The Models of Proton Assisted and the Unassisted Formation of CGC Base Triplets

Chitrani Medhi†

Chemistry Department, Gauhati University, Guwahati-781014, India

Received August 7, 2001

The triple helix is formed by combining a double and a single strand DNAs in low pH and dissociates in high pH. Under such conditions, protonation of cytosine in the single strand is necessary for triplex formation where cytosine-guanine-cytosine (CGC+) base triplet stabilizes the triple helix. The mechanism of CGC+ triplet formation from guanine-cytosine (GC) and a protonated cytosine (C+) shows the importance of N3 proton. Similarly in the case of CGC (unprotonated) triplet, the donor acceptor H-bond at N3 hydrogen of the cytosine analog (C) initiates the interaction with GC. The correspondence between the two models of triplets, CGC+ and CGC, unambiguously assigned that protonation at N3 cytosine in low pH to be the first step in triplet formation, but a donor acceptor triplet (CGC) can be designed without involving a proton in the Hoogsteen H-bond. Further, the bases of cytosine analogue also show the capability of forming Watson Crick (WC) H-bonds with guanine.

INTRODUCTION

The triple helix nucleic acid is known for its usefulness in gene expression and therapeutic strategies.[1−6] It is formed by the interaction of a double helix DNA with single strand nucleic acid where the guanine-cytosine (GC) sequence recognizes adenine-thymine (AT) sequence in the double helix with less affinity for other sequences. In the process, CGC+ and TAT sequences are formed adjacent to each other in the triple helix.[7]

Experimentally, the triple helix has been observed only in low pH and dissociates to double and single strand nucleic acids in high pH.[8−12] The observations apparently indicate the importance of cytosine protonation in the single strand of DNA in triplex formation, and the requirement of cytosine protonation has been analyzed experimentally.[13−16] Moreover the previous works were mainly restricted to the influence of pH on the global equilibrium process, and protonation of individual bases within a sequence has not been analyzed either theoretically or experimentally.[17,21,22]

From the geometrical feature of triple helix, the stability in low pH is essentially due to the Hoogsteen H-bonds, and experimental observation shows that these bonds are lost in high pH with concomitant conversion of triple to double helix.[2,21] Thus the intermolecular forces of Hoogsteen H-bonds in CGC+ might be important for understanding the stability of triple helix.

Again, the stability of triple helix only in low pH gives rise to the question of its relevance under physiological condition when the stability of triplex is less. To enhance the stability in high pH the nucleosides are designed to replace the protonated cytosine so that the Hoogsteen H-bonds are of donor acceptor type but not through proton.[2,12] Thus we report here the various steps of CGC triplet formation in two different ways, one with protonated cytosine (C+) and the other with a base of cytosine analog (C). The base of cytosine analogue is further analyzed to assess their ability for H-bonding as Watson Crick GC so that these bases can gain practical applicability in intracellular gene therapy.

The H-bonds between the bases in double and triple helixes are stabilized by intermolecular forces, which are believed to contribute from electrostatic interaction with less contribution from delocalization or dispersion interaction. On the other hand, theoretical studies on various base pairs in double helix including base stacking have been reported.[22,23] Therefore, in the present study we use ab initio calculation with polarization function in the basis set for understanding the intermolecular interaction in base triplets.[22,23]

METHOD

Theoretical studies on the triplet formation are carried out by using ab initio method with STO-3G, 6-31G, and 6-31G* basis functions.[26] The model of Watson Crick GC base pair was constructed from the available geometrical parameters of the bases in the crystal structure of DNA.[24] The constructed geometry of GC is further optimized to get a reasonable structure as found in B DNA where the sugar−sugar distance (C1−C1) is approximately 10.7 Å.[23,24] Then the model of GC and optimized structure of protonated cytosine was used to construct the CGC+ triplet (Figure 1). The interaction energies at different intermolecular distances of GC and C+ are computed by ignoring geometrical relaxation in the bases during interaction. In the major groove of DNA, A1 and A2 positions of GC base pair are particularly engaged in Hoogsteen H-bonds (Figure 3). The H-bonding ability of protonated cytosine at these two sites were estimated for different intermolecular distances without considering either interaction due to stacking of vertical bases or the effect of change in configuration from sugar backbone.

Similarly the geometry of donor acceptor triplet (CGC) (Figure 2) has been constructed using a non-natural base (C) in place of protonated cytosine. Both the geometries of CGC+ and CGC are optimized with STO-3G basis sets, and we use these structures for evaluating energies with 6-31G and 6-31G* basis functions. We further analyzed few non-

† Corresponding author e-mail: chitrani@satyam.net.in.

