Crystallographic Studies on Damaged DNAs: III. N^4-Methoxycytosine Can Form Both Watson-Crick Type and Wobbled Base Pairs in a *B*-Form Duplex[1]

M. Tofazzal Hossain,* Toshiyuki Chatake,* Takaaki Hikima,* Masaru Tsunoda,* Tomoko Sunami,* Yoshihito Ueno,† Akira Matsuda,† and Akio Takénaka*,2

*Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226-8501; and †Graduate School of Pharmaceutical Sciences, Hokkaido University, Nishi 6, Kita 12, Kita-ku, Sapporo 060-0812

Received April 2, 2001; accepted May 30, 2001

To investigate the mutation mechanism of purine transition in DNA damaged with methoxyamine, a DNA dodecamer with the sequence d(CGCGAATTmo⁴CGCG), where mo⁴C is 2′-deoxy-N^4-methoxycytidine, has been synthesized and its crystal structure determined. Two dodecamers form a *B*-form duplex. Electron density maps clearly show that one of the two mo⁴C residues forms a pair with a guanine residue of the opposite strand, the geometry being the canonical Watson-Crick type, and that the other mo⁴C residue forms a wobble pair with the opposite guanine residue. These two pairings are ascribed to the tautomerization of the methoxylated cytosine moieties between the *amino* and *imino* forms.

Key words: damaged DNA, N^4-methoxycytosine, mutagenesis, Watson-Crick type pairing with non-complementary bases, X-ray structure.

In all organisms, complementary base-pair formation between adenine and thymine, and between guanine and cytosine is used as the absolute rule for ensuring the high accuracy of storage, expression and replication of genetic information. However, violation of this rule can occur when DNA bases are damaged by certain chemicals called mutagens (*1*). Such a mutagen is methoxyamine, which attacks adenine and cytosine moieties, converting them to N^6-methoxyadenine and N^4-methoxycytosine, respectively (*2, 3*). Crystallographic studies on DNA dodecamers containing N^6-methoxyadenine have revealed that methoxylated adenine (mo⁶A) residues form base-pairs with cytosine, as well as with thymine residues of the opposite strand, with Watson-Crick geometry (*4, 5*). On the other hand, N^4-methoxycytosine residues induce purine transitions (G→A or A→G) during replication. When 2′-deoxycytidine 5′-triphosphate (dCTP) is methoxylated (mo⁴dCTP), *Escherichia coli* DNA polymerase I incorporates mo⁴dCTP at both the A and G residues of the template DNA strand (*6*). When a cytosine residue of the template DNA strand is methoxylated, adenosine-5′-triphosphate (ATP) is incorporated by RNA polymerase at the site opposite the methoxylated cytosine (mo⁴C) (*7*). Similar reactions will occur in the case of DNA replication.

Several structural studies on oligonucleotides containing mo⁴C have been reported. The X-ray structure of a hexamer containing mo⁴C showed that mo⁴C is in the *imino* tautomeric state with a *syn* methoxy group relative to N^3, and forms a wobble pair with a guanine residue (*8*). In this duplex, the hexamer adopts a *Z* conformation. Such a conformation and wobble pairing are not compatible with DNA polymerase binding because the enzyme only accepts Watson-Crick base pairs in a *B*-form duplex (*9*). On the other hand, an NMR study of a DNA heptamer suggested that two structurally different geometries of mo⁴C:G base-pairs are in equilibrium (*10*). The postulated geometry was a Watson-Crick type and the other one was a wobble type. To confirm the NMR results and to visualize the exact interaction geometry of mo⁴C with G in a *B* form duplex, we synthesized a Dickerson-Drew type DNA dodecamer containing 2′-deoxy-N^4-methoxycytidine with the sequence d(CGCGAATTmo⁴CGCG). The crystal structure has been investigated by X-ray analysis.

