# Triplex Formation at Single-Stranded Nucleic Acid Target Sites of Unrestricted Sequence by Two Added Strands of Oligonucleotides: A Proposed Model

## Tina L. Trapane* and Paul O. P. Ts'o*

*Contribution from the Department of Biochemistry, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, Maryland 21205*

**Abstract:** By using the standard purine nucleosides, guanosine and adenosine, and the pyrimidine C-nucleosides, pseudoisocytidine and pseudouridine, as complements on a probe strand, it is possible to construct a regular Watson—Crick helix with a single-stranded target sequence having *any arrangement* of the four naturally-occurring bases found in nucleic acids. The major groove of this helix will have a unique configuration of hydrogen-bonding sites on the probe strand for each of these four base pairs. By using this duplex as a framework, an ensemble of recognition patterns composed of base triads may be constructed. In these patterns, either a homopyrimidine or homopurine third strand binds in the major groove of the duplex formed by the target and probe strands. Ten distinct geometries, or motifs, are shown, each one consisting of four isomorphic base triads built upon recognition of C, G, A, or U(T) residues in the target strand. Four motifs contain pyrimidines as residues on the third strand which base pair to the second strand through specific hydrogen-bonding interactions, four motifs involve purines which base pair to the second strand through donor—acceptor sites located on their six-membered ring, and two motifs utilize purines binding to the second strand at sites located on both their five- and six-membered rings. For base triads or for base-pairing interactions which involve the common bases found in nucleic acids, most of the hydrogen-bonding patterns have been previously recognized. In order to maintain specific hydrogen bonding and to construct isomorphous triads, the use of several nonstandard bases is proposed. A subset of the base triads may also be used to design oligonucleotides which may bind as a third strand to naturally-occurring homopyrimidine—homopurine double-stranded target sites such as those found in DNA. In addition, another set of four bases which have Watson—Crick complementarity to the target-strand bases and which provide alternative patterns of donor—acceptor pairs for third-strand interactions can be proposed for use on the second strand. With the palette of eight second-strand bases, four possible permutations of (target strand)—(second strand) interactions are shown. For each of the four residues on the target strand, any single permutation presents unique hydrogen-bonding patterns for third-strand binding which may occur according to one of the ten triad motifs.

## Introduction

The formation of triple-stranded helices, or triplexes, by the naturally-occurring nucleic acids is a well-known phenomenon. Triplex formation, in general, has been limited in its scope by several important structural requirements. Shown in Figure 1 are the standard Watson—Crick (W·C) base pairs, C·G and T·A.[1] Recognition of any sequence in one of the strands is achieved by formation of specific sets of hydrogen bonds (H-bonds) with complementary bases in the other strand. The bases in each pair generally reside in a plane which is approximately perpendicular to the double-helix axis. As shown, N-glycosyl linkages between sugars and bases are on opposite sides of the lower edge of the base-pair plane. This arrangement places backbones of the two strands closest to each other at this edge of the base pairs giving rise to major and minor grooves in a double-stranded helix. Because of steric and electrostatic considerations, interaction of a third nucleic acid strand with a double-stranded target site occurs almost exclusively within the major groove with certain sequence restrictions. Pyrimidine bases provide only one H-bonding site at C4 in the major groove of a W·C-type duplex, whereas purine bases afford two sites at C6 and N7 (shown by arrows in Figure 1). In addition, the methyl group on thymidine residues in DNA double-helices can

sterically block H-bonding interactions within the major groove. The donor—acceptor patterns in the major groove for A and G are distinct from one another, therefore, sequence-specific recognition through formation of pairs of H-bonding interactions may occur only at purine bases. Regular placement of a third strand in the major groove requires that all purine complements of the W·C base pairs reside on the same strand. Thus, triplex formation (including addition of two probe strands to a single-stranded target) has been limited in general to homopyrimidine·homopurine (pyr·pur) duplex sites. Here we propose a model, based upon isomorphic base-pairing geometries and novel H-bonding recognition schemes, in which various triple-stranded helices may be generated at single-stranded target sites of *unrestricted* sequence by addition of two oligonucleotide probes.

## Model

**Nomenclature.** The atom numbering system for standard purine and pyrimidine bases used in the proposed triad motifs is the standard IUPAC-IUB nomenclature[2] (Figure 2a,b). For nonstandard nucleosides, atom numbering follows that for standard bases, with covalent attachment to the sugar (R) at N9 for purines and at N1 for pyrimidines. Pyrimidine bases with C-glycosyl linkages (Figure 2c), where attachment to the

---

* Address correspondence to either author.
® Abstract published in *Advance ACS Abstracts,* October 15, 1994.
(1) Watson, J. D.; Crick, F. H. C. *Nature* **1953,** *171,* 737—738.

(2) IUPAC-IUB Joint Commission on Biochemical Nomenclature. *Eur. J. Biochem.* **1983,** *131,* 9—15.

10438  *J. Am. Chem. Soc., Vol. 116, No. 23, 1994*
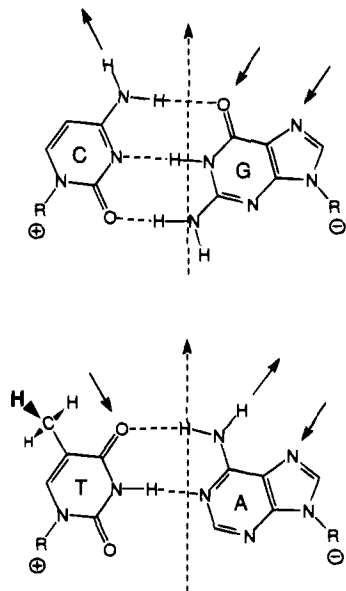
*Trapane and Ts'o*



**Figure 1.** Base-pairing schemes C·G (upper) and T·A (lower) having Watson—Crick H-bonding patterns. H-Bonds are indicated by dashed lines. Backbone orientations (⊕ and ⊖) of the two antiparallel strands are indicated at the glycosyl bond where sugar moieties are represented by R. Pseudodyad axes are shown as dashed arrows, and H-bond donor—acceptor sites in the major groove are indicated by solid arrows.
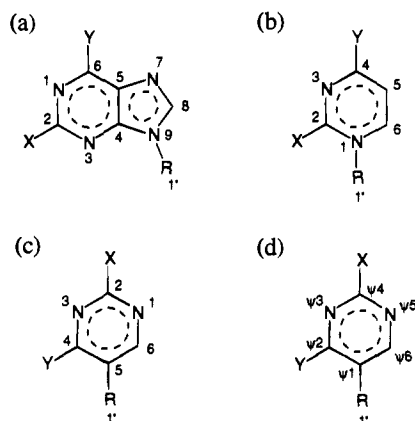


**Figure 2.** Atom numbering for ring positions on heterocyclic bases. Standard (IUPAC-IUB) system for purines (a) and pyrimidines (b). Standard (c) and proposed (d) system for pyrimidine C-nucleoside bases. Only nitrogen atom positions for parent ring structures are shown, all other atoms in the ring are carbon. For heterocyclic bases used in the model, the pendant groups X and Y can be oxygen (keto function), amino (NH₂), or proton (H). In the schemes, only protons that can participate in H-bonding are explicitly shown.

sugar is at C5, have a numbering system that places the pertinent H-bonding substituent at C4 (Y) closest to the minor groove side of a W·C base-pairing interaction rather than in the major groove as for a standard pyrimidine. This relationship is reversed for the H-bonding substituent at C2 (X). In order to avoid confusion due to these differences in numbering, an alternate system (Figure 2d) for pyrimidine C-nucleosides (pseudopyrimidines) will be used in development of the model. This adapted numbering system allows the position of H-bonding sites on pseudopyrimidines to be geometrically equivalent to standard bases in relationship to the glycosyl bond which links base to sugar. Therefore, pseudouridine (see Figure 4) will have H-bonding acceptors at $\psi$O2 and $\psi$O4 and donors at $\psi$N3 and $\psi$N5. Note that this pattern is exactly the same as for the standard uridine nucleoside except for an additional donor site at $\psi$N5.

The use of several other base analogues is also proposed in the model. Some analogues have the same pendant chemical groups as one of the common bases, but the positions of these groups have been exchanged. Specifically, the two positions on pyrimidine bases are C2 and C4 (or $\psi$C2 and $\psi$C4), and the two positions on purine bases are C2 and C6. Interchange of functional groups at these positions results in bases that are H-bonding isomers of the normal base. Therefore, the trivial name given to such analogues will be the same as that for the usual base appended with the prefix "iso" (abbreviations are given the italicized prefix "*i*"). For example, the nucleoside of 2-ketopurine has a carbonyl acceptor at C2 and a proton at C6. The keto and proton groups in this base are interchanged from those normally found in the inosine (I) nucleoside, therefore this analog is named isoinosine (*i*I). The pseudopyrimidine bases are a notable exception to the H-bonding isomer terminology. By virtue of the glycosyl linkage at $\psi$C1 rather than at $\psi$N5 (Figure 2d), these bases become H-bonding isomers of their parent nucleosides in a geometrical sense. Specifically, pseudocytidine ($\psi$C) has amino and keto H-bond donor—acceptor groups at $\psi$C2 and $\psi$C4, respectively. These functions are interchanged geometrically from those found on the standard cytosine base; therefore, the H-bonding isomer of pseudocytidine is pseudoisocytidine ($\psi i$C), which will have the *same* donor—acceptor pattern at $\psi$C2 and $\psi$C4 as the cytosine base (see Figure 6a for schemes of these examples). Other base analogues result from modifications within the heterocyclic ring system itself. These modifications entail substitution of carbon for nitrogen (or vice versa) at a particular position. Ring systems which have been thus modified have their own particular atom numbering and nomenclature; however, the overall geometries remain isosteric to those of standard bases. In this proposal, the atom numbering system adopted for use with analogues of this type will be according to the isosteric position on the standard pyrimidine or purine ring system. Trivial names for these analogues will be the same as for the usual base with a prefix indicating the position (or positions) at which a nitrogen atom has been deleted ("deaza") and substituted by carbon, or where a nitrogen atom has been introduced ("aza") in substitution for carbon. Single-letter abbreviations will indicate ring modifications in a left superscript as "*n*C" for deaza or "*n*N" for aza, representing carbon at position *n* or nitrogen at position *n*, respectively. For example, the heterocyclic base analog of adenosine having nitrogen at N9 substituted by carbon is given the name 9-deazaadenosine ($^9$C A). This purine C-nucleoside analog will have a C-glycosyl bond between C9 of the heterocyclic base and C1' of the sugar.

