Rapid Communication

J. Biochem. **140**, 759–762 (2006)
doi:10.1093/jb/mvj213

DNA Octaplex Formation with an *I*-Motif of Water-Mediated A-Quartets: Reinterpretation of the Crystal Structure of d(GCGAAAGC)

Yoshiteru Sato[1], Kenta Mitomi[1], Tomoko Sunami[1], Jiro Kondo[2] and Akio Takénaka[1,*]
[1]*Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, 4259 Nagatsuda, Midori-ku, Yokohama 226-8501 and* [2]*Institut de Biologie Moléculaire et Cellulaire du CNRS, Université Louis Pasteur, 15 rue René Descartes, Strasbourg F-67084, France*

Received September 24, 2006; accepted October 17, 2006

The crystal structure of the tetragonal form of d(gcGAAAgc) has been revised and reasonably refined including the disordered residues. The two DNA strands form a base-intercalated duplex, and the four duplexes are assembled according to the crystallographic 222 symmetry to form an octaplex. In the central region, the eight strands are associated by *I*-motif of double A-quartets. Furthermore, eight hydrated-magnesium cations link the four duplexes to support the octaplex formation. Based on these structural features, a proposal that folding of d(GAAA)$_n$, found in the non-coding region of genomes, into an octaplex can induce slippage during replication to facilitate length polymorphism is presented.

Key words: A-quartet, DNA octaplex, non-coding DNA, VNTR, X-ray analysis.

Human genome analyses have revealed that non-coding regions occupy more than 98% of the genome (*1*). VNTR (Variable number tandem repeat) is a family of non-coding sequences organized into long repetitive units, and it is dispersed throughout the genome of all vertebrates (*2*). Polymorphism of VNTR occurs as a consequence of mutational processes such as replication slippage and/or unequal sister chromatid exchange (*3*). Thus, the analysis of VNTR patterns has been employed as a method of DNA fingerprinting.

A VNTR immediately adjacent to the human pseudo-autosomal telomere contains the G-rich sequence d(ccGA[G]$_4$Agg)[1] (*4*), and exhibits a high degree of length polymorphism. The repetitive unit is repeated eight times, suggesting that this VNTR forms a specific structure. To examine the structural property of the sequence, DNA fragments including several analogues were prepared. X-Ray analysis revealed that DNA fragments with the sequence d(gcGA[G]$_1$Agc) (hereafter G1) form an octaplex with *I*-motif of G-quartets, in which four base-intercalated duplexes are assembled to interact at the central bases (*5*). Furthermore, we found that DNA fragments with the sequence d(gcGA[A]$_1$Agc) (hereafter A1), mutated at the 5th residues from G to A, also form base-intercalated duplexes. In the previous report (*6*), however, it was difficult to ascertain the formation of the octaplex due to disordering at the 5th residues. In the present study, the A1 crystal structure (tetragonal form) was revised. In the re-refined crystal structure, an octaplex structure with *I*-motif of double water-mediated A-quartets has been found. In this paper, the details of the octaplex structure will be described by comparing with that of the *I*-motif of double G-quartets, and its biological significance will be discussed.

The crystal structure of the tetragonal form of A1 (*6*) was previously refined with the space group *I422*. Only the 5th residue was assumed to be disordered, with two different conformers, **1** and **2**. For the present study, the crystallographic two-fold symmetry between the two strands of the duplex was initially released to refine (with *I222*) the disordered structure, and then the *I422* symmetry was applied in the final refinement. The X-ray diffraction patterns, which were collected in the previous study, were reprocessed with the space groups *I222* and *I422* using the program *HKL2000* (*7*). Statistics of data reprocessing with *I422* are summarized in Table 1.

Since the conformer **2** of A_5 in the previous structure did not fit to the electron density map [see the Supplementary Figure (*a*)], it was replaced with a new conformer, **3**, which was built on the $2|F_o| - |F_c|$ map with the program *QUANTA* (Accelrys Inc.). Under the *I*222 symmetry, four duplexes combined with the two strands containing conformers (**1-3***, **3-1***, **1-1*** and **3-3***)[2] are possible around the two-fold axis. However, the **1-1*** and **3-3*** combinations were ruled out because the distances between C4′ and C4′* and between O4′ and O4′* of the A_5 and A_5* residues were abnormally shorter than the lower limits of the van der Waals interactions. The two alternative structures containing **1-3*** and **3-1*** were separately refined using the programs *CNS* (*8*) and *Refmac5* (*9*) through a combination of crystallographic conjugate gradient minimization and *B*-factor fitting techniques. The refined **1-3*** and **3-1*** structures were symmetric to each other, and their *R* values were almost the same (19.7% and 19.8%). In the latter stages of refinement, half of the duplex structure, which

*To whom correspondence should be addressed. Tel: +81-45-924-5709, Fax: +81-45-924-5748, E-mail: atakenak@bio.titech.ac.jp

[1]Lowercase characters indicate that they can form a Watson-Crick G:C or C:G pair when the two fragments are aligned in an anti-parallel fashion.

