FEBS Letters 491 (2001) 184–187 FEBS 24634

Tetraplex formation by the progressive myoclonus epilepsy type-1 repeat: implications for instability in the repeat expansion diseases

Tapas Saha, Karen Usdin*

Section on Genomic Structure and Function, Laboratory of Molecular and Cellular Biology, National Institute of Diabetes and Kidney Diseases, Building 8, Room 202, National Institutes of Health, 8 CENTER DR MSC 0830, Bethesda, MD 20892-0830, USA

Received 1 November 2000; revised 30 January 2001; accepted 30 January 2001

First published online 9 February 2001

Edited by Ned Mantei

Abstract The repeat expansion diseases are a group of genetic disorders resulting from an increase in size or expansion of a specific array of tandem repeats. It has been suggested that DNA secondary structures are responsible for this expansion. If this is so, we would expect that all unstable repeats should form such structures. We show here that the unstable repeat that causes progressive myoclonus epilepsy type-1 (EPM1), like the repeats associated with other diseases in this category, forms a variety of secondary structures. However, EPM1 is unique in that tetraplexes are the only structures likely to form in long unpaired repeat tracts under physiological conditions.

Key words: Repeat expansion; DNA structure; Instability

1. Introduction

Progressive myoclonus epilepsy of Unverricht–Lundborg type (EPM1) is an autosomal recessive disorder characterized by myoclonus, seizures and cerebellar ataxia. This disorder results from mutations in the cystatin B (CSTB) gene which encodes a non-caspase cysteine protease inhibitor [1]. The most common mutation associated with EPM1 is an expansion of a tandem array of the dodecamer d(C_4GC_4GCG)·d($CGCG_4CG_4$) in the cystatin B promoter [1–5].

Pathogenesis caused by expansion of a specific tandem array is a characteristic feature of the repeat expansion diseases, a group of neurological and neuromuscular disorders. The repeats associated with all other diseases in this group form various secondary structures [6–25], as do a number of other hypervariable sequences [21,26,27]. Here we demonstrate that the EPM1 dodecamer repeat also forms secondary structures. These include hairpins comprised entirely of non-Watson–Crick base pairs, tetraplexes, and *i*-motif structures, but tetraplexes are the most likely to prevail in long repeat tracts under physiological conditions. The fact that all disease associated repeats form secondary structures lends support to the idea that these structures play a role in expansion.

*Corresponding author. Fax: (1)-301-402 0053.
E-mail: ku@helix.nih.gov

2. Materials and methods

2.1. Oligonucleotides

The oligodeoxyribonucleotides used in this study were EPM1T (5′-GTACGAATT(C_4GC_4GCG)$_4$TAGGGCTCGAGTCAACGTAACACTTT-3′) and EPM1B (5′-GTACGAATT($CGCG_4CG_4$)$_4$TAGGGCTCGAGTCAACGTAACACTTT-3′). They were synthesized using standard phosphoramidite chemistry (Life Technologies, Gaithersburg, MD, USA), and purified by denaturing gel electrophoresis on 20% polyacrylamide gels. Their concentrations were determined from the UV absorbance measured at room temperature using the molar extinction coefficients 9517 M^{-1} cm^{-1} for EPM1T and 10789 M^{-1} cm^{-1} for EPM1B.

2.2. Circular dichroism

Circular dichroism measurements were made using a Jasco J-715 A spectropolarimeter with a thermostatted cuvette holder and 1 cm pathlength quartz cuvettes. The samples were dissolved in 10 mM Na-cacodylate at the indicated pH. KCl, NaCl, or LiCl were added where indicated to a final concentration of 100 mM. CD spectra were obtained for each oligonucleotide starting at 10°C and increasing in 5°C increments. After complete denaturation at the end of the experiment, the samples were allowed to reanneal for 24 h and the CD melting profiles remeasured. The data from three independent sets of measurements were averaged after adjustment to correct for the spectral contributions of the buffer and cuvette. Spectras are plotted as $\Delta\varepsilon[\varepsilon(L-R)]$ in units of M^{-1} cm^{-1}.

2.3. Chemical modification

Chemical modification of the EPM1B oligodeoxyribonucleotide was carried out as previously described [26]. The resultant autoradiograph was then scanned and analyzed using Image Gauge V3.0 (Fuji Photo Film Co, Ltd, Japan). The peak height was normalized to the peak height of bases flanking the repeat (indicated by the asterisks in Figs. 3 and 4).