2′-Deoxy-N^4-methoxycytidine was incorporated at the 9th position of the Dickerson-Drew type DNA dodecamer by the reported method (*11*). Droplets were prepared by mixing 2 μl of a 1.25 mM DNA solution and 2 μl of the reservoir solution. Suitable crystals were obtained at 4°C by the hanging drop vapor diffusion method when the reservoir solution contained 28% (v/v) 2-methyl-2,4-pentanediol, 36 mM magnesium acetate, 9 mM spermine tetrahydrochloride, and 20 mM sodium cacodylate (pH 7.0). Crystals were transferred to a cryoprotectant solution containing 40% (v/

[1] This work was supported in part by a grant for the RFTF (97L00503) from the Japanese Society for the Promotion of Science, by the Structural Biology Sakabe Project, and by a Grant-in-Aid for Scientific Research (No. 12480177) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

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

Abbreviations: dATP, 2′-deoxyadenosine 5′-triphosphate; TTP, thymidine 5′-triphosphate; dGTP, 2′-deoxyguanosine 5′-triphosphate; dCTP, 2′-deoxycytidine 5′-triphosphate; mo⁴C, 2′-deoxy-N^4-methoxycytidine; mo⁴dCTP, 2′-deoxy-N^4-methoxycytidine 5′-triphosphate; A, 2′-deoxyadenosine residue; T, thymidine residue; G, 2′-deoxyguanosine residue; C, 2′-deoxycytidine residue; ATP, adenosine 5′-triphosphate.

© 2001 by The Japanese Biochemical Society.

v) 2-methyl-2,4-pentanediol and then flash frozen in liquid N_2. X-ray diffraction data were collected at 100 K with the Sakabe-Weissenberg camera (*12*) with synchrotron radiation (λ = 1.00 Å) at the Photon Factory (BL-18b and BL-6b) in Tsukuba. For data collection, two crystals were used in different orientations to compensate for blind regions in reciprocal space. Diffraction patterns at 1.6 Å resolution were separately processed with the program DENZO (*13*), and intensity data were scaled and merged, and finally converted to independent structure factors with the programs SCALA, AGROVATA, and TRUNCATE in the CCP4 suite (*14*). Initial phases were derived by molecular replacement with the program AMoRe (*15*) using the atomic coordinates of the Dickerson-Drew type DNA dodecamer d(CGCGAATTCGCG) (*16*) as a probe. The molecular structure was constructed and modified on a graphics workstation by inspecting $|F_o|-|F_c|$ omit maps at every nucleotide residue with the program QUANTA (Molecular Simulation Inc.). According to the hydrogen-bonding scheme observed in the $|F_o|-|F_c|$ maps, one of the two mo^{4}C residues was assumed to adopt an *amino* form with the methoxy group in the *anti* conformation, and the other mo^{4}C residue was assigned as an *imino* form with a *syn* methoxy group. The structure was refined with the program CNS (*17*) through a combination of rigid body, simulated annealing, crystallographic conjugate gradient minimization refinements, and *B*-factor refinement, followed by interpretation of omit maps at every nucleotide residue. During refinement no restraints were applied between paired nucleotides. One magnesium cation, which is octahedrally coordinated by six water molecules, was found. A total of 110 peaks were assigned as water molecules. Crystal data, and statistics of data collection and structure determination are summarized in Table I. The atomic coordinates have been deposited in the Protein Data Bank (PDB) with the ID code of 1I3T.

TABLE I. **Crystal data, data collection, and structure determination.**

Space group	$P2_12_12_1$
Unit cell (Å)	a=25.2, b=40.9, c=63.7
Asymmetric unit (duplex)	1
Resolution (Å)	100–1.6
Unique reflections	9,068
Completeness (%)	97.6
in the outer shell (%)	93.9(1.66–1.6 Å)
R_{merge}[a] (%)	8.0
Structure refinement	
Resolution range (Å)	10–1.6
Used reflection	8,895
R-factor[b] (%)	22.3
R_{free}[c] (%)	25.6
DNA atoms	490
Water molecules	110
Magnesium atom	1
R.m.s. deviation from ideal geometry	
Bond lengths (Å)	0.004
Bond angles (deg.)	0.9
Improper angles (deg.)	1.4
Average B-factors (Å^2)	
DNA	22.1
Waters	34.4
Magnesium atom	19.1

[a]R_{merge} = 100 × $\sum_{hklj}|I_{hklj}-\langle I_{hkl}\rangle|/\sum_{hklj}\langle I_{hkl}\rangle$. [b]$R$-factor=100 × $\sum||F_o|-|F_c||/\sum|F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively. [c]Calculated using a random set containing 10% of the observations that were not induded during refinement (*18*).

Figure 1 shows an overview of two chains (a and b), which are associated to form a right-handed double helix. The root-mean-square deviation from the original Dickerson-Drew type duplex (*21*) is 0.92 Å for all non-hydrogen atoms except the methoxy groups. The two representative helical parameters (rise and displacement) and the sugar puckers (pseudorotation phase angle) fluctuate around average values close to those of the typical *B*-form DNA, like those of the original dodecamers, as shown in Fig. 1b. Their patterns plotted along the nucleotide sequences are very similar to each other, indicating that methoxylation causes no significant changes in the overall DNA conformation.

Interestingly, the present X-ray analysis has revealed that the two mo^{4}C residues form different types of base pairs with the opposite guanine residues in the same duplex. Figure 2 (a and b) show these base pairs with a $2F_o-F_c$ electron density map. At the mo$^4C_{a9}$ residue, the geometry of the mo$^4C_{a9}$:G_{b4} base pair is very close to that of the normal canonical Watson-Crick C:G pair (Fig. 2c). To form the three hydrogen bonds, $N^4\cdots O^6$, $N^3\cdots N^1$, and $O^2\cdots N^2$, the mo$^4C_{a9}$ residue should be in the *amino* tautomeric state. For this pairing, the methoxy group adopts an *anti* conformation with respect to the N^3 atom around the N^4-C^4 bond. At the other site, the mo$^4C_{b9}$ residue forms a base pair with G_{a4} in a wobble geometry through the $N^3\cdots O^6$ and $O^2\cdots N^1$ hydrogen bonds. Taking the donor/acceptor relationship into consideration, it is noted that the mo$^4C_{b9}$ residue should be chemically in the *imino* form. This is supported by the facts that N^4-methoxycytidine derivatives exist predominantly as the *imino* form (*24*), and even in the crystalline state (*25*). The methoxy group adopts a *syn* con-formation with respect to the N^3 atom around the N^4-C^4 bond. To compensate for a missing hydrogen bond, a water molecule links the O^2 atom of mo$^4C_{b9}$ and the N^2 atom of G_{a4} through two hydrogen bonds.

The present analysis has shown that N^4-methoxycytosine residues can form two types of pairs with the opposite guanine residues in a *B*-form DNA duplex, through tautomerization between the *amino* and *imino* forms. The existence of the two types of base pairs is consistent with the results of a NMR study on a DNA heptamer (*10*). The wobble type pairing is similar to that found in a *Z*-form hexamer (*8*). The Watson-Crick type pairing of the mo^{4}C residue is the first X-ray structure indicating that the methoxylated cytosine residue still has the ability to form this type of pairing with a guanine residue, just like an unmodified cytosine residue. It may be expected that the two geometries of pairing will have different effects on the DNA conformation. As can be seen in Fig. 1b, however, the local helical parameters and the sugar puckers apparently show no significant difference at the two sites (mo$^4C_{a9}$ and mo$^4C_{b9}$). As compared with the original dodecamer (*21*), the positions of the bases in the Watson-Crick type mo$^4C_{a9}$:G_{b4} pair are almost the same as those of the unmodified bases (see Fig. 2c). In the wobble base pair, the mo$^4C_{b9}$ residue is rotated by 15° toward the major groove, and the G_{a4} residue is rotated by 10° toward the minor groove. These movements slightly affect the phosphate conformations. Since the environments of the two mo^{4}C:G pairs are also different, the two pairing modes might be induced and stabilized by the surrounding water structures.