[2]Asterisk represents the counter strand of the duplex around the two-fold axis.

Table 1. **Statistics of data processing and structure refinement.**

Data processing[a]	
Resolution (Å)	50–1.58
Observed reflections	68,468
Unique reflections	3,359
Completeness (%)	99.9
in the outer shell (%)	99.4
R_{merge}[b] (%)	6.3
Structure refinement	
Resolution range (Å)	8.0–1.60 ($F_o > 3\sigma$)
Used reflections	2,764
R-factor[c] (%)	19.0
R_{free}[d]	21.1
Number of DNA atoms	164
Number of ions	1.25 Mg^{2+}
Number of water molecules	72
R.m.s. deviation	
Bond lengths (Å)	0.013
Bond angles (°)	2.1
Improper angles (°)	1.5

[a]The programs *MOSFLM* and *HKL2000* were used respectively for the previous and present data processings. There are no changes in the unit cell dimensions regardless of the program used. However, the R_{merge} value in the present processing is slightly higher due to different computing techniques between the two programs. [b]$R_{merge} = 100 \times \sum_{hklj} | I_{hklj} - \langle I_{hklj} \rangle | / \sum_{hklj} \langle I_{hklj} \rangle$. [c]$R$-factor $= 100 \times \sum || F_o | - | F_c || / \sum | F_o |$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively. [d]Calculated using a random set containing 10% of observations that were not included throughout refinement (*10*).

was averaged between **1-3*** and **3-1***, was refined with the space group *I422*. Statistics of the structure refinements are summarized in Table 1. The electron density maps around the A_5 and A_5* residues are shown in Supplementary Figure, produced with the program *O* (*11*). All local helical parameters including torsion angles and pseudorotation phase angles of ribose rings were calculated using the program *3DNA* (*12*). Figures 1 and 2b were drawn with the program *RASMOL* (*13*).

The final R and R_{free} values for the revised structure are reduced to 19.0% and 21.1%, respectively (refer to Table 1 and Supplementary Table 1 for comparisons). The A_5 and A_5* residues are well fitted to the electron density map [see Supplementary Figure (b)]. Moreover, the atomic distances between C4′ and C4′* and between O4′ and O4′* of the A_5 and A_5* residues, which were questionable in the previous structure, are improved (C4′...C4′: from 2.9 to 3.6 Å and O4′...O4′: from 2.5 to 2.8 Å) to suit the van der Waals interaction ranges. Although the ribose-phosphate backbones of the A_5 residues are disordered, their adenine bases are not and occupy almost the same position, as shown in Supplementary Figure (c).

The local helical parameters and the sugar puckers of the revised structure are given in the Supplementary Table 2 (a). The overall structure of the present A1 duplex is almost the same as the previous one; there are no significant differences in the helical parameters. Only the sugar puckers of the A_5 residues are changed from C2′-*endo* to C1′-*exo* in one strand and to C3′-*exo* in the other [Supplementary Table 2 (b)]. These sugar conformations are still in the range of B-form conformations.

The two strands, **1** and **3***, form a base-intercalated duplex similar to those described previously (*6*). When the four duplexes are generated according to the crystallographic 222 symmetry, however, it has been found that they form an octaplex at the central part, as shown in Fig. 1a. Only the central A_5 residues are asymmetric. The helical axis of the octaplex is parallel to the crystallographic *c* axis.

Figure 1, b–f, shows the interactions among the four duplexes around the helical axis. In the octaplex, the eight A_5 bases are close together to form two quartets with water-mediated hydrogen-binding networks (hereafter each is designated as water-mediated A-quartet). The two water-mediated A_5-quartets stack on each other. Above and below the double A_5-quartets, the eight A_4 residues also form two other A-quartets by water-mediated hydrogen binding. In each of the water-mediated A_4-quartets, the adenine bases are shifted in a radial direction from the central helical axis to make a central space in which other water molecules are disordered.