3. Results and discussion

3.1. The C-rich top strand of the EPM1 promoter forms an unstable intermolecular protonated complex

The CD spectra obtained for EPM1T, an oligonucleotide containing four repeats of the top strand of the EPM1 promoter, are shown in Fig. 1. At 35°C and a pH of 7.0, the CD spectrum has a positive peak at 282 nm and a negative peak at 249 nm. The positive peak shifts by about 3 nm to 285 nm when the pH is dropped to 6.5, and shifts by an additional 2 nm to 287 nm at pH 6.0. This red shift is characteristic of the formation of protonated complexes [28], most likely *i*-motif structures. An *i*-motif is a four-stranded DNA structure consisting of two pairs of parallel DNA strands held together by $C{\cdot}C^+$ pairs. The four-stranded structure arises by intercalation of the $C{\cdot}C^+$ pairs on the two paired DNA strands [29–31]. Similar structures have been reported for the C-rich strands of the Fragile X triplet repeat [15], and the hypervariable repeat

0014-5793/01/$20.00
PII: S0014-5793(01)02190-1

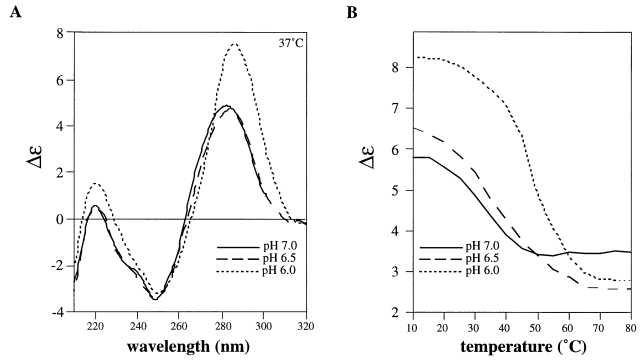

Fig. 1. The CD spectra of an oligodeoxyribonucleotide containing four repeats of the top strand of the EPM1 promoter (EPM1T). A: CD spectra measured at 35°C at the indicated pHs in the presence of 100 mM KCl. B: Melting profile of EPM1T measured at 287 nm in 100 mM KCl at the indicated pHs.

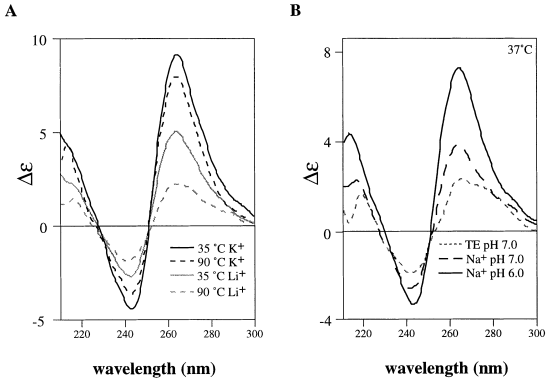

Fig. 2. The CD spectra of an oligodeoxyribonucleotide containing four repeats of the bottom strand of the EPM1 promoter (EPM1B). A: Effect of different cations on the CD spectra of EPM1B. The CD spectra were measured at pH 7.0 at a EPM1B strand concentration of 900 nM in the presence of either 100 mM KCl or LiCl at the indicated temperatures. B: CD spectra of EPM1B at 37°C in either TE buffer (pH 7.0) or 10 mM sodium cacodylate at the indicated pH.

in the human insulin promoter [32]. The red shift characteristic of *i*-motif structure is seen even in samples that have been cooled for <15 min after denaturation (data not shown). This suggests that the *i*-motif structure forms rapidly. However, since the T_m of this structure is 32.5°C at pH 7.0 in 100 mM KCl (Fig. 1B), it is unlikely to be stable at physiologically reasonable conditions.

3.2. The G-rich bottom strand of the EPM1 promoter forms acid stabilized intermolecular G_4-tetraplexes

The CD spectrum of EPM1B, which contains four repeats from the bottom strand of the EPM1 promoter, shows a positive peak at 263 nm and a negative peak at 242 nm at 35°C (Fig. 2A, black solid line). This spectrum is characteristic of parallel G_4-tetraplexes [33–35]. In such a structure all the Gs are *anti* and interact with three other guanine bases via Hoogsteen hydrogen bonds, creating a series of cyclic G-quartets that comprise the four-stranded helix.

