Formation of intramolecular triplex in homopurine-homopyrimidine mirror repeats with point substitutions

B.P.Belotserkovskii, A.G.Veselkov, S.A.Filippov[1,†], V.N.Dobrynin[1], S.M.Mirkin* and M.D.Frank-Kamenetskii
Institute of Molecular Genetics, Academy of Sciences of USSR, Moscow 123182 and
[1]M.M.Shemyakin Institute of Bioorganic Chemistry, Moscow, USSR

Received July 16, 1990; Revised and Accepted October 13, 1990

ABSTRACT

We have used two-dimensional gel electrophoresis to study the structural transition to the triplex H form of sequences 5'-AAGGGAGAAXGGGGTATAGGGGYAA-GAGGGAA-3' where X and Y are any DNA bases. The transition was observed at acid pH under superhelical stress. For X = Y = A or X = Y = G the sequences corresponded to homopurine-homopyrimidine mirror repeats (H-palindrome) which are known to adopt the H form under acid pH and superhelical stress. We have shown that the H form is actually formed for all X and Y, though in cases other than X = Y = A and X = Y = G the transition requires larger negative superhelical stress. Different substitutions require different superhelicity levels for the transition to occur. Theoretical analysis of the data obtained made it possible to estimate the energy cost of triplex formation due to all possible mismatched base triads.

INTRODUCTION

Homopurine-homopyrimidine sequences are known to adopt at acid pH and/or negative supercoiling an unusual DNA structure, the H form (for review see (1)). A major element of the H form is the triplex formed by half of the purine strand and the hairpin formed by the pyrimidine strand (see Fig.1). The second half of the purine strand remains unstructured, which determines the hypersensitivity of H-DNA to single stranded-specific nucleases and chemical reagents attacking single-stranded DNA (2–8).

In principle, two isomeric forms of H-DNA may exist, where the triplex is formed either at the 3'- or at the 5'-end of the purine strand (1,2). Chemical probing indicated that the isoform which carries the triplex at the 3'-end of the purine strand usually prevails (3,4,5,7). However, Htun and Dahlberg reported the formation of the other isomer at low negative superhelicity (6).

In H-DNA, the triplex consists of T·A*T and C·G*C$^+$ base-triads (2). These triads are isomorphous, which guarantees triplex formation regardless of the sequence of the triads. This also leads to sequence requirements for H-DNA: the most favorable sequence should be the homopurine-homopyrimidine mirror repeat (1). Any deviation from the mirror symmetry would lead to the formation of non-canonical base-triads and, consequently, destabilize the triplex (and the H form). That this is actually the case was shown by Mirkin et al. (9) who studied the sequence:

5'-AAGGGAGAAXGGGGTATAGGGGYAAGAGGGAA-3'

Because in H-DNA several central nucleotides are looped out in both strands, four central nucleotides of the sequence were chosen arbitrarily as TATA.

It was shown that for X = Y = G and X = Y = A the transition to the H form was facile. By contrast, deviations from the mirror symmetry led either to an increased negative superhelicity of the transition (X = A, Y = G) or to a complete failure to detect it (X = G, Y = A). Using a more sensitive method of chemical probing, Voloshin et al. (3) were able to detect the H form extrusion even for the case of X = G, Y = A. Also, in all the four cases studied only the isomeric form was extruded which carried the triplex at the 3'-end of the purine strand. Other cases of formation of H-DNA in non-strictly-H-palindromic sequences were also demonstrated by chemical and enzymatic probing (4,5–8).

The data on 2-D gel electrophoresis (9) indicated that different substitutions might affect stability in different ways. To study this problem in greater detail we cloned within the pUC19 plasmid all 16 possible variants of the above sequence and studied them by 2-D gel electrophoresis. Quantitative analysis of these data made it possible to estimate, for the first time, the energy difference between different mismatched triads.

MATERIALS AND METHODS

Oligonucleotides

We synthesized oligonucleotides that follow, as described in (9):

A) 5'-CCCTATACCCX$_C$TTCTCCCTTG-3'
B) 5'-AATTCAAGGGAGAAXGGGG-3'
C) 5'-GATCCTTCCCTCTTY$_C$C-3'
D) 5'-TATAGGGGYAAGAGGGAAG-3'

* To whom correspondence should be addressed
* Present address: Department of Genetics, University of Illinois at Chicago, 808 South Wood Street, Chicago, IL 60612, USA
† Deceased

X and Y were either A, G, T or C. X_C and Y_C were nucleotides complementary to X and Y respectively.

