Jamison, C. S., & Degen, S. J. F. (1991) *Biochim. Biophys. Acta 1088*, 208-216.
Kobayashi, K., Lin, L.-W., Yeadon, J. E., Klickstein, L. B., & Smith, J. A. (1989) *J. Biol. Chem. 264*, 8892-8899.
Kok, K., Osinga, J., Carritt, B., Davis, M. G., van der Hout, A. H., van der Veen, A. Y., Landsvater, R. M., de Leij, L. F. M. H., Berendsen, H. H., Postmus, P. E., Poppema, S., & Buys, C. H. C. M. (1987) *Nature 330*, 578-581.
Kovacs, G., Erlandsson, R., Boldog, F., Ingvarsson, S., Muller-Brechlin, R., Klein, G., & Sumegi, J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 1571-1575.
Kozak, M. (1986) *Cell 44*, 283-292.
Magnusson, S., Petersen, T. E., Sottrup-Jensen, L., & Claeys, H. (1975) in *Proteases and Biological Control* (Reich, E., Rifkin, D. B., & Shaw, E., Eds.) pp 123-149, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499-560.
Mount, S. M. (1982) *Nucleic Acids Res. 10*, 459-472.
Nakamura, T., Nishizawa, T., Hagiya, M., Seki, T., Shimonishi, M., Sugimura, A., Tashiro, K., & Shimizu, S. (1989) *Nature 342*, 440-443.
Naylor, S. L., Johnson, B. E., Minna, J. D., & Sakaguchi, A. Y. (1987) *Nature 329*, 451-454.
Naylor, S. L., Marshall, A., Hensel, C., Martinez, P. F., Holley, B., & Sakaguchi, A. Y. (1989) *Genomics 4*, 355-361.
Okajima, A., Miyazawa, K., & Kitamura, N. (1990) *Eur. J. Biochem. 193*, 375-381.
Queen, C., & Korn, L. J. (1984) *Nucleic Acids Res. 12*, 581-599.
Sager, R. (1989) *Science 246*, 1406-1412.
Seizinger, B. R., Rouleau, G. A., Ozelius, L. J., Lane, A. H., Farmer, G. E., Lamiell, J. M., Haines, J., Yuen, J. W. M., Collins, D., Majoor-Krakauer, D., et al. (1988) *Nature 332*, 268-269.
Sharp, P. A. (1981) *Cell 23*, 643-646.
Tashiro, K., Hagiya, M., Nishizawa, T., Seki, T., Shimonishi, M., Shimizu, S., & Nakamura, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 3200-3204.
Tomlinson, J. E., McLean, J. W., & Lawn, R. M. (1989) *J. Biol. Chem. 264*, 5957-5965.
von Heijne, G. (1983) *Eur. J. Biochem. 133*, 17-21.
Watson, M. E. E. (1984) *Nucleic Acids Res. 12*, 5145-5164.
Welch, H. M., Darby, J. K., Pilz, A. J., Ko, C. M., & Carritt, B. (1989) *Genomics 5*, 423-430.
Whang-Peng, J., Bunn, P. A., Jr., Kao-Shan, C. S., Lee, E. C., Carney, D. N., Gazdar, A. F., & Minna, J. D. (1982) *Cancer Genet Cytogenet. 6*, 119-134.
Zbar, B., Brauch, H., Talmadge, C., & Linehan, M. (1987) *Nature 327*, 721-724.

Sequence Specificity in Triple-Helix Formation: Experimental and Theoretical Studies of the Effect of Mismatches on Triplex Stability

Jean-Louis Mergny,‡ Jian-Sheng Sun, Michel Rougée, Thérèse Montenay-Garestier, Francisca Barcelo,§ Jacques Chomilier, and Claude Hélène*
Laboratoire de Biophysique, INSERM U201, CNRS UA481, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75005 Paris, France

Received April 25, 1991; Revised Manuscript Received July 16, 1991

ABSTRACT: The specificity of a homopyrimidine oligonucleotide binding to a homopurine–homopyrimidine sequence on double-stranded DNA was investigated by both molecular modeling and thermal dissociation experiments. The presence of a single mismatched triplet at the center of the triplex was shown to destabilize the triple helix, leading to a lower melting temperature and a less favorable energy of interaction. A terminal mismatch was less destabilizing than a central mismatch. The extent of destabilization was shown to be dependent on the nature of the mismatch. Both single base-pair substitution and deletion in the duplex DNA target were investigated. When a homopurine stretch was interrupted by one thymine, guanine was the least destabilizing base on the third strand. However, G in the third strand did not discriminate between a C·G and an A·T base pair. If the stretch of purines was interrupted by a cytosine, the presence of pyrimidines (C or T) in the third strand yielded a less destabilizing effect than purines. This study shows that oligonucleotides forming triple helices can discriminate between duplex DNA sequences that differ by one base pair. It provides a basis for the choice of antigene oligonucleotide sequences targeted to selected sequences on duplex DNA.

Sequence-specific recognition of nucleic acids is essential for the regulation of cellular functions including replication, transcription, and translation. In most cases, regulation of gene expression in living organisms is achieved by specific nucleic acid binding proteins. In a limited number of cases it has been demonstrated that nucleic acids could also play a regulatory role [see Hélène and Toulmé (1990) for a review]. Homopyrimidine oligodeoxynucleotides can be targeted to a homopurine–homopyrimidine tract of double-helical DNA (Le Doan et al., 1987; Moser & Dervan, 1987; Lyamichev et al., 1988). They form a local *triple helix*, a structure that has been first discovered for homopolynucleotides (Felsenfeld et al., 1957; Stevens & Felsenfeld, 1964). Homopyrimidine oligodeoxynucleotides could control transcription, using what we call the "antigene" strategy (Hélène & Toulmé, 1990). Intermolecular triplex formation occurs upon binding of the third oligopyrimidine strand to the major groove of double-

* Author to whom correspondence should be addressed.
‡ J.-L.M. was supported by a financial grant from the Institut de Formation Supérieure Biomédicale and Rhône-Poulenc.
§ Permanent address: Universitat de les Illes Balears, Dpto. de Biologia i Ciències de la Salut, Palma de Mallorca, Spain.

