PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE

Structural, electronic and energetic consequences of epigenetic cytosine modifications

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

P. Cysewski[a] and R. Oliński[b]

The hydrogen bonding patterns of cytosine and its seven C5-modifed analogues paired with canonical guanine were studied using the *first principle* approach. Both global minima and biologically relevant conformations were studied. The former resulted from full gradient geometry optimizations of hydrogen bonded pairs, while the latter were obtained based on 125 d(GpC) dinucleotides found in PDB database. The obtained energetic, electronic and structural data lead to the conclusion that the epigenetically relevant modification of cytosine may have serious consequences on hydrogen bonding with guanine. First of all, the significant substituent effects were observed for such trends as charges on sites involved in hydrogen bonding and the total intermolecular interaction energy or electron densities at bond critical points. Moreover, the molecular orbital polarization contribution resulting from energy decomposition expressed in terms of absolutely localized molecular orbitals exhibited an inverse linear correlation with frozen density contributions. A substituent effect on the amount of charge transfer from pyrimidine toward guanine was also observed. The increase of intermolecular interactions of guanine with modified cytosine is associated with increase of an electro-donating character of C5-substituent. However, only pairs involving 5-methylcytosine are more stable than ones formed by canonical cytosine. Furthermore, the energy differences observed for global minima remain important also for a broad range of displacement and angular parameters defining pairs conformation in model d(GpC) dinucleotides. Due to the sensitivities of intermolecular interactions to mutual arrangements of monomers the modification of cytosine at C5 site can significantly alter the actual energy profiles. Consequently it may be anticipated that the modified dinucleotides will adopt different conformations than a standard G-C pair in B-DNA double helix.

Introduction

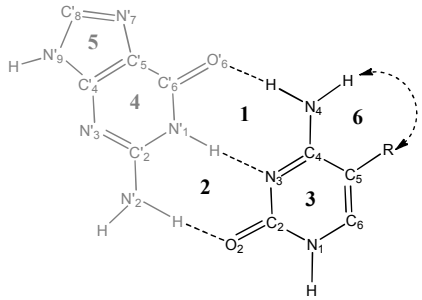
Cytosine methylation, usually at d(CpG) dinucleotides, is one of the most important epigenetic modifications with a profound impact on gene repression, cellular identity and organismal fate[1]. The occurrence of the d(CpG) dinucleotide is generally low in the bulk of genome but largely increasing in specific regions described as CpG islands (CGIs), which are DNA regions of about several hundred base pairs composed of at least 50% of GC content. Cytosine methylation may also occur in a non–CpG context (CpH, where H=A,T,C), but this kind of dinucleotides is rare in most mammalian cell types[1,2]. The DNA methylation has been studied as a stable epigenetic modification for decades[1-3]. However, an increasing body of recent experimental evidences implicates active DNA demethylation, which involves enzymatic oxidation of 5-methylcytosine (C5m) with subsequent formation of its hydroxylated form (5-hydroxymethylcytosine, C5hm) and other derivatives such as 5-formylcytosine (C5for) and 5-carboxyliccytosine (C5ac). This stands for the key event in epigenetic reprogramming[4]. Then base excision repair pathway is induced by involvement of TDG glycosylase to replace C5for and C5ca with cytosine to demethylate DNA[5]. It is worth noting that C5hm is not the direct substrate for thymine-DNA glycosylase. The results of a plethora of studies have confirmed the pivotal role of C5hm in active DNA demethylation while the role of the other derivatives is still obscure[6]. Epigenetic modifications are recognized by different proteins such as methyl CpG-binding proteins (MeCP1 and MeCP2), transcription factors, chromatin remodellers and DNA repair enzymes[7]. MeCP2 binding protein, in addition to binding mCpG, may also binds hmC what predominantly occurs at CpA nucleotides. Interestingly, mutation in the gene encoding MeCP2 may be directly linked with neurological disorders such as Rett syndrome[8-10]. However, the mechanisms by which the epigenetic modifications exert their effects are poorly recognized. Other kinds of C5 modification are inflammation-mediated DNA base modifications such as C5Br and C5Cl, which were identified in cellular DNA[11,12]. Interestingly, some data suggest that the aforementioned modifications, when present in DNA, are able to mimic C5m[13]. Moreover, 5-fluorocytosine is the oldest known antifungal agent[14]. It is possible that subtle differences of hydrogen bonding patterns