Cloning

1 7μg of oligonucleotides A and D were kinased by polynucleotide kinase of phage T4 in a standard buffer with 10 mM ATP. Then the pairs of oligonucleotides (A + B) and (C + D) were annealed in 1 SSC buffer. Double-strand oligonucleotides were mixed in equimolar amounts and ligated with the pUC19 plasmid, linearized with Eco R1 and Bam H1 restriction enzymes.Then the strain JM83 was transformed by the resulting DNA. The DNA of selected clones was sequenced by the Maxam-Gilbert procedure in the usual way.

Two-dimensional gel electrophoresis

DNA samples with a wide distribution over topoisomers were prepared as usual (2). The plasmid carrying the insert 5′-AA-GGGAGAAXGGGGTATAGGGGYAAGAGGGAA-3′ was denoted pXY32 (9). Two-dimensional gel electrophoresis in 1.5% agarose gel was conducted in the first direction in 200 mM of sodium citrate buffer, pH 4.17, at 20°C, at the field tension 2 v/cm for 20 hours. Then the gel was saturated for 3−4 hours with TAE buffer containing 3 μg/ml of chloroquine. This buffer was used for electrophoresis in the second direction at 20°C, 2 v/cm for 20 hours.

S1-mapping

The reaction with S1-nuclease was performed in 100 μl of the mock S1 buffer: 0.25 M NaCl, 30 mM NaAc (pH4.2), 30 μM $ZnSO_4$. To this solution, 2 μg DNA and 1 unit of S1-nuclease were added. The reaction was performed for 2 min at 20°C.

The digestion was stopped by phenol extraction and ethanol precipitation.Then DNA was cut by the Hind III restriction enzyme and labelled either the 3′-end with [α^{32}P] dCTP or the 5′-end with [γ^{32}P] ATP. After labelling, the DNA was cut by the Bgl I restriction enzyme,the fragment in question was extracted and analyzed in a 8% polyacrylamide sequencing gel.

RESULTS

Fig.2 shows that the pAA32 plasmid, which carries the H-palindromic sequence (X = Y = A), exhibits a drop of mobility, which reflects the B-H transition at the 6th topoisomer at pH 4.17.

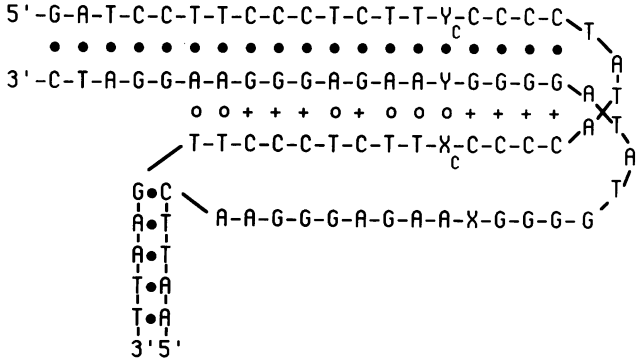

Fig. 1. H-DNA adopted by the insert of plasmid pXY32. The isoform shown corresponds to the triplex at the 3′-end of the purine strand. The $Y_C \cdot Y^* X_C$ triad is formed. The filled circles show the Watson-Crick base pairs, open circles show the Hoogsteen AT pairs, the + symbols show the GC^+ Hoogsteen pairs.

At the same time, the pAC32 plasmid exhibits a similar transition (with the same mobility drop) at the 9th topoisomer. The mobility drop is clearly associated with the insert because we could not observe any mobility drop in the parental pUC19 plasmid under the same conditions (data not shown).

The transition is clearly pH-dependent. Indeed, we could not observe any transition in the pAC32 plasmid at pH 5.0 by 2-D gel electrophoresis (data not shown). We therefore conclude that in both cases presented in Fig.2, as well as in all other 14 cases studied by us, for which the data are not shown, the inserts undergo transitions into the H form.

