Very Stable Mismatch Duplexes: Structural and Thermodynamic Studies on Tandem G·A Mismatches in DNA

Susanne Ebel,‡ Andrew N. Lane,§ and Tom Brown*,‡

Edinburgh Centre for Molecular Recognition, Department of Chemistry, Edinburgh University, Kings Buildings, West Mains Road, Edinburgh EH9 3JJ, U.K., and Laboratory of Molecular Structure, National Institute of Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.

Received June 17, 1992; Revised Manuscript Received September 21, 1992

ABSTRACT: We have used ultraviolet melting techniques to compare the stability of several DNA duplexes containing tandem G·A mismatches to similar duplexes containing tandem A·G, I·A, and T·A base pairs. We have found that tandem G·A mismatches in 5′-Y-G-A-R-3′ duplexes are more stable than their I·A counterparts and that they are sometimes more stable than tandem 5′-Y-T-A-R-3′ sequences. This is not the case for tandem G·A mismatches in other base stacking environments, and it suggests that tandem G·A mismatches in 5′-Y-G-A-R-3′ sequences have a unique configuration. In contrast to tandem 5′-G-A-3′ mismatches, tandem 5′-A-G-3′ mismatches were found to be unstable in all cases examined.

Base-pair mismatches are occasionally incorporated into DNA during replication, and if they are not removed by proofreading and repair enzymes, they can give rise to substitution mutations (Friedberg, 1985). Isolated mismatches in synthetic oligonucleotide duplexes have been studied by a variety of methods such as NMR spectroscopy (Patel et al., 1987), X-ray crystallography (Kennard, 1987), and ultraviolet duplex melting techniques (Werntges et al., 1986; Tibanyenda et al., 1984; Gralla et al., 1973; Aboul-ela et al., 1985), but there is little information on multiple misinsertions at a single locus. G·A mispairs are particularly significant as they are removed from genomic DNA less efficiently than other mismatches (Fersht et al., 1982). The enzymic recognition of mispaired bases in DNA depends upon structural and thermodynamic factors (Steitz et al., 1987), and in the case of G·A mismatches the situation is complex as stability and conformation vary according to the base stacking environment (Brown et al., 1989; 1990a,b; Gao et al., 1988; Leonard et al., 1990; Lane et al., 1991). As part of a long-term study on the characteristics of base-pair mismatches in DNA, we now report the results of detailed studies on tandem and multiple G·A mismatches in DNA duplexes.

MATERIALS AND METHODS

DNA Synthesis and Purification. Oligonucleotides were synthesized and purified by standard methods (Beaucage & Caruthers, 1981; Brown & Brown, 1991). Synthesis was carried out on an ABI 380B DNA synthesizer, and HPLC purification was performed on a Gilson HPLC system using a Brownlee Aquapore reversed-phase octyl (C8) column (25 cm × 10 mm) and a gradient of 0% to 20% acetonitrile in 0.1 M ammonium acetate buffer over a period of 30 min.

Melting Curves. Absorbance vs temperature melting curves were measured at 260 nm on a Perkin-Elmer Lambda 15 ultraviolet spectrophotometer equipped with a Peltier block and controlled by an IBM PS2 microcomputer. A heating rate of 0.9 °C min^{-1} was used, and all melting curves were measured in triplicate. Data were collected and processed using the PECSS2 software package (Perkin-Elmer Ltd.). Absorbance versus temperature data were converted to the