[a] Address here.
[b] Address here.

originating from the alteration of electronic densities and the energetics of intermolecular interactions between canonical base pairs and C5 modified analogues may have some influence on the structural properties of DNA, which in turn may affect the biological function of the C5-containing DNA duplexes. Therefore, the aim of this study was to characterize the substituent effects of C5-modification on the hydrogen bonded complexes of guanine and cytosine. In the first part the global minima of pairs formed by guanine with cytosine and its C5-modifed analogues (C5m, C5hm, C5ac, C5for and additionally the halogen derivatives C5F, C5Cl and C5Br) were characterized. Then model pairs in conformations found in B-DNA d(GpC) dinucleotides were used for a scan of intermolecular interaction energies (IIE).

Methods

Seven complexes formed between guanine and cytosine and the cytosine derivatives which are defined in Fig.1 have been examined. The ωB97XD density functional was invoked using the 311++G(2df,2pd) basis set implemented in GAUSSIAN package[15] to fully optimize the geometries of these complexes with the keyword "Int=Ultrafine". This approach takes advantage of accounting for dispersion interactions corrections[16]. The obtained geometries were used for computation of the Hirshfeld-type and NBO-derived atomic charges. Additionally cSAR index[17,18] was calculated for characteristics of electrostatic properties of the substituents. The values for intermolecular interaction energies (IIE) were computed including corrections for basis superposition error according to counterpoise approach[19,20]. Besides, the rings and hydrogen-bond critical points were characterized based on the Quantum Theory of Atoms in Molecules[21] using AIMAll software[22]. In order to characterize more precisely the contributions to pair stabilization an energy decomposition analysis was performed based on absolutely-localized orbitals method (ALMO)[23] implemented in QChem package[24]. In this approach the total intermolecular binding energy is decomposed into the three components, namely frozen density contribution (FRZ), the polarization term (POL) and the charge-transfer (CT) portion of IIE[25]. In the second part of the paper an energy scan was performed for characteristics of intermolecular interactions within canonical and modified G-C pairs in conformations resembling B-DNA structure. For this purpose the hydrogen bonding patterns of diverse G-C pairs were analysed by collecting different conformations observed in biochemical systems. Thus, the nucleic acid database (NDB)[26] was searched for B-DNA records comprising non-modified structures of double strands, without any ligands or mismatches. Conformations of d(GpC) were expressed in terms of six parameters defining mutual arrangement of hydrogen bonded pairs. Due to the accepted standard[27] it is sufficient to provide values of *Shear*, *Stretch*, *Stagger*, *Buckle*, *Propeller* and *Opening* for univocal characteristics of dinucleotides. The first three quantities represent displacement parameters and the other three define angles between molecular planes. For the purpose of structural analysis the X3DNA program[28] was used enabling direct analysis of deposited PDB files. In total, 125 pairs of G-C for configurational hyperspace characteristics were considered. Based on the obtained values the diversity of G-C pairs occurring in B-DNA can be documented. Having the spectrum of possible values a model of d(GpC) dinucleotides was constructed with monomers geometries coming from quantum chemistry optimizations. This step is indispensable since the direct use of PDB geometries is not recommended[29]. For this purpose the library of X3DNA was locally modified and used for generation of the span of structures encompassing experimentally observed values of all six parameters. Then, a series of IIE computations was performed for precise characteristics of structure to energy relationships. Such an approach of scanning of the configurations hyperspace was successfully used for heterogeneity of stacking interactions in DNA[30-32], non-additivities in dinucleotide interactions[33] or energetic consequences of oxidation of telomere repeat unit[33,34].



code	name	σ_p
C	cytosine	0.00
C5m	5-methylcytosine	-0.17
C5hm	5-hydroxymethylcytosine	0.00
C5ac	5-carboxylcytosine	0.45
C5for	5-formylcytosine	0.42
C5F	5-fluorocytosine	0.06
C5Cl	5-chlorocytosine	0.23
C5Br	5-bromocytosine	0.23

Fig.1. Schematic representation of structure of guanine paired with cytosine and its C5 modified analogues. The values of Hammett constant[36] characterizing each substituent are provided in the last column.