0006-2960/91/0430-9791$02.50/0

stranded DNA, in a parallel orientation with respect to the homopurine strand of the duplex. Recognition of the sequence is achieved, without denaturation of the double-stranded helix, through Hoogsteen hydrogen bonding of thymine with A·T and protonated cytosine with G·C Watson–Crick base pairs (Rajagopal & Feigon, 1989; Santos et al., 1989; Pilch et al., 1990a).

The demonstration that homopyrimidine oligodeoxynucleotides can bind specifically to homopurine–homopyrimidine tracts of duplex DNA via intermolecular triplex formation has opened a new field of study aimed at designing sequence-specific DNA-binding reagents. Oligonucleotide-directed triple-helix formation may also be used to design tools for molecular and cellular biology, for example artificial endonucleases that can be used to map genes on chromosomes (Moser & Dervan 1987; Hélène et al., 1989; Strobel & Dervan, 1990) or reagents for site-directed mutagenesis. These oligonucleotides could artificially control gene expression by competing for protein binding sites on DNA (François et al., 1989; Maher et al., 1989). All these applications require a precise knowledge of the sequence specificity of triple-helix formation in order to ensure a high specificity of biological effects (Van Vlijmen et al., 1990). The presence of secondary binding sites involving slightly different sequences could seriously hamper the use of this approach. It should be noted that purine-rich oligonucleotides can also form triple helices at DNA sequences that are mostly homopurine–homopyrimidine tracts (Cooney et al., 1988; Beal & Dervan, 1991). Such triplexes can inhibit transcription (Cooney et al., 1988; Orson et al., 1991). Here we will discuss the sequence specificity in the recognition of homopurine–homopyrimidine sequences by homopyrimidine oligonucleotides. The goal of the present theoretical and experimental studies was to get a better understanding of the energetic and structural characteristics of triple-helix formation underlying the sequence-specific recognition of double-stranded DNA by oligonucleotides.

The following convention will be used for the assignment of the strands involved in triple-helix formation: the homopyrimidine strand engaged in Watson–Crick base pairing will be called strand I, the homopurine strand having both Watson–Crick and Hoogsteen base pairing will be called strand II, and the homopyrimidine strand that binds to strand II via Hoogsteen pairing will be called strand III. The nucleotides involved in the base triplet at a mismatched site will be called X, Y, and Z in strand I, II, and III respectively. This triplet will be called X·Y*Z.

Materials and Methods

Chemicals. All unmodified oligonucleotides were synthesized by Institut Pasteur (Paris) on the 0.2-μmol scale. They were dissolved in 200 μL of bidistilled water, precipitated with 5 volumes of ethanol in the presence of 0.2 M sodium acetate, and washed with ethanol. The pellet was lyophilized, resuspended in bidistilled water, and then gel-filtered on a Sephadex Quick spin G25 column. The oligonucleotide was then reprecipitated and resuspended in bidistilled water. Purity and concentration of all oligonucleotides were estimated by UV absorption at 20 °C, with use of the sequence-dependent absorption coefficients given by Cantor and Warshaw (1970). Some oligonucleotides were also purified by reverse-phase HPLC, with a linear gradient (5–30%) of acetonitrile. All other products were Sigma commercial reagents.

UV Absorption. Thermal denaturation profiles were obtained with a Kontron Uvikon 820 or 940 spectrophotometer, with use of quartz cuvettes of 1-cm optical path length. The six-sample cell holder was thermostated by a circulating liquid (70% water/30% glycerol). Temperature of the liquid was controlled with a Haake water bath, and temperature of one of the cuvettes was measured with a thermoresistance. The temperature of the bath was increased or decreased at the rate of 0.1 deg/min, thus allowing complete thermal equilibrium of the six cells between two measurements. Reading of absorbance at 260 nm was performed every 8 min. All strand concentrations were 1 μM, leading to a total absorbance around 0.5 for a triplex. All melting curves concerning samples in a 1 M salt buffer were reversible, giving the same profile independently of whether the temperature was raised from 0 to 60 °C or decreased from 60 to 0 °C. For samples in 0.1 M salt buffer, a reproducible hysteresis was observed, leading to different melting profiles with increasing or decreasing temperatures. This hysteresis phenomenom, a consequence of slow kinetics of triplex formation, was observed for triplexes involving 11–22 base triplet oligonucleotides at low ionic concentration. It disappeared at 1 M NaCl concentration due to an increase in the association rate for triplex formation (M. Rougée et al., unpublished observations).

Molecular Modeling. Conformational energy minimization was performed with the JUMNA program (Lavery, 1988) using helicoidal coordinates, particularly suitable for nucleic acids. The net charge on each phosphate group was set to −0.5e, equally shared by the two ionized oxygens, to mimic counterion screening effects. Explicit counterions were included in the ratio of one sodium per phosphate group. No water molecules were taken into account in this calculation. Their effect was simulated by use of a sigmoidal distance-dependent dielectric function (Lavery et al., 1986).

The triple helix was constructed from a library of nucleotides by specifying the desired helicoidal parameters in agreement with the Cambridge convention (Dickerson et al., 1989). Helicoidal coordinates were derived from atomic ones taken from fiber diffraction data of an A-like DNA triple helix (Arnott et al., 1976). They were repeatedly rotated around the helical axis by the twist angle (30.2°) and translated by the rise per base pair (3.24 Å) to generate consecutive triplets. These structural data allow the generation of poly(dT)·poly(dA)·poly(dT) homopolymers only. However, the use of helicoidal coordinates instead of atomic ones enabled us to substitute any nucleotide easily. The mismatched nucleotide was first carefully docked by inspection on a 3D graphic monitor to be in a good hydrogen-bonding orientation with respect to strand II prior to minimization. The glycosidic angles and the backbone coordinates were initially kept constant relative to those of strand III whenever possible, to give a similar backbone conformation regardless of the mutation. The helicoidal variables, the sugar as well as the phosphate backbone variables, were locked at fixed values. Typically, the first step in the minimization procedure consisted of locking sugar and helicoidal parameters for all nucleotides, so that the backbones alone were released. In the second step each base triplet was constrained to move as a fixed entity except the mismatched nucleotide, which was free to evolve. In the third step only inclination and tilt angles were locked except for the mismatched nucleotide; finally, all conformational parameters were minimized without constraints. Minimization was performed by successively decreasing the number of constraints until the energy convergence criterion was reached, i.e., a difference between two steps equivalent to 10^{-5} kcal/mol. No cutoff distance function was used except for hydrogen bonds, which was set at 6 Å. We have checked that the results were not significantly affected by the order under which the con-

