Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 3840–3844, May 1992
Biochemistry

Elucidation of the sequence-specific third-strand recognition of four Watson–Crick base pairs in a pyrimidine triple-helix motif: T·AT, C·GC, T·CG, and G·TA

KYONGGEUN YOON*†, CHERYL A. HOBBS*, JULIE KOCH*, MARK SARDARO‡, ROSTYSLAW KUTNY‡, AND ALEXANDER L. WEIS§¶

Departments of *Molecular Biology, ‡Molecular Characterization, and §Medicinal Chemistry, Sterling Winthrop Pharmaceuticals Research Division, 25 Great Valley Parkway, Malvern, PA 19355

Communicated by I. Tinoco, Jr., January 24, 1992

ABSTRACT **We report a specific pattern of recognition by third-strand bases for each of the four Watson–Crick base pairs within a pyrimidine triple-helix motif as determined by PAGE: T·AT, C·GC, T·CG, and G·TA. Our recognition scheme for base triplets is in agreement with previous studies. In addition, we identified another triplet, T·CG, under physiological conditions, in which formation of triple helix was observed at equimolar ratios of the third strand and duplex target. Although different nearest-neighbor effects are expected, this finding extends the base-recognition code to all 4 base pairs in double-stranded DNA under physiological conditions. Base-composition analysis of putative triplex species provided independent evidence for the formation of triplex and confirmed the base-recognition code determined by PAGE. Moreover, the formation of triplex, as detected by gel electrophoresis, was seen to be an all-or-none phenomenon, dependent upon a single-base mismatch among 21 nucleotides. This result suggests a high specificity for the recognition of double-stranded DNA by a third strand. In addition, we report the surprising finding that triplex stability depends on the length and sequence of the target duplex DNA.**

Approaches at gene regulation via triple-helix formation have been hindered by limited information regarding specific third-strand base recognition of all 4 Watson–Crick base pairs. Current information on triplex base recognition is limited to homopurine or homopyrimidine sequences. Pyrimidine oligonucleotides bind in the major groove of Watson–Crick double-stranded DNA (1–3). Specificity is due to Hoogsteen hydrogen bonding, in which thymine in the third strand recognizes A·T base pairs in double-stranded DNA (T·AT triplet) and protonated cytosine in the third strand recognizes G·C base pairs (C^+·GC triplet) (1–6). Purine oligonucleotides in the third strand have also been shown to bind to purines in duplex DNA (A·AT and G·GC triplets) (7–10). Recently, several nonstandard base triplets (C^+·AT, U·GC, G·GC, and A·AT) as well as standard base triplets (T·AT and C^+·GC) were determined by using a combinatorial approach (11). This method involves the identification of DNA or RNA sequences for their ability to recognize double-helical DNA via the synthesis of random RNA sequences and subsequent selection for triple-helix formation. To target oligonucleotides to any given sequence of the genome, it is important to determine the specificity of the base recognition of the third strand for all four combinations of Watson–Crick base pairs.

We examined the relative affinities of natural bases for all four Watson–Crick base pairs within a pyrimidine triple-helix motif, in which 1 base of the third strand involved in duplex recognition is flanked by 10 T·AT triplets on either side. We report the detection of triplex formation involving one specific base in the third strand for each Watson–Crick base-pair combination with PAGE: T·AT, C·GC, T·CG, and G·TA. Although different nearest-neighbor effects are expected, this finding extends specific recognition within a pyrimidine-triplex motif to all 4 possible base pairs in double-stranded DNA. Moreover, the all-or-none phenomenon seen in triple-strand formation contingent upon a single-base mismatch among 21 bases, suggests a high specificity of base recognition in triple-stranded DNA.

MATERIALS AND METHODS

Deoxyoligonucleotides. All deoxyoligonucleotides were purchased from Midland Certified Reagent (Midland, TX). The concentration of oligonucleotides was determined spectrophotometrically. The reported extinction coefficient for poly(dA) [ε_{257} = 8600 cm^{-1}·(mol of base/liter)$^{-1}$] was used for A_{21}, $A_{10}XA_{10}$, and $CGA_{10}XA_{10}GC$, where X = A, G, C, or T (12). The extinction coefficient for poly(dT) [ε_{265} = 8700 cm^{-1} (mol of base/1)$^{-1}$] was used for T_{21}, $T_{10}YT_{10}$, and $GCT_{10}YT_{10}CG$, where Y = A, G, T, or C (13).