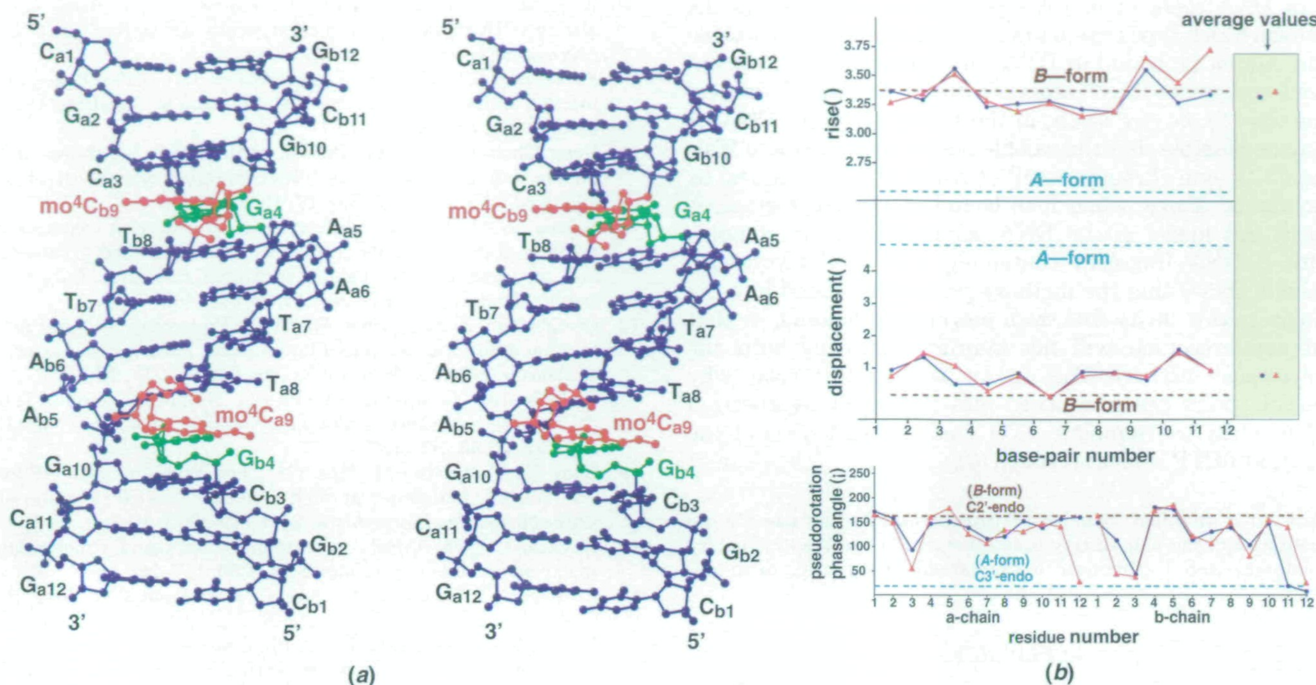

Fig. 1. **A stereo-view of the present DNA dodecamer with the sequence d(CGCGAATTmo⁴CGCG) (*a*), drawn with the program MOLSCRIPT (*19*), and the two local helical parameters and sugar puckering, calculated with the program NUPARAM (*20*) (*b*).** In (*a*), the mo⁴C residues are shown in red and their counter residues in green. The nucleotides are numbered from the 5′ end independently in the two strands, a and b. In (*b*), the helical rises (top), displacements (middle) and pseudorotation phase angles (bottom) of the present (△) and original (○) (*21*) dodecamers are plotted along the nucleotide sequences. Large local changes at the 3rd residue may be ascribed to the G:G interaction between the two duplexes and to a magnesium ion bound in the major groove as a common feature of Dickerson-Drew type dodecamers (*22*).