The G_3 and A_6* residues form sheared pairs, and thus place the guanine bases inside of the octaplex, where the four G_3s are linked together by water molecules. At one end of the octaplex, the G_1 and C_8* residues, as well as the C_2 and G_7* residues form the Watson-Crick type pairs. The phosphate groups of the four G_1 residues, and likewise of the four C_2s, are projected inside the octaplex and interact to each other also by water networks. These structural features are the same at the other end of the octaplex due to the crystallographic symmetry.

As shown in Fig. 1 (c and d), the eight hydrated magnesium cations are bound specifically to the eight G_7 residues, the O6 and N7 atoms of which are directly hydrogen bonded to the cations. In addition, the cations are bridged to the N2 atom of G_3* in the same duplex and to the phosphate oxygen atom of G_3* in the adjacent duplex, through hydrogen bonds. The octaplex formation is thus stabilized by the magnesium linkages.

The G1 and A1 DNA fragments both form octaplex structures, suggesting the stability of octaplex formation. Here it is interesting to compare the G1 and A1 octaplex structures (see Fig. 1, f and g). In the G1 octaplex, the central double G-quartets are tightly formed through the two direct hydrogen bonds (N1-H...O6 and N2-H...N7). In addition, the G1 octaplex is stabilized by three potassium cations. One is located at the center of the space between the double G-quartets and is surrounded by the eight O6 atoms of guanine bases. The two remaining cations are bound to the four O6 atoms above and below the double G-quartets, respectively. In contrast, the A1 octaplex is stabilized by the magnesium linkages, which are found between the two duplexes. Furthermore, the A1 octaplex is swollen at the central region, in which several water molecules form hydrogen-bonded networks to stabilize the octaplex formation. These differences are reflected in the diagonal C1′...C1′ distances at the X_5 residues (21.6 Å for A1 and 16.2 Å for G1) and at the A_4 residues (23.1 Å for A1 and 19.2 Å for G1). The swelling of the central space of the A1 octaplex may allow the ribose-phosphate backbone conformation to disorder.

The central sequence of A1, d(GAAA), is repeated in human (*14*), canine (*15*), *Meloidogyne artiellia* (*16*) and *Oryza sativa* (*17*) genomes and exhibit length