Triplex Formation and Electrophoretic Analysis. Duplex was made by combining equimolar amounts of oligonucleotides in a 0.15 M NaCl/10 mM $MgCl_2$/5 mM Tris acetate (pH 7.0) buffer, heating at 70°C for 15 min, and slow cooling to room temperature. Triplex DNA was made by adding an equimolar amount of the third strand to the duplex and then incubating at various temperatures overnight. The concentration of each strand was 14 μM in a total vol of 10 μl. Electrophoresis was done at 4°C, 23°C, and 37°C with 12% polyacrylamide gels (acrylamide/bisacrylamide, 19:1; 20 × 20 × 0.15 cm) in a 50 mM Tris borate/5 mM $MgCl_2$ (pH 8.3) buffer. In general, electrophoresis was at a constant 10 V/cm with a current of 20–30 mA. After electrophoresis, gels were stained by using a Bio-Rad silver-staining kit according to manufacturer's specifications.

Triplex DNA Isolation for HPLC Analysis. The formation of triplex was as described above. The concentration of each strand was 81 μM in a total vol of 60 μl. The entire DNA sample was loaded on a 12% polyacrylamide gel and separated according to mobility. After gel electrophoresis, DNA bands were visualized by UV shadowing and dissected from the gel. Gel slices were cut into smaller pieces and placed in 1 ml of buffer containing 0.1× standard saline citrate and 2 mM EDTA (pH 8.0) for 2 hr with gentle rocking. The buffer was removed and stored at 4°C. Gel slices were incubated overnight in an additional 1 ml of buffer. The combined 2 ml of sample was filtered through a Microfilter of 0.45-μm-pore size (Schleicher & Schuell).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: RP, reverse phase.

†To whom reprint requests should be addressed.
¶Present address: Cancer Therapy and Research Center, Department of Research, 8121 Data Point Drive, San Antonio, TX 78229.

Strand Separation of Triplex by Reverse-Phase (RP)-HPLC. A Waters 600E HPLC system (Millipore) was used in conjunction with a Kratos Spectroflow 783 variable-wavelength detector (Applied Biosystems) set at 260 nm. For strand separation of triplex, the DNA sample was injected onto a RP-18 column (5-μm-particle size, 150 × 4.6 mm i.d.) from Supelco. DNA was eluted with a binary gradient at a flow rate of 1.0 ml/min. The mobile phase consisted of a mixture of solvents A and B: solvent A was 0.01 M triethylamine acetate at pH 7.5; solvent B was 80% (vol/vol) acetonitrile/20% (vol/vol) solvent A. Elution was accomplished by using a linear gradient up to 20% solvent B in 30 min.

Base-Composition Analysis. Base composition was analyzed on a Hewlett–Packard 1090M HPLC system with a diode-array detector fixed at 260 nm. DNA was digested to completion by a mixture of snake venom phosphodiesterase and bacterial alkaline phosphatase (Pharmacia), at 1 unit/ml and 8 units/ml, respectively, in 15 mM $MgCl_2$/33 mM Tris·HCl, pH 7.5, buffer for 16 hr at 37°C (14). After digestion, protein was removed by ethanol precipitation. Nucleosides were reconstituted in 50 μl of HPLC-grade water and injected onto a RP-18 column (Spheri 5; 5-μm-particle size, 250 × 4.6 mm i.d.) from Applied Biosystems. The mobile phase consisted of a mixture of solvents C and D: solvent C was 0.1 M triethylamine acetate, pH 7.0; solvent D was 90% acetonitrile plus 10% solvent C. Elution was done by using an initial 10-min isocratic period in 100% solvent C followed with a linear gradient to 10% solvent D in 30 min with a flow rate of 1.0 ml/min. Data were recorded and integrated for peak height and area using a Perkin–Elmer–Nelson analytical chromatography system (Perkin–Elmer). For standards, each nucleoside was analyzed by using conditions identical to those described above. Three different amounts of each nucleoside—0.3, 3, and 30 nmol—assayed in triplicate, were used to generate a standard curve. Base composition was calculated by normalization of each peak calibrated against the standard curve.

RESULTS

Gel Mobility and Stoichiometry of Triplex. To establish the stoichiometry of triplex DNA, mixtures of different molar ratios of the deoxyoligonucleotides A_{21} and T_{21} were incubated at 4°C, separated by gel electrophoresis according to mobility, and visualized by silver staining. The mixture containing an equimolar ratio of A_{21} and T_{21} was detected as a duplex, whereas the mixtures containing a molar excess of T_{21} with respect to A_{21} resulted in the conversion of duplex to a complex of slower mobility, as shown in Fig. 1*a*. The mobility of this complex was invariant among mixtures containing a molar excess of T_{21}, indicating a constant stoichiometry of 2:1, $(2T{\cdot}A)_{21}$. In contrast, mixtures containing molar ratios in which A_{21} exceeded T_{21} did not form any higher-molecular-weight complex, as shown in Fig. 1*b*. Thus, under the conditions of this experiment, a T·AT triplet was seen, but an A·AT triplet was not seen.