Finally, designation of the three interacting nucleic acid strands and nucleoside residues contained on each must be made unambiguous. The nucleic acid strand containing the sequence of natural bases which is the intended binding site of the added oligonucleotides will be termed the "target strand". The oligomer designed to bind as a Watson—Crick complement to the target strand will be alternately termed the "second strand" or "W·C-probe strand". The oligomer designed to bind in the major groove of the duplex formed by the target and second strands will be termed the "third strand". A "base pair" is the (approximate) coplanar association of two bases through formation of at least *two* adjacent H-bonds. A "base triad" is formed when one member of a base pair has an additional H-bonded association with another base at an alternate interface. Therefore, a base triad consists of three bases having two, approximately planar, base-pairing interactions among them. In the text, target-strand bases will be indicated first (bold type) followed by the complementary base in the second strand
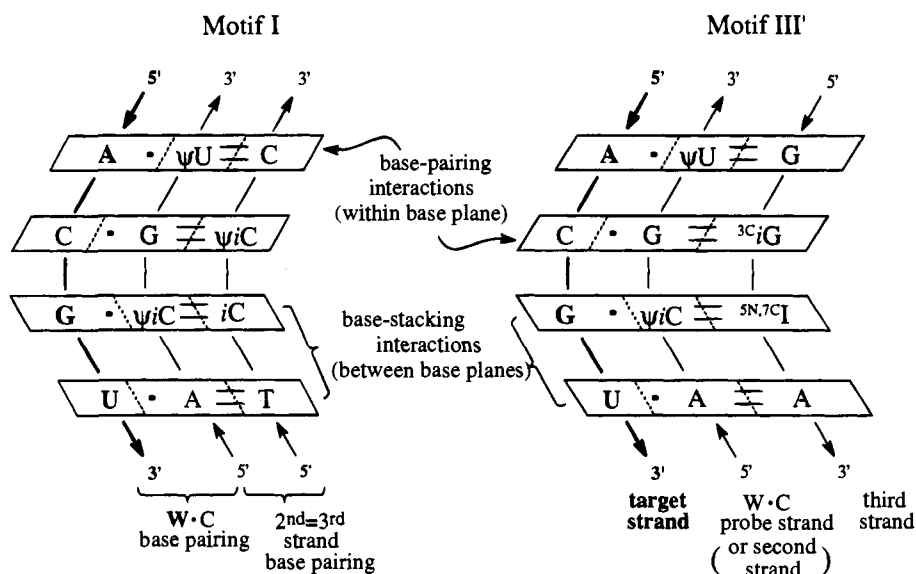
**Figure 3.** Schematic representations of two proposed triple helices. In the schemes, each horizontal parallelogram represents a base triad. Abbreviations are in bold type for target-strand bases and in regular type for probe-strand bases. Each W·C base-pairing interaction is represented by a bullet (·) and each pairing interaction between second- and third-strand bases is represented by the symbol =. Dotted lines within parallelograms delineate the approximate area occupied by purine and pyrimidine bases. Residues in each strand are connected by internucleoside linkages which are represented by generally vertical bars. The directions (5′→3′) of these "backbones" are shown by arrows at the upper and lower edges of triad stacks. These schemes diagram triplexes which might be formed at a single-stranded target site having the sequence 5′-ACGU-3′. The second strand in each scheme has W·C complementarity to the target strand and unique pairs of H-bond donor−acceptor sites in the major groove of the resulting duplex for third-strand binding. On the left, a homopyrimidine third strand binds sequence specifically to the second strand in a parallel orientation, while on the right, a homopurine third strand binds sequence specifically to the second strand in an antiparallel orientation.

separated by a bullet (·), a symbol representing W·C base pairing (e.g., U·A). Base triads are designated by the motif number (Roman capital) followed by a colon, and then the W·C base pair followed by the third-strand base separated by the symbol =, which represents the pair of specific H-bonds between bases on the second and third strands (e.g. I:U·A=T). *This symbol does not represent equivalence.* Shown in Figure 3 are schematics of two types of triple helices proposed by the model which illustrate these definitions. Exact H-bonding patterns and triad geometries for these triplexes will be given in the section on triad motifs.

**W·C Base Pairing between Target and Probe Strands.** The four W·C-type base pairs which form the foundation for the entire set of base-triad motifs are shown in Figure 4. Recognition of pyrimidine bases in the target strand by purine bases in the second strand gives the standard base pairs C·G and U·A. Recognition of purine bases in the target strand, however, is made by pyrimidine bases on the W·C-probe strand which have a C-nucleoside glycosyl bond. This results in the base pairs G·ψiC and A·ψU. Note that the H-bonding patterns in these "pseudo W·C" base pairs are exactly as for the standard G·C and A·U pairs. The advantage of using pseudopyrimidine bases is that now there exists a pair of H-bonding sites in the major groove of the double helix at the pyrimidine base analog on the second strand. An additional site is provided at ψN5 of each pyrimidine C-nucleoside base at a position approximately isomorphous to N7 of purines in the C·G and U·A pairs. Examination of the four pairing schemes reveals a unique pattern of donor−acceptor sites on each base of the W·C-probe strand. Specifically, C·G has two acceptors, G·ψiC has two donors, U·A has a donor and an acceptor (as viewed from the major groove), and A·ψU has an acceptor and a donor.

**Structural Requirements for Construction of Triad Motifs.** Using the unique patterns of H-bonding sites on the second strand it is now possible to develop a series of base-triad motifs. Recognition of H-bonding patterns for bases on the second strand is made by specific base-pairing interactions with either

pyrimidine or purine nucleosides on a third strand. Each motif consists of four triads built upon the four W·C base pairs shown in Figure 4 and is constructed according to the following guidelines.

(1) All nucleosides on the third strand are assumed to have the *anti* configuration of the sugar−base glycosyl linkage. This places furanose rings of the sugars away from bulky substituents on the bases. In addition, when all nucleosides are specified as having the same glycosyl configuration, the relative orientations (parallel *vs* antiparallel) of any two backbones in the helix may be established by comparing the relative orientations of their nucleosides.[3] The sequence of residues for the three strands in any helix formed according to these motifs may then be ascertained. Assignment of relative strand orientation is achieved by superimposing residues on separate strands along their sugar→base bonds without lifting the structures from the plane of the paper. Subsequently, the directions of the N(C)9−C4 bond of purines and/or the N(C)1−C2 bond of pyrimidines are compared. (These bonds point toward the W·C binding "face" of the bases.) The selected backbones have the same orientations (parallel strands) if these bonds face in the same direction and have opposite orientations (antiparallel strands) if these bonds face in opposite directions. Assignment of third-strand orientation, therefore, can be made in reference to either strand in the W·C duplex.

(2) A pair of H-bonding interactions must be made at adjacent donor−acceptor sites between each base on the second strand and the complementary base on the third strand. Pyrimidine bases possess two sets of adjacent sites at C4−N3 and N3−C2, whereas purine bases have three at C6−N1, N1−C2 and C6−N7. The motifs may be separated into three general classes

(3) We note that other rules have been employed to specify relative strand orientation in nucleic acid interactions; for examples, see: (a) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; pp 119−122. (b) Westhof, E. *Nature* **1992**, *358*, 459−460. However, because of the use of pyrimidine C-nucleosides on the second strand, such rules do not apply to this model.
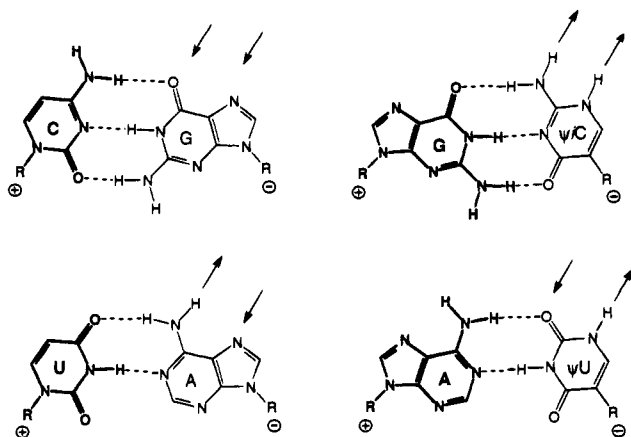
**Figure 4.** Proposed recognition schemes for the four naturally-occurring residues, **C**, **G**, **U** or **A**, that may occur at a single-stranded nucleic acid target site. G, $\psi$iC, A, and $\psi$U residues, respectively, on the second strand form base pairs having Watson—Crick complementarity with each of the four target bases. Target- and second-strand bases in each pair are on the left and right, respectively, with sugar—base (glycosyl) linkages and minor grooves oriented downward. Strand orientation is indicated at C1', where R⊕ and R⊖ indicate backbone directions of opposite polarity. Arrows indicate pairs of H-bonding sites on the second-strand bases which are available in the major groove for sequence-specific interactions with a third strand.

according to geometries of second=third-strand base-pairing interactions. Class A motifs include triplexes that are formed using homopyrimidine third strands. Class B and C motifs include triplexes that are formed using homopurine third strands. Purine bases have a six-membered ring that can be analogous to the pyrimidine bases in terms of H-bonding geometries; therefore, Class B motifs include triplexes formed by homopurine third strands which base pair via substituents found only on their six-membered ring. Class C motifs include triplexes formed by homopurine third strands which H-bond to the bases on the second strand via donor—acceptor sites found on both their five- and six-membered rings.