10.1021/ci0100752 CCC: $22.00
Published on Web 04/11/2002

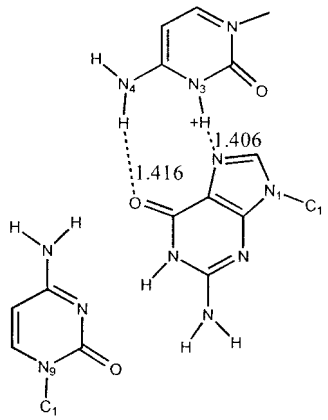

Figure 1. Structure of CGC+ triplet with optimized Hoogsteen bonds (bond lengths in angstrom).

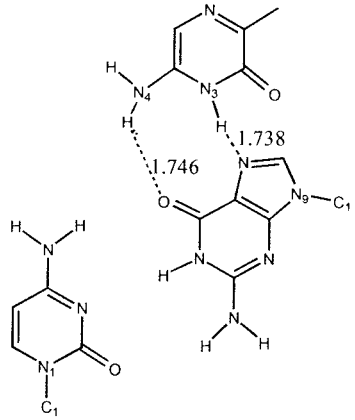

Figure 2. Structure of CGC triplet with optimized Hoogsteen bonds (bond lengths in angstrom).

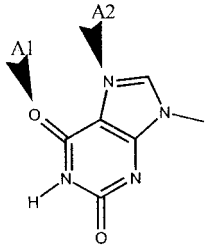

Figure 3. Hoogsteen hydrogen bonding sites of guanine in GC base pair.

natural bases that can form Watson Crick base pair (Figure 8a,b) with guanine. The models of the base pairs are constructed from individual bases and followed by geometry optimization with the restriction that the optimized geometries of the bases could accommodate in B DNA (sugar−sugar distance ~11 Å, ∠C1C1N1 and ∠C1C1N9 less than 90°). Geometry optimization is carried out with STO-3G basis sets, and the energies with higher basis sets are computed for the optimized geometry.

RESULTS

In both the CGC+ and CGC triplets, the two sites A1 and A2 of GC base pair (Figure 3) are used in Hoogsteen H-bonding with N3H+ or N3H and N4H of bases (C+ and C). The minimum interaction energies of protonated cytosine and GC through the optimized Hoogsteen H-bond lengths are computed (Table 1), and the corresponding geometrical features are shown in Figures 1 and 2. To analyze the

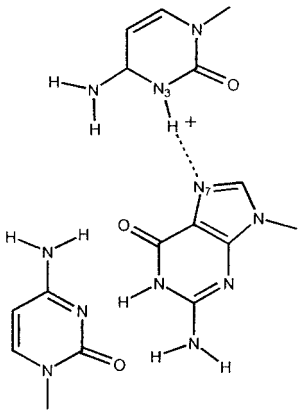

Figure 4. Unrelaxed geometry of CGC+at the minimum distance between GC and C+.

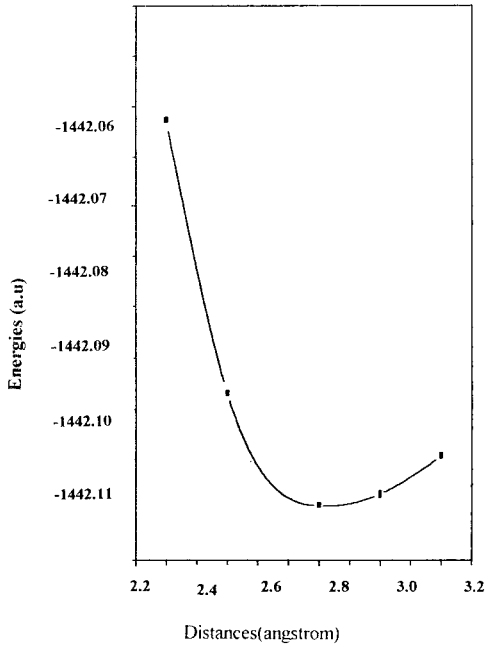

Figure 5. Plot of total energies versus the distances between GC and C+.