Specificity and Stability for Recognition of Double-Stranded DNA by a Third Strand. Third-strand base-specific interactions with all 4 Watson–Crick base pairs were examined. Deoxyoligonucleotides, 21 nucleotides in length ($A_{10}XA_{10}$ and $T_{10}YT_{10}$, where X·Y = A·T, G·C, C·G, or T·A), were mixed at an equimolar ratio to form each of the four combinations of duplex. An equimolar amount of the third strand ($T_{10}ZT_{10}$, where Z = T, A, G, or C) was added to each duplex combination to achieve a stoichiometric conversion of duplex to triplex at equilibrium, at near physiological conditions:

$$\begin{array}{l} T_{10}\ Z\ T_{10} \\ A_{10}\ X\ A_{10} \\ T_{10}\ Y\ T_{10}. \end{array}$$

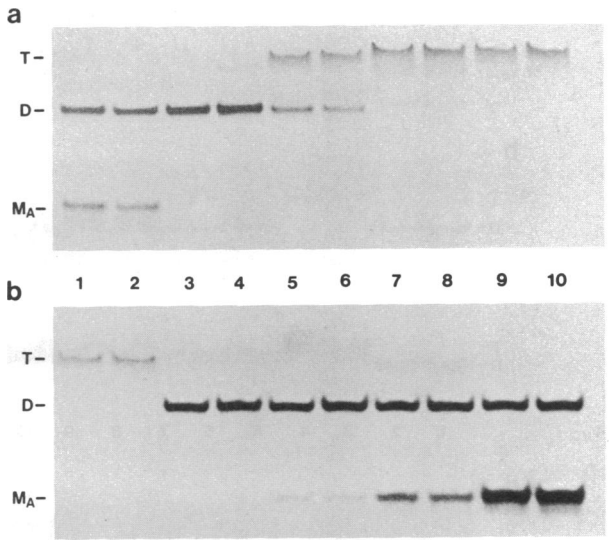

FIG. 1. (*a*) Silver-stained 12% polyacrylamide gel showing complexes of different mobilities formed by combining a fixed amount of A_{21} with increased amounts of T_{21}. Ratios of A_{21} to T_{21} are 1:0.5 (lanes 1, 2); 1:1 (lanes 3, 4); 1:1.5 (lanes 5, 6); 1:2 (lanes 7, 8); and 1:3 (lanes 9, 10). The species are labeled triplex (T), duplex (D), or single-stranded A_{21} (M_A) according to mobilities. Single-stranded T_{21} was not detectable by silver staining but was detected by a UV shadowing technique. The formation of complex and gel electrophoresis were done as described. (*b*) Silver-stained 12% polyacrylamide gel showing complexes formed by combining a fixed amount of T_{21} with increased amounts of A_{21}. Ratios of A_{21} to T_{21} are 0.5:1 (lanes 1, 2); 1:1 (lanes 3, 4); 1.5:1 (lanes 5, 6); 2:1 (lanes 7, 8); and 3:1 (lanes 9, 10).

The different molecular species—monomer, duplex, and triplex—were separated by PAGE and detected by silver staining, as described. To measure the stability of the triplex containing variation at a single base (Z) among 21 bases, gel electrophoresis was done at three temperatures—4°C, 23°C, and 37°C.

At 4°C, all 16 combinations of equimolar mixtures (Z·XY) produced a species that migrated as a higher-molecular-weight complex as compared with duplex, shown in Fig. 2 *a*, *d*, *g*, and *j*. The $T_{10}{\cdot}A_{10}T_{10}$ interaction on both sides of Z appears sufficient to stabilize triplex formation at 4°C, regardless of the identity of Z. In contrast, distinct species of triplex were detected for only one specific combination of Z·XY at 23°C for each Watson–Crick base pair (XY). These combinations of Z·XY were T·AT, C·GC, T·CG, and G·TA, as shown in Fig. 2 *b*, *e*, *h*, and *k*. Moreover, the specificity of the base recognition at 23°C was to the extent that the complex registered as either a triplex or a duplex in an all-or-none fashion dependent on the variation at a single base (Z) among 21 bases. At 37°C, all complexes migrated to the duplex position, except for $T_{21}{\cdot}A_{21}T_{21}$, shown in Fig. 2 *c*, *f*, *i*, and *l*. Under these gel-electrophoresis conditions, the T·AT triplet appears the most stable, consistent with previous findings (5, 6).