(3) Bases in the third strand must exhibit tautomeric forms existing at neutral or physiological pH which satisfy the required H-bonding schemes for base pairing. Some preferred tautomeric forms of the common bases, especially purines and their analogues, do not fulfill this requirement. In order to obtain the desired donor—acceptor patterns, several nonstandard purine analogues having unusual heterocyclic ring systems are proposed. Although a majority of these nucleoside derivatives have not actually been tested experimentally for their ability to participate in triplex formation, the rationale for choosing each one of the proposed analogues is addressed with respect to their intended role.

(4) The triads within each motif should be geometrically isomorphous. Specifically, this requirement means that the overall dimensions and configurational relationships of the three bases within the plane of any triad should be structurally comparable to those of the other triads within the motif. This structural requirement provides a uniform conformation of the base triads along the triple helix so that stacking interactions may be facilitated.[4] Since the target strand may contain any sequence of bases, the W·C pairing interactions in the triads will have the conventional pseudodyad symmetry of such base pairs. Superposition of W·C pyr·pur and pur·pyr base pairs (Figure 5) allows comparison of positions of H-bond sites on
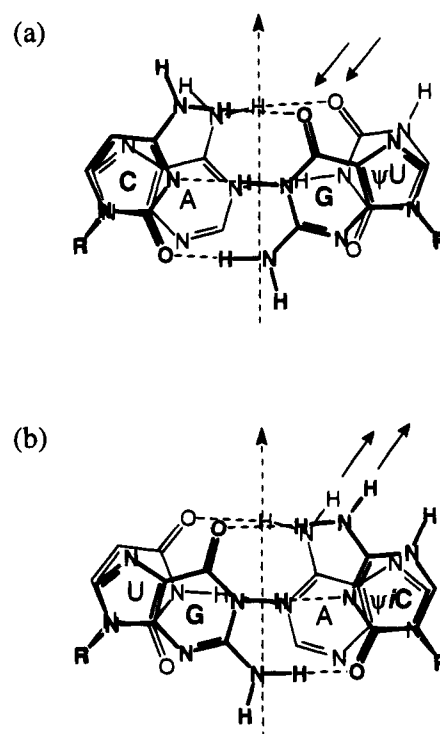
(4) (a) DeVoe, H.; Tinoco, I. *J. Mol. Biol.* **1962**, *4*, 500—517. (b) Ts'o, P. O. P. *Basic Principles in Nucleic Acid Chemistry;* Academic Press: New York, 1974; Vol. I, pp 453—584.

(a)



(b)



**Figure 5.** Superposition of (a) the base pairs C·G and A·$\psi$U and (b) the base pairs G·$\psi$iC and U·A. Pseudodyad axes are represented by dashed arrows. Solid arrows show displacement toward the center of the major groove for H-bonding substituents at C6 on purine bases in comparison to H-bonding substituents on $\psi$C4 of pyrimidine C-nucleoside bases.

the W·C-probe strand in the major groove. Indeed, positions of $\psi$N5 for $\psi$iC and $\psi$U are geometrically similar to N7 of G and A. However, exocyclic keto and amino groups at $\psi$C4 of pyrimidine C-nucleoside bases are displaced further from the pseudodyad axis than these same groups at C6 on the purine bases. As a result of this displacement, third-strand residues which are involved in base pairing to purines on the W·C-probe strand may be oriented more toward the W·C dyad axis than those pairing to the pseudopyrimidines. Such accommodation might be necessary in order to optimize H-bonding geometries. Hydrogen bonds are considered to be "soft" bonding interactions as they can be formed over a range of angular and distance constraints. Therefore, the eventual configuration of the three bases in any triad will be determined within the context of the helix. More importantly, due to differences in shape of pyrimidine and purine bases, the third strand should be limited to one type only. This requirement is important to the formation of triple-stranded helices in order to ensure regular positioning of the third-strand backbone and to optimize stacking interactions between nearest-neighbor triads. If the third strand were to contain both pyrimidine and purine residues, there would be a loss of uniformity in the helical structure for two reasons. First, the base-stacking interactions between neighboring triads would not be regular, as the surface area of triads containing one or the other type of base in the third strand will not be equivalent. Second, positioning of the glycosyl bond and, therefore, the sugar—phosphate backbone will be irregular for nucleoside residues having differently sized bases. These distortions are likely to cause helix instability, as a triple helix should have a regular and presumably stable configuration along the helical axis for all three strands.

**Triad Motifs.** Schemes of the proposed base-triad motifs for Classes A, B, and C are shown in Figures 6, 7, and 8, respectively. These schemes demonstrate H-bonding patterns

**Table 1.** Base Triad Motifs. [Abbreviations are as defined in the nomenclature section. Directions of H-bond donors (X−H → Y) are shown for second−third-strand pairing interactions. Backbone orientations (⊕ or ⊖) of the three strands are given in parentheses below each motif number. Pairing interactions between the second and third strands which involve standard bases are noted.]

**Class A - *Pyrimidine third strand***

| | | | | |
|---|---|---|---|---|
| **I**<br>(⊕ · ⊖=⊖) | C·G⁻ψiC | G·ψiC⁻iC | (T)U·A⁻T(U)<br>(a) | A·ψU⁻C |
| **I′**<br>(⊕ · ⊖=⊕) | C·G⁻ψiC | G·ψiC⁻iC | (T)U·A⁻C<br>(b) | A·ψU⁻T(U) |
| **II**<br>(⊕ · ⊖=⊖) | C·G⁻ψC | G·ψiC⁻C | (T)U·A⁻iC | A·ψU⁻T(U) |
| **II′**<br>(⊕ · ⊖=⊕) | C·G⁻ψC | G·ψiC⁻C | (T)U·A⁻T(U)<br>(c) | A·ψU⁻iC |

**Class B - *Purine third strand - recognition at C6-N1-C2 face***

| | | | | |
|---|---|---|---|---|
| **III**<br>(⊕ · ⊖=⊖) | C·G⁻³CiG | G·ψiC⁻⁵N,⁷CI | (T)U·A⁻G(I)<br>(d) | A·ψU⁻A |
| **III′**<br>(⊕ · ⊖=⊕) | C·G⁻³CiG | G·ψiC⁻⁵N,⁷CI | (T)U·A⁻A<br>(e) | A·ψU⁻G(I) |
| **IV**<br>(⊕ · ⊖=⊖) | C·G⁻G<br>(f) | G·ψiC⁻⁵N,⁷CiI | (T)U·A⁻iA | A·ψU⁻³,⁷CiI |
| **IV′**<br>(⊕ · ⊖=⊕) | C·G⁻G<br>(g) | G·ψiC⁻⁵N,⁷CiI | (T)U·A⁻³,⁷CiI | A·ψU⁻iA |

**Class C - *Purine third strand - recognition at C6-N7 face***

| | | | | |
|---|---|---|---|---|
| **V**<br>(⊕ · ⊖=⊖) | C·G⁻⁹CA | G·ψiC⁻I | (T)U·A⁻A<br>(h) | A·ψU⁻⁹CI |
| **V′**<br>(⊕ · ⊖=⊕) | C·G⁻⁹CA | G·ψiC⁻I | (T)U·A⁻⁹CI | A·ψU⁻A |

| | |
|---|---|
| (a) Donohue #25, Hoogsteen | (e) Donohue #3, found in tRNA |
| (b) Donohue #27 | (f) Donohue #10, found in G-tetrad |
| (c) Donohue #26, reverse Hoogsteen | (g) Donohue #11, found in tRNA |
| (d) Donohue #28, A(*syn*)·G(*anti*) | (h) Donohue #1, found in poly(A⁺) and in tRNA |

and geometries of base-pairing interactions required by this model for formation of triple-helical complexes at single-stranded target sites without restriction of sequence. In these schemes, several non-W·C base-pairing symmetry relationships may be recognized. A discussion of symmetry elements in the model is contained in the supplementary material. A listing of triads in each motif is given in Table 1, wherein base-pairing interactions involving standard bases and their occurrence in natural nucleic acid structures are noted.

Class A motifs (Table 1 and Figure 6a−d) utilize pyrimidine third-strand residues designed for specific binding to bases on the W·C-probe strand. Motif I (Figure 6a) is patterned after the most well-known base triads, U·A=U and T·A=T,[5] which are analogous to I:U·A=T. Here the third-strand base, thymidine, accepts and donates H-bonds at O4 and H3 to the second-strand base, adenosine, at H6 and N7, respectively. This type of base pairing was first identified by Hoogsteen in crystals of 1-methylthymine and 9-methyladenine.[6] The third strand in

triple helices containing only this triad has been shown to be parallel to the second strand.[7] The remaining three triads in this motif are geometrically patterned after the I:U·A=T triad and are constructed by identifying pyrimidine bases which possess the requisite H-bond donor−acceptor pairs at C4 and N3 or at isosteric positions in pyrimidine C-nucleosides. Recognition of the W·C-probe strand at G requires two H-bonding donors on the third-strand base as provided by ψiC and shown in the triad I:C·G=ψiC.[8] Note that the tautomeric form of ψiC is now different from that required for W·C pairing by this base. Recognition of the ψiC base on the W·C-probe strand requires two H-bond acceptors on a third-strand base. The pyrimidine base analog, isocytosine (iC),[9] has acceptor sites in the correct positions as shown in the triad I:G·ψiC=iC. Completing this first motif, the triad I:A·ψU=C involves recognition of ψU in the W·C-probe strand by C in the third strand which provides a donor and acceptor (again, as viewed

(5) Felsenfeld, G.; Davies, D. R.; Rich, A. *J. Am. Chem. Soc.* **1957**, *79*, 2023−2024.