Results and discussion

The whole project was divided into two parts. In the first one some properties of the studied pairs in their global minimum conformations were characterized. After geometry optimization a series of computations was performed for characterizing detailed electronic and energetic properties. Special attention has been paid to the substituent effects on hydrogen bonding alterations as a consequence of substituent attachment to C5 atom of cytosine. In the second part the realistic arrangements of nucleotides were used for scanning the conformational hyperspace in closer proximity to biochemically relevant structures. This was done by application of conformations characterizing an experimentally available span of all six coordinates defining mutual arrangements of

monomers forming pairs within d(GpC) duplex in B-DNA conformation.

Global minima of G-C* pairs

One of the fundamental building blocks of B-DNA double helix comprises guanine-cytosine pair stabilized by three hydrogen bonds. In the case of epigenetic alterations of cytosine at C5 centre the formal bonding pattern remains unchanged as it is presented in Fig.2. However, such local modifications can alter the coding abilities of cytosine analogues due to the different nature of substituents. As it is documented in Fig.1 among seven epigenetically important cytosine derivatives there are ones enriched with a quite strong electro-donation and electro-withdrawing character of substituent, which can be quantified by values of Hammett constants, σ_p. The optimized structures of all pairs defined in Fig.1. were used for detailed electronic, energetic and structural characteristics of the consequences of cytosine C5 substitution. The values of Hammett constant provided in Fig.1. define the diversities of the electro-donating character of the studied substituents. As was presented in Fig.1. each pair comprises three internal heterocyclic rings constituting purine and pyrimidine heterocyclic moieties. In all the analysed pairs there are also two additional rings formed by hydrogen bonding patterns. The existence of all these rings was confirmed by electron density analysis and location of ring critical points on electron density surfaces. Although all cytosine derivatives possess an additional intramolecular ring involving amino group of cytosine and substituent, this ring was excluded from the detailed analysis.

It is well known that hydrogen-bonded nucleic acid base pairs substantially contribute to the structure and stability of nucleic acids[37-41] Base pairing via hydrogen bonding is categorized as resonance assisted H-bonds[39] and since nucleobases possess large dipole moments and are highly polarizable the electrostatic interactions dominate in DNA[34,39]. However, it is documented that the orbital interaction component can be of the same order of magnitude as the electrostatic component[41,42]. The substituent effect of halogen atom at C8 centre of purine and C6 positions pyrimidine has relatively small effects by reducing the hydrogen-bond-accepting and increases the hydrogen-bond-donating capabilities of a DNA base[42].