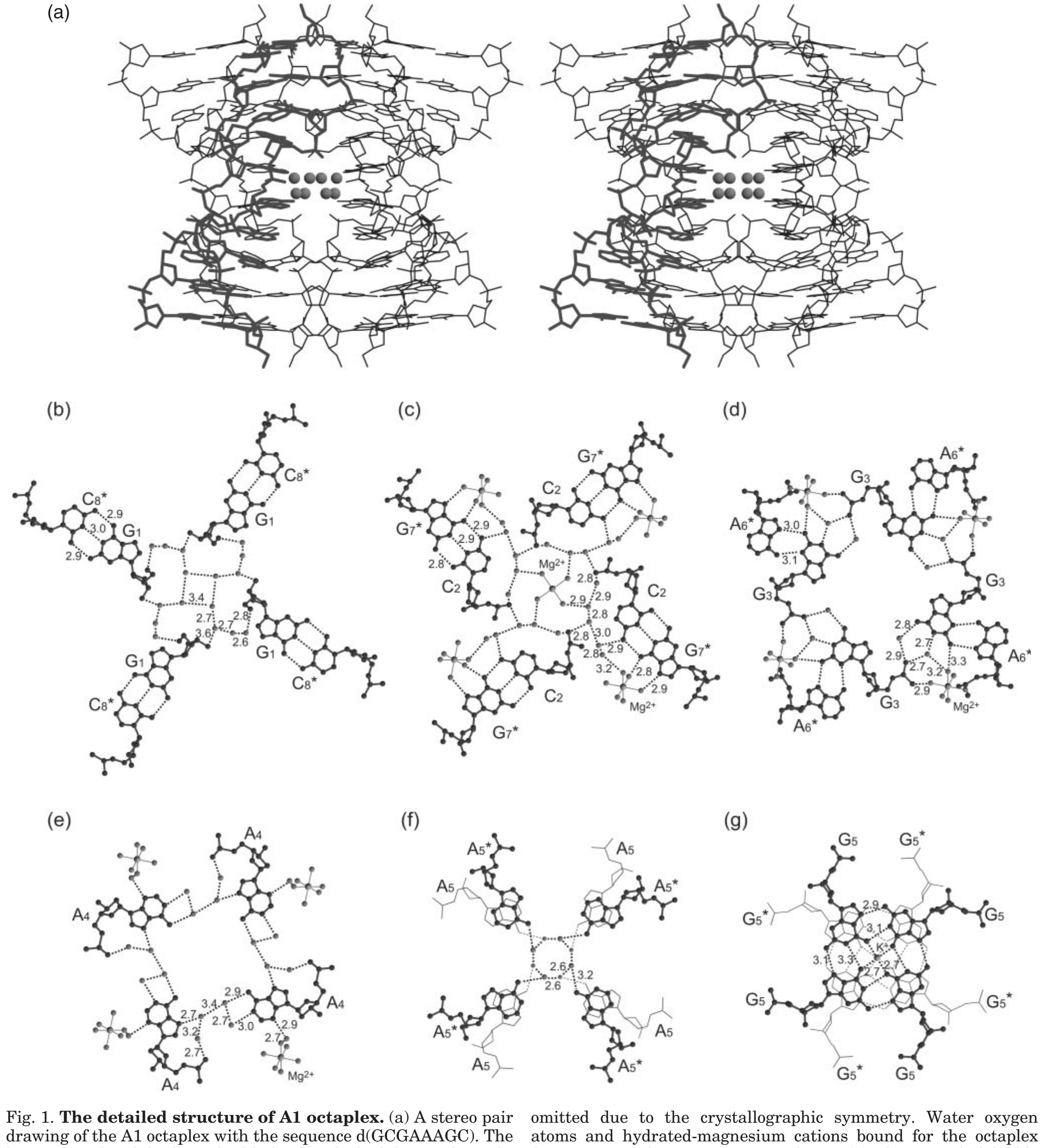

Fig. 1. **The detailed structure of A1 octaplex.** (a) A stereo pair drawing of the A1 octaplex with the sequence d(GCGAAAGC). The base-intercalated duplex is highlighted with thick lines. The eight central A_5 residues form an *I*-motif of double water-mediated A-quartets. Gray spheres represent water oxygen atoms participating in the stabilization of octaplex formation. (b–f) Detailed structures inside of the octaplexes. The nucleotides of the 1st to 5th residues in the A1 octaplex are drawn. The remaining parts are omitted due to the crystallographic symmetry. Water oxygen atoms and hydrated-magnesium cations bound for the octaplex formation are drawn with gray spheres. (g) The *I*-motif of double G-quartets found in the G_1 octaplex (*5*) is shown to compare with the *I*-motif of double water-mediated A-quartets of the A1 octaplex. A potassium cation (gray) is bound to stabilize the G-quartet formations. Broken lines indicate possible hydrogen bonds. Values indicate hydrogen bond distances (Å).

polymorphism (*14–16*). This short sequence is essential in the formation of the present octaplex, and is also already known to form a stable hairpin structure under low magnesium concentration conditions (*18*, *19*). Using the two structural motifs, we constructed an octaplex model for d(GAAA) repeats, as shown in Fig. 2, a and b. It is possible to speculate that when a long GAAA repeated region is folded into a cluster like an octaplex, it can induce sequence

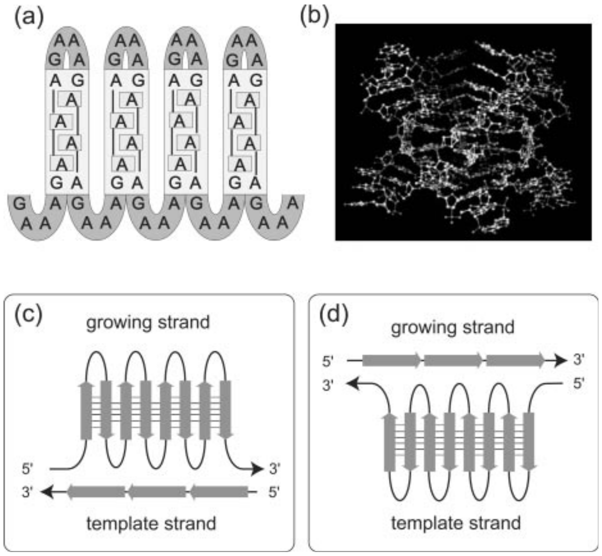

Fig. 2. **A possible folding of (GAAA)$_n$ repeats.** A schematic model constructed by alternately combining the d(GAAA) fragments that form the central part of the A1 octaplex and the stable hairpin structures of d(GAAA) (a), and its computer model (b). A slippage mechanism explaining repeat increase in the growing strand (c) and repeat decrease in the template strand (d).

slippage during DNA replication. As illustrated in Fig. 2, c and d, the number of repeats could increase when slippage occurs on the growing strand, or decrease when slippage occurs on the template strand. It is plausible that such slippages frequently occur to regulate transcription and translation of genes according to its number of repeats (*20*). To confirm the validity of our hypotheses, more extensive and intensive investigations including the structural analysis of GAAA repeats are required.