(6) Hoogsteen, K. *Acta Crystallogr.* **1963**, *16*, 907−916.

(7) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645−650.

(8) Ono, A.; Ts'o, P. O. P.; Kan, L.-S. *J. Org. Chem.* **1992**, *57*, 3225−3230.

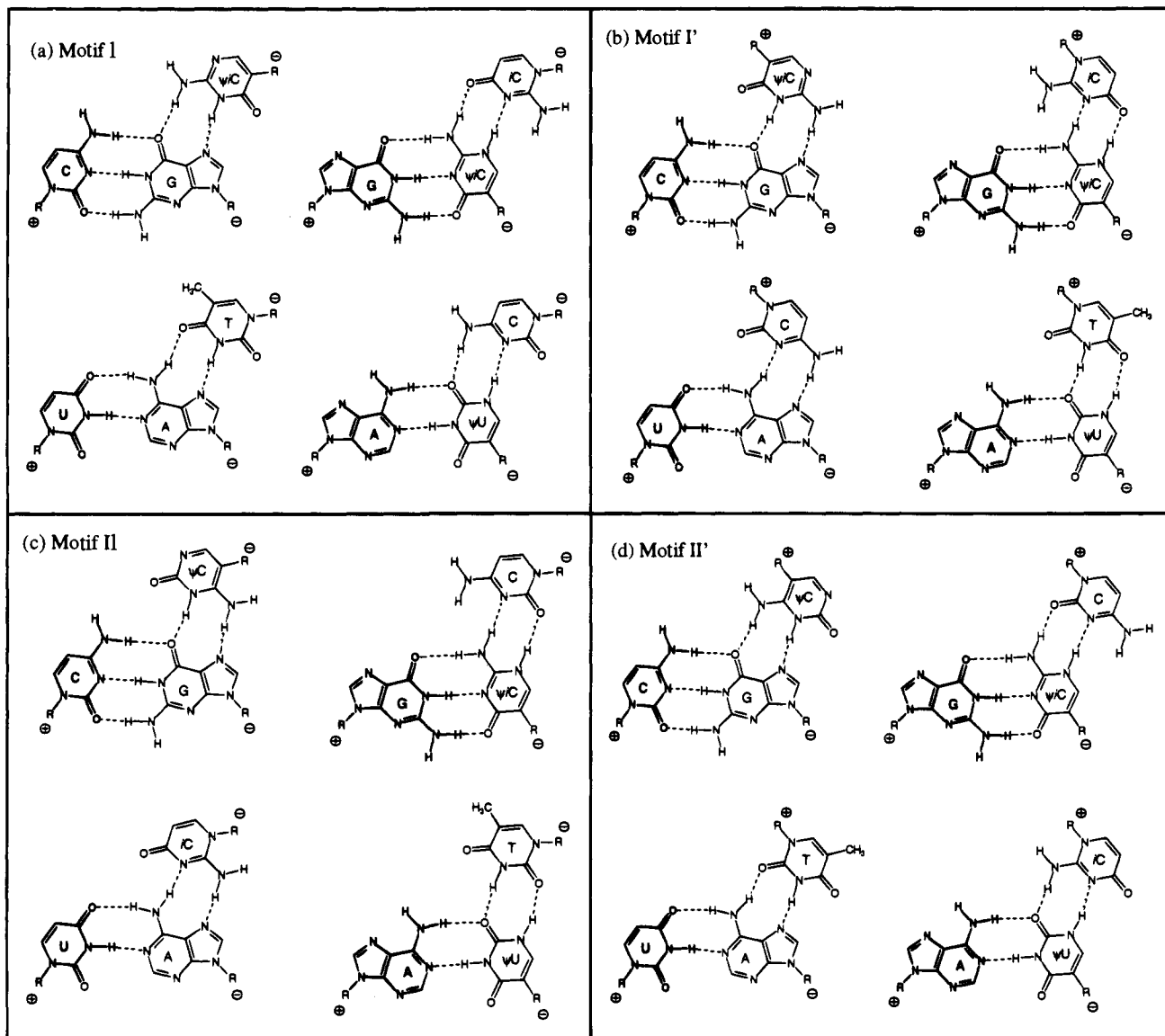(9) Switzer, C.; Moroney, S. E.; Benner, S. A. *J. Am. Chem. Soc.* **1989**, *111*, 8322−8323.

**Figure 6.** Class A motifs for pyrimidine residues on the third strand base-pairing to residues on the second strand. Each base triad is composed of a target-strand base (bold face, lower left), a W•C-probe-strand, or second-strand, base (lower right) and a third-strand base (upper right). Strand orientations (⊕ or ⊖) are indicated at the glycosyl linkages to each sugar C1' (R).

from the major groove of the W•C helix). Triple-stranded helices containing both I:T•A=T and I:A•$\psi$U=C base triads have now been demonstrated experimentally.[10]

As shown in Figure 6b, Motif I' is related to Motif I by inversion of the orientation of the third-strand backbone. Similar to Motif I, recognition of bases on the W•C-probe strand is made by donor—acceptor sites at C4 and N3 of pyrimidine bases on the third strand. Therefore, base pairings involving two donors or two acceptors (I':C•G=$\psi$iC or I':G•$\psi$iC=iC) can be constructed by inverting the orientation of the third-strand base (thereby the orientation of the backbone) for Motif I and repositioning in the base plane to make the desired H-bond contacts. Now the substituent at N3 (or $\psi$N3) of the third-strand base interacts with the site on the W•C-probe strand which is closest to the center of the major groove, and the substituent at C4 (or $\psi$C4) interacts at the site closest to the glycosyl bond of the second-strand base. Such an inversion of bases in Motif I for the U•A=T or A•$\psi$U=C triads will result in mispairing of H-bond donor—acceptor pairs. However, an interchange of

third-strand bases allows for correct H-bond contacts to be made, resulting in the triads I':U•A=C and I':A•$\psi$U=T.

Construction of triads for Motif II (Figure 6c) involves recognition of bases on the W•C-probe strand by H-bonding functions at N3 and C2 of pyrimidine bases on the third strand. As for Motif I, the third strand has the same orientation as the W•C-probe strand. A tautomer of the pyrimidine C-nucleoside base analog, pseudocytidine ($\psi$C),[11] has the correct arrangement of H-bond donors for recognition of the W•C-probe strand at G residues, resulting in the base triad II:C•G=$\psi$C. With two acceptor sites at N3 and C2, cytidine may base pair with $\psi$iC residues in the second strand to form the triad, II:G•$\psi$iC=C. The H-bonding isomer of cytidine, isocytidine, now provides a donor—acceptor pair for recognition of A residues in the W•C-probe strand to form the triad II:U•A=iC. The triad II:A•$\psi$U=T (U) results from thymidine (or uridine) residues on the third strand forming specific H-bonds to $\psi$U of the W•C-probe strand within the geometric context of the motif. Motif II' (Figure 6d) is related to Motif II by similar rules interconverting Motifs

(10) Trapane, T. L.; Christopherson, M. S.; Roby, C. D.; Ts'o, P. O. P.; Wang, D. *J. Am. Chem. Soc.* **1994**, *116*, 8412—8413.

(11) Pankiewicz, K. W.; Hirota, K.; Matsuda, A.; Watanabe, K. A. *Carbohydr. Res.* **1984**, *127*, 227—233.
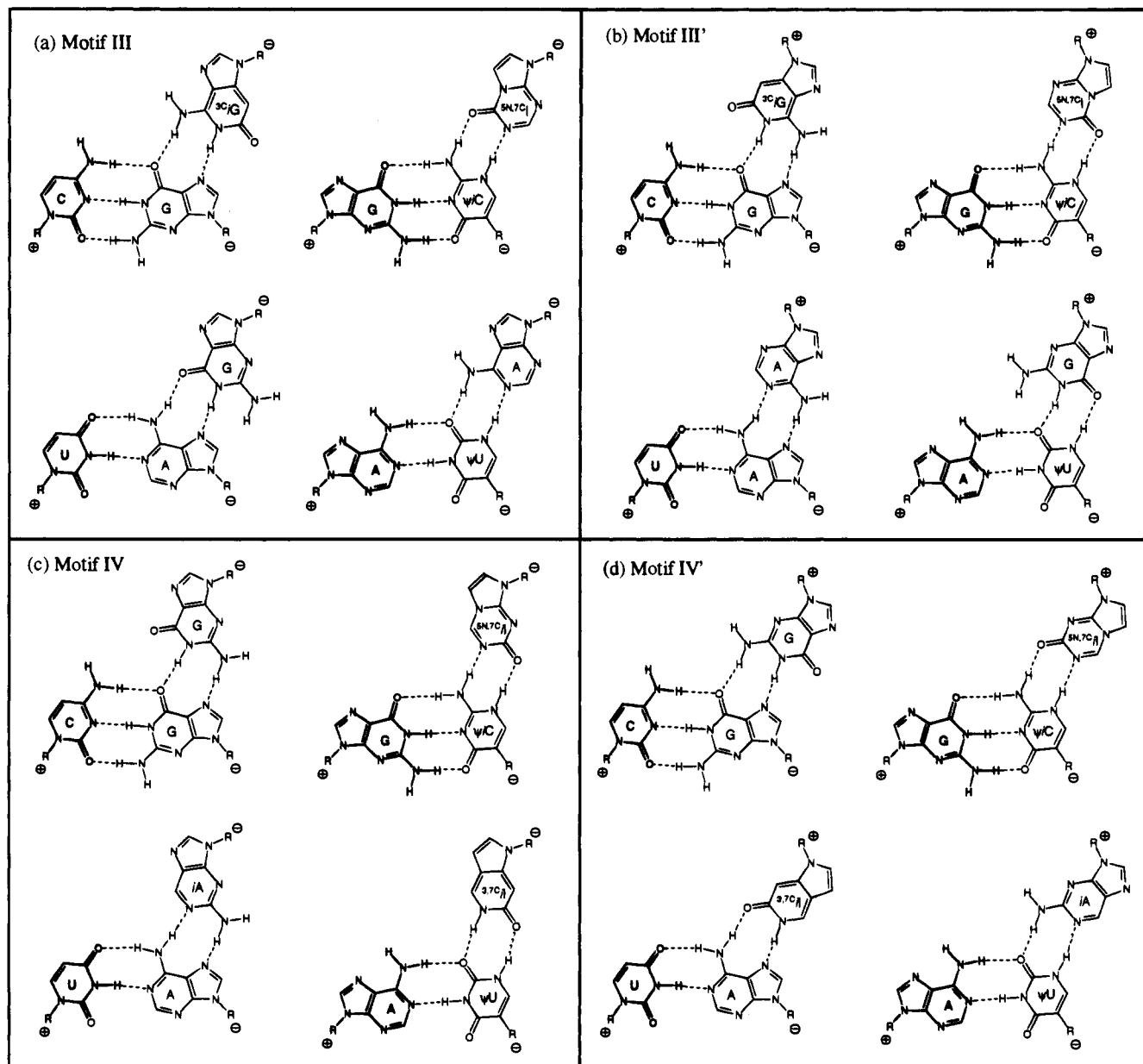
**Figure 7.** Class B motifs for purine residues on the third strand base-pairing via H-bond donor—acceptor sites on their six-membered-ring face to residues on the second strand. The representation is analogous to Figure 6.