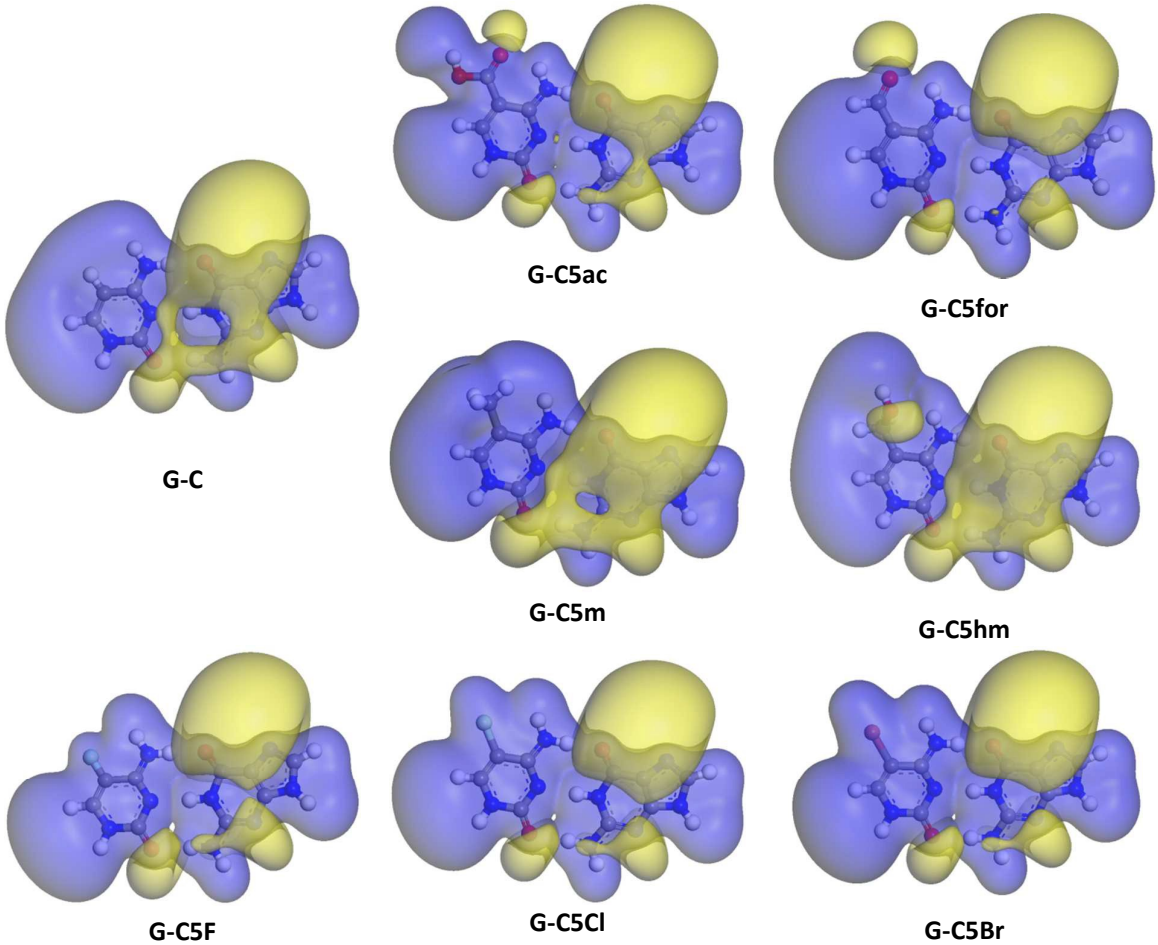


Fig.2. Representation of isosurface of electrostatic potential (±0.02 e/Å3) found in optimized pars of guanine paired with cytosine and its epigenetically important modifications. Positive values are in blue colour, while yellow represent regions with negative values.

Journal Name

ARTICLE

Since the electrostatic interactions are crucially important in characterizing intermolecular interactions between nucleobases[43,44] and their derivatives[45] the maps of electrostatic surface potential (ESP) seem to be important in the studied pairs. As one could expect the modification of cytosine at C5 position has significant consequences on the electrostatics of pairs with guanine. The electronegative region found in canonical G-C pair originating from C8, N7 and N3 centres of guanine encompasses almost the whole purine molecule including the hydrogen bonding area. On the contrary the space surrounding cytosine is characterized by a positively electrostatic field. This region of G-C pair, which is exposed out of the axis of the double helix, is the most important area from the perspective of interactions of DNA with other chemicals defining recognition patterns also for repair system agents. As one can see the ESP maps corresponding to global minima geometries of the studied pairs can be significantly affected by epigenetic modifications. Indeed, replacement of hydrogen atom with electro-accepting groups results in significant alterations of ESP maps, especially in the case of C5for and C5ac. The presence of additional electronegative centres breaks down the recognition pattern typical for G-C at groove side. Interestingly, similar effect is not so clear in the case of halogen-substituted cytosine, for which the Hammett constants have much smaller values compared to carboxyl or formyl groups. This is immediately reflected in ESP contours. In the case of C5hm an additional negative centre is present in the region responsible for stacking interactions. It might be expected then that intermolecular interactions within d(GpC5hm) dinucleotide will be much more strongly altered with respect to canonical dinucleotide. Due to electro-donating character the methyl substituent if introduced at C5 centre of cytosine leads to alterations of charge distribution around hydrogen bonding motives suggesting a rise of pairs stabilization. Interestingly, despite these changes in ESP the computed values of Fukui Function (FF) distributions proved that modification of cytosine does not lead to significant changes in all three FF indices. For all analysed pairs it is expected that the electrophilic attack takes place mainly on the guanine side since 89% of the total atomic FF(-) index is located on this site. This value is almost unchanged irrespective of the type modification imposed on cytosine. The probability of both nucleophilic and radical attack is almost equal for both constituents of the analysed pairs. Furthermore, the summarized values of FF(+) and FF(0) indices are not sensitive to cytosine substitution. It would be interesting to see if other properties are substituent dependent.