In contrast to the *i*-motif formed by the top strand, the G_4-tetraplex is extremely stable: very little change is seen in the spectra produced in the presence of either 20 mM or 100 mM KCl between 10°C and 94°C (Fig. 2A, black dotted line, and data not shown). However in 100 mM LiCl, the amplitude of both the positive peak and the negative one are strongly reduced at both 35°C and 90°C (solid and dashed gray lines in Fig. 2A respectively). This is consistent with the known properties of other G_4-tetraplexes and results at least in part from the ability of Na^+ and K^+, but not Li^+, to stabilize tetraplexes via the formation of coordination complexes with the eight carbonyl oxygens in two adjacent G-quartets [36]. At low [Na^+], the amplitude of the 263 nm peak is significantly higher at pH 6.0 than it is at pH 7.0 (Fig. 2B), indicating that this structure is stabilized by protonation. This is consistent with a tetraplex containing a mixture of G_4-tetrads and hemiprotonated C·C^+ pairs.

3.3. The G-rich but not the C-rich strand forms stable intrastrand structures at 37°C

Electrophoresis of dilutions of each strand of the promoter was carried out in 10% polyacrylamide gels at either pH 7.0 or pH 8.0 at 37°C. Under these conditions EPM1B migrated as a single species with a more rapid mobility than expected from its molecular weight (data not shown). This mobility shift is characteristic of the formation of intrastrand structures.

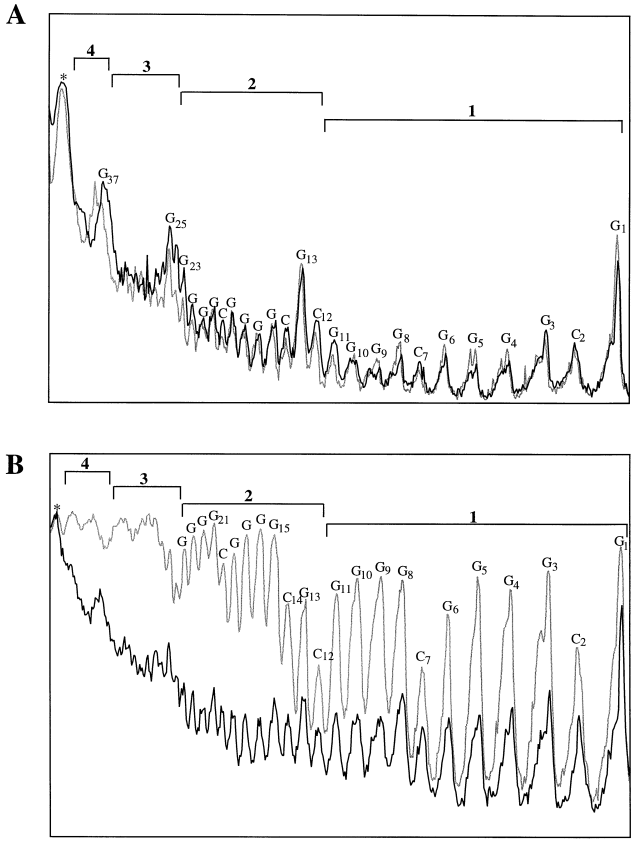

Fig. 3. Chemical modifications of the EPM1B oligodeoxyribonucleotide in the presence and absence of KCl. A: The DEPC reactivity of EPM1B. The gray line is the DEPC modification profile seen in the absence of K^+, the black line is that seen in the presence of that cation. Each repeat is demarcated by the square brackets above the scan, with the repeats being numbered from the 5′ end of the oligonucleotide. The peak outside the repeat region that was used to normalize the two scans is indicated by an asterisk. B: The DMS reactivity of EPM1B. The gray line is the DMS modification profile seen in the absence of K^+, the black line is that seen in the presence of that cation. Each repeat is demarcated by the square brackets above the scan, with the repeats being numbered from the 5′ end of the oligonucleotide. The peak outside the repeat region that was used to normalize the two scans is indicated by an asterisk.

EPM1T showed no such anomalous mobility at this temperature. This suggests that the top strand does not form intrastrand structures that are stable at physiological pHs and temperatures. We therefore focused exclusively on the nature of the intrastrand structures formed by the G-rich strand.

3.4. Individual strands of the G-rich repeat form hairpins in the absence of KCl and a G_4-tetraplex when KCl is present

We treated the G-rich oligodeoxyribonucleotide with either diethylpyrocarbonate (DEPC) or dimethylsulfate (DMS) in the presence and absence of K^+. Both DMS and DEPC react with the N^7 of Gs. DMS reacts strongly with all Gs as long as the N^7 position is not involved in a hydrogen bond. In contrast, Gs that are well stacked within a helix are much less reactive with DEPC than extrahelical or loop Gs even if they are unpaired. DMS also modifies unpaired Cs albeit less effectively.