Table I: Sequence of the Oligonucleotides Used in Experiments of Thermal Dissociation and Half-Dissociation Temperature Values Obtained for All X·Y*Z Triplets[a]

(A) Sequence of the Oligonucleotides

31 mer (double-stranded)

```
5' AATCTCCTCCTTTATACTCTTCXTTTTTCAA 3'
3' TTAGAGGAGGAAATATGAGAAGYAAAAAGTT 5'
      3' CTCTTCZTTTTTC 5'
          13 mer
```

(B) Half-Dissociation Temperature Values [T_m (°C)]

XY	Z[b] A	C	G	T
TA	23.5	16.5	21	37
CG	26	38	29	25
AT	17.5	11	26	15.5
GC	18	23	20.5	22.5
no X, no Y	21	23	21	28.5

[a] The T_m for a target with a single base-pair deletion (no X, no Y) is also given. [b] Z is the base on the third strand in the 13 mer shown above.

straints were released except in the first steps of minimization where strong close contacts could exist. In order to check that the energy-minimized models were stable, several starting structures were used. When the mismatched nucleotides were docked in a syn conformation an unfavorable intrastrand interaction of strand III was obtained due to the lack of base stacking at both anti–syn and syn–anti steps. Therefore, only anti conformations were considered. The values given in Tables II and III are the lowest energies obtained when starting from different structures.

In order to get a better insight into the interactions between the nucleotides in the miminized triplexes, the interaction energies were decomposed into terms of interstrand energies (i.e., I–II, I–III, II–III) and intrastrand ones (i.e., I–I, II–II, III–III) as well as energies between specific nucleotides or groups of atoms (i.e., base–base, sugar–base, phosphate–base, etc.). The total of interstrand interactions between strand III and the Watson–Crick double helix was characterized by the sum of I–III and II–III. The interaction energies were expected to display substantial alterations at and around the sites of mismatched nucleotides. One should keep in mind that only free energies would be relevant to describe the stability of triplexes and thus could be directly compared to the thermal dissociation experimental results; however, the use of interaction energies is an acceptable and useful tool. Nevertheless, extreme care should be taken in interpreting such results, because the hydration effect, for instance, was not correctly treated in this simple model. Structural properties of the minimized triplexes were analyzed by use of the program CURVES (Lavery & Sklenar, 1989), which uses an algorithm to calculate helicoidal coordinates of any irregular, multistranded nucleic acid segment with respect to an optimized, global helical axis.

These computations were performed on a Silicon Graphics 4D/120 workstation. Typical CPU time per step was about 18 s, and each computation needed about 2000 iterations. The molecules were visualized with the help of the INSIGHT II software commercialized by BIOSYM, fully interfaced with JUMNA in our laboratory.

RESULTS

Thermal Dissociation Studies by UV Absorption

Effect of a Single Mismatch on the Stability of a 13-Base-Long Triple Helix. The substrate chosen for oligonucleotide binding was a synthetic DNA fragment of 31 base pairs containing a 13-base-pair homopurine–homopyrimidine

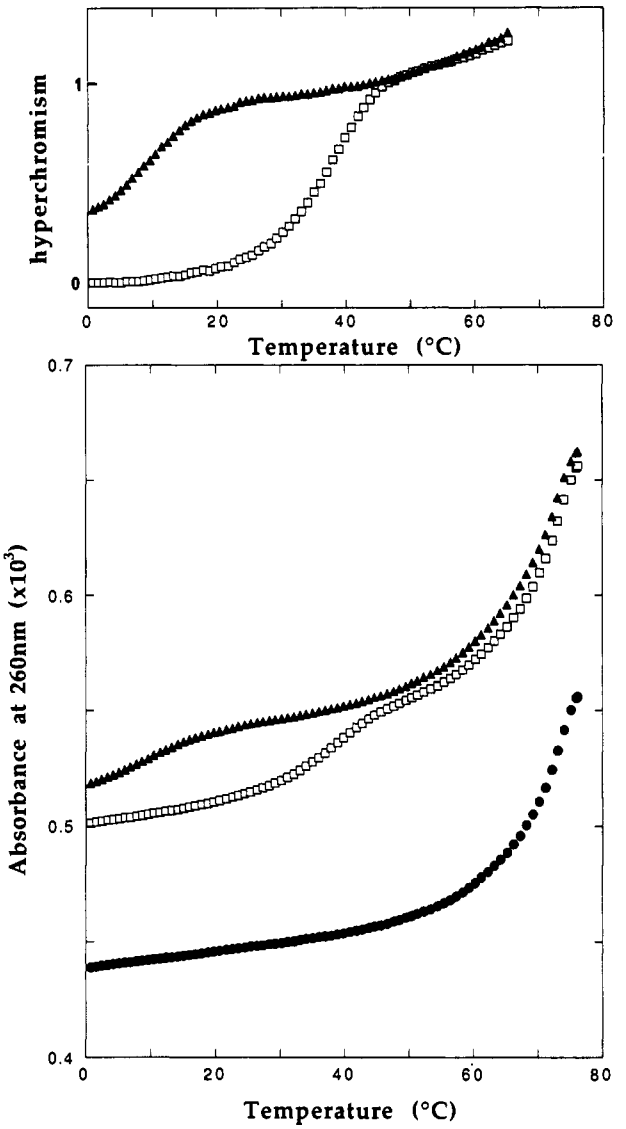


FIGURE 1: (Bottom) Examples of absorbance as a function of temperature for a duplex alone (●), a duplex with a fully complementary 13-base-long oligonucleotide (X·Y*Z = T·A*T) (□), or the most destabilizing mismatch (X·Y*Z = A·T*C) (▲). All experiments were performed in a pH 5.6 buffer containing 1 M salt and 100 mM cacodylate. All melting curves were reversible in these experimental conditions. (Top) Melting profile after subtraction of the contribution from the duplex-to-single strand transition.

sequence (Table I), thus allowing a 13-base-long pyrimidine oligonucleotide to form a triple helix. It was necessary to use a high salt concentration (0.1 M sodium cacodylate, 1 M NaCl) and a relatively low pH (5.6) to be able to observe the triplex-to-duplex transition in the case of a non fully complementary homopyrimidine oligonucleotide. The longer DNA duplex allowed us to easily distinguish between the triplex-to-duplex and the duplex-to-single strand transition.