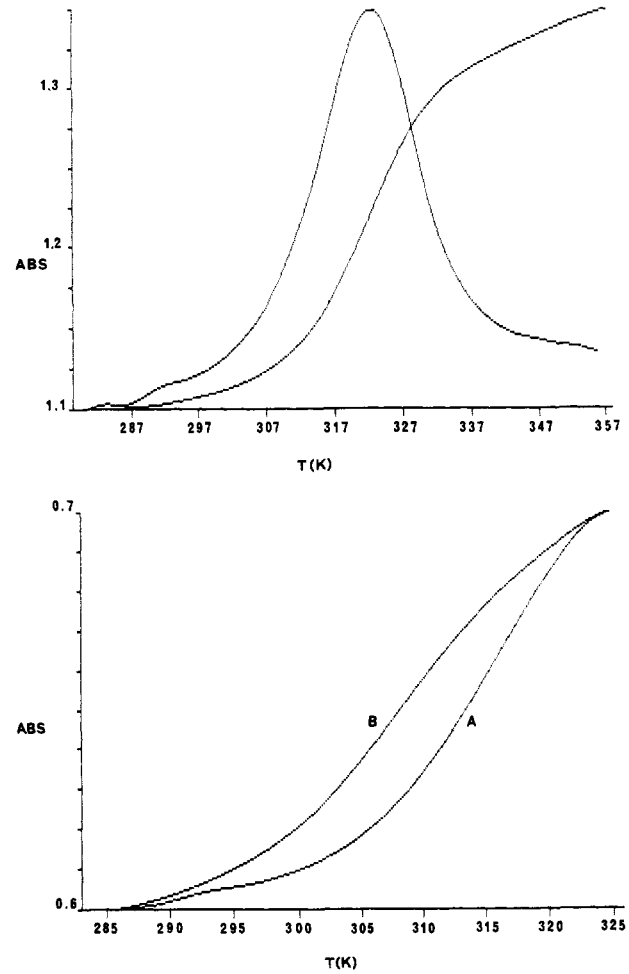

FIGURE 1: Ultraviolet melting curves. (A, top) Melting curve and first derivative for the $d(CCAGTACTGG)_2$ duplex. The melting temperature (323 K) is defined as the maximum point of the first derivative curve. (B, bottom) Superposition of the ultraviolet melting curves of (duplex A) d(CCACGAATGG)·d(CCATGAGTGG); (duplex B) d(CCACIAATGG)·d(CCATIAGTGG). The tandem G·A mismatch (duplex A) is significantly more stable (T_m = 316 K) than the tandem I·A mismatch (duplex B) (T_m = 309 K).

* Author to whom correspondence should be addressed.
‡ Edinburgh University.
§ National Institute of Medical Research.

0006-2960/92/0431-12083$03.00/0

first derivative, and the melting temperature was defined as the temperature at which the curve attained its maximum point (Figure 1A). Almost all the melting temperatures determined in this way were reproducible to within 0.15 K, and the remainder were reproducible within 0.4 K. The concentrations of certain samples were determined as follows: the oligonucleotides were converted to free nucleosides by enzymic digestion using phosphodiesterase 1 (type VII, *Crotalus atrox* venom, 1 unit/mg) and alkaline phosphatase (type III, *Escherichia coli*, 31 units/mg). The concentrations of the original single strands were then calculated using the known base compositions of the oligonucleotides and the published extinction coefficients of the individual nucleosides at 260 nm (Gait & Sproat, 1984). The hypochromicity factor (F) was defined as the ratio of the absorbance of the oligonucleotide single strand at 70 °C to that of the fully digested oligonucleotide at 37 °C, giving the following results: d(GTGAACTT) $F = 0.9$; d(GTTAACTT) $F = 0.9$; d(GTTAAC) $F = 0.7$; d(GTGAAC) $F = 0.7$. Hypochromicity factors of other samples were estimated to be $F = 0.8$. Thermodynamic data for d(GTGAACTT) and d(GTTAACTT) were obtained from concentration-dependent melting studies. The melting curves and first derivatives for the above compounds showed a single transition (Figure 1A,B). This and the detailed NMR study in the following paper strongly support a two-state model for the melting transitions. One particular sequence, d(CCACIAGTGG), produced a biphasic melting curve. At present, we have no explanation for this particular anomaly, but the sequence is currently being investigated in detail by high-field NMR methods. The data points for d(GTGAACTT)$_2$ and d(GTTAACTT)$_2$ were fitted to a van't Hoff plot of T_M^{-1} vs ln C_T, and thermodynamic parameters were calculated using the following equations (Santa Lucia et al., 1991a):

$$1/T_M = (R/\Delta H°) \ln C_T + \Delta S°/\Delta H°$$

(C_T = total concentration of single strands)