Voloshin et al. (3) showed that plasmid pAA32, pAG32, pGA32 and pGG32 form H-DNA in which 3′-half of the purine chain participates in the triplex whereas the 5′-half is unstructured. We believe that the same is true in all the cases we examined. This conclusion stems from the S1-probing of plasmids pGG32, pTT32, pAC32, pCA32, pGT32, pTG32, pGC32, pCG32, pTC32, pCC32, pTA32, pAT32. Fig.3 shows as an example the S1-fine-mapping pattern for the pTT32 plasmid. One can see clear-cut digestion of the 5′-half of the purine strand and the center of the pyrimidine strand (lanes 4 and 7 respectively). 2-D gel electrophoresis shows that the mobility drop corresponded to 3 superhelical turns in all cases, and the area of transition contained no double spots, which Htun and Dahlberg considered as an indication of the presence of both H-DNA isoforms (7).

If our conclusion is true, the formula of non-canonical triads in H-DNA in plasmid pXY32 is $Y_C \cdot Y^* X_C$ (where X_C and Y_C are bases complementary to X and Y).

Our results are summarized in Table 1. Along with the name of the plasmid and the type of base-triad, we present the value

$$\Delta\tau_{XY} = \tau_{XY} - \tau_{AA'}$$

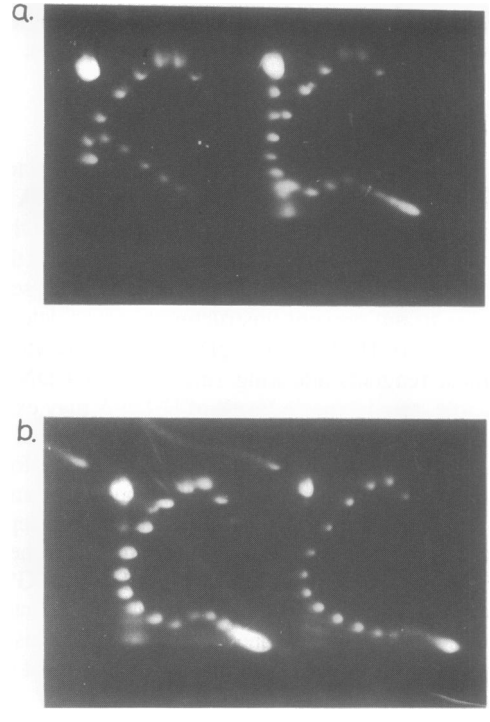

Fig. 2. a. Two-dimensional gel electrophoresis patterns for plasmids pAA32 (left) and pAC32 (right). b. Two-dimensional gel electrophoresis pattern for pTG32 (left) and pTA32 (right).

where τ is the number of superhelical turns which corresponds to a half-transition of the plasmid. The $\Delta\tau_{XY}$ value corresponds to the pXY32 plasmid, the $\Delta\tau_{AA}$ value to the pAA32 plasmid.

Table 1 shows that any mismatched base-triad makes triplex formation more difficult. However, different triads affect the transition in different ways. While for pAG32 and pTG32 $\Delta\tau$ = 2, for pTC32 $\Delta\tau$ = 4.

To determine the energy cost of the mismatched triads, let us use the expression for the energy of DNA supercoiling (10):

$$\Delta G = 10NRT\sigma^2 = 10RT\gamma^2\tau^2/N.$$

Here N is the plasmid length (in b.p.), σ is superhelical density, γ is the number of base pairs per one turn of the double helix under given ambient conditions, R is the gas constant, T is absolute temperature. The energy difference for H forms in the pXY32 and pAA32 plasmids is:

$$\Delta E = 10RT\gamma^2 \{ [\tau_{XY}^2 - (\tau_{XY} - \frac{m}{\gamma})^2] - [\tau_{AA}^2 - (\tau_{AA} - \frac{m}{\gamma})^2] \}/N = 20RTm\gamma\Delta\tau_{XY}/N.$$

where m is the insert length (in b.p.). For m = 32, N = 2700 γ = 10.5, T = 293 K one obtains: $\Delta E = 1.45\ \Delta\tau_{XY}$ (kcal/mol.).

We conclude that the energy difference between canonical and non-canonical base-triads may vary between 3 and 6 kcal/mol. This estimation is correct as long as the above formula for superhelix energy is valid.