To circumvent any possibility of strand exchange between $T_{10}YT_{10}$ and $T_{10}ZT_{10}$ in the triplex formation, a longer target duplex was combined with the 21-nucleotide third strand. Deoxyoligonucleotides, 25 nucleotides in length ($\begin{matrix}CGA_{10}XA_{10}GC \\ GCT_{10}YT_{10}CG\end{matrix}$, where X·Y = A·T, G·C, C·G, and T·A), were mixed at an equimolar ratio to form each of the four duplex combinations. An equimolar amount of the third strand ($T_{10}ZT_{10}$, where Z = T, A, G, or C) was added to each combination of the longer target duplex:

$$\begin{array}{l} T_{10}\ Z\ T_{10} \\ CGA_{10}\ X\ A_{10}GC \\ GCT_{10}\ Y\ T_{10}CG. \end{array}$$

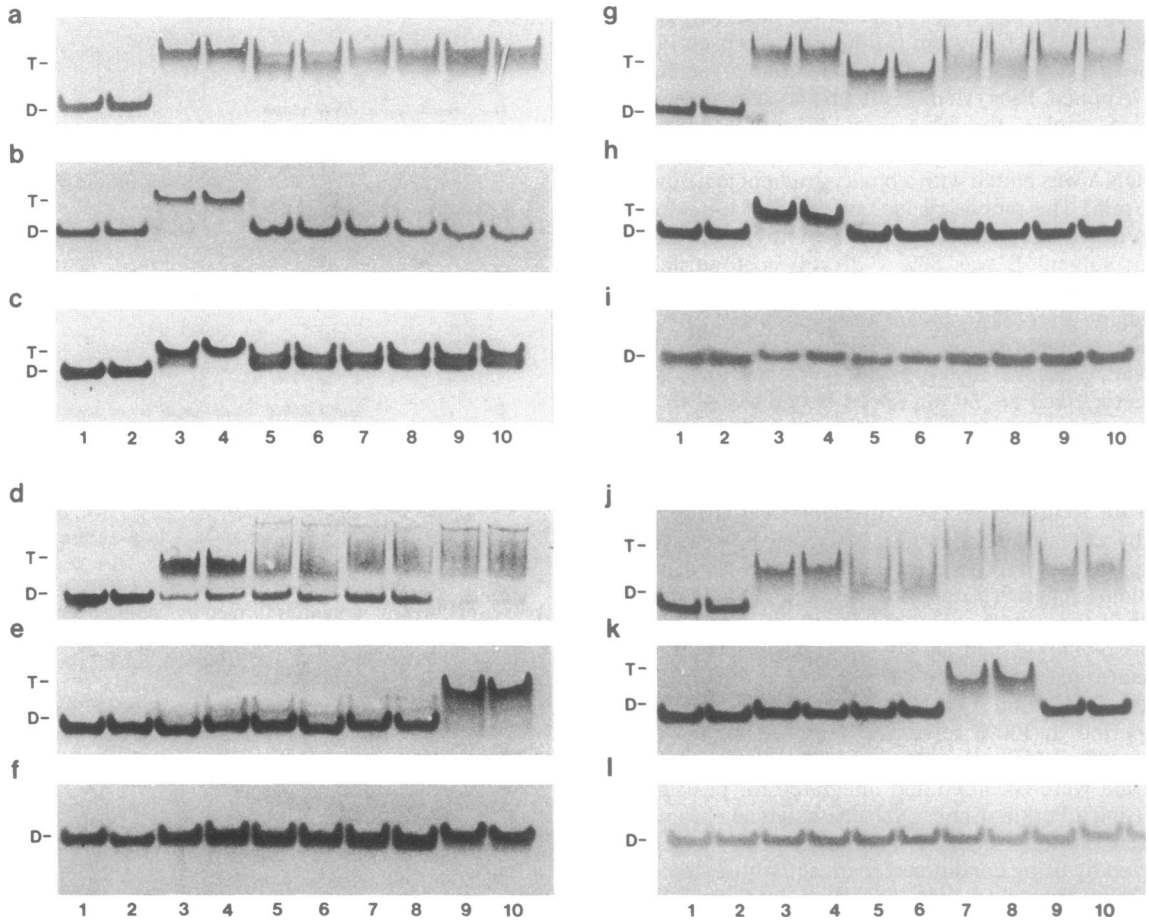

FIG. 2. Temperature profile of triplex formation between a duplex, $\frac{A_{10}XA_{10}}{T_{10}YT_{10}}$, and an equimolar amount of a third strand, $T_{10}ZT_{10}$, in which Z = T, A, G, or C, analyzed on a 12% polyacrylamide gel. (*a*, *b*, and *c*) Gels run at 4°C, 23°C, and 37°C, respectively, in which X·Y = A·T. (*d*, *e*, and *f*) Gels run at 4°C, 23°C, and 37°C, in which X·Y = G·C. (*g*, *h*, and *i*) Gels run at 4°C, 23°C, and 37°C, in which X·Y = C·G. (*j*, *k*, and *l*) Gels run at 4°C, 23°C, and 37°C, in which X·Y = T·A. The species exhibiting different mobilities are labeled as either triplex (T) or duplex (D). Lanes 1 and 2 contain only the duplex. The remaining lanes are as follows: Z = T (lanes 3, 4); Z = A (lanes 5, 6); Z = G (lanes 7, 8); and Z = C (lanes 9, 10). The formation of complex and electrophoresis were done as described.