Substituent effect on charges distributions

The first aspect studied here refers to charges distribution on centres involved in hydrogen bonding. Two measures were used for quantifying substituent effects, namely Hammett constants and additionally cSAR index. Although these two measures are not independent since there is observed modestly correlation (R^2=0.874) between the two but they can provide two complementary and natural interpretations of substituent effects. According to chemical intuition the inverse trend is to be expected and with higher values of cSAR stronger electro-donating properties are to be expected. As was documented in Fig.3 substituents connected to C5 centre with cytosine introduce quite systematic changes on centres involved in hydrogen bonding. Interestingly, among existing three hydrogen bonds, both vicinal HB are much more prone to substituents then the central one.

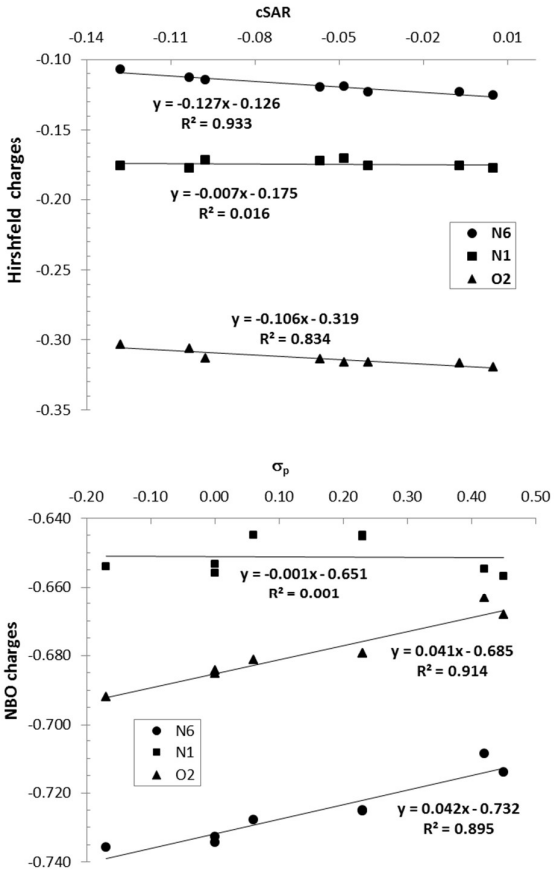


Fig.3. The charge distributions on donor and acceptor centres of modified cytosine the Hirshfeld charges as a function of cSAR parameter (top) and the NBO charges as a function of Hammett constant (bottom).

The presented correlations in Fig.3 provide consistent substituent effect expressed either via Hammett constant or cSAR index. The immediate and univocal conclusion states that the more the electro-donating group is present on C5 site the lower the charge resting on the N6 and O2 atoms. Interestingly, no such effect is observed for N1 atom irrespective of the type of cytosine modification.