The atomic coordinates have been deposited in the Protein Data Bank (PDB) with the ID code 2DZ7. We thank M. Suzuki and N. Igarashi for facilities and help during data collection and E. C. M. Juan for proofreading of the original manuscript. This work was supported in part by Grants-in-Aid for Scientific Research (Nos. 12480177 and 14035217) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

REFERENCES

1. International Human Genome Sequencing Consortium (2004) Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931–945
2. Jeffreys, A.J., Wilson, V., and Thein, S.L. (1985) Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**, 67–73
3. Jeffreys, A.J., Royle, N.J., Wilson, V., and Wong, Z. (1988) Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. *Nature* **332**, 278–281
4. Inglehearn, C.F. and Cooke, H.J. (1990) A VNTR immediately adjacent to the human pseudoautosomal telomere. *Nucleic Acids Res.* **18**, 471–476
5. Kondo, J., Adachi, W., Umeda, S., Sunami, T., and Takénaka, A. (2004) Crystal structures of a DNA octaplex with *I*-motif of G-quartets and its splitting into two quadruplexes suggest a folding mechanism of eight tandem repeats. *Nucleic Acids Res.* **32**, 2541–2549
6. Sunami, T., Kondo, J., Hirao, I., Watanabe, K., Miura, K., and Takénaka, A. (2004) Structures of d(GCGAAGC) and d(GCGAAAGC) (tetragonal form): a switching of partners of the sheared G·A pairs to form a functional G·A×A·G crossing. *Acta Crystallogr.* **D60**, 422–431
7. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode in *Methods in Enzymology* (Carter, C.W. and Sweet, R.M., eds.) Vol. 276, pp. 307–326, Academic Press, New York
8. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T., and Warren, G.L. (1998) *Crystallography and NMR system*: a new software suite for macromolecular structure determination. *Acta Crystallogr.* **D54**, 905–921
9. Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr.* **D53**, 240–255
10. Brünger, A.T. (1992) Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475
11. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110–119
12. Lu, X.J. and Olson, W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121
13. Sayle, R.A. and Milner-White, E.J. (1995) RASMOL: biomolecular graphics for all. *Trends Biochem. Sci.* **20**, 374
14. Williams, S., Hayes, L., Elsenboss, L., Williams, A., Andre, C., Abramson, R., Thompson, J.F., and Milos, P.M. (1997) Sequencing of the cholesteryl ester transfer protein 5′ regulatory region using artificial transposons. *Gene* **197**, 101–107
15. Shibuya, H., Collins, B.K., Collier, L.L., Huang, T.H., Nonneman, D., and Johnson, G.S. (1996) A polymorphic (GAAA)$_n$ microsatellite in a canine Wilms tumor 1 (WT1) gene intron. *Anim. Genet.* **27**, 59–60
16. De Luca, F., Reyes, A., Veronico, P., Di Vito, M., Lamberti, F., and De Giorgi, C. (2002) Characterization of the (GAAA) microsatellite region in the plant parasitic nematode *Meloidogyne artiellia*. *Gene* **293**, 191–198
17. McCouch, S.R., Teytelman, L., Xu, Y., Lobos, K.B., Clare, K., Walton, M., Fu, B., Maghirang, R., Li, Z., Xing, Y., Zhang, Q., Kono, I., Yano, M., Fjellstrom, R., DeClerck, G., Schneider, D., Cartinhour, S., Ware, D., and Stein, L. (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res.* **9**, 199–207
18. Hirao, I., Nishimura, Y., Tagawa, Y., Watanabe, K., and Miura, K. (1992) Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effect on DNA sequencing. *Nucleic Acids Res.* **20**, 3891–3896
19. Sunami, T., Kondo, J., Hirao, I., Watanabe, K., Miura, K., and Takénaka, A. (2004) Structure of d(GCGAAAGC) (hexagonal form): a base-intercalated duplex as a stable structure. *Acta Crystallogr.* **D60**, 90–96
20. Nakamura, Y., Leppert, M., O'Connell, P., Wolff, R., Holm, T., Culver, M., Martin, C., Fujimoto, E., Hoff, M., Kumlin, E., and White, R. (1987) Variable number of tandem repeat (VNTR) markers for human gene mapping. *Science* **235**, 1616–1622