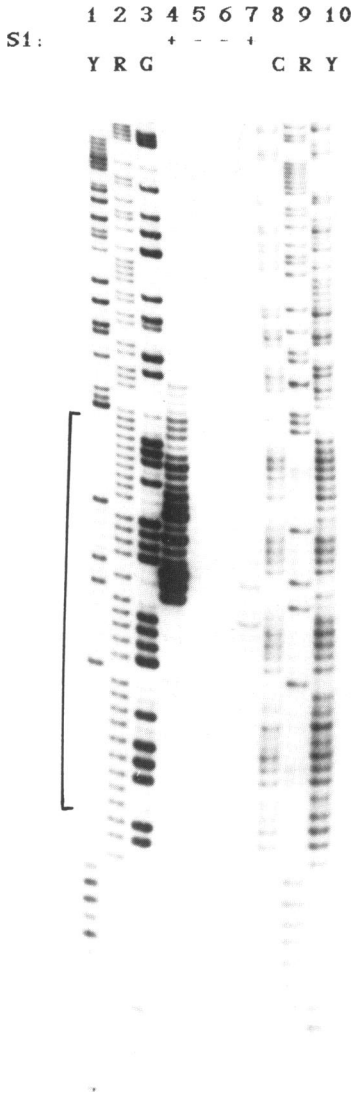

Fig. 3. Mapping of S1-nuclease cleavage sites of the plasmid pTT32 at pH 4.2. Lanes 1–5 correspond to the Pu-rich strand (the 5'-end at the top). Lanes 6–10 correspond to the Py-rich strand (the 5'-end at the botton). Lanes 1,2,3,8,9,10 are Maxam-Gilbert sequencing reaction products. Lanes 4 and 7 are nicking patterns of the Pu- and Py-rich strands. Lanes 5 and 6 correspond to the mock S1 of the Pu- and Py-rich strands. The insert is indicated by the bracket. For experimental conditions see Materials and Methods.

DISCUSSION

We attribute the free energy difference between the H forms formed in plasmids pXY32 and pAA32 entirely to energy difference between the mismatched and canonical base-triads, $Y_C \cdot Y^* X_C$ and $T \cdot A^* T$ (the asterisk means Hoogsteen or mismatched-Hoogsteen pairing). Along with the assumption that only one isomer (with the triplex at the 3'-end) of the H form extrudes, this attribution means that we neglect all other energy differences. Specifically, we neglect the fact that the B-H transition is associated with the melting out of the canonical base pair XX_C. The possible error may be estimated as (12):

$$\Delta\tau = (\Delta G_{GC} - \Delta G_{AT})N/20RTm\gamma = (T_{GC} - T_{AT})\Delta H_{AT}N/20RTm\gamma T_{AT'}$$

where ΔH_{AT} is the melting enthalpy of the AT pair, T_{AT} and T_{GC} are melting temperatures of the AT and GC homopolymers. For T_{AT} = 340 K, T_{GC} = 380 K, ΔH_{AT} = 6.9 kcal/mol one obtains $\Delta\tau$ = 0.6. This effect is within the experimented error of our measurements. It by no means explains the observed $\Delta\tau_{XY}$ values. Indeed, if this effect significantly contributed in the $\Delta\tau_{XY}$ values one could expect plasmids pGC32 and pCG32 to exhibit larger $\Delta\tau_{XY}$ values than plasmids pTT32, pAT32 and pTA32. Table 1 shows that that is not the case.

We therefore conclude that the difference in the free energy of formation of H-DNA in different inserts reflects the energy difference in the incorporation of various mismatched triads into

Table 1. The $\Delta\tau_{XY}$ values for different plasmids pXY32.

Plasmid	Triad	$\Delta\tau$XY
pAA32	T·A*T	0
pGG32	C·G*C	0
pAG32*	C·G*T	2
pTG32	C·G*A	2
pAC32	G·C*T	2,5
pGT32	A·T*C	2,5
pGC32	G·C*C	3
pCG32	C·G*G	3
pCT32	A·T*G	3,5
pAT32	A·T*T	3,5
pTA32	T·A*A	3,5
pTT32	A·T*A	3,5
pCC32	G·C*G	3,5
pGA32	T·A*C	3,5
pCA32	T·A*C	3,5
pTC32	G·C*A	4