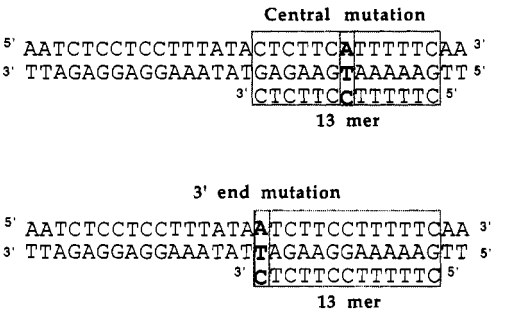
Triple-helix formation, as previously reported, is associated with a hypochromism at 260 nm (Sun et al., 1989; Pilch et al., 1990a,b; Plum et al., 1990; Shea et al., 1990; Xodo et al., 1990). Thus, thermal dissociation studies allowed us to determine half-dissociation temperatures after subtraction of the duplex-to-single strand transition curve from the global melting curves (Figure 1). Under our experimental conditions, a triple helix with 13 base triplets had a T_m of either 37 °C (X·Y*Z = C·G*C$^+$) or 38 °C (X·Y*Z = T·A*T) for a totally complementary target and a T_m between 11 and 29 °C for a target with a single base pair substitution (see Table I). The highest destabilization was obtained with a C opposite an A·T base

pair (A·T*C); the smallest was obtained with G opposite C·G (C·G*G). When the target sequence contained only purines on one strand and pyrimidines on the other strand, the C·G*G triplet was the least destabilizing as compared to the usual base triplets C·G*C and T·A*T. The largest effect was obtained with C opposite T·A (T·A*C). When the purine stretch was interrupted by one T in the center, A·T*G was the least destabilizing while C opposite A·T (A·T*C) had the strongest effect among all mismatches as stated above. When the stretch of purines was interrupted by a C, the presence of pyrimidines (C or T) in strand III gave a less destabilizing effect than purines. The melting curve for the T·A*T triplet was fit to a simple bimolecular equilibrium model (duplex + single strand = triplex). A ΔH value of -80 ± 10 kcal/mol was estimated. The large uncertainty arises from the difficulty in obtaining good base lines before and after the triplex-to-duplex transition. When the concentration was raised from 1 to 9 μM, the T_m increased by 5 °C, in agreement with the estimated ΔH value. The difference in ΔH between the mismatched and canonical triplets falls within the uncertainty limits, as an increase in concentration from 1 to 9 μM for several of these triplets gave rise to a similar increase in T_m ($\approx$5 °C). A much broader concentration range (at least 100-fold) would be needed to obtain reliable ΔH values. The T_m values reported in Table I were obtained at a concentration of 1 μM. Therefore, at the T_m the association constant is 2×10^6 M^{-1} ($\Delta G \approx -9$ kcal/mol). Using the estimated ΔH value for the T·A*T triplet, it is possible to calculate $K_{T\cdot A^*T}$ at the T_m of the mismatched triplets. At 29 °C, the difference in binding free energy between T·A*T (or C·G*C) and C·G*G triplet is 1.8 ± 0.34 kcal/mol. At 26 °C, the T·A*T triplet is more stable than the A·T*G triplet by 2.5 ± 0.4 kcal/mol. At 11 °C, the T·A*T triplet is stabilized by 5.9 ± 0.8 kcal/mol as compared to the most destabilizing mismatch A·T*C.

The thermal stability of a longer triplex, formed with 22 triplets, containing a point mutation in a central position, was also investigated in a buffer with a lower ionic concentration (sodium cacodylate 10 mM, 0.1 M NaCl) and a slightly higher pH (pH 6). With this 22 mer, and under these experimental conditions, whatever the central X·Y*Z triplet, a hysteresis of the melting curves was observed, giving a temperature of half-dissociation T_d (i.e., starting with a preformed triplex and increasing the temperature) higher than the temperature of half-association T_a (starting at high temperature with a 22 mer homopyrimidine third strand unbound to its target). This phenomenom was the consequence of very slow association and dissociation kinetics, which will be described in detail in another paper. All the 16 X·Y*Z triplets, resulting from all X·Y*Z combinations could be classified in roughly the same order by their T_d as presented above for the 13 mer.

Effect of the Location of the Mutation on Triple-Helix Stability. To investigate the effect of position of a given mutation in a triplex, a double-stranded 31 mer with a C·G to A·T mutation at the end of the homopurine–homopyrimidine stretch was synthesized. The transformation of a C·G*C to an A·T*C triplet was chosen as it was the most destabilizing mutation (11 °C instead at 38 °C) in the center of the sequence (Table I). This mutation at the end of the triplex led to a much smaller decrease in triplex stability, as the T_m was reduced by only 7.5 °C (Figure 2) as compared to that of the matched triplet.

The positional effect of mismatches on destabilization of triplex structures is therefore similar to what was previously observed with duplexes (Wallace et al., 1981). A higher discrimination between matched and mismatched sequences



	Site of mutation		C.G*C+
	Central	3' end	
Tm (°C)	11	30.5	38

FIGURE 2: Effect of mismatch location on triplex stability. (Top) Structure of the mismatched oligonucleotides used in this experiment. (Bottom) T_m values obtained for these triple-stranded structures.

was observed when the mutation occurred in the center of the target sequence rather than at its ends. This effect is due to the fact that a terminal triplet interacts with only one neighbor in contrast to a central triplet, which interacts with two neighbors. Consequently, the cooperativity between base triplets is strongly perturbed by a mismatch in the center of the sequence.

Effect of Base-Pair Deletion in the Double-Stranded Target. The effect of a deletion of the XY base pair in the double helix was also studied. The target in this case was a 30-base-pair-long double-stranded oligonucleotide, with a 12-base-pair-long homopurine–homopyrimidine stretch. In order for a 13-mer homopyrimidine oligonucleotide to be fully hybridized, the Z base on strand III must be looped out.