The modification profile for EPM1B in the absence of KCl is shown as the gray traces in Fig. 3A and B, and the profile for EPM1B in the presence of KCl is shown as black traces in Fig. 3A and B. In the absence of KCl, the Gs in the repeat region of the EPM1B oligonucleotide are not uniformly reactive with DEPC suggesting that some Gs are base paired or otherwise stacked within the DNA helix while others are not. The first G in each of the four repeats is DEPC reactive (G_1, G_{13}, G_{25}, and G_{37}), while the remaining Gs are much more weakly reactive. In contrast, under these conditions, all the Gs in the repeat are uniformly reactive with DMS (gray line in Fig. 3B). The periodicity of the DEPC protection cannot be explained by a tetraplex-like structure since all the Gs are accessible to methylation by DMS. A single large hairpin would not account for the DEPC reactivity either since this would produce a pattern of reactivity where the Gs in the central loop and the 5′ and 3′ flanks would be reactive. Rather our data suggest that under these conditions the G-rich strand of the EPM1 unstable repeat forms a pair of hairpins each containing two repeat units (Fig. 4). If these hairpins contained a mixture of G·C and G·G pairs we would

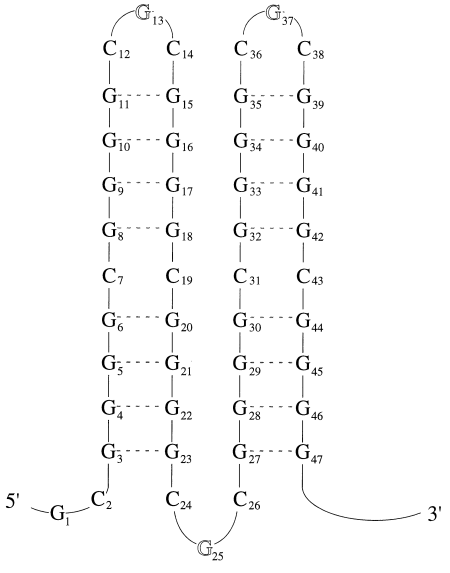

Fig. 4. Model for the structure formed by the bottom strand of the EPM1 repeat in the absence of K^+. The G·G pairs are indicated by dotted lines. The DEPC-reactive Gs within the structure are shown in outline font.

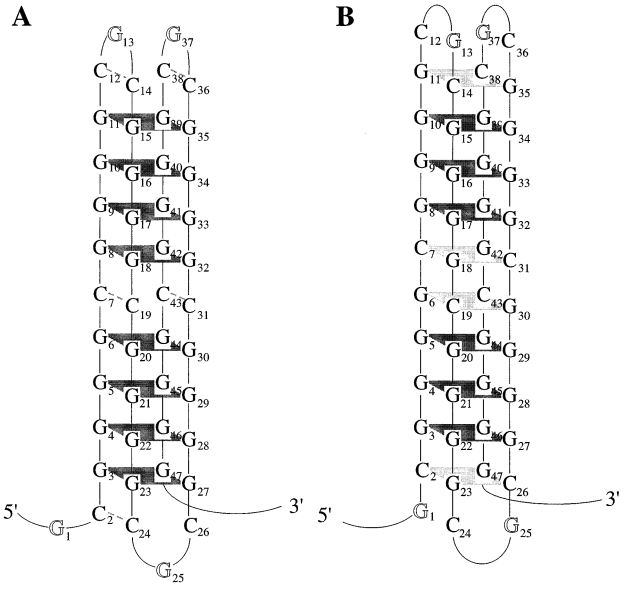

Fig. 5. Two different models for the structure formed by the bottom strand of the EPM1 repeat in the presence of K^+. G·G·G·G tetrads are shown as dark gray parallelograms, G·C·G·C tetrads as light gray parallelograms and base pairs by dotted lines. The DEPC-reactive Gs are shown in outline font.

expect to see differences between the Gs in their DMS reactivity since Gs in G·C base pairs are completely reactive with DMS while each G in a G·G pair has a 50% chance of being an N^7 donor. Since the Gs all show similar levels of DMS reactivity, our data suggest that all of the stem Gs are involved in G·G base pairs, and that the hairpins contain no Watson–Crick base pairs.