The complex was subjected to gel analysis as described above. Distinct species of triplex were detected at 4°C for one specific combination of Z·XY for each Watson–Crick base pair (X·Y), depicted in Fig. 3 *a*, *e*, and *g*. However, as shown in Fig. 3*c*, both G·GC and C·GC triplets were observed under these experimental conditions. The pattern of third-strand base recognition of the 25-base-pair (bp) duplex target was similar to that determined for the 21-bp duplex target and led to the conclusion regarding specificity of third-strand binding to a duplex target summarized in Table 1. Once again, the complex registered as either a triplex or a duplex in an all-or-none phenomenon dependent on a single-base mismatch among 21 bases.

It is interesting to note that triplexes containing duplex targets of two different lengths exhibited differential stability. All 16 combinations of complex containing a 21-base single-stranded DNA and the 25-bp duplex target existed as duplex at 23°C, shown in Fig. 3 *b*, *d*, *f*, and *h*. In contrast, four combinations containing the 21-bp duplex target existed as triplex at 23°C, as shown in Fig. 2. Destabilization of the triplex containing a longer duplex target was also observed at 4°C: only 4 combinations containing the longer duplex registered as triplex, whereas all 16 combinations containing the shorter duplex formed a triplex at 4°C. This unexpected result suggests that triplex stability depends on the length or sequence of the target duplex DNA. A triplex containing a 25-bp duplex was less stable than that containing only a 21-bp duplex.

Base-Composition Analysis of Triplex Isolated from Gel Electrophoresis. To provide direct evidence that the species with slower electrophoretic mobilities were actually triple-stranded DNA, DNA bands were isolated from polyacrylamide gels and analyzed for base composition. Four putative triplexes were characterized in this way (triplex of $T_{10}ZT_{10}/CGA_{10}XA_{10}GC/GCT_{10}YT_{10}CG$, where Z·XY = T·AT, C·GC, T·CG, and G·TA; see Table 2). Each species of isolated DNA was subjected to RP-HPLC for strand separation, which resulted in three peaks, indicating that the complex consisted of three DNA strands. Fig. 4*a* shows one example in which peaks 1, 2, and 3 correspond to $CGA_{10}CA_{10}GC$, $GCT_{10}GT_{10}CG$, and T_{21}, respectively. Because the two thymine-rich strands did not completely separate under the RP-HPLC conditions used, they were combined for base-composition analysis. The DNA contained in the peaks was completely hydrolyzed into nucleosides by enzymatic digestion, as described. The hydrolysis product was again subjected to RP-HPLC and separated according to the mobility of each nucleoside (Fig. 4*b*). Base composition was calculated by the integration of each peak normalized against a known amount of each nucleoside used as a standard. Results of base-composition analysis of the four putative triplexes are summarized in Table 2. In general, the experimentally determined base composition agreed well with the predicted base composition for all four species examined. This result confirmed that the slower mobility species seen by gel electrophoresis was, indeed, triplex.