Commentary. The transition point was defined as the topoisomer number corresponding to the half-transition. Since the width of the transition corresponded to about one superturn, the accuracy of the $\Delta\tau_{XY}$ value was estimated within ±0.5 superturns. The data for pAG32 were taken from (9) and corresponded to pH 4.3 and 200 mM Na^+ in sodium citrate buffer.

the triplex. The variation in stability of non-canonical triads clearly reflects their different structures. For some mismatched triads, for example C·G*A+ (13) and C·G*G (4), models of hydrogen bonds were suggested. The question whether this kind of hydrogen bonding exists in the H-form triplex needs further investigation. However, in most cases the structural features of the non-canonical triads remain obscure. Note that Griffin and Dervan (14) previously reported about a facile formation of the A·T*G triad. According to our data in Tabl.1, this mismathed triad is not among the most favorable.

In conclusion, non-canonical (mismatched) triads can be incorporated into the triplex. The energy cost of this incorporation is within the range of 3–6 kcal/mol for different mismatched triads. On the one hand this means that deviations from homopurine-homopyrimidine pattern is tolerated to some extent in the H form. On the other hand, this shows that even a single mismatched triad significantly disfavours the triplex, as compared with a canonical one. This makes very promising the idea of a sequence-specific recognition of homopurine-homopyrimidine region in duplex DNA by pyrimidine oligonucleotides via triplex formation.

The energy cost for mismatched triads is within the same range as the energy cost for mismatched base pairs (15,16). This indicates that the sequence requirements for triplex formation are similar to the sequence requirements for complementary recognition in a duplex. Although both modes of recognition exhibit similar sequence-specificity, the great potential of the triplex mode of recognition stem from the fact that it does not need DNA unwinding. People have recently started utilizing triplex recognition for sequence-specific labelling of duplex DNA in genome mapping (17). Even more attractive, though much less developed, is the idea of 'gene drugs', which presupposes the delivery, via triplex formation, to specific sites in the genome of drugs tagged to oligonucleotides.

REFERENCES

1. Frank-Kamenetskii,M.D. (1990) in: 'DNA Topology and Its Biological Effects'. pp.185–215, ed by N.R.Cozzarelli and J.C.Wang, Cold Spring Harbor Lab. Press.
2. Lyamichev,V.I., Mirkin,S.M. and Frank-Kamenetskii,M.D. (1986) J. Biomol. Struct. Dyn. 3, 667–669.
3. Voloshin O.N., Mirkin,S.M., Lyamichev,V.I., Belotserkovskii,B.P. and Frank-Kamenetskii,M.D.(1988) Nature 330, 475–476.
4. Kohwi,Y. and Kohwi-Shigematsu,T. (1988) Proc. Natl. Acad. Sci. USA. 85, 3781–3785.
5. Hanvey,J.C., Klysik,J. and Wells R.D. (1988) J. Bol.Chem. 263, 7386–7396.
6. Htun,H. and Dahlberg,J.E. (1988) Science 243, 1571–1576.
7. Htun,H. and Dahlberg,J.E.(1989) Science 243, 1791–1796.
8. Jonston,B.H. (1988) Science 241, 1800–1804.
9. Mirkin,S.M., Lyamichev,V.I., Drushlyak,K.N., Dobrynin,V.N., Filippov,S.A. and Frank-Kamenetskii,M.D. (1987) Nature 330, 495–497.
10. Horowitz,D.S. and Wang,J.C. (1984) J. Mol. Biol. 173, 75–91.
11. Frank-Kamenetskii,M.D. and Vologodskii,A.V. (1984) Nature 307, 481–482.
12. Kozyavkin,S.A., Mirkin,S.M. and Amerikyan,B.R. (1986) Biopolimers and Cell (Russian) 2, 124–129.
13. Johnson,D. and Morgan,A.R. (1978) Proc. Natl. Acad. Sci. USA 75, 1637–1641.
14. Griffin,L.C. and Dervan,P.B. (1989) Science 245,967–970.
15. Aboul-ela,F., Koh,D. and Tinoco,I,Jr. (1985) Nucl. Acids Res. 13, 4811–4824.
16. Perelroyzen,M.P. and Vologodskii,A.V. (1988) Nucl.Acid Res. 16, 4693–4704.
17. Moores,J.C. (1990) Strategies Mol. Biol. 3, 23–24 and 29 (Stratagene Cloning Systems Publ.).