mechanism of CGC+ formation we have studied the interaction of GC and C+ at different intermolecular distances. It is observed that the GC base pair does not form Hoogsteen bonds at a distance greater than 3.3 Å, and interaction starts at this distance at N3H+ of protonated cytosine with A2 of guanine of GC; however, the other H-bond at A1 forms at a smaller distance (2.2 Å). The plot of total energies, calculated with 6-31G* basis sets versus R1, is shown in Figure 5, and the unrelaxed geometry (Figure 4) corresponding to the minimum energy of the plot indicates only one Hoogsteen bond, R2. From this observation, it is possible to understand the H-bonding abilities at A1 and A2 sites of GC with N3H+ and N4H of protonated cytosine in CGC+ (Figure 1). To differentiate the effect of proton at the these sites of GC, we have taken another model of triplet (CGC) using the non-natural base of a cytosine analog (C) instead of the protonated cytosine where Hoogsteen H-bonds are formed by donor acceptor interaction but not through proton. The minimum interaction energies of its optimized Hoogsteen hydrogen bonds are calculated with different basis sets (Table 1), and the correlation plot is shown in Figure 6. The two plots (Figures 5 and 6) are accounted as an ultimate test for

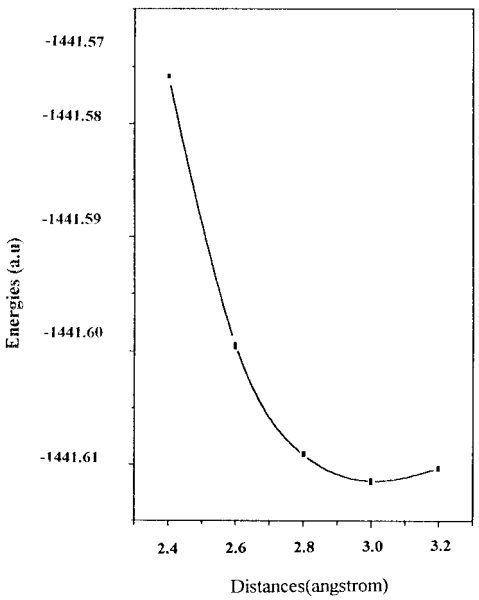

Figure 6. Plot of energies versus distances between GC and C.

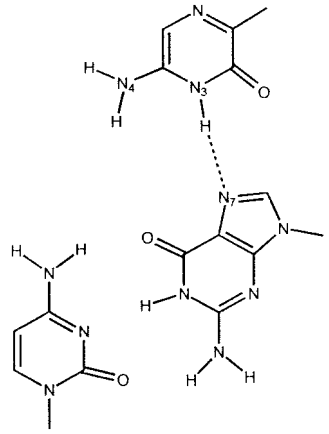

Figure 7. Unrelaxed geometry of CGC at the minimum distance between GC and C.

differentiating the H-bonding abilities of protonated cytosine and a non-natural base at A1 and A2 of GC. The geometrical features of triplets (Figures 4 and 7) corresponding to the minimum energies in the plots reveal only one Hoogsteen bond in each triplet which are N3H+...N (in CGC+) and NH....N (in CGC) bonds. Therefore, these bondings are apparently important in triplet formation. The energies of the relaxed geometries of CGC+ and CGC are calculated from the optimized structures where sugar−sugar distance of the GC base pair in the triplet is kept fixed at 10.7 Å approximately by assuming that the geometry of GC changes insignificantly and only the Hoogsteen H-bonds relax in the process (Table 2).

Interestingly the two H-bonds are of almost equal lengths in the relaxed structures (Figures 1 and 2). Therefore, from the geometrical viewpoint, the approach of the molecules, GC and C+, is substantially initiated by R2 bond, and the other bond R1 becomes operational only at a small intermolecular distance. Thus the interaction of C+ at A2 is more preferable than A1. The energy of the relaxed structure with both R1 and R2 Hoogsteen bonds is slightly lower than the unrelaxed structure which shows only one Hoogsteen bond, R2 (Figures 1 and 4). Hence the N4H...O bond (R1) contributes less significantly to the stability of CGC+, which

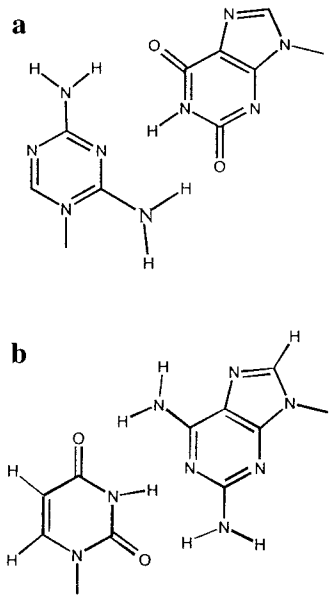

Figure 8. (a) Geometry of Mod1 and (b) geometry of Mod2.