Substituent effect on intermolecular interaction energies

The noticed substituent effects on charges distributions has also systematic influence on pair stabilization energies. Indeed, as is shown in Fig.4a the stronger the electro-accepting character of the substituent the lower the values of IIE. This trend is of a linear character and the strongest intermolecular interactions were noticed for G-C5m pairs. This intermolecular complex is about 0.6 kcal/mol more stable than canonical G-C pair. All pairs formed by other derivatives have been found to be less stable compared to the canonical one. Interestingly, the observed trend for cytosine seems to be different than one noticed[46] for other pyrimidines. Indeed, the role of the introduction of a methyl substituent at the pyrimidine ring on hydrogen-bonding has been proven to be quite small in the case of thymine and uracil pairing with adenine[46]. The stabilization of Watson–Crick type complexes of T with A are supposed to be about 1% decreased in comparison with T-A pair. The moderate electron-donating capacity of the methyl group can be attributed to this change. However, in the case of cytosine derivatives, the presence of amino group instead of oxygen center leads to the opposite effect since the electro-donating character of the methyl group strengthens the already strong amine group electro-donation. The analysis of optimized U-T and T-A pairs at the same level of theory as C-G pairs revealed the origin of such an inverse trend. First of all the presence of an additional hydrogen bond allows for much stronger interactions and electron densities relaxation. This can be expressed in values of frozen density and polarization contribution to the total binding energy. In the case of T-A and U-A pairs the FRZ terms are equal to -2.01 and -1.95kcal/mol, respectively while POL are equal to -6.13 and -6.19kcal/mol, respectively. Both these terms have much stronger stabilizing contribution to C-G pair and are equal to FRZ=-9.04 and POL=-11.71kcal/mol, respectively. These contributions are also more attractive if expressed per one hydrogen bond. Also base-to-base charge transfer is much stronger in the case of cytosine pairs than thymine and uracil. The CT contribution in the former case equals -9.21kcal/ and for T-A and U-A are only about -6.85 and -6.98 kcal/mol. Thus, the substituent effects in general is to be considered as pyrimidine-type dependent. The consequences of cytosine modification on the binding energy are quite consistent as it can be inferred from Fig.4a. The observed linear trend suggests that in the case of cytosine derivatives the more electro-accepting character of substituent the less stabilisation of WC pairs is to be expected. In Fig.4b the relative values of the contributions to the boding energy were presented with respect of optimized G-C pair. According to decomposition analysis the highest contribution to IIE comes from EPOL term and covers about 39.4% of the total IIE. The other two contributions, ECT and EFRZ, share the rest of binding energy in almost equal amount, namely 31.0% and 29.6%, respectively. Thus, it is evident that in all the cases studied here the non-charge transfer contributions to the binding energies are a more dominant portion of the IIE. Interestingly, the values of these energy contributions are almost independent of the nature of substituent. In this context it is worth noticing that the observed trend of IIE with change of Hammett constant values originates from interdependency of the FRZ and POL. Indeed, the correlation is as high as R^2=0.933 and suggests the rise of polarization with the lowering of frozen density contribution. This compensation is visible in the substituent effects on IIE.

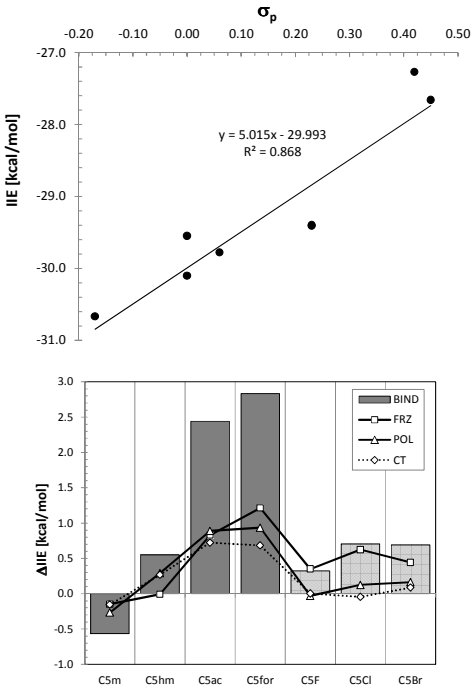


Fig.4. Energetic characteristics of global minima of pairs formed by guanine and C5 modified cytosine. The substituent effect on IIE is plotted in panel (a) and relative values of binding (ΔIIE) with respects of canonical nucleic acid G-C pair are presented in panel (b).

Substituent effect on hydrogen bonding properties

The observed inverse relationship between electro-donation of C5 substituent and pairs stabilities provokes a question about the sensitivities of each of the hydrogen bonds formed between guanine and cytosine analogue. For quantifying this effect the electron densities at bond-critical points, ρ_{BCP}, were computed based on Bader formalism[21]. It is commonly accepted that stronger hydrogen bonds are characterized by the higher values of ρ_{BCP}. In Fig.5 the obtained values of electron densities characterizing hydrogen bonding patterns were plotted as a function of IIE. The general trend of IIE expressed as a function of total ρ_{BCP}, summed over all three HBs, is fairly linear (R^2=0.971). This confirms the notion that stronger intermolecular interactions are associated with an increase of electron density in the hydrogen bonding region.