The pattern of DEPC reactivity does not change in the presence of KCl (black line in Fig. 3A). However, in the presence of this cation all the Gs are strongly protected from reaction with DMS (black line in Fig. 3B). This indicates that in the presence of physiological concentrations of KCl, these Gs are obligatory N^7 donors. The Cs also are much less reactive with DMS in the presence of K^+ than they are in its absence, suggesting that some hydrogen bonding occurs in the presence of this cation. The protection of all the Gs from modification by DMS in the presence of KCl is consistent with the formation of an intrastrand tetraplex in which all the Gs in the stem of the tetraplex are obligatory N^7 donors. Two types of tetrads containing Gs have been demonstrated to occur in different tetraplex-forming sequences: G·G·G·G tetrads [37–40] and G·C·G·C tetrads [20,41,42]. In both of these tetrads, the N^7 of the constituent guanines are involved in hydrogen bonds. Two different structures are thus possible in the EPM1 repeat, one in which the guanines are involved in G·G·G·G tetrads with the Cs interacting with one another (Fig. 5A), and one containing a mixture of G·G·G·G tetrads and G·C·G·C tetrads (Fig. 5B). Our chemical modification data are consistent with both of these structures. However, the fact that in the absence of K^+ this sequence forms a hairpin that contains no G·C pairs, together with the observation that the intermolecular parallel-stranded version of the tetraplex is stabilized by protonation (Fig. 2B), lends some support to the first structure (Fig. 5A). A similar role for protonation is seen in the tetraplex formed by the G-rich strand of Fragile X repeat [12,21,43].

Whatever the molecular details, this structure is extremely stable since EPM1B retains its anomalously rapid mobility during electrophoresis through polyacrylamide gels even when electrophoresis is carried out at 65°C (data not shown). This is consistent with the stability we have observed for the interstrand tetraplexes formed by the EPM1 repeat, and with the reported stability of other intermolecular and intramolecular tetraplexes [12,35].

The CGG repeat associated with Fragile X syndrome and FRAXE mental retardation forms both tetraplexes [12,19,21,43] and hairpins containing a mixture of G·G and C·G pairs that are also quite stable at physiological pHs [6,7,12,14]. The EPM1 promoter is thus unique among the repeat expansion diseases thus far described in having tetraplexes as the predominant structures formed by long repeat tracts at physiological temperatures, pHs, and ionic strengths. The folded nature of intrastrand tetraplexes suggests that they may be able to affect DNA replication and repair in similar ways to those proposed for the hairpins formed by other disease associated repeats (see [44] for review). In addition, it is possible that their extraordinary stability confers upon them unique properties [12,21]. Our demonstration of the formation of tetraplexes by the EPM1 repeat shows that the ability to form such structures is indeed a common property of all hypervariable sequences, and lends credence to the hypothesis that these structures are responsible, at least in part, for expansion.

Acknowledgements: The authors wish to thank Dr. Peter McPhie for the use of his spectropolarimeter and for his guidance in making the CD measurements. We also wish to thank Drs. Anthony V. Furano, Nancy Nossal, Daman Kumari and Ed Grabczyk for their careful reading of this manuscript.