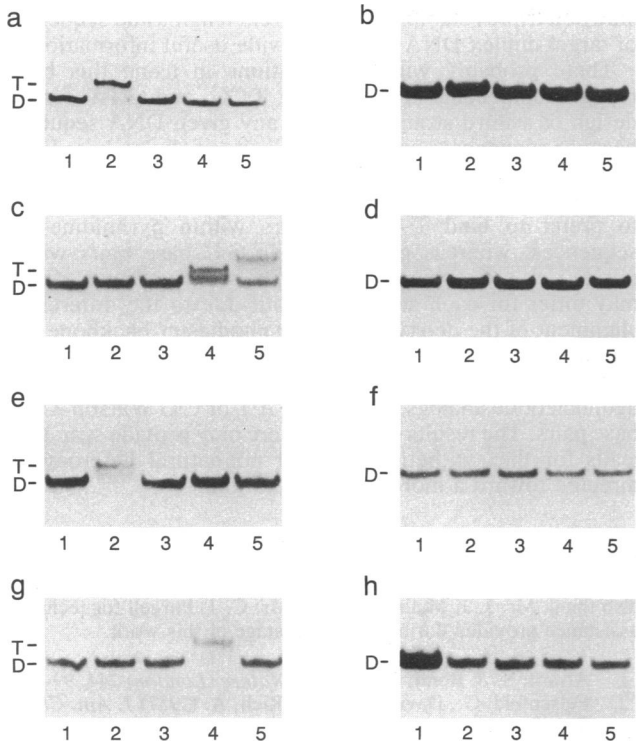

FIG. 3. Temperature profile of triplex formation between a duplex, $\frac{CGA_{10}XA_{10}GC}{GCT_{10}YT_{10}CG}$, and an equimolar amount of a third strand, $T_{10}ZT_{10}$, in which Z = T, A, G, C, analyzed on a 12% polyacrylamide gel. (*a* and *b*) Gels run at 4°C and 23°C, respectively, in which X·Y = A·T. (*c* and *d*) Gels run at 4°C and 23°C in which X·Y = G·C. (*e* and *f*) Gels run at 4°C and 23°C in which X·Y = C·G. (*g* and *h*) Gels run at 4°C and 23°C in which X·Y = T·A. The species exhibiting different mobilities are labeled as either triplex (T) or duplex (D). Lanes 1 contains only the duplex. The remaining lanes are as follows: Z = T (lanes 2); Z = A (lanes 3); Z = G (lanes 4); Z = C (lanes 5). The conditions for formation of complex and electrophoresis are described.

DISCUSSION

We have identified a base-recognition scheme for triplex formation as T·AT, C·GC, T·CG, and G·TA. The specificity of base recognition in triple-helix formation for all four Watson–Crick base pairs was shown as an all-or-none phenomenon under physiological conditions by using an equimolar ratio of the third strand and duplex target. In general, our recognition scheme for base triplets agreed with previous studies. Two well-established base-recognition schemes for triplex formation, T·AT and C·GC, were confirmed (1–6). Furthermore, our results agree with the G·TA triplet determined by entirely different approaches, using thymidine/EDTA/Fe(II) for DNA cleavage (15, 16) and UV melting methods (17). An unusual feature of the base-recognition scheme elucidated here is the T·CG triplet, which was seen under physiological conditions. Both C·CG and T·CG triplets have been preferentially stabilized in the formation of intramolecular triple helix at pH 4.2, as measured by the energy required for superhelical stress (18). Recently, the preference for a pyrimidine in the third strand for C·G base-pair recognition was attributed to lower steric hindrance (17). A purine in the third strand exhibited an unfavorable interaction with the bulky amino group at position 4 of cytosine in double-stranded DNA, as compared with a pyrimidine (17). Our results indicate that T·CG was preferentially stabilized over C·CG under physiological conditions. Moreover, a RNA triplet involving interaction of two pyrimidines, U·CG, was found conserved phylogenetically in group I introns, providing independent evidence for occurrence of a T·CG triplet *in vivo* (19). According to our observations, thymine exhibited degenerate recognition of both A·T and C·G Watson–Crick base pairs. The T·AT triplet was more stable than the T·CG triplet, as determined by gel electrophoresis at 37°C (Fig. 2).

Table 1. Triplex detection by silver staining

W–C duplex (X·Y)	Middle base (Z) of the third strand			
	T	C	A	G
A·T	Triplex			
G·C		Triplex		
C·G	Triplex			
T·A				Triplex

Triplexes containing a 21-bp or 25-bp duplex target were detected in an all-or-none phenomenon at 23°C and 4°C, respectively, dependent on the variation of a single base Z. W–C, Watson–Crick.

Base-composition analysis has shown conclusively that the slower-mobility species detected by gel electrophoresis, in each case, was composed of three strands of DNA. Although the complex was postulated to be a triplex according to its stoichiometry and mobility, these experiments demonstrate by base-composition analysis that the complex is triplex. The close agreement between the experimentally determined and predicted base composition of all four species examined provided independent evidence for the formation of triplex and confirmed the base-recognition code.

One surprising finding from these experiments is the all-or-none phenomenon seen for triplex formation as detected by gel electrophoresis, dependent upon a single-base mismatch among 21 bases. This finding reflects a high specificity for triplex-base recognition. The free energy difference between matched and mismatched triplets can vary between 3 and 6 kcal/mol for intramolecular triplex formation at pH 4.2 (18), similar to that of mismatched duplex base pairs (20, 21). Recently, a single mismatch in the third strand was shown to cause destabilization of 3.2–4.0 kcal/mol, equivalent to the corresponding values for duplex DNA and RNA, indicating that triple helix formation is as specific as duplex formation (22). High selectivity of triple-helix formation was also seen in the binding of a circular oligonucleotide to single-stranded

Table 2. Base-composition analysis of four triplexes isolated from a 12% polyacrylamide gel.