The strongest hydrogen bond N-H…O' seems to be formed between the donor centre of pyrimidine (N4-H) and the acceptor centre of guanine (O6'). This bond is fairly independent of the substituent type since ρ_{BCP} values remain on a similar level for all analysed complexes. The second-stable hydrogen bond is formed between heterocyclic centres of cytosine analogue (N3) and guanine (N1-H). The weakest bond is formed between oxygen acting as acceptor on the pyrimidine side (O2) and nitrogen atom of amino group (N'2) playing the donor role. The latter two bonds show modest correlation of ρ_{BCP} with the energetics of the system. Although the use of bond-critical points for estimating hydrogen-bonding interactions can in some cases pose a serious challenge[47] linear relationships between electron density at bond-critical point and stabilization energy of many hydrogen bonded complexes[48] were observed including nucleobases[49]. Even more, relationships between IIE and of ρ_{BCP} were used for classification of various types of hydrogen-bonded interactions[50].

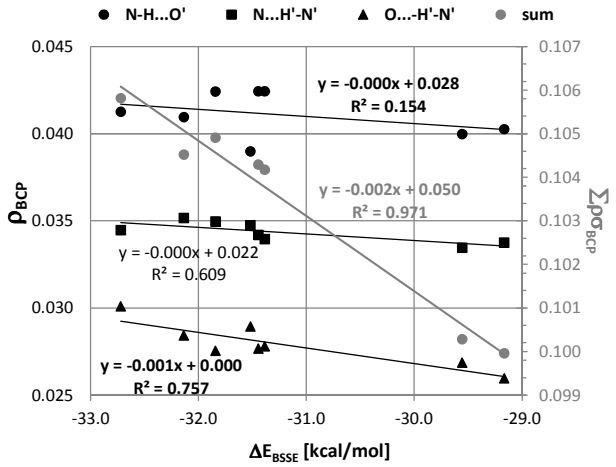


Fig.5. The correlation between electron densities of bond critical points of HB as a function of IIE of pairs formed between canonical guanine and modified cytosine.

Substituent effect on charge transfer

Intermolecular complexes stabilized by hydrogen bonds often allow for significant charge transfer between monomers. Although in the studied complexes this is not the most significant contribution it still is crucial for understanding of the nature of bonding patterns. The ALMO obtained during energy decomposition analysis can be used as a reference basis set for representing of bonding between molecules in terms of orbitals called complementary occupied-virtual pairs[51]. This model of orbital interactions can be used not only for conceptual description of intermolecular bonding but it can also quantify the amount of bidirectional charge transfer from and to interacting monomers. The results of decomposition of the total charge-transfer into complementary occupied-virtual pairs typically expressed as portion of electron charge dislocation leads typically to four contributions characterizing intra and inter CT of both components. Interestingly taking into account only intermolecular CT from pyrimidine to purine, a linear trend of the charge transfer as a function of Hammett constant values is observed, which is presented in Fig.6. This trend agrees with chemical intuition since it is expected that the more accepting nature of C5 substituent the stronger depletion of the ring π-electron delocalization what in turn reduces the ability of the charge transfer during formation of the intermolecular hydrogen bonds from the pyrimidine moiety. Interestingly, the opposite effect of charge transfer from guanine to modified cytosine is not associated with the parallel linear trend. This suggests that the observed substituent trend is of local character imposed on cytosine by C5 modifier.

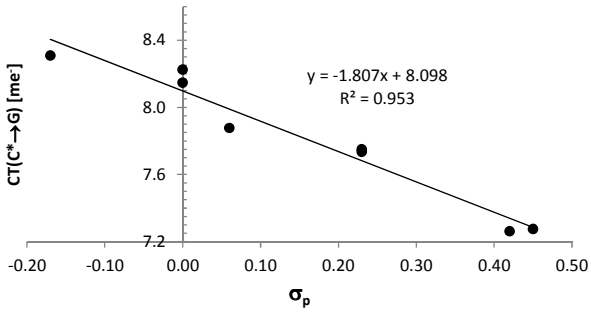


Fig.6. The substituent effect on the charge transfer in hydrogen bonded pairs formed between canonical guanine and modified cytosine.

Model d(GpC) in B-DNA conformations

The global minima obtained offer optimization, although interesting from the chemical point of view, need not be very informative in the context of biochemical mechanisms involving DNA functionalities. Indeed, in real recognition patterns[31,52] only a portion of the available energy is used in biochemical systems. Thus, more realistic conformations are to be considered. According to the details provided in methodology part the IIE scans were performed by altering only one variable. Based on these computations one can acquire energetic diversity of formed pