, F. and Debelak, H. 2000. The N⁸-(2'-deoxyribofuranoside) of 8-aza-7-deazaadenine: A universal nucleoside forming specific hydrogen bonds with the four canonical DNA constituents. *Nucl. Acids Res.* 28:3224-3232.
- Seela, F. and Kaiser, K. 1986. Phosphoramidites of base-modified 2'-deoxyinosine isosteres and solid-phase synthesis of d(GCI*CGC) oligomers containing an ambiguous base. *Nucl. Acids Res.* 14:1825-1844.
- Switzer, C.Y., Moroney, S.E., and Benner, S.A. 1993. Enzymatic recognition of the base pair between isocytidine and isoguanosine. *Biochemistry* 32:10489-10496.
- Tanaka, K. and Shionoya, M. 1999. Synthesis of a novel nucleoside for alternative DNA base pairing through metal complexation. *J. Org. Chem.* 64:5002-5003.
- Ueno, Y., Mikawa, M., and Matsuda, A. 1998. Synthesis and properties of oligodeoxynucleotides containing 5-[N-[2-[N,N-bis(2-aminoethyl)amino]ethyl]carbamoyl]-2'-deoxyuridine and 5-[N-[3-[N,N-Bis(3-aminopropyl)amino]propyl]carbamoyl]-2'-deoxyuridine. *Bioconjugate Chem.* 9:33-39.
- Uesugi, S., Oda, Y., Ikehara, M., Kawase, Y., and Ohtsuka, E. 1987. Identification of I-A mismatch base-pairing structure in DNA. *J. Biol. Chem.* 262:6965-6968.
- Voegel, J.J. and Benner, S.A. 1994. Nonstandard hydrogen bonding in duplex oligonucleotides. The base pair between an acceptor-donor-donor pyrimidine analog and donor-acceptor-acceptor analog. *J. Am. Chem. Soc.* 116:6929-6930.
- Wagner, R.W., Matteucci, M.D., Lewis, J.G., Gutierrez, A.J., Moulds, C., and Froehler, B.C. 1993. Antisense gene inhibition by oligonucleotides containing C-5 propyne pyrimidines. *Science* 260:1510-1513.
- Wu, Y., Ogawa, A.K., Berger, M., McMinn, D.L., Schultz, P.G., and Romesberg, F.E. 2000. Efforts toward expansion of the genetic alphabet: Optimization of interbase hydrophobic interactions. *J. Am. Chem. Soc.* 122:7621-7632.
- Yu, H., Eritja, R., Bloom, L.B., and Goodman, M.F. 1993. Ionization of bromouracil and fluorouracil stimulates base mispairing frequencies with guanine. *J. Biol. Chem.* 268:15935-15943.
- Zhang, P., Johnson, W.T., Klewer, D., Paul, N., Hoops, G., Davison, V.J., and Bergstrom, D.E. 1998. Exploratory studies on azole carboxamides as nucleobase analogs: Thermal denaturation studies on oligodeoxyribonucleotide duplexes containing pyrrole-3-carboxamide. *Nucl. Acids Res.* 26:2208-2215.

Contributed by Donald E. Bergstrom
Purdue University
West Lafayette, Indiana