$$R/\Delta H° = \text{slope} \qquad \Delta S°/\Delta H° = \text{intercept}$$

$$\Delta G°(298\ \text{K}) = RT \ln C_T \quad \text{and}$$

$$\Delta G°(298\ \text{K}) = \Delta H° - T\delta S°$$

Melting temperatures of all octamers and hexamers were determined in a buffer consisting of 0.1 M NaCl, 0.1 M NaH_2PO_4, 20 mM sodium cacodylate, and 1.0 mM EDTA adjusted to pH 7 by addition of sodium hydroxide. The pH-dependent ultraviolet melting studies on d(GTGAACTT)$_2$ were carried out in buffers of the same salt composition adjusted to pH 5.6, 6.0, 6.5, 7.0, or 7.9 as appropriate. Melting temperatures of decamers were determined in 0.1 M NaCl, 0.01 M NaH_2PO_4, 20 mM sodium cacodylate, and 1.0 mM EDTA at pH 7.

RESULTS AND DISCUSSION

We have carried out ultraviolet melting experiments and high-field NMR studies (see the following paper in this issue) which indicate that in certain cases tandem G·A mismatches are as stable as tandem Watson–Crick T·A base pairs. We first observed this in an NMR study of the sequence d(GTGAACTT), which forms a very stable self-complementary hexamer duplex d(GTGAACTT)$_2$ (paired bases underlined), with 3′ single-stranded dTT ends (Table I) T_m values were measured over a 20-fold concentration range in order to determine thermodynamic parameters (Aboul-ela, 1985) for this unusual mismatched duplex and its Watson–Crick counterpart d(GTTAACTT)$_2$. The unexpected mismatch stability is not due to the 3′-TT extensions as it also occurs in the hexamer d(GTGAAC)$_2$ (Table I).

Studies on a series of decamer duplexes with central 5′-N-G-A-N-3′ sequences established that the phenomenon is sequence-dependent and predictable. If the tetramer 5′-Y-G-A-R-3′ is inserted into a palindromic DNA sequence, the resulting duplex has a much higher melting temperature than that of 5′-Y-A-G-R-3′ sequence (Tables I and IIa,d), and in all but one case the T_m is comparable to that of the equivalent Watson–Crick 5′-Y-T-A-R-3′ sequence (Tables Ia,b and

Table I

(a) Duplex Stability

5′-seq-3′	type	T_m (K)	GTGAACTT at same concn	concn (μM)
GTGAACTT	Y-G-A-R	298.5		28
GTAGACTT	Y-A-G-R	279.0	298.5	28
GTIAACTT	Y-I-A-R	282.0	298.5	25
GTTAACTT	Y-T-A-R	301.0	300.5	42
GTGAAC	Y-G-A-R	293.0	300.5	45
GTTAAC	Y-T-A-R	288.0	301.5	53

(b) Thermodynamic Parameters for GTGAACTT and GTTAACTT

	$\Delta H°$ (kJ/mol)	$\Delta S°$ (J/mol·K)	$\Delta G°$ (kJ/mol)
GTTAACTT	−188.6	−543.1	−26.7
GTGAACTT	−189.8	−547.9	−26.5

Table II [a]

5′-seq-3′	type	T_m (K)
(a)		
CCACGAGTGG	Y-G-A-R	325
CCACAGGTGG	Y-A-G-R	<278
CCACIAGTGG	Y-I-A-R	<305 (biphasic)
CCACTAGTGG	Y-T-A-R	323.0
(b)		
CCAGGACTGG	R-G-A-Y	319.5
CCAGAGCTGG	R-A-G-Y	308.5
CCAGIACTGG	R-I-A-Y	322.5
CCAGTACTGG	R-T-A-Y	323.0
(c)		
CCAAGATTGG	R-G-A-Y	305.5
CCAAAGTTGG	R-A-G-Y	<283
CCAAIATTGG	R-I-A-Y	305.5
CCAATATTGG	R-T-A-Y	314.5
(d)		
CCATGAATGG	Y-G-A-R	308.0
CCATAGATGG	Y-A-G-R	<278
CCATIAATGG	Y-I-A-R	<283
CCATTAATGG	Y-T-A-R	313.5
(e)		
CCAAGAATGG/ CCATGATTGG	R-G-A-R/ Y-G-A-Y	298.0
CCAAIAATGG/ CCATIATTGG	R-I-A-R/ Y-I-A-Y	301.5
(f)		
CCAAGAGTGG/ CCACGATTGG	R-G-A-R/ Y-G-A-Y	310.0
CCAAIAGTGG/ CCACIATTGG	R-I-A-R/ Y-I-A-Y	310.5
CCAATAGTGG/ CCACTATTGG	R-T-A-R/ Y-T-A-Y	317.5
(g)		
CCACGAATGG/ CCATGAGTGG	Y-G-A-R/ Y-G-A-R	316.0
CCACIAATGG/ CCATIAGTGG	Y-I-A-R/ Y-I-A-R	309.0
CCACTAATGG/ CCATTAGTGG	Y-T-A-R/ Y-T-A-R	317.0