Table 1. Binding Energies (in au) of CGC+ as a Function of Basis Sets

	total energies (au) of			
basis sets	GC	C+	CGC+	BE (au)
6-31G	−1009.6110	−431.8385	−1441.5119	0.062
6-31G*	−1010.0654	−432.0245	−1442.12831	0.038

Table 2. Binding Energies of CGC as a Function of Basis Sets

	total energies (au) of			
basis sets	GC	C	CGC	BE (au)
6-31G	−1009.6110	−431.4170	−1441.0668	0.039
6-31G*	−1010.0654	−431.6132	−1441.6929	0.014

on the other hand supports the experimentally observed instability at high pH when the N3H+..N bond (R2) cleaves due to deprotonation.

Moreover, it has been reported that the donor−acceptor triplet is stable in high pH, and the present theoretical study on donor acceptor CGC triplet is taken to be a model under such condition. It has been observed that CGC triplet formation occurs in the similar fashion as CGC+ but without a proton in Hoogsteen H-bonds.

The geometrical feature corresponding to the minimum energy in the plot is shown in Figure 7 in which only one N3H..N bond is operated in the triplet. Similar to CGC+, we observed very small differences in total energies of relaxed and unrelaxed structures; both Hoogsteen bonds are involved in the stability of the relaxed structures (Figure 2). Thus in the process of triplet formation either through proton or donor acceptor bond, R2 plays a more important role than R1. However the calculated binding energy of CGC+ is lower than CGC and has shorter Hoogsteen H-bonds, but with such low binding energies of both triplets it is not possible to comment on the significance of these triplets for understanding the stability of the triple helix.

The complementarity of H-bonding in both the triplets (CGC and CGC+) suggested that the non-natural base used in CGC triplet could also form full Watson Crick (WC) base pairs with guanine.[2] Such base pairs (Figure 8a,b) could have

Table 3. Comparative Structural Features of the Watson Crick (WC) GC Base Pair with Donor Acceptor Model of Non-Natural Base Pairs

base pairs	length between sugars	R1(Å)	R2(Å)	R3(Å)	angles C1C1N1, C1C1N9 (deg)
GC	10.69	2.53	2.64	2.72	65, 67
Mod1	10.91	2.74	2.69	2.75	60, 67
Mod2	10.66	2.67	2.69	2.69	60, 67

Table 4. Computed Energies of Watson Crick Base Pairs of Guanine and Bases (Cytosine and Its Analogues)[a]

base pairs	STO-3G	6-31G	6-31G*
GC	−997.217	−1009.611	−1010.065
Mod1	−997.230	−1009.612	−1010.065
Mod2	−997.235	−1009.625	−1010.066

[a] Energies (au) with basis sets.

advantages in intracellular gene therapy when the triplex stability under physiological pH is taken care of. We report here such base pairs, and their structures are compared with that of GC (Table 3). We impose certain restrictions to the geometry optimization so that these base pairs can be incorporated in B-DNA. In all cases we have fixed the sugar−sugar distance approximately 10.7 Å, and angles ∠C1C1N1 and ∠C1C1N4 are taken less than 90°. The constructed models of these base pairs are very similar to GC (WC) but have longer H-bonds and maintained perfect linearity. The computed energies of the optimized geometries (restricted to accommodate in B DNA) are not very different, hence energetically compatible (Table 4). The base pair, mod2, acquires lower energy than mod1 and GC. Thus, the non-natural bases might be suitable to incorporate in B DNA directly, which on the other hand can be considered as an approach to stabilize triple helix in high pH.

CONCLUSION

On the basis of the present study, it is obvious that protonation at N3 of cytosine plays a predominant role in the triplet formation, while the dissociation of triplex to doublex observed experimentally at high pH is due to the cleavage of the N3H+...N bond expected after deprotonation. Similar observations are found in the formation of donor acceptor CGC triplet where the N3H..N bonding initiates the interaction of GC and C (non-natural base). Thus the stabilities of both the triplets are strongly dependent upon N3H+..N and NH..N bonds. The geometrical features of CGC+ and CGC are found to be complementary.

Further it is shown that like cytosine, the non-natural bases can form WC bonding with guanine, and the use of such bases might be advantageous to compound the utilities of oligonucleotides used in gene therapy.

