Computational analysis of triplex formation of oligonucleotides: protonated and 5-methylated py-pu-py motif*

LEE Imshik, BAI Chunli (白春礼)**, WANG Chen (王 璟) and WANG Xinwen (汪新文)
(Institute of Chemistry, Chinese Academy of Sciences, Beijing 100080, China)

Received May 29, 1996; revised August 10, 1996

Abstract A triple helix composed of three dodecanucleotides, 5'-d(TC)₆, 5'-(AG)₆ and 3'-d(TC)₆ with each containing a methylcytosine or a protonated-cytosine, was examined in terms of various interactions. Computational analysis of the stability of triplex formation was consistent with the experimental observations. The Hoogsteen base pairing interaction was increased through substitution of 5-methylcytosine for cytosine and/or protonation. The interaction between water molecules and DNA triplex was visualized in terms of the end-to-end distance change of each strand. The chain length was reduced by about 7.5%—16.2% due to the screening effect. In addition, as the screening effect increases, Hoogsteen base pairs appear to be stable in triplex formation. Our computational results also show that the effect of protonated- and methylated-cytosine on the stability of triplex formation was due to the change of the electrostatic interactions rather than the surface stacking interaction.

Keywords: triplex DNA, computational analysis, stability, screening effect.

The stability of triple helical DNA is known to be related to the protonation or substitution of 5-methylcytosine for cytosine in pyrimidine strands. The improved binding stability of methylcytosine relative to that of cytosine in triple helix formation was first reported by Lee et al.[1]. The protonation of the N3 atom is also required for cytosine to form a stabler DNA triple helix[2,3,4]. The general models for DNA triple helices were determined from X-ray fiber diffraction[5] and NMR[6].

A triple oligonucleotide helix has promising potency in biological and biochemical applications because of its sequence-specific recognition. Recently, oligonucleotide-directed triplex formation is showing promise as a tool for site-specific cleavage of DNA, target sitting on RNA, inhibition of DNA-protein binding, and inhibition of gene expression[6—14]. For instance, the methylcytosine at CG pairing sequences has been associated with activation of gene expression in eukaryotes[12]. Brossallina and Toulme[14] proved that a hairpin oligonucleotide contains a sequence which can be used as a single strand probe to form a triplex DNA via interaction with a double helix target and vice versa.

An oligo pyrimidine-purine-pyrimidine DNA (py-pu-py DNA) triple helix system involves an antiparallel Watson-Crick-paired py-pu system containing a collinear Hoogsteen-paired pyrimidine strand which is bound to the purine target. Such a formation of triple helix by the nucleotides, 5'-d(AG)_n, 3'-d(TC)_n and 5'-d(TC)_n, could be influenced by the modified bases of 5-methylcytosine or protonated-cytosine.

* Work supported by the Chinese Academy of Sciences.
** To whom correspondence should be addressed.
Modifying the distribution of the hydrophobic and hydrophilic faces become a factor causing the chain to fold and to moderate the binding molecules to DNA. The hydrophilic sites (sugar-phosphate backbone) can stabilize water molecules by forming hydrogen bonds, while the hydrophobic groups (bases) lessen their interaction with water molecules by altering the normal structure of water molecules in the vicinity of the base sites. Both effects compete in aqueous solution. In order to increase the hydrophobic interaction, methyl groups attached to cytosine are preferable, because a methyl group is hydrophobic in general.

In the present work, a triple helix consisting of three dodecanucleotides 3'-d(TC)₆ (strand II), 5'-(AG)₆ (strand I) and 5'-d(TC)₆ (strand III), with each containing the TAT triad and CGC triad (fig. 1), was chosen for examining the effect of 5-methylcytosine and of protonated-cytosine on the formation of a triple helix in aqueous solution. The data obtained from a minimized three-dimensional triple helix were compared with the empirical data. An investigation of a minimized conformation reveals several interesting features. Strand-strand interactions, as well as inter- and intra-strand base stacking energy, were interpreted in our computation. Base-specified stacking was also analyzed in the work.

1 Molecular modeling and computational methods

The coordinates of a triple stranded triplex DNA, 3’-d(TC)₆ ⊕ 5’-(AG)₆ ⊙ 5’-(TC)₆ (where ⊕ indicates the Watson-Crick base pair and ⊙ the Hoogsteen base pair), were modified by replacing bases from the A-DNA structure determined by X-ray diffraction[5]. Fig. 1 shows those triads used in this work. A starting structure was constructed by keeping the general features of triple helical stem consisting of Watson-Crick base pairs and Hoogsteen base pairs which was displayed on an SGI 4D310 workstation. Models of CGC, CGC+, CGmC, mCGC, and mCGmC were built by using the optimized bases under AMBER force field in order to prevent model irregularities from producing an artificial substitution of a methyl group in geometry.

Fig. 1. A model of triads CGC, TAT, CGC+, CGmC, mCGC and mCGmC, and that of CGC triplex.