[a] Oligonucleotide concentrations were 16–18 μM throughout.

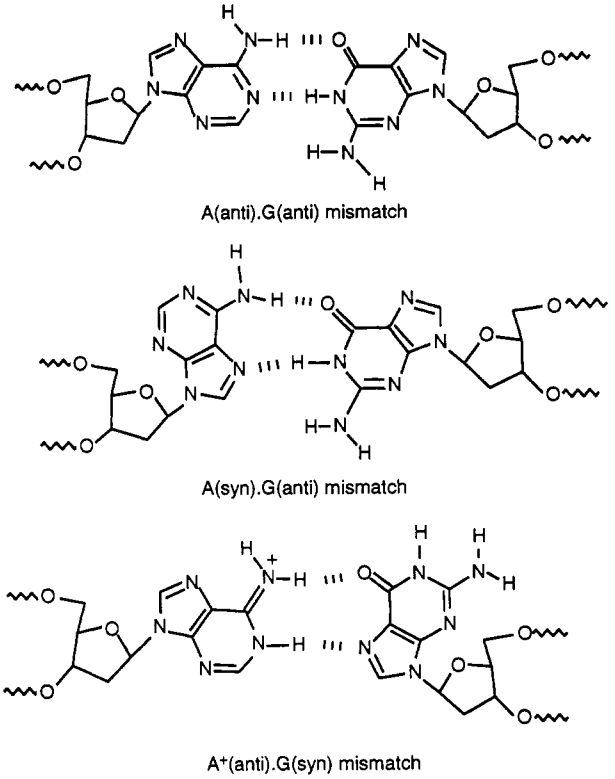

FIGURE 2: Structures of G·A mismatches in the imino pairing form. (Top) A(anti)·G(anti) base pair as observed in the crystal state (Privé et al., 1987) and in solution at neutral pH (Gao & Patel, 1988; Lane et al, 1991). (Center) A(syn)·G(anti). (Bottom) A^+(anti)·G(syn) observed at low pH with protonation of Adenine N1 (Gao & Patel, 1988; Lane et al., 1991). Wavy lines represent connection to the remainder of the molecule.

IIa,d,g). Other flanking base sequences, 5′-R-G-A-Y-3′ or 5′-R-G-A-R-3′/5′-Y-G-A-Y-3′, yield tandem G·A mismatches which are generally *less* stable than their Watson–Crick counterparts (Table IIb,c,e,f). Tandem A·G mismatches are always very unstable (Tables Ia and IIa–d), regardless of the base stacking environment.

Deoxyinosine is an analogue of deoxyguanosine that lacks the 2-amino group and as such is a useful structural probe. It has been used frequently in DNA to evaluate the importance of minor groove interactions of G·C base pairs with proteins and drug molecules. Substitution of deoxyinosine for deoxyguanosine in 5′-Y-G-A-R-3′ sequences produces relatively unstable tandem I·A mismatches (Tables Ia and IIa,d,g and Figure 1B), indicating that the 2-amino group of guanine is crucial to the stability of tandem G·A base pairs in DNA duplexes in this base stacking environment. In contrast, deoxyinosine substitution in 5′-R-G-A-Y-3′ sequences is not destabilizing (Table IIb,c,e,f), and it is therefore clear from these inosine substitution experiments that 5′-Y-G-A-R-3′ duplexes differ from 5′-R-G-A-Y-3′ duplexes in the structural properties of the G·A mispairs.