As shown in Table I (lower lane), whatever Z was, a triplex-to-duplex transition was still observed. The amplitude of the transition led to the conclusion that the six bases on both sides of Z were hybridized, forming a bulge with a protuberant Z base. The stability of these triplexes depended on Z; a purine was less favorable than a cytosine, or a thymine, probably because pyrimidines are less bulky and/or purines tend to stack more with neighboring bases in the third strand. The smallest difference in terms of thermal stability between a fully matched 13-base-triplet sequence and the one with 12 base triplets (8.5 °C) was observed with Z = T.

Molecular Modeling Study and Energy Minimization

The effects of a single mismatch involving the Hoogsteen-bonded strand on triple-helix stability were investigated by minimizing the conformational energy for various sequences. For reasons of acceptable computation time, the length of triple helix was limited to seven triplets, where the mismatched nucleotide involved in the Hoogsteen-bonded strand was located in the central position. The end effect was checked to be negligible, if any. Preliminary studies showed that the explicit counterions included in the calculation played an important role in order to obtain self-consistent esults.

The sequence of the triple helix studied by energy minimization modeling is shown in Table II, where the central part of strand III has the sequence $^{5'}$TTTZCTT$^{3'}$ around the mismatch site. The Z nucleotide could be adenine, neutral or protonated cytosine, guanine, thymine, or an abasic residue. The effects of nearest neighbors on both sides of the mismatch site were also studied by molecular modeling. They will be called TZC, TZT, CZT, and CZC triplexes where the se-

Table II: Decomposition of the Interaction Energies (kcal/mol) of Different Triplexes Where the Mismatched Nucleotide Involved in the Hoogsteen-Bonded Strand Was Located in the Central Position[a]

7 mer

5' TTCXTTT 3'
3' AAGYAAA 5'
3' TTCZTTT 5'

strands	C+	C	T	A	G	none
(1) Triplexes with T·A*Z Triplets						
I–II (WC)	−118.0	−118.0	−119.0	−118.8	−119.4	−118.1
II–III (H)	−141.4	−143.1	−161.6	−140.0	−151.5	−138.3
I–I	−70.8	−72.6	−72.9	−72.9	−71.4	−73.0
II–II	−67.0	−68.1	−67.4	−66.8	−64.8	−68.1
III–III	−77.0	−84.7	−61.8	−64.4	−56.7	−48.1
I–III	−16.6	−0.8	3.1	1.1	5.3	3.9
total	−490.9	−487.4	−479.5	−461.8	−458.6	−441.7
(2) Triplexes with A·T*Z Triplets						
strands	G	C+	A	T	C	none
I–II (WC)	−119.3	−120.7	−116.6	−119.4	−117.4	−119.9
II–III (H)	−137.6	−134.5	−131.4	−138.5	−127.5	−134.0
I–I	−69.2	−66.5	−73.5	−68.9	−69.1	−71.9
II–II	−66.3	−63.7	−65.1	−65.7	−65.7	−67.7
III–III	−63.8	−52.0	−63.6	−44.5	−65.4	−47.5
I–III	−4.0	−19.9	−7.1	−10.3	−0.1	3.5
total	−460.2	−457.4	−457.3	−447.2	−445.3	−437.6
(3) Triplexes with C·G*Z Triplets						
strands	C+	C	A	G	T	none
I–II (WC)	−135.3	−133.8	−135.1	−136.9	−136.3	−135.4
II–III (H)	−183.2	−158.3	−151.5	−153.2	−147.7	−144.2
I–I	−73.2	−75.7	−74.4	−74.1	−74.0	−74.0
II–II	−65.5	−64.6	−64.6	−64.5	−66.0	−67.9
III–III	−62.6	−67.1	−69.9	−64.4	−55.8	−46.9
I–III	−0.3	3.4	4.0	3.5	2.0	4.8
total	−520.1	−496.1	−491.5	−489.6	−477.8	−463.6
(4) Triplexes with G·C*Z Triplets						
strands	C+	C	T	G	A	none
I–II (WC)	−134.5	−136.4	−136.7	−136.9	−136.4	−134.0
II–III (H)	−140.3	−142.9	−152.9	−140.2	−132.1	−137.9
I–I	−72.2	−71.4	−71.1	−70.1	−67.8	−70.6
II–II	−62.0	−67.8	−68.7	−65.5	−62.9	−67.3
III–III	−72.8	−74.9	−60.9	−55.6	−43.5	−41.7
I–III	−11.3	0.3	3.1	−8.7	−14.5	−2.4
total	−493.2	−493.0	−487.2	−476.9	−457.2	−453.9

[a] I–I, II–II, and III–III designate the intrastrand energies of strands I, II, and III respectively. I–II, I–III, and II–III are the interstrand energies between different strands. None indicates triplets with an abasic residue in strand III.

quence of the central part of strand III was 5'TTTZCTT3', 5'TTTZTTT3', 5'TTCZTTT3', and 5'TTCZCTT3', respectively. Only the energetic and structural properties of the TZC triplexes will be described in some details. Other cases (TZT, CZT) will be mentioned when the effects of nearest neighbors are evidenced.

Energetic and Structural Characteristics. The interaction energies of the TZC triplexes are listed in Table II. As expected, introduction of an abasic residue in strand III caused the largest destabilization of the triplexes compared to the standard triplex. This destabilization was mainly localized at the abasic site and mainly due to two interaction energy terms: (i) the interstrand interaction energies between strands II and III engaged in Hoogsteen base pairing due to the lack of hydrogen bonds at the abasic site and (ii) the intrastrand energies within strand III due to the loss of base–base interactions (part of stacking energy), as well as sugar–base and phosphate–base interactions. The backbone of strand III remained unchanged with respect to the T·A*T or C·G*C+ triplet.