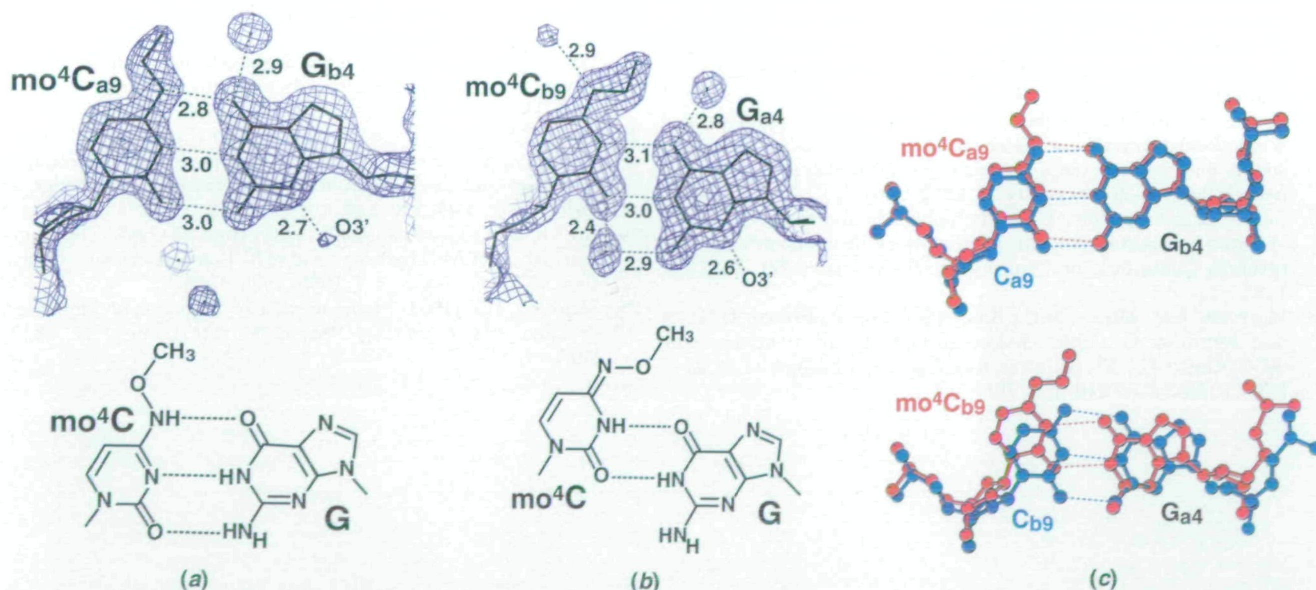

Fig. 2. **Final $2|F_o|-|F_c|$ electron density maps for the mo⁴C_{a9}:G_{b4} (*a*) and mo⁴C_{b9}:G_{a4} (*b*) base pairs, contoured at the 1.3σ level with the program O (*23*).** Broken lines indicate possible hydrogen bonds (distances in Å). The corresponding chemical structures are shown at the bottom. **Movements of bases, going from a Watson-Crick pairing to a wobble pair, are shown in (*c*).** For comparison, the unmodified C:G base pairs (*21*), drawn in blue, are superimposed on the mo⁴C_{a9}:G_{b4} (top) and mo⁴C_{b9}:G_{a4} (bottom) base pairs in red.

In DNA replication, DNA polymerase only accepts the Watson-Crick type base pairs in a *B* form duplex. A wobble pair cannot be bound in DNA polymerase, because the enzyme recognizes the O^2 atom of cytosine and the N^3 atom of guanine (*9*), or *vice versa*, in the minor groove of DNA. It may be possible that the wobble pair is converted to a Watson-Crick one through small rotation of the residues, as mentioned above, when it is bound in the polymerase. A structural model of the DNA polymerase (*9*) in complex with a DNA fragment containing a N^4-methoxycytosine residue shows that the methoxy group is extruded into the major groove cavity free from polymerase binding, so that the methoxylation will not interfere sterically with the polymerase activity. This is an additional reason why mo^4C:G pairs are introduced into the newly synthesized DNA when the template C is methoxylated (*7*) and the reactant dCTP is methoxylated (*6*).

We wish to thank N. Sakabe, M. Suzuki, and N. Igarashi for the facilities and the help during data collection at the Photon Factory (Tsukuba), and T. Simonson for proofreading the original manuscript.