	Z·XY*							
	T·AT		C·GC		T·CG		G·TA	
	E	*T*	*E*	*T*	*E*	*T*	*E*	*T*
Peak 1†								
Cytosine	1.1	2	1.8	2	2.7	3	1.9	2
Guanine	2.1	2	3.1	3	2.0	2	2.0	2
Thymine	0.8	0	0.0	0	0.0	0	1.0	1
Adenine	21.0	21	20.1	20	20.3	20	20.2	20
Peaks 2, 3‡								
Cytosine	2.0	2	3.7	4	2.1	2	2.0	2
Guanine	2.3	2	1.8	2	3.5	3	3.3	3
Thymine	41.7	42	40.4	40	40.7	41	39.5	40
Adenine	0.0	0	0.0	0	0.2	0	1.2	1

E, Experimentally determined number of each nucleoside contained in peaks as described (the error level, ±0.2 base per strand of DNA, was measured from the base-composition analysis of six DNA strands assayed, in triplicate); *T*, theoretical number of each nucleoside contained in peaks.

*The triplex $\frac{CG\ A_{10}X\ A_{10}GC}{GC\ T_{10}Y\ T_{10}CG}$ was analyzed for Z·XY.

†Peak 1 contained the adenine-rich strand separated by RP-HPLC.

‡Peaks 2 and 3 contained thymine-rich strands of 21 and 25 bases, respectively, and were combined for base-composition analysis.

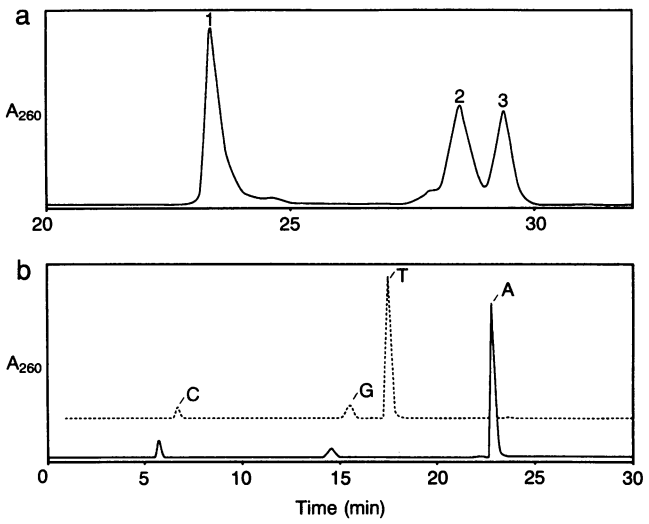

FIG. 4. (*a*) Strand separation of triplex by RP-HPLC. The complex was formed by incubating a duplex, $\frac{CGA_{10}CA_{10}GC}{GCT_{10}GT_{10}CG}$, and an equimolar amount of a third strand, T_{21}, at 4°C which were separated by PAGE. The species with slower mobility was isolated from a 12% polyacrylamide gel, purified, and subjected to HPLC analysis as described. The HPLC profile exhibited three peaks, indicating an equimolar composition of the three strands corresponding to $CGA_{10}CA_{10}GC$ (peak 1), $GCT_{10}GT_{10}CG$ (peak 2), and T_{21} (peak 3). The position of each strand was determined by HPLC elution profiles exhibited upon separate injections of each strand (data not shown). (*b*) Base-composition analysis of DNA isolated from RP-HPLC. Because the two thymine-rich strands did not completely separate under the HPLC conditions described, DNAs eluted in peaks 2 and 3 (*a*) were combined for base-composition analysis. DNA contained in these peaks was subjected to complete hydrolysis into nucleosides and to RP-HPLC analysis. The different nucleosides eluted at the positions marked as C, G, T, or A. The nucleoside elution profile of peak 1 in *a*, corresponding to the adenine-rich sequence $CGA_{10}CA_{10}GC$, is indicated by the solid line; the dotted line represents the nucleoside elution profile of combined peaks 2 and 3, shown in *a*, corresponding to the thymine-rich sequences $GCT_{10}GT_{10}CG$ and T_{21}.

DNA (23, 24). The all-or-none phenomenon observed in our studies for triple-helix formation suggests that the difference in melting temperature (ΔT_m) due to a single-base mismatch in a triplex might be as substantial as that in duplex formation. The possibility exists that the observed all-or-none effect, contingent on a single-base mismatch, could come from the fact that gel electrophoresis conditions did not necessarily reflect the equilibrium situation in solution. Nevertheless, the high specificity of base recognition in a triplex structure is advantageous to a triplex approach toward gene regulation by targeting a gene in a sequence-specific manner.

Another finding of our studies is that the stability of triplex appears to depend on the length or sequence of the target DNA. Triplex containing a longer target duplex DNA was qualitatively less stable. Current models of pyrimidine/purine/pyrimidine triplexes propose that the third strand is Hoogsteen base-paired in the major groove formed by Watson–Crick base-pairing of two strands in an A′-DNA-like conformation (25, 26). Because triplex formation involves a structural change of the target duplex DNA, it is likely to cost more free energy to convert the longer and more stable 25-bp duplex to a triple-stranded configuration than would be required for similar conversion of a 21-bp duplex. A detailed study on triplex stability with different lengths and sequences of target duplex DNA should provide useful information.