I and I'. It should be noted that triad II':U·A=T involves an A=T H-bonding scheme of a type also found in crystals of thymine and adenine known as reverse Hoogsteen base pairing.[6]

Class B triad motifs (Table 1 and Figure 7a—d) are constructed using purine bases on the third strand in a manner analogous to development of Class A motifs. In this class, only the six-membered (or pyrimidine) ring is involved in base pairing to residues on the W·C-probe strand. Motifs III (Figure 7a) and III' (Figure 7b) utilize H-bond donor—acceptor sites at C6 and N1 of bases on the third strand in parallel and antiparallel orientations, respectively, to bases on the second strand. The nucleoside analog isoguanosine may be considered as a logical choice of a residue having two H-bond donors at these positions for recognition of G residues in the W·C-probe strand. However, this derivative is quite difficult to synthesize.[12] In addition, competing tautomeric forms place the exchangeable proton at O2 or N3 instead of at N1 where it is required in the H-bonding scheme proposed here.[9] An analog having the desired H-bond

donors at C6-N1 might be 3-deazaisoguanosine ($^{3C}iG$) as shown in triads III and III':C·G=$^{3C}iG$. This purine ring derivative belongs in the family of imidazo[4,5—c]pyridines and may be synthesized by methods similar to other 3-deazapurine analogues.[13] A purine base on the third strand having H-bond acceptors at both C6 and N1 also requires use of an analog designed to have these particular features. Naturally-occurring nucleosides having an acceptor at C6 (keto function) are guanosine and inosine; however, the standard tautomeric form of these bases places an imido proton at N1. An inosine analog, 5-aza-7-deazainosine ($^{5N,7C}I$), has acceptor sites in the appropriate positions as shown in triads III and III':G·$\psi i$C=$^{5N,7C}I$. This purine nucleoside may be obtained by methods used in synthesizing other imidazo[1,2—a]-s-triazines.[14] The pur=pur base pair in triad III:U·A=G(I) has been found in certain mispairing interactions within double-stranded DNA of

(12) Mantsch, H. H.; Goia, I.; Kezdi, M.; Bârzu, O.; Dânsoreanu; Jebeleanu, G.; Ty, N. G. *Biochemistry* **1975**, *14*, 5593—5601.

(13) Revankar, G. R.; Gupta, P. K.; Adams, A. D.; Dalley, N. K.; McKernan, P. A.; Cook, P. D.; Canonico, P. G.; Robins, R. K. *J. Med. Chem.* **1984**, *27*, 1389—1396.

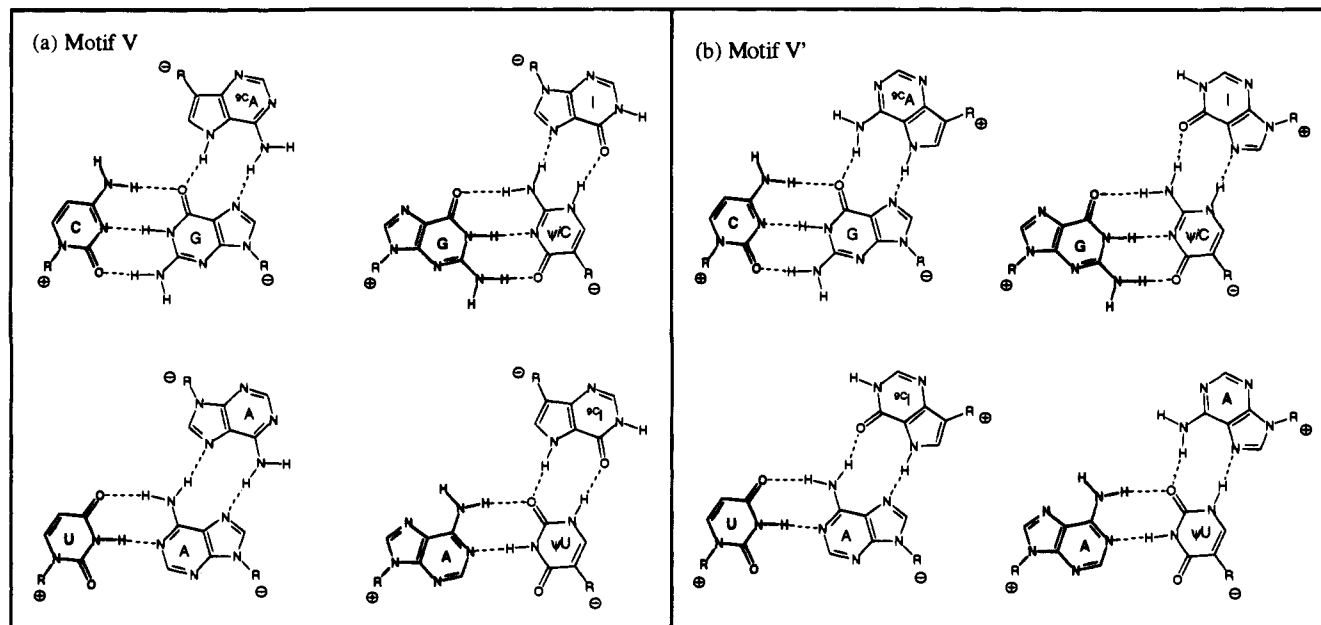(14) Rosemeyer, H.; Seela, F. *J. Org. Chem.* **1987**, *52*, 5136—5143.

**Figure 8.** Class C motifs for purine residues on the third strand base-pairing via H-bond donor—acceptor sites on their five-membered-ring face to residues on the second strand. The representation is analogous to Figure 6.

A(*syn*)=G(*anti*) type.[15] Here, the exocyclic amino function at C2 on the third-strand base, guanosine, is not required for recognition and may lead to adventitious pairing interactions. The base, hypoxanthine, possesses appropriate H-bonding sites without the function at C2; therefore, its nucleoside, inosine, may be a favorable substitute in this case. The remaining triad in Motif III requires recognition of ψU in the W·C-probe strand by a purine base on the third strand. Adenosine has an N1—C6 acceptor—donor pattern for this purpose (triad III:A·ψU=A). Motifs III and III′ are interconverted by similar rules as for motifs in Class A. It should be noted that the triad III′:U·A=A is a feature found among tRNA tertiary interactions.[16]

Motifs IV (Figure 7c) and IV′ (Figure 7d) involve H-bonding at N1 and C2 of purine bases on the third strand, in parallel and antiparallel orientations, respectively, to the second strand. Guanine readily provides two donors at these positions for recognition of G residues on the W·C-probe strand at the C·G base pair. The G=G interactions proposed for triads IV:C·G=G and IV′:C·G=G have been demonstrated for G-tetrad[17] and tRNA[16] structures, respectively. A nonstandard purine base having two acceptor sites at N1 and C2 is required to recognize ψiC on the W·C-probe strand at the G·ψiC base pair. The heterocyclic ring derivative, 5-aza-7-deaza-2-ketopurine ($^{5N,7C}i$I), may meet this requirement. This purine analog belongs in the same class of purine ring derivatives as 5-aza-7-deazainosine, which was proposed for recognition of ψiC on the second strand in Motifs III and III′. It is the H-bonding isomer of $^{5N,7C}$I; therefore, the ring structure of the base analog proposed for use in Motifs IV and IV′ may be given the trivial name of 5-aza-7-deazaisoinosine. Two possible base-pairing schemes involving this analog are shown in triads IV and IV′:G·ψiC=$^{5N,7C}i$I. A commonly used base analog having an N1 acceptor and a C2 donor is 2-aminopurine (*i*A), which can be viewed as a H-bonding isomer of adenine. The exocyclic amino function

at C2 of isoadenosine also makes it isosteric with guanine at this position of the six-membered ring. This analog has been synthesized by a number of methods,[18] and the base triad as shown in IV:U·A=*i*A has been proposed to be the specific tertiary interaction responsible for the activity of a mutant ribozyme from *Tetrahymena*.[18b] Isoadenosine on the third strand may also form a base pair with ψU in the W·C-probe strand to give the triad IV′:A·ψU=*i*A. A purine base on the third strand having an N1 donor and a C2 acceptor for H-bonding to second-strand ψU and A residues in parallel and antiparallel orientation is required to complete Motifs IV and IV′, respectively. Xanthine is a familiar base having these H-bonding groups. However, xanthosine nucleosides are notorious for their ability to base pair through many types of H-bonding interactions other than those required here. The 3,7-dideaza derivative of inosine belongs to the class of pyrrolo[3,2−*c*]pyridine analogues[19] and possesses H-bonding donor—acceptor groups similar to inosine at its six-membered-ring face. This base analog, however, has an exocyclic keto acceptor at C6 rather than at C2. Therefore, the proposal is made to use a H-bonding isomer of inosine, 3,7-dideazaisoinosine ($^{3,7C}i$I), as shown in triads IV:A·ψU=$^{3,7C}i$I and IV′:U·A=$^{3,7C}i$I.