Three forms of the G·A mismatch have been identified (Figure 2) in X-ray crystal structures and NMR studies of DNA duplexes (Gao & Patel, 1988; Leonard et al., 1990; Lane et al., 1991; Brown et al., 1986; Prive et al., 1987; Kan et al., 1983), and none of these requires direct participation of the guanine 2-amino group in interbase hydrogen bonds. Thus, all three base pairs can be formed between inosine and adenine, and they are all consistent with the melting results for the G·A mismatches found in 5′-R-G-A-Y-3′ sequences.

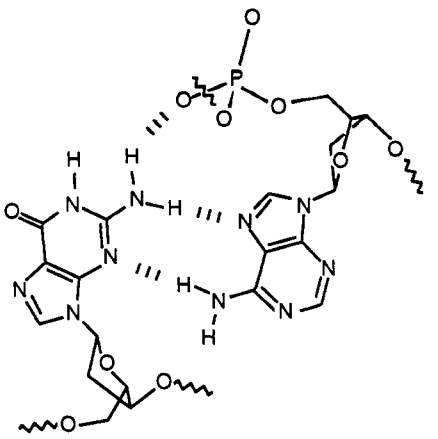

FIGURE 3: Structure of the G·A mismatch with the amino pairing. This drawing is based on the structure proposed by Li et al. (1991) for a tandem G·A mismatch.

Table III

5′-seq-3′	type	T_m (K)	concn (μM)
CGAGTGAACGAG	$(Y\text{-}G\text{-}A\text{-}R)_3$	321.5	17
CGAGTGAGTGAG	$(G\text{-}A\text{-}G\text{-}T)_n$	297.0	17
GAGTGAACGA	terminal G-A	319.0	17
CCTGAGAAGG	Y-G-A-G-A-R	<278	16

None of the base pairs in Figure 2 can be invoked to explain the destabilizing effect of deoxyinosine substitution in 5′-Y-G-A-R-3′ duplexes. The protonated G(syn)·AH^+ (anti) base pair (Figure 2C) can be discounted on additional grounds, as we have shown that the stability of the 5′-Y-G-A-R-3′ sequence in the $d(GTGAACTT)_2$ duplex is invariant in the pH range 5.5–8.0. However, the destabilization caused by inosine substitution is consistent with the tandem G·A mismatch (Figure 3) which has recently been proposed (Li et al., 1991). In this case, the guanine 2-amino group is involved in direct interbase hydrogen bonding and is therefore absolutely required for the formation of a stable base pair. The instability of I·A base pairs relative to G·A base pairs under certain conditions has been observed previously (Wilson et al., 1988).

The above observations led us to investigate the possibility of incorporating large numbers of mismatches in DNA duplexes (Table III). Longer tracts of alternating G·A base pairs are strongly destabilizing, but it is possible to construct a stable 12-mer duplex containing six G·A mismatches provided that they are of the form 5′-Y-G-A-R-3′. A stable duplex forms even when the tandem G·A mismatches are separated by tandem G·T mismatches. Terminal tandem G·A mismatches are also quite stable, and this is illustrated by the surprisingly high T_m (46 °C) of the sequence GAGTGAACGA, which can form a duplex with a 60% G·A mismatch content!

The stability of tandem mismatches in RNA duplexes has recently been examined (Santa Lucia et al., 1990, 1991b), and in some cases a slight increase in stability for G·A versus A·G mispairs and certain structural differences were observed. However, substitution of inosine for guanine had the effect of increasing duplex stability, even for 5′-Y-G-A-R-3′ sequences (Santa Lucia et al., 1991b). The authors measured melting transitions at a salt concentration (1.0 M) much higher than in our studies (0.1–0.2 M). The relatively low salt concentrations in our work were chosen to mimic physiological conditions, as it has been noted previously (Leonard et al., 1990) that buffer concentration can influence G·A mismatch configuration. One explanation for the different trends in the

two studies is that a direct comparison between mismatch configuration in RNA and DNA duplexes may not be possible as the former generally adopt the A-conformation and the latter the B-form. It has been shown that tandem I·A mismatches in RNA can adopt a different hydrogen-bonding pattern to tandem G·A mismatches in the same base-stacking environment (Santa-Lucia et al., 1992).