There probably will be limitations in using this base-recognition scheme (T·AT, C·GC, T·CG, and G·TA) for the design of a third strand to target any given DNA sequence because sequences surrounding the site of triplex-base interactions appear to be important in base recognition as well as third-strand polarity. For example, guanine has been shown to prefer to bind T·A base pairs within pyrimidine-rich sequences, whereas guanine binds G·C base pairs within purine-rich sequences (6, 8, 15). Specificity of base triplets may differ for each structural motif due to the differential alignment of the deoxyribosylphosphodiester backbone and the heterocyclic bases along the major groove. The degenerate recognition by thymine requires the design of deoxyoligonucleotide analogs specific for A·T or C·G Watson–Crick base pairs. The results described here may provide structural leads for the synthetic design of nonnatural heterocycles directed toward a more general solution.

We thank Dr. D. H. Turner for discussions and critical reading of the manuscript. We thank Drs. J. J. Farrar, L. I. Kruse, and B. N. Jones for support of this work and Dr. J. R. Fresco for advice. We also thank Mr. J. J. Malinowski and Mr. C. J. Purcell for technical assistance provided during the initial stage of this work.

1. Arnott, S. & Bond, P. J. (1973) *Nature (London)* **244,** 99–101.
2. Felsenfeld, G., Davies, D. R. & Rich, A. (1957) *J. Am. Chem. Soc.* **79,** 2023–2024.
3. Hoogsteen, K. (1959) *Acta Crystallogr.* **12,** 822–823.
4. Morgan, A. R. & Wells, R. D. (1968) *J. Mol. Biol.* **37,** 63–80.
5. Moser, H. E. & Dervan, P. B. (1987) *Science* **238,** 645–650.
6. Letai, A. G., Palladino, M. A., Fromm, E., Rizzo, V. & Fresco, J. R. (1988) *Biochemistry* **27,** 9109–9112.
7. Broitman, S. L., Im, D. D. & Fresco, J. R. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 5120–5124.
8. Beal, P. A. & Dervan, P. B. (1991) *Science* **251,** 1360–1363.
9. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J. & Hogan, M. E. (1988) *Science* **241,** 456–459.
10. Orson, F. M., Thomas, D. W., McShan, W. M., Kessler, D. J. & Hogan, M. E. (1991) *Nucleic Acids Res.* **19,** 3435–3441.
11. Pei, D., Ulrich, H. D. & Schultz, P. G. (1991) *Science* **253,** 1408–1411.
12. Chamberlin, M. J. (1965) *Proc. Fed. Am. Soc. Exp. Biol.* **24,** 1446–1457.
13. Riley, M., Mailing, B. & Chamberlin, M. J. (1966) *J. Mol. Biol.* **20,** 359–389.
14. Eadie, J. S., McBride, L. J., Efcavitch, J. W., Hoff, L. B. & Cathcart, R. C. (1987) *Anal. Chem.* **165,** 442–447.
15. Griffin, L. C. & Dervan, P. B. (1989) *Science* **245,** 967–971.
16. Horne, D. A. & Dervan, P. B. (1991) *Nucleic Acids Res.* **19,** 4963–4965.
17. Mergny, J. L., Sun, J. S., Rougee, M., Montenay-Garestier, T., Barcelo, F., Chomilier, J. & Helene, C. (1991) *Biochemistry* **30,** 9791–9798.
18. Belotserkovskii, B. P., Veselkov, A. G., Filippov, S. A., Dobryniun, V. N., Mirkin, S. M. & Frank-Kamenetskii, M. D. (1990) *Nucleic Acids Res.* **18,** 6621–6624.
19. Michel, F., Ellington, A. D., Couture, S. & Szostak, J. W. (1990) *Nature (London)* **347,** 578–580.
20. Aboul-ela, F., Koh, D. & Tinoco, I., Jr. (1985) *Nucleic Acids Res.* **13,** 4811–4824.
21. Perelroyzen, M. P. & Vologodskii, A. V. (1988) *Nucleic Acids Res.* **16,** 4693–4704.
22. Roberts, R. W. & Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9397–9401.
23. Kool, E. T. (1991) *J. Am. Chem. Soc.* **113,** 6265–6266.
24. Prakash, G. & Kool, E. T. (1991) *J. Am. Chem. Soc.* **114,** 1161–1163.
25. Rajagopal, P. & Feigon, J. (1989) *Nature (London)* **339,** 637–640.
26. Sklenar, V. & Feigon, J. (1990) *Nature (London)* **345,** 836–838.