Class C motifs (Table 1 and Figure 8a,b) utilize H-bond donor—acceptor sites at C6 and N7 of purine nucleosides on the third strand. Here, the five-membered (or imidazole) rings of purine bases are involved in recognition of bases on the second strand; therefore, interactions are said to occur at the five-membered-ring "face" of purines on the third strand. Since there is only one pair of adjacent H-bond donor—acceptor sites on this face of the purine ring, only two motifs in this class may be constructed in terms of base-pairing interactions between second and third strands. Motif V (Figure 8a) involves second and third strands in parallel orientation, and Motif V′ (Figure 8b) should involve second and third strands having antiparallel orientation. However, Motif V′ possesses a structural characteristic which may cause it to violate the first rule set forth for construction of triads, that is, all of the bases in each strand

(15) Brown, T.; Hunter, W. N.; Kneale, G.; Kennard, O. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2402—2406.

(16) Kim, S.-H. *Adv. Enzymol.* **1987**, *46*, 279—315.

(17) (a) Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proc. Natl. Acad. Sci. U.S.A.* **1962**, *48*, 2013—2018. (b) Zimmerman, S. B.; Cohen, G. H.; Davies, D. R. *J. Mol. Biol.* **1975**, *92*, 181—192. (c) Kang, C. H.; Zhang, X.; Ratliff, R.; Moyzis. R.; Rich, A. *Nature* **1992**, *356*, 126—131.

(18) (a) McLaughlin, L. W.; Leong, T.; Benseler, F.; Piel, N. *Nucleic Acids Res.* **1988**, *16*, 5631—5644. (b) Doudna, J. A.; Szostak, J. W.; Rich, A.; Usman, N. *J. Org. Chem.* **1990**, *55*, 5547—5549.

(19) Seela, F.; Bourgeois, W. *Heterocycles* **1987**, *26*, 1755—1760.

should have the *anti* configuration of the glycosyl bond. For triads in Motif V', the distance between sugar—base linkages in nucleoside residues of the second and third strands is quite close (see Figure 7b). Nucleosides on the second strand must have the specified *anti* configuration in order to form a W·C duplex; therefore, sugars on this strand will always be oriented toward the major groove wherein lies the third strand. If all third-strand residues in Motif V' were also to have the *anti* glycosyl configuration, sugars on this strand would be oriented in the direction of the second strand. The space between glycosyl bonds of the second- and third-strand residues shown in Figure 7b appears too narrow to accommodate the backbones of both strands. For steric reasons, therefore, it may not be possible to bind a homopurine third strand in an antiparallel orientation to the second strand according to the geometry of Motif V'. However, it is possible that, for Motif V', all purine residues in the third strand can adopt a *syn* glycosyl configuration. In this case, sugars of the third-strand backbone will be oriented away from the groove between the second and third strands. Therefore, in order for sequence-specific recognition between second and third strands to occur according to pairing geometries of the motif, the third strand should have *the same* orientation as the second strand. We include the description of geometries for triads in Motif V' in order to complete all possible base-pairing interactions between second and third strands.

Inosine on the third strand provides the desired adjacent acceptor sites on its five-membered-ring face for recognition of pseudoisocytidine residues on the W·C-probe strand as shown in triads V and V':G·$\psi$iC=I. With a donor at C6 and an acceptor at N7, adenosine on the third strand can pair with itself (parallel interaction) or with $\psi$U (antiparallel interaction) on the second strand as shown in triads V:U·A=A and V':A·$\psi$U=A, respectively. The pur=pur base-pairing interaction in the former triad is the same as that first observed in the parallel double helix formed by poly(A$^+$)[20] and later in a U·A=A triad found in tRNA$^{Phe}$.[16] Naturally-occurring purine bases at physiological conditions do not have a donor proton at N7 on the imidazole ring. Therefore, another type of purine analog is proposed for use on the third strand to complete the remaining triads in this class. Substitution at position 9 of carbon for nitrogen (9-deaza) in a standard purine system increases the valency of the five-membered ring by one, thereby providing a proton at the isosteric equivalent of N7. These analogues are members of the 5*H*-pyrrolo[3,2−*d*]pyrimidine system of heterocyclic bases, and the C9 glycosyl derivatives of such bases are also C-nucleosides. The 9-deaza analog of adenosine ($^9$C A) has been synthesized[21] and can provide the desired H-bond donors as shown in triads V and V':C·G=$^9$C A. Likewise, 9-deazainosine ($^9$C I)[22] has the appropriate acceptor—donor pattern at C6—N7 to complete this class of motifs as shown in triads V:A·$\psi$U=$^9$C I and V':U·A=$^9$C I. It should be mentioned that the 8-oxopurine analogues are another set of derivatives which may provide H-bond donors at N7. In fact, 8-oxoadenosine has been shown to be a suitable replacement for protonated cytidine recognition of the C·G base pair in pyr·pur=pyr types of triplexes.[23] The H-bonding pattern and geometry of the base triad proposed for these systems is analogous to the V:C·G=$^9$C A triad. For several reasons, however, the 9-deaza analogues may be more favorable

candidates for third-strand purine bases having N7 donor capability. The 8-oxo group adds an unnecessary acceptor site at C8, and there exists the possibility of an unfavorable tautomeric form in which the N7 proton is shifted to form an 8-hydroxy group. Most importantly, because of the bulky group at C8, 8-oxopurine analogues are found quite frequently in the *syn* configuration for the glycosyl linkage.[24] Since purine bases, in general, have a relatively low barrier for rotation between the *anti* and the *syn* glycosyl configurations, incorporation of 8-oxopurine nucleosides into a third strand may result in partial or total change of the preferred glycosyl bond configuration for nucleosides in this strand. Complete *anti* to *syn* conversion could readily occur due to cooperative effects, as purine bases tend to maximize their stacking interactions, even in a single-stranded state. In such a case, the orientation of the third strands for class C motifs would have to be reversed in order to maintain the correct specificity for the target W·C duplex. (As previously mentioned, this might be a desired structural feature in Motif V'.) Finally, 9-deazapurine analogues are C-nucleosides and should have increased chemical stability. The C1'−C9 glycosyl bond is much less susceptible to acid hydrolysis in purine type ring systems than is the C1'−N9 glycosyl bond.

**Double-Stranded Target Sequences.** A subset of the proposed base-triad motifs may be utilized to recognize naturally-occurring double-stranded target **pyr·pur** sequences. The base pairs C·G and T·A found in DNA are equivalent to the first and third W·C pairs in each motif (Table 1). Therefore, it may be possible to form triple-stranded helices at double-stranded target sites by interaction of a single oligomer probe designed to bind as a third strand according to one of the ten motifs presented here. For example, it has been demonstrated that a synthetic oligonucleotide probe containing $\psi$iC and T residues may bind sequence specifically to a **pyr·pur** target sequence at neutral pH, forming a triple-stranded complex according to Motif I.[8] In addition, early and recent literature reports several examples of pyr·pur=pur triplexes formed with homopolymer sequences. The C·G=G triplex first reported by Lipsett[25] may form according to either Motif IV or IV'. Recent studies on intramolecular complexes, which force the two purine strands to be antiparallel, seems to indicate that Motif IV' is the preferred G=G interaction.[26] Triple helices having only U·A=A or T·A=A base triads have also been observed,[27] although the specific A=A base-pairing schemes in these systems have not been determined. Several other investigations have reported the formation of pyr·pur=pur triplexes by third strands consisting of mixed adenosine and guanosine residues.[28] It is important to note, however, that none of the proposed Class B or C motifs can be used to construct a recognition motif in which a homopurine second strand containing both G and A can be bound by a third strand utilizing only G and A residues. In other words, a third strand comprised of only G and A residues cannot be expected to form a triple-stranded complex with a naturally-occurring **pyr·pur** target sequence according to the structural requirements presented here; viz., the *anti*

(20) Rich, A.; Davies, D. R.; Crick, F. H. C.; Watson, J. D. *J. Mol. Biol.* **1961**, *3*, 71−86.

(21) Lim, M.-I.; Klein, R. S. *Tetrahedron Lett.* **1981**, *22*, 25−28.

(22) Lim, M.-I.; Klein, R. S.; Fox, J. J. *Tetrahedron Lett.* **1980**, *21*, 1013−1016.

(23) (a) Young, S. L.; Krawczyk, S. H.; Matteucci, M. D.; Toole, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10023−10026. (b) Miller, P. S.; Bhan, P.; Cushman, C.; Trapane, T. L. *Biochemistry* **1992**, *31*, 6788−6793.

(24) Kouchakdjian, M.; Veeraiah, B.; Shibutani, S.; Eisenberg, M.; Johnson, F.; Grollman, A. P.; Patel, D. J. *Biochemistry* **1991**, *30*, 1403−1412.