REFERENCES

1. Singer, B. and Kusmierek, J.T. (1982) Chemical mutagenesis. *Annu. Rev. Biochem.* **52**, 655–693
2. Budowsky, E.I., Sverdlov, E.D., Spasokukotskaya, T.N., and Koudelka, J. (1975) Mechanism of the mutagenic action of hydroxyamine VIII. Functional properties of the modified adenosine residues. *Biochim. Biophys. Acta* **390**, 1–13
3. Kochetkov, N.K. and Budowsky, E.I. (1969) The chemical modification of nucleic acids. *Prog. Nucleic Acid Res. Mol. Biol.* **9**, 403–438
4. Chatake, T., Ono, A., Ueno, Y., Matsuda, A., and Takénaka, A. (1999) Crystallographic studies on damaged DNAs I. An N^6-methoxyadenine residue forms a Watson-Crick base pair with a cytosine residue in a *B*-DNA duplex. *J. Mol. Biol.* **294**, 1215–1222
5. Chatake, T., Hikima, T., Ono, A., Ueno, Y., Matsuda, A., and Takénaka, A. (1999) Crystallographic studies on damaged DNAs II. N^6-methoxyadenine can present two alternate faces for Watson-Crick base-pairing, leading to pyrimidine transition mutagenesis. *J. Mol. Biol.* **294**, 1223–1230
6. Reeves, S.T. and Beattie, K.L. (1985) Base pairing properties of N^4-methoxydeoxycytidine 5′-triphosphate during DNA synthesis on natural templates, catalyzed by DNA polymerase I of *Escherichia coli*. *Biochemistry* **24**, 2262–2268
7. Singer, B. and Spengler, S. (1981) Ambiguity and transcriptional errors as a result of modification of exocyclic amino groups of cytidine, guanosine, and adenosine. *Biochemistry* **20**, 1127–1132
8. Meervelt, L.V., Moore, M.H., Kong Thoo Lin, P., Brown, D.M., and Kennrad, O. (1990) Molecular and crystal structure of d(CGCGmo^{4}CG): N^4-methoxycytosine:guanine base-pairs in Z-DNA. *J. Mol. Biol.* **216**, 773–781
9. Kiefer, J.R., Mao, C., Braman, J.C., and Beese, L.S. (1998) Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature* **391**, 304–307
10. Fazakerley, G.V., Gdaniec, Z., and Sowers, L.C. (1993) Base-pair induced shifts in the tautomeric equilibrium of a modified DNA base. *J. Mol. Biol.* **230**, 6–10
11. Kong Thoo Lin, P., and Brown, D.M. (1989) Synthesis and duplex stability of oligonucleotides containing cytosine-thymine analogues. *Nucleic Acids Res.* **17**, 10373–10383
12. Sakabe, N. (1991) X-ray diffraction data collection system for modern protein crystallography with a Weissenberg camera and an imaging plate using synchroton radiation. *Nucl. Instrum. Methods Phys. Res. Ser. A* **303**, 448–463
13. Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode: Macromolecular crystallography Part A. *Methods Enzymol.* **276**, 307–326
14. Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. Sect. D* **50**, 760–763
15. Shui, X., McFail-Ison, L., Hu, G.G., and Williams, L.D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry* **37**, 8341–8355
16. Navaza, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr. Sect. A* **50**, 157–163
17. 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 & NMR System: a new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D* **54**, 905–921
18. Brünger, A.T. (1992) Free-R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475
19. Kraulis, P.J. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950
20. Bansal, M., Battacharyya, D., and Ravi, B. (1995) NUPARAM and NUCGEN: Software for analysis and generation of sequence dependent nucleic acid structures. *Comput. Appl. Biosci.* **13**, 281–287
21. Sines, C.C., McFail-Isom, L., Howerton, S.B., VanDerveer, D., and Williams, L.D. (2000) Cations mediate B-DNA conformational heterogeneity. *J. Am. Chem. Soc.* **122**, 11048–11056
22. Tsunoda, M., Karino, N., Ueno, Y., Matsuda, A., and Takénaka, A. (2001) Crystallization and preliminary X-ray analysis of a DNA dodecamer containing 2′-deoxy-5-formyluridine; what is the role of magnesium cation in crystallization of Dickerson-type DNA dodecamers? *Acta Cystallogr. Sect. D* **57**, 345–348
23. Jones, T.A., Zou, J.Y., Cowman, 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 Cystallogr. Sect. A* **47**, 110–119
24. Brown, D.M., Hewlins, M.J.E., and Schell, P. (1968) The tautomeric state of *N*(4)-hydroxy- and of *N*(4)-amino-cytosine derivatives. *J. Chem. Soc. Sect. C* **1968**, 1925–1929
25. Meervelt, L.V. (1991) Structure of 3′,5′-di-*O*-acetyl-N^4-methoxycytosine. *Acta Crystallog. Sect. C (Cr. Str. Comm.)* **47**, 2635–2637