REFERENCES AND NOTES

(1) Moser, H. E.; Dervan P. B. *Science* **1987**, *238*, 645−650.
(2) van Krosigk, U.; Benner, S. A. *J. Am. Chem. Soc.* **1995**, *117*, 5361−5362.
(3) Sun, J. S.; Helene, C. *Curr. Opin. Struct. Biol* **1993**, *3*, 345−356.
(4) Cheng, Y. K.; Pettitt, M. *Prog. Biophys. Mol. Biol.* **1992**, *58*, 225−257.
(5) Frank-Kamenetskii, M. D. *Method Enzymol.* **1992**, *211*, 180−191.
(6) Mirkin, S. M.; Lyamichev, V. I.; Drushlyak, K. N.; Dobrynin, V. N.; Flippov, S. A.; Frank-Kamenestskii, M. D. *Nature* **1987**, *330*, 495−497.
(7) Griffin, L. C.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 967−971. Mergney, J.; Sun, J.; Rougee, M.; Montenay, G. T.; Barcelo, F.; Chromilier, M.; Helene, C. *Biochemistry* **1991**, *30*, 9791−9798.
(8) Singleton, S. F.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 6957.
(9) Cooney, D. A.; Thuong, N. T.; Helene C. *J. Am. Chem. Soc.* **1991**, *113*, 1457−1458.
(10) De los Santos, C.; Rosen, M.; Patel, D. *Biochemistry* **1989**, *28*, 7282−7289.
(11) Moser, H. E.; Dervan, P. B. *Science* **1987**, *238*, 645−650.
(12) (a) Rajagopal, P.; Feigon, J. *Nature* **1989**, *339*, 637−640. (b) Rajagopal, P.; Feigon, J. *Biochemistry* **1989**, *28*, 7859−7870. (c) Sklender, V.; Feigon, J. *Nature* **1990**, *345*, 836−838. (d) Voegel, J. J.; Benner, S. A. *J. Am. Chem. Soc.* **1994**, *116*, 6929−6930. (e) Plum, G. E.; Breslauer, K. J. *J. Mol. Biol* **1995**, *248*, 679−695. (f) Osborne, S. E.; Cain, R. J.; Glick, G. D. *J. Am. Chem. Soc* **1997**, *119*, 1171−1182.
(13) Lyamichev, V. I.; Mirkin, S. M.; Frank-Kamenetskii, M. D.; Cantor, C. R. *Nucleic Acids Res.* **1988**, *16*, 2165−2178.
(14) Povsic, T. J.; Dervan, P. B. *J. Am. Chem. Soc.* **1989**, *111*, 3059−3061.
(15) Meher, L. J., III; Dervan, P. B.; Wold, B. *Biochemistry* **1990**, *29*, 8820−8826.
(16) Plum, G. E.; Park, Y. W.; Singleton, S. F.; Dervan, P. B.; Breslauer, K. J. *Proc. Natl. Acad. Sci.* **1990**, *87*, 9436.
(17) Fossella, J. A.; Kim, Y. J.; Shih, H.; Richards, E. G.; Fresco, J. R. *Nucleic Acids Res.* **1993**, *21*, 4511−4515.
(18) Pilch, D. S.; Brousseau, R.; Schafer, R. H. *Nucleic Acids Res.* **1990**, *18*, 5743−5750.
(19) Manzini, G.; Xodo, X. E.; Gasparotto, D.; van der Marel, G. A.; van Boom, J. H. *J. Mol. Biol.* **1990**, *213*, 833−843.
(20) Pilch, D. S.; Levensen, C.; Shafer, R. H. *Biochemistry* **1991**, *30*, 6083−6087.
(21) Leitner, D.; Schrodinger, W.; Weisz, K. *J. Am. Chem. Soc.* **1998**, *120*, 7123−7124.
(22) (a) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302. (b) Hobza, P.; Sponer, J.; Polasek M. *J. Am. Chem. Soc.* **1995**, *117*, 792−798. (c) Sponer, J.; Leszozynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590−5596.
(23) Price, S. L.; Fabrizio Lo celso; Treichel, J. A.; Goodfellow, J. M.; Umrania, Y. *J. Chem. Soc., Faraday Trans.* **1993**, *89*, 3407−3417.
(24) (a) Voet, D.; Rich, A. *Prog. Nucleic Acid Res. Mol. Biol.* **1970**, *10*, 183. (b) Soyfer, V. N.; Potaman, V. N. *Triple helical Nucleic Acids*; Springer-Verlag: New York, 1996.
(25) (a) Escude, C.; Francois, J. C.; Sun, J. S.; Ott, G.; Sprinzl, M.; Helene, C. *Nucleic Acids Res.* **1993**, *21*, 5547−5553. (b) Holland, J. A.; Hoffmann, D. W. *Nucleic Acids Res.* **1996**, *24*, 2841−2848.
(26) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*; Gaussian, Inc.: Pittsburgh, PA, 1995.

CI0100752