(25) Lipsett, M. N. *J. Biol. Chem.* **1964**, *239*, 1256−1260.

(26) (a) Chen, F.-M. *Biochemistry* **1991**, *30*, 4472−4479. (b) Radhakrishnan, I.; de los Santos, C.; Patel, D. *J. Mol. Biol.* **1991**, *221*, 1403−1418.

(27) (a) Broitman, S. L.; Im, D. D.; Fresco, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 5120−5124. (b) Trapane, T. L.; Ts'o, P. O. P. *Biophys. J.* **1992**, *61*, 2437.

(28) (a) Cooney, M.; Czernuszewicz, G.; Postel, E. H.; Flint, S. J.; Hogan, M. E. *Science* **1988**, *241*, 456−459. (b) Pilch, D. S.; Levenson, C.; Shafer, R. H. *Biochemistry* **1991**, *30*, 6081−6087. (c) Trapane, T. L.; Kan, L.-S.; Reynolds, M.; Hogrefe, R.; Ts'o, P. O. P. *J. Biomol. Struct. Dyn.* **1991**, *8*, a229.

**Table 2.** Based Triads Patterned after Naturally-Occurring pur=pur Interactions That May Be Used To Form Triple Helices at Naturally-Occurring, Homopyrimidine—Homopurine (pyr•pur), Double-Stranded Target Sites by Addition of Purine Third Strands. [Directions of H-bond donors for base-pairing interactions between purine bases in the target duplex and purine bases in the third strand are shown. Strand orientations are as indicated in Table 1]

| Triad Motif | Base Triads | | Strand Orientation | Natural Paradigm |
|---|---|---|---|---|
| III | $C•G_-^{-3C}iG$ | $T•A_-^-G(I)$ | $\oplus•\ominus=\ominus$ | A(*syn*)•G(*anti*) |
| III' | $C•G_-^{-3C}iG$ | $T•A_-^-A$ | $\oplus•\ominus=\oplus$ | tRNA |
| IV | $C•G_-^-G$ | $T•A_-^-iA$ | $\oplus•\ominus=\ominus$ | G-tetrad |
| IV' | $C•G_-^-G$ | $T•A_-^{-3,7C}iI$ | $\oplus•\ominus=\oplus$ | tRNA |
| V | $C•G_-^{-9C}A$ | $T•A_-^-A$ | $\oplus•\ominus=\ominus$ | poly(A$^+$), tRNA |

configuration for the sugar—base torsion angles, a pair of specific H-bonds for each second=third-strand interaction formed by stable tautomers of purine bases, and isomorphic triad geometry along the triplex. As mentioned in the development of triads containing purine third strands, there are well-documented examples of pur=pur interactions involving G and A. These examples are found in DNA double helices (III: U•A=G), tRNA tertiary interactions (III':U•A=A, IV':C•G=G and V:U•A=A), G-tetrads found at telomere and recombination sites (IV:C•G=G) and homopolymers (V:U•A=A). On the basis of these natural paradigms, five sets of **pyr•pur**=pur motifs can be proposed which should form triple helices with structural uniformity (Table 2). However, if the aforementioned structural requirements are to be kept, each set must incorporate on the third strand one naturally-occurring purine and one purine analog as recognition elements for **G** and **A** residues in the target duplex.

**Alternative Bases for the Second-Strand Probe.** During development of the schemes for motifs in Class C, it became obvious that residues in the second strand, which are involved in W•C base pairing to bases in the target strand, would not have to be limited to the four bases initially chosen as complements. The 9-deazapurine analogues of guanosine ($^{9C}$G)[29] and of adenosine ($^{9C}$A) may form W•C H-bonded base pairs to C and U, respectively, yet present different recognition patterns in the major groove due to the H-bond donor at N7 (Figure 9). However, the C•$^{9C}$G base pair now has an acceptor—donor pattern at C6—N7 which is equivalent to the A•ψU base pair, and the U•$^{9C}$A pair with two donors becomes equivalent in the major groove to G•ψiC (compare Figures 4 and 9). It now becomes necessary to identify alternative pyrimidine complements to target-strand **G** and **A** residues which have H-bonding patterns in the major groove distinct from C•$^{9C}$G and U•$^{9C}$A base pairs. There are only four possible permutations of H-bonding patterns involving a pair of donor—acceptor sites, and the original group of four W•C base pairs (Figure 4) includes one of each permutation. For bases on the W•C-probe strand, these donor—acceptor pairs are found at C6—N7 for purines and at ψC4—ψN5 for pseudopyrimidines. The base on the target strand will determine the H-bonding substituent on its complement which is located nearest to the W•C dyad axis (see Figure 5). Specifically, as target **C** and **A** bases have an exocyclic amino function at C4 and C6, respectively, their

complementary bases (G and U) must have a keto function at C6 and C4, respectively. This situation is reversed for target **U** and **G** bases in that their complements (A and C) have exocyclic amino functions at the H-bonding site closest to the center of the major groove. Therefore, purine bases on the second strand may be modified to change their donor—acceptor capability at N7 in order to generate permutations in H-bonding patterns for recognition by third-strand bases. Likewise, pyrimidine bases on the second strand may be modified to change the H-bonding character at ψN5. A cytosine analog which will have an acceptor site at ψN5 is the triazine derivative 5-azacytosine ($^{5N}$C).[30a] The nucleotide of this base analog has been incorporated into double-stranded DNA via template-directed enzymatic synthesis and has been shown to specifically base pair with guanosine.[30b] The remaining permutation in H-bonding patterns on the W•C-probe strand is provided by the ψC pyrimidine analog. This C-nucleoside may form an appropriate W•C base pair with target-strand A residues and has the required acceptor site at ψN5. With these four alternative bases added to the original palette of second-strand complements, it is now possible to propose four permutations of second-strand recognition schemes having W•C complementarity to the four naturally-occurring target-strand bases and, at the same time, having *unique* pairs of donor—acceptor sites for third-strand interactions. These permutations are listed in Table 3 and are arranged by the H-bonding patterns presented in the major groove on the second-strand base. This table also gives proposed third-strand bases necessary for recognition of each pattern according to geometries of the ten base-triad motifs. In order to design a particular triplex, the structural requirements outlined above must be fulfilled for triads according to a specific motif. For example, if permutation number 3 (Table 3) of bases was chosen for the second strand to construct a triple-stranded complex patterned after pyr•pur=pur interactions found in tRNA, then either Motif III', IV' or V could be used. Specifically, triads in Motif III' will be A•ψC=$^{3C}$iG, G•ψiC=$^{5N,7C}$I, U•A=A, and C•$^{9C}$G=G(I), triads in Motif IV' will be A•ψC=G, G•ψiC=$^{5N,7C}$iI, U•A=$^{3,7C}$iI, and C•$^{9C}$G=iA, and triads in Motif V will be A•ψC=$^{9C}$A, G•ψiC=I, U•A=A, and C•$^{9C}$G=$^{9C}$I. Of course, with the eight bases available for use as W•C complements it is possible to design second strands which have the same, or degenerate, patterns of H-bonds in the major groove for two

(29) Girgis, N. S.; Michael, M. A.; Smee, D. F.; Alaghmandan, H. A.; Robins, R. K.; Cottam, H. B. *J. Med. Chem.* **1990**, *33*, 2750—2755.

(30) (a) Beisler, J. A.; Abbasi, M. M.; Kelley, J. A.; Driscoll, J. S. *J. Carbohydr., Nucleosides, Nucleotides* **1977**, *4*, 281—299. (b) Hayashibara, K. C.; Verdine, G. L. *Biochemistry* **1992**, *31*, 11265—11273.
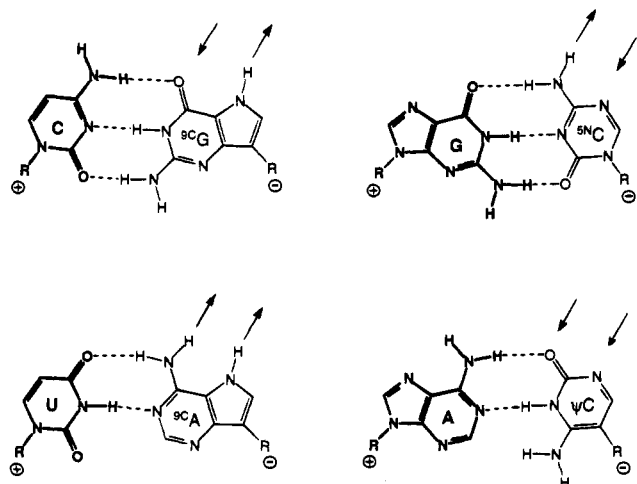
**Figure 9.** Alternative Watson—Crick base pairs formed by the four target-strand bases and four bases on the second strand which have H-bonding patterns in the major groove that are permuted from those shown in Figure 4.

target-strand bases. There are twelve such degenerate permutations and, under certain circumstances, a controlled degeneracy in second=third-strand base-pair recognition may be desirable. However, use of such second strands may decrease specificity of third-strand interactions with the intended target; therefore, design of such triple-stranded complexes will not be discussed here. We also note that there are 640 possible nearest-neighbor base interactions among the triads proposed here.

## Discussion

An exhaustive compilation of possible pairing interactions between the four naturally-occurring bases (A, C, G, and T) in standard tautomeric forms was made more than 30 years ago by Donohue.[31] For the most part, the Donohue base pairings which occur within the present model (indicated in Table 1) have been identified in various nucleic acid structures as noted above. The only pairing interaction which has yet to be verified is the A=C interaction shown in triad I':U·A=C. The ten motifs presented here are comprehensive in that geometries of base triads involving a pair of specific H-bonds between second and third strands have been thoroughly considered. Experimental investigations may determine whether or not any or all of these triad motifs can be used to design actual triplexes.[10] Calculations which address possible helical configurations and stabilities for these proposed structures are currently in progress.