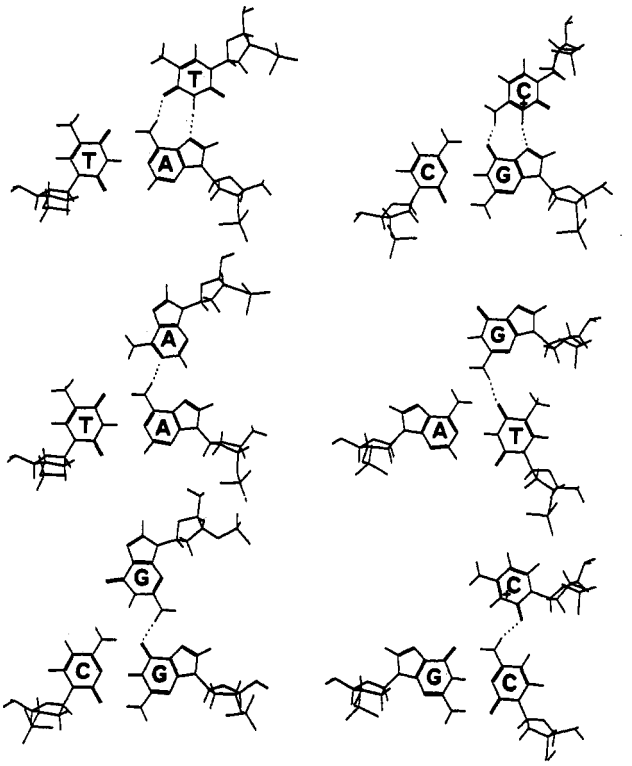


FIGURE 3: Some matched or mismatched triplets observed in the minimized triplexes: T·A*T triplet (upper left); C·G*C+ (upper right); T·A*A (middle left); A·T*G (middle right); C·G*G (lower left); and G·C*C+ (lower right). Hoogsteen-like hydrogen bonds are shown by dashed lines.

The total interaction energy should give an indication of the affinity of strand III for the Watson–Crick base pairs. All triplexes with mismatched nucleotides involved an unfavorable replacement of one triplet, and a less favorable interaction energy was expected. The interaction energies of the minimized triplexes fully confirmed this assumption except for the triplets T·A*C+ and T·A*C (see paragraph below). Note that the energies of the triplets T·A*Z (A·T*Z) and C·G*Z (G·C*Z) should be compared only with that of the canonical triplet T·A*T or C·G*C+, respectively, due to the difference of interaction between A·T (T·A) and G·C (C·G) base pairs in the duplex. The results are consistent with the interaction between strands II and III (Hoogsteen pairing) being affected by mismatched bases. For this interaction, the two canonical triplexes have the strongest interaction in all cases. However, the interaction between strands I and III cannot be neglected in some cases. The perturbation induced by a single mismatched nucleotide in the triple helix seems to affect slightly the interactions of Watson–Crick base pairs already involved in the triplex (see rows I–I, I–II, and II–II in Table II) as compared to those of abasic sequences. Some mismatched nucleotides in strand III can form not only one but even several hydrogen bonds and not only with the bases in strand II as expected but also with bases in strand I as well (Table III). Some of these interactions are illustrated in Figure 3.

As mentioned above, the total interaction energy of the triplets T·A*C+ and T·A*C is lower than that of the standard triplet T·A*T. This is in conflict with our experimental results obtained by measurements of thermal dissociation curves. As expected, the lack of Hoogsteen hydrogen bonding is reflected in a loss of energy (18–20 kcal) in strand II–III interactions: a T·A*T triplet has much more favorable II–III interaction energy than any other T·A*Z triplet. However, this is more than compensated by intrastrand interactions of strand III and

Table III: Structural Characteristics of a Point Mutation in the Minimized Triple Helix of Various Triplets X·Y*Z at the Mismatched Site[a]

XY	Z				
	A	C	C+	G	T
TA	H2N6(A)–N1(A) 2.0 Å, 10.0°	H2N6(A)–N3(C) 2.1 Å, 22.0°	O4(T)–HN3(C+)[e] 2.1 Å, 14.0°	H2N6(A)–O6(G) 2.0 Å, 17.0°	N7(A)–HN3(T) 2.1 Å, 10.0° H2N6(A)–O4(T) 1.9 Å, 17.0°
AT	–[b]	–[b]	–[b]	O4(T)–H2N2(G) 2.0 Å, 21.0° –[c]	–[b]
CG	–[d]	N7(G)–H1N4(C) 2.4 Å, 32.0°	N7(G)–HN3(C+) 1.9 Å, 7.0° O6(G)–H1N4(C+) 2.0Å, 21.0°	O6(G)–H1N2(G) 1.9 Å, 18.0°	N7(G)–HN3(T) 2.2 Å, 22.0°
GC	–[b]	H2N4(C)–N3(C) 2.0 Å, 13.0°	H2N4(C)–O2(C+) 2.0 Å, 7.0°	–[b]	H2N4(C)–O4(T) 2.0 Å, 25.0°

[a] Observed hydrogen bonds between acceptor and donor atoms that belong to different nucleotides (indicated in the parentheses) are given. Length of hydrogen bonds is defined by the distance separating the hydrogen atom in the donor group from the oxygen or nitrogen atom in the acceptor group. Angles of hydrogen bonds are measured by the angle of N-H···O or N-H···N. [b] No hydrogen bond and the mismatched nucleotide is out of strand III. [c] The base G has an important tilt resulting in a weak hydrogen bond between the donor (H1N2) of guanine and the acceptor O4 of thymine in strand I, which is above or below the X·Y*G triplet. [d] No hydrogen bond but the base A is stacked within strand III. [e] Hydrogen bond across the thymine in strand I and the protonated cytosine in strand III.

interstrand interactions between strand I and III, especially in the case of the triplet T·A*C+. Because protonated cytosine was not able to adopt a suitable hydrogen-bonding orientation with respect to the adenine, it was constrained to move into the major groove resulting in an unusually strong phosphate–base electrostatic interaction between strands I and III. But even with neutral cytosines, the mismatched cytosines were constrained in an extrahelical position on strand III, leading to a strong intrastrand interaction of strand III due to base–phosphate as well as base–base proximity on the 3′-side of the mismatched site. It seems that this overestimated interaction is due to the approximations in accounting for hydration effect (Cieplak et al., 1987). It can be seen that when Z = C or C+ in the mismatched triplets G·C*Z, the intrastrand interaction of strand III was large compared to the triplexes with other mismatched triplets G·C*Z. Detailed energy analysis revealed that this is due to enhanced intrahelical base stacking, which can explain why the cytosines are the least destabilizing nucleotides of all other mismatched triplets G·C*Z.