References

[1] Pennacchio, L.A. et al. (1996) Science 271, 1731–1734.
[2] Lafreniere, R.G. et al. (1997) Nat. Genet. 15, 298–302.
[3] Lalioti, M.D. et al. (1997) Am. J. Hum. Genet. 60, 342–351.
[4] Lalioti, M.D., Scott, H.S., Buresi, C., Rossier, C., Bottani, A., Morris, M.A., Malafosse, A. and Antonarakis, S.E. (1997) Nature 386, 847–851.
[5] Virtaneva, K. et al. (1997) Nat. Genet. 15, 393–396.
[6] Chen, X. et al. (1995) Proc. Natl. Acad. Sci. USA 92, 5199–5203.
[7] Gacy, A.M., Goellner, G., Juranic, N., Macura, S. and McMurray, C.T. (1995) Cell 81, 533–540.
[8] Mariappan, S.V., Catasti, P., Chen, X., Ratliff, R., Moyzis, R.K., Bradbury, E.M. and Gupta, G. (1996) Nucleic Acids Res. 24, 784–792.
[9] Mariappan, S.V. et al. (1998) J. Biomol. Struct. Dyn. 15, 723–744.
[10] Mariappan, S.V., Silks 3rd, L.A., Bradbury, E.M. and Gupta, G. (1998) J. Mol. Biol. 283, 111–120.
[11] Mitas, M., Yu, A., Dill, J., Kamp, T.J., Chambers, E.J. and Haworth, I.S. (1995) Nucleic Acids Res. 23, 1050–1059.
[12] Usdin, K. and Woodford, K.J. (1995) Nucleic Acids Res. 23, 4202–4209.
[13] Zheng, M., Huang, X., Smith, G.K., Yang, X. and Gao, X. (1996) J. Mol. Biol. 264, 323–336.
[14] Mitas, M., Yu, A., Dill, J. and Haworth, I.S. (1995) Biochemistry 34, 12803–12811.
[15] Yu, A. et al. (1997) Biochemistry 36, 3687–3699.
[16] Yu, A., Dill, J., Wirth, S.S., Huang, G., Lee, V.H., Haworth, I.S. and Mitas, M. (1995) Nucleic Acids Res. 23, 2706–2714.
[17] Yu, A., Dill, J. and Mitas, M. (1995) Nucleic Acids Res. 23, 4055–4057.
[18] Nadel, Y., Weisman-Shomer, P. and Fry, M. (1995) J. Biol. Chem. 270, 28970–28977.
[19] Fry, M. and Loeb, L.A. (1994) Proc. Natl. Acad. Sci. USA 91, 4950–4954.
[20] Kettani, A., Kumar, R.A. and Patel, D.J. (1995) J. Mol. Biol. 254, 638–656.
[21] Usdin, K. (1998) Nucleic Acids Res. 26, 4078–4085.
[22] Sakamoto, N., Chastain, P.D., Parniewski, P., Ohshima, K., Pandolfo, M., Griffith, J.D. and Wells, R.D. (1999) Mol. Cell 3, 465–475.
[23] Hanvey, J.C., Shimizu, M. and Wells, R.D. (1988) Proc. Natl. Acad. Sci. USA 85, 6292–6296.
[24] Grabczyk, E. and Usdin, K. (2000) Nucleic Acids Res. 28, 2815–2822.
[25] Gacy, A.M. et al. (1998) Mol. Cell 1, 583–593.
[26] Weitzmann, M.N., Woodford, K.J. and Usdin, K. (1998) J. Biol. Chem. 273, 30742–30749.
[27] Weitzmann, M.N., Woodford, K.J. and Usdin, K. (1997) J. Biol. Chem. 272, 9517–9523.
[28] Antao, V.P. and Gray, D.M. (1993) J. Biomol. Struct. Dyn. 10, 819–839.
[29] Leroy, J.L., Gueron, M., Mergny, J.L. and Helene, C. (1994) Nucleic Acids Res. 22, 1600–1606.
[30] Ahmed, S., Kintanar, A. and Henderson, E. (1994) Nat. Struct. Biol. 1, 83–88.
[31] Leroy, J.L., Gehring, K., Kettani, A. and Gueron, M. (1993) Biochemistry 32, 6019–6031.
[32] Catasti, P., Chen, X., Deaven, L.L., Moyzis, R.K., Bradbury, E.M. and Gupta, G. (1997) J. Mol. Biol. 272, 369–382.
[33] Guo, Q., Lu, M. and Kallenbach, N.R. (1993) Biochemistry 32, 3596–3603.
[34] Lu, M., Guo, Q. and Kallenbach, N.R. (1993) Biochemistry 32, 598–601.
[35] Chen, F.M. (1992) Biochemistry 31, 3769–3776.
[36] Zimmerman, S.B., Cohen, G.H. and Davies, D.R. (1975) J. Mol. Biol. 92, 181–192.
[37] Wang, Y. and Patel, D.J. (1993) Structure 1, 263–282.
[38] Torigoe, H., Kamiya, M., Shindo, H. and Sarai, A. (1995) Nucleic Acids Symp. Ser. 34, 199–200.
[39] Scaria, P.V., Shire, S.J. and Shafer, R.H. (1992) Proc. Natl. Acad. Sci. USA 89, 10336–10340.
[40] Sen, D. and Gilbert, W. (1990) Nature 344, 410–414.
[41] Bouaziz, S., Kettani, A. and Patel, D.J. (1998) J. Mol. Biol. 282, 637–652.
[42] Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R.A. and Patel, D.J. (1998) J. Mol. Biol. 282, 619–636.
[43] Chen, F.M. (1995) J. Biol. Chem. 270, 23090–23096.
[44] Usdin, K. and Grabczyk, E. (2000) Cell Mol. Life Sci. 57, 914–931.