The use of pyrimidine C-nucleoside residues on the second strand is structurally essential to the proposed model. These analogues provide an extra H-bonding site in the major groove of a W·C-type duplex for third-strand recognition which can alleviate the restriction of triplex formation to pyr·pur target sites. One of the most common modified nucleosides found in the RNA of virtually all organisms is the pyrimidine C-nucleoside, pseudouridine. A survey of the motifs reveals that all of the standard bases can pair with $\psi$U through formation of two H-bonds; $\psi$U=A pairs are proposed in Motifs III and V', a $\psi$U=C pair is proposed in Motif I, a $\psi$U=G(I) pair is proposed in Motif III', and $\psi$U=T(U) pairs can be proposed in Motifs I' and II. The donor—acceptor pattern of $\psi$U in the major groove of a duplex in which it is W·C base paired to adenosine will be different from that of A or G. This unique property of the pseudouridine C-nucleoside may enable it to be used as a specific recognition site in biological systems for

tertiary or higher-order interactions. The pseudoisocytidine C-nucleoside, on the other hand, has an apparent contradiction in regard to the requirement for specific base pairing in that it has two distinct tautomeric forms, allowing it to base pair via a W·C H-bonding scheme (G·$\psi$iC) or as a third-strand base (Motifs I and I':C·G=$\psi$iC). The dual base-pairing capability of pseudoisocytidine has been demonstrated in oligonucleotide systems containing this pyrimidine C-nucleoside in substitution for cytidine.[8] A crystal structure of the related pyrimidine base, isocytosine, shows that both of these tautomeric forms are stable and that they may coexist in a H-bonded complex.[32] A C5 glycosyl linkage for isocytosine results in the pyrimidine C-nucleoside analog, pseudoisocytidine ($\psi$iC), which is of interest here. It is anticipated that tautomerization of the $\psi$iC base may be determined within the context of its base-pairing interactions.

According to the current model, substitutions for particular bases may be made in any motif as long as specific H-bonding patterns are maintained and the new triad remains geometrically compatible to other triads in the motif. For instance, third-strand purine bases in class B do not involve H-bonding interactions at N7. Therefore, it may be advantageous to synthesize some or all of Class B third-strand purine residues as 7-deaza analogues (derivatives of tubercidin) in order to avoid unwanted interactions at this face of the third-strand bases. In addition, certain bases proposed for use on the third strand have exocyclic amino functions which are not directly involved in base-pairing interactions to the second strand. These amino groups are found on third-strand iC residues in Motifs I—I', C residues in Motifs II—II', and G residues in Motifs III—III' and may adversely affect the proposed H-bonding schemes in that they provide an alternative site for base-pairing interactions. More importantly, if the third-strand bases in question associate with second-strand residues according to the proposed triads, these amino functions will be unable to readily involve both of their donor protons in stable H-bonds. This will be due to proximity of the other bases in the triad and to general exclusion of solvent water molecules from the hydrophobic core of stacked bases in the helix. Therefore, these nonessential amino groups may carry an energy penalty in terms of available proton donors which cannot participate in H-bonding at the expense of forming the desired pairing interaction. As mentioned previously, inosine may be substituted for guanosine in the third strands of Motifs III—III'. For the same reason, pyrimidine analogues of isocytidine and cytidine which lack this group may be used in Motifs I—I' and II—II', respectively. Keto functions which are not utilized in base-pairing interactions should not pose the same problems expected to be encountered with amino functions. The carbonyl group does not require the presence of a donor proton, and in some cases the positions of these keto functions are important in determining the conjugation of the heterocyclic system for appropriate H-bonding patterns located at specific ring positions (e.g. third-strand C and G residues in Motifs I—I' and IV—IV', respectively). Concerning base substitutions on the second strand, the most obvious would be to use 2-aminoadenosine (2,6-diaminopurine) as the W·C complement for target-strand uridine residues. This base pair will form with an additional H-bond between uridine O2 and 2-aminoadenosine N2. Conversely, the amino function at $\psi$C2 in pseudocytidine is not required for recognition of target-strand adenosine residues and may adversely affect its ability to base pair in a W·C motif (Figure 9).

Chemistry and stereostructure of the sugar moieties and internucleoside backbone linkages of the oligomer probes,

(31) (a) Donohue, J. *Proc. Natl. Acad. Sci. USA* **1956**, *42*, 60—65. (b) Donohue, J.; Trueblood, K. M. *J. Mol. Biol.* **1960**, *2*, 363—371.

(32) McConnell, J. F.; Sharma, B. D.; Marsh, R. E. *Nature* **1964**, *203*, 399—400.

**Table 3.** Permutations of Possible Second-Strand Recognition Schemes Having Watson—Crick Complementarity to the Four Naturally-Occurring Target-Strand Bases and Unique Hydrogen-Bonding Patterns for Third-Strand Recognition (Upper Section). [The directions of H-bond donor—acceptor sites exhibited by the second-strand bases in the major groove of the duplex are shown to the right of each base pair. Third-strand bases which are complementary to these H-bond recognition patterns are dictated by choosing one of the geometrically isomorphous Triad Motifs I—V' (lower section)]

| Permutation Number | Watson—Crick Base Pairs | | | |
|---|---|---|---|---|
| 1 | C·G⁻ | G·ψiC⁻ | (T)U·A⁻ | A·ψU⁻ |
| 2 | C·G⁻ | (T)U·⁹ᶜA⁻ | G·⁵ᴺC⁻ | A·ψU⁻ |
| 3 | A·ψC⁻ | G·ψiC⁻ | (T)U·A⁻ | C·⁹ᶜG⁻ |
| 4 | A·ψC⁻ | (T)U·⁹ᶜA⁻ | G·⁵ᴺC⁻ | C·⁹ᶜG⁻ |

| Triad Motif | ⇄ | ⇉ | ⇄ | ⇄ | Strand Orientation |
|---|---|---|---|---|---|
| **Class A** | | | | | |
| I | ψiC | iC | T(U) | C | ⊕·⊖=⊖ |
| I' | ψiC | iC | C | T(U) | ⊕·⊖=⊕ |
| II | ψC | C | iC | T(U) | ⊕·⊖=⊖ |
| II' | ψC | C | T(U) | iC | ⊕·⊖=⊕ |
| **Class B** | | | | | |
| III | ³ᶜiG | ⁵ᴺ,⁷ᶜI | G(I) | A | ⊕·⊖=⊖ |
| III' | ³ᶜiG | ⁵ᴺ,⁷ᶜI | A | G(I) | ⊕·⊖=⊕ |
| IV | G | ⁵ᴺ,⁷ᶜiI | iA | ³,⁷ᶜiI | ⊕·⊖=⊖ |
| IV' | G | ⁵ᴺ,⁷ᶜiI | ³,⁷ᶜiI | iA | ⊕·⊖=⊕ |
| **Class C** | | | | | |
| V | ⁹ᶜA | I | A | ⁹ᶜI | ⊕·⊖=⊖ |
| V' | ⁹ᶜA | I | ⁹ᶜI | A | ⊕·⊖=⊕ |

though important topics, are considerations outside of this proposal. However, a few issues regarding these important components of nucleic acid helices should be mentioned. In terms of the *anti—syn* configuration about the glycosyl bond, the C-nucleosides will most likely have a reduced energetic barrier for rotation due to the longer C—C bond (~1.54 Å) as compared to the C—N bond (~1.48 Å). In addition, the preferred pseudorotation angle of the sugar moiety is expected to be dependent, in part, on properties of the attached bases; therefore, some nucleosides of the unusual heterocyclic bases may adopt sugar configurations quite different from those exhibited by standard purine and pyrimidine nucleosides.[33] These structural considerations will affect eventual placement of backbones for the three strands within the helix and can be addressed for each base analog both experimentally and through model building.

The model proposed here involves the binding of two synthetic probe strands to a sequence in a single-stranded target nucleic acid. Therefore, this approach may provide a possible means to recognize and bind naturally-occurring sequences of interest. Some aspects of this proposed application should be considered. It can be readily seen that the duplex formed by

W·C-probe strand binding to the target sequence will have *novel* H-bonding patterns in its major groove that do not exist in natural duplexes. Therefore, formation of a (target strand)·-(second strand) duplex will generate a site for third-strand binding which is *unique* among cellular duplex sequences. Even though it will have a homopurine or homopyrimidine sequence, the third-strand oligomer probe should not have any sustained linear complementarity to cellular sequences, either single- or double-stranded, except at the site created by actual binding of the second-strand oligomer probe to its target. The second strand will contain a mixture of purines and pyrimidines complementary to the heterologous sequence of the target site at its Watson—Crick binding face. This strand also possesses complementary H-bonding sites for the third-strand probe at an alternative face. It remains to be determined, therefore, whether or not the W·C-probe strand will associate with the third strand by itself (in the absence of the target strand) and what will be the mechanism of binding of these two synthetic strands to the target sequence.

In conclusion, a possible means of forming triple-stranded complexes at single-stranded target sites of any sequence has been presented. Structure building through the formation of specific H-bonds and base-triad geometries is the foundation

(33) For an example, see: Saran, A.; Chatterjee, C. L. *Int. J. Quantum Chem.* **1984**, *25*, 743—752.

of this approach. Successful triplex formation according to any one of the motifs will provide a major advance in terms of nucleic acid—nucleic acid recognition schemes. The knowledge gained from such efforts may then be very well put to use for the design of other recognition schemes which do not necessarily have to be limited to double- or triple-stranded complexes.

**Supplementary Material Available:** A listing and discussion of the twofold symmetry elements found in the base triads (5 pages). This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.