Nearest-Neighbor Effect. An exhaustive computation of all the TZT, TZC, CZC, and CZT sequences on strand III triplexes revealed nearest-neighbor effects. Minor effects were observed in the TZT series. When a C·G*T triplet was flanked on both sides by T·A*T, the interaction energy of this triplet was lower than that of the triplets C·G*A and C·G*G. This exception is ascribed to the increase of intrastrand energies in strand III, due to van der Waals interaction between the stretch of methyl groups at position 5 of thymines in this strand.

Other effects occurred when a nucleotide involved in a mismatched triplet could make a hydrogen bond between strands I and III, for instance when a triplet A·T*G was flanked on both sides by T·A*T. A hydrogen bond between the acceptor O4 of thymine on strand I and the NH2 of guanine on strand III was observed, leading to a tilt of guanine, as this bond involves two bases on different triplets. On the contrary, if the triplet A·T*G was flanked on both sides by C·G*C+ (i.e., series of CZC triplexes), no GT hydrogen bond was possible and the tilt of guanine was consequently reduced.

Discussion

We have studied the effect of a single mutation on triplex stability. The central base pair, called XY on the double-stranded target, could be any of T·A, C·G, A·T, or G·C base pairs. The first two possibilities do not interrupt the homopurine sequence: a non-interrupted triplex with no mismatch can be formed, provided that the corresponding base in strand III is a T in the first case and a C in the second case. These two oligonucleotides gave similar thermal dissociation profiles from the double helix under our experimental conditions, with a T_m of 37–38 °C. Moreover, the enthalpies of triplex formation, derived from the slope of the denaturation curve, were almost identical (−80 kcal/mol), suggesting that at pH 5.6, a C·G*C+ triplet is as stable as a T·A*T triplet. At higher pH a T·A*T triplet should be more stable than C·G*C, as protonation of the cytosine on strand III is required to form two Hoogsteen hydrogen bonds. All T_m values with any of the 14 other X·Y*Z combinations were significantly lower.

The main conclusion of molecular modeling studies of the triple helix in which a mismatched nucleotide is located in the central position of the Hoogsteen-bonded strand can be summarized as follows: when the mismatched nucleotide (Z) is a purine, the destabilizing effect is less important than for a pyrimidine except when the Watson–Crick base pair is G·C with C located in the homopurine strand, in which case a pyrimidine is favorable (Table II). These observations are in very good agreement with the results obtained by thermal dissociation experiments (Table I). In a recent paper, Frank-Kamenetskii and co-workers (Belotserkovslii et al., 1990) have also shown, using two-dimensional gel electrophoresis, that a mismatched triplet G·C*Z where Z is a pyrimidine requires less negative superhelical stress to adopt an H-DNA intramolecular triplex structure than when Z is a purine. This exception is the consequence of steric hindrance due to the bulky amino group at position 4 of cytosine in strand II and the mismatched purine nucleotide in strand III. A mismatched purine nucleotide can no longer adopt any Hoogsteen-like orientation and is expelled into the major groove at the expense of the stacking energy of strand III. On the other hand, a pyrimidine can be well stacked within strand III and a hydrogen bond can be formed between H_2N4 of cytosine in strand II and the acceptor groups of cytosine (Table III) or thymine in strand III.

On the contrary, almost all mismatched nucleotides are expelled out of strand III in the case of triplets A·T*Z. This is due to the presence of the methyl group at position 5 of thymine in the triplet. This is probably why the T_m values and the interaction energies are lower than those of any other mutations. Only guanine in the third strand escapes this rule because only a mismatched guanine can make a hydrogen bond

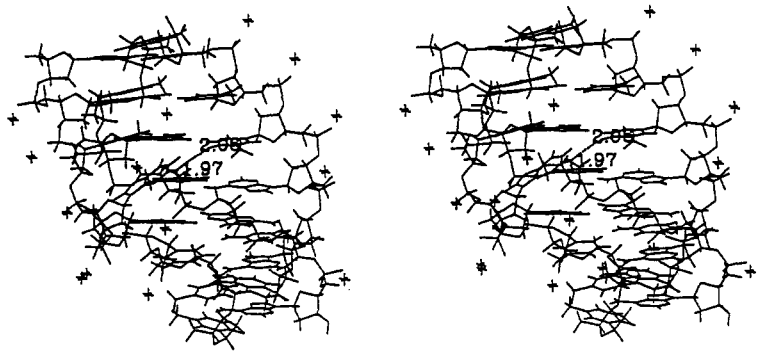


FIGURE 4: Stereoview of the triplex with a mismatched A·T*G triplet at the central position. The mismatched guanine is tilted out of the mean plane of the triplet A·T*G forming the hydrogen bond (1.97 Å) shown in Figure 3 (middle right) and another hydrogen bond (2.06 Å) with the acceptor O4 of the thymine in strand I, which is above the A·T*G triplet. This tilt is reduced and the I–III strand hydrogen bond is abolished when the A·T*G triplet is flanked on both sides by standard C·G*C$^+$ triplets.

between its amino group and the O4 of thymine (Table III), leading to the least unfavorable mutation in this series. However the guanine base is quite tilted out of the average plane of the triplet A·T*G, forming loose hydrogen bonds with O4 of thymines of strand I, which are above or below the A·T*G triplet (Figure 4). A similar feature was also observed for other mismatched triplets (C·G*G, etc.). This type of I–III strand hydrogen bonding is dependent upon the flanking triplets (see Results). The triplet A·T*G is one among the less unfavorable mismatches but the T_m is still significantly lower (11–12 °C) with an A·T*G triplet than with a T·A*T or C·G*C$^+$ triplet. Therefore, if a T interrupts a homopurine sequence then a G in strand III will give less destabilization of the triplex than any other base as previously reported by Griffin and Dervan (1989). However, this triplet is much less stable than that formed with an uninterrupted homopurine sequence. In addition, from the selectivity point of view, a G-containing strand III will not distinguish a C·G from an A·T base pair in the duplex sequence (see Table I): the C·G*G base triplet is even more stable than the A·T*G triplet. The discrepancy between our results and those of Griffin and Dervan might be due to nearest-neighbor effects [T·A*T on the 5′-side and C·G*C$^+$ on the 3′-side in our studies, T·A*T on both sides in Griffin and Dervan (1989)].