Stable G·A mismatches of the type shown in Figure 3 may occur in duplex regions of ribozymes, ribosomes, and messenger RNA molecules. If so, this would permit the N(1) and O(6) atoms of guanine and the N(1) atom of adenine to participate in non-base-pairing interactions that might stabilize tertiary structures.

In the following paper, we present NMR and CD data to investigate the structure of tandem G·A mismatches in DNA. The effects of unpaired terminal T·T bases on DNA duplex stability are also discussed.

REFERENCES

Aboul-ela, F., Koh, D., & Tinoco, I., Jr. (1985) *Nucleic Acids Res. 13*, 4811–4824.

Beaucage, S. L., & Caruthers, M. H. (1981) *Tetrahedron Lett. 22*, 1859–1862.

Brown, T., & Brown, D. J. S. (1991) *Oligonucleotides and Analogues, A Practical Approach* (Eckstein, F., Ed.) pp 1–23, IRL Press, Oxford.

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

Brown, T., Leonard, G. A., Booth, E. D., & Chambers, J. (1989) *J. Mol. Biol. 207*, 455–457.

Brown, T., Leonard, G. A., Booth, E. D., & Kneale, G. (1990a) *J. Mol. Biol. 212*, 437–440.

Brown, T., Booth, E. D., & Leonard, G. A. (1990b) in *Molecular Mechanisms in Bio-Organic Processes* (Bleasdale, C., & Golding, B. T., Eds.) Royal Society of Chemistry Press, London.

Fersht, A. R., Knill-Jones, J. W., & Tsui, W. C. (1982) *J. Mol. Biol. 156*, 37–51.

Friedberg, E. C. (1985) *DNA Repair*, W. H. Freeman & Co., New York.

Gait, M. J., & Sproat, B. S. (1984) in *Oligonucleotide Synthesis, A Practical Approach*, Chapter 4, pp 83–116, IRL Press, Oxford and Washington, D.C.

Gao, X., & Patel, D. J. (1988) *J. Am. Chem. Soc. 110*, 5178–5182.

Gralla, J., & Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301–319.

Kan, L.-S., Chandrasegaran, S., Ulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.

Kennard, O. (1987) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D., Eds.) Vol. 1, pp 25–52, Springer Verlag, Berlin, Heidelberg, and New York.

Lane, A. N., Brown, D. J. S., Jenkins, T. C., & Brown, T. (1991) *Biochem. J. 279*, 269–281.

Leonard, G. A., Booth, E. D., & Brown, T. (1990) *Nucleic Acids Res. 18*, 5617–5623.

Li, Y., Zon, G., & Wilson, W. D. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 26–30.

Patel, D. J., Shapiro, L., & Hare, D. (1987) in *Nucleic Acids and Molecular Biology* (Eckstein, F., & Lilley, D., Eds.) Vol. 1, pp 70–84, Springer Verlag, Berlin, Heidelberg, and New York.

Privé, G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.

Santa Lucia, J., Jr., Kierzek, R., & Turner, D. H. (1990) *Biochemistry 29*, 8813–8819.

Santa Lucia, J., Jr., Kierzek, R., & Turner, D. H. (1991a) *Biochemistry 30*, 8242–8251.

Santa Lucia, J., Jr., Kierzek, R., & Turner, D. H. (1991b) *J. Am. Chem. Soc. 113*, 4313–4322.

Santa Lucia, J., Jr., Kierzek, R., & Turner, D. H. (1992) *Science 256*, 217–219.

Steitz, T. A., Beeze, L., Freemont, P. S., Friedman, J. M., & Sanderson, M. R. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 465–471.

Tibanyenda, M., DeBruin, S. H., Hasnoot, C. A. G., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1984) *Eur. J. Biochem. 139*, 19–27.

Werntges, H., Steger, G., Riesner, D., & Fritz, H. J. (1986) *Nucleic Acids Res. 14*, 3773–3790.

Wilson, D. W., Dotrong, M.-H., Zuo, E. T., & Zon, G. (1988) *Nucleic Acids Res. 16*, 5137–5151.

Registry No. A, 73-24-5; G, 73-40-5; T, 65-71-4; I, 58-63-9.