The perturbation induced by a single mismatched nucleotide in triplexes is limited to nearest neighbors of the triplet. The flanking triplets on both sides of the mismatch site may have some minor effect on the relative stability and on the conformation around the mismatched triplet. General characteristics of the triplex conformation around the mismatched site show that the triplex tends to preserve the global backbone conformation of the triple helix at the expense of hydrogen bonding at the mismatched site. Hydrogen bonds could still be made only if they do not strongly affect the regular pattern of the phosphate–sugar backbone. For example, the triplet C·G*G could form two hydrogen bonds involving N7(G)-H1N2(G) and O6(G)-HN1(G), as it does in the homopolymers poly(dC)·poly(dG)·poly(dG). However, this conformation is not isomorphous with the remaining triplets; therefore only one hydrogen bond is made between O6(G) and H1N2(G), together with an important tilt to release the constraints on the backbone.

In conclusion, both experimental and theoretical results obtained in the present study demonstrate that triple-helix formation is highly sequence-specific, at least as high as double-helix formation. Replacement of a single base pair in the center of a homopurine–homopyrimidine sequence is highly detrimental to triple-strand binding. As expected a mismatch at the ends of the homopurine–homopyrimidine sequence is less destabilizing. These results are in agreement with kinetics studies (to be reported elsewhere), which show that nucleation of triple-helix formation requires about 5 base triplets and that the remaining base triplets are formed in a highly cooperative way. A base triplet mismatch in the center of the target sequence disrupts the cooperativity of interactions between neighboring base triplets, thereby destabilizing the triplex structure. These results show that oligonucleotides targeted to double-stranded DNA sequences could be used to control transcription and replication processes in a highly sequence-specific manner. This so-called "antigene strategy" (Hélène & Toulmé, 1990) opens new possibilities for artificial control of gene expression.

ACKNOWLEDGMENTS

We thank Dr. R. Lavery for helpful advice on JUMNA.

REFERENCES

Arnott, S., Bond, P. J., Selsing, E., & Smith, P. J. C., (1976) *Nucleic Acids Res. 3*, 2459–2470.

Beal, P. A., & Dervan, P. B. (1991) *Science 251*, 1360–1363.

Belotserkovskii, B. P., Veselkov, A. G., Filippov, S. A., Dobrynin, V. N., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1990) *Nucleic Acids Res. 18*, 6621–6624.

Cantor, C. R., & Warshaw, M. M. (1970) *Biopolymers 9*, 1059–1077.

Cieplak, P., Rao, S. N., Hélène, H., Montenay-Garestier, T., & Kollman, P. A. (1987) *J. Biomol. Struct. Dyn. 5*, 361–382.

Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456–459.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekman, S., Hunter, N. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Soumpasis, D. M., Tung, O. S., von Kitzing, E., Wang, A. H.-J., & Zhurkin, V. B. (1989) *J. Biomol. Struct. Dyn. 6*, 627–634.

Durland, R. H., Kessler, D. J., Duvic, M., & Hogan, M. (1990) in *Molecular Basis of Specificity in Nucleic Acids–Drug Interactions* (Pullman, B., & Jortner, J., Ed.) pp 565–578, Kluwer Academic Publishers, Dordrecht, The Netherlands.

Felsenfeld, G., Davies, D. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023.

François, J. C., Saison-Behmoaras, T., Thuong, N. T., & Hélène, C. (1989) *Biochemistry 28*, 9617–9619.

Griffin, L. C., & Dervan, P. B. (1989) *Science 245*, 967–971.

Hélène, C., & Toulmé, J. J. (1990) *Biochim. Biophys. Acta 1049*, 99–125.

Hélène, C., Thuong, N. T., Saison-Behmoaras, T., & François, J. C. (1989) *Trends Biotechnol. 7*, 310–315.

Lavery, R. (1988) in *Structure and Expression, Vol. 3: DNA Bending and Curvature* (Olson, W. K., Sarma, R. H., Sarma, M. H., & Sundaralingam, M., Eds.) pp 191–211, Adenine Press, Guilderland, NY.

Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655–667.

Lavery, R., Sklenar, H., Zakrzewska, K., & Pullman, B. (1986) *J. Biomol. Struct. Dyn. 3*, 929–1014.

Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J., & Hélène, C. (1987) *Nucleic Acids Res. 15*, 7749–7760.

Lyamichev, V. L., Mirkin, S. M., Frank-Kamenetskii, M. D., & Cantor, C. R. (1988) *Nucleic Acids Res. 16*, 2165–2178.

Maher, L. J., III, Wold, B., & Dervan, P. B. (1989) *Science 245*, 725–730.

Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.

Orson, F. M., Thomas, D. W., McShan, W. M., Kessler, D. J., & Hogan, M. E. (1991) *Nucleic Acids Res. 19*, 3435–3441.

Pilch, D. S., Brousseau, R., & Shafer, R. H., (1990a) *Nucleic Acids Res. 18*, 5743–5750.

Pilch, D. S., Levensen, C., & Shafer, R. H., (1990b) *Proc. Natl. Acad. Sci. U.S.A. 87*, 1942–1946.

Plum, G. E., Park, Y. W., Singleton, S. F., Dervan, P. B., & Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9436–9440.

Rajagopal, P., & Feigon, J. (1989) *Biochemistry 28*, 7859–7870.

Santos, C. D. L., Rosen, M., & Patel, D. (1989) *Biochemistry 28*, 7282–7289.

Shea, R. G., Ng, P., & Bischofberger, N. (1990) *Nucleic Acids Res. 18*, 4859–4866.

Stevens, C. L., & Felsenfeld, G., (1964) *Biopolymers 2*, 293–314.

Strobel, S. A., & Dervan, P. B., (1990) *Science 249*, 72–75.

Sun, J. S., François, J. C., Montenay-Garestier, T., Saison-Behmoaras, T., Roig, V., Chassignol, M., Thuong, N. T., & Hélène, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9198–9202.

Van Vlijmen, H. W. T., Ramê, G. L., & Pettitt, B. M., (1990) *Biopolymers 30*, 517–532.

Wallace, R. B., Johnson, M. J., Hirose, T., Miyake, T., Kawashima, E. H., & Itakura, K. (1981) *Nucleic Acids Res. 9*, 879–894.

Xodo, L. E., Manzini, G., & Quadrifoglio, F. (1990) *Nucleic Acids Res. 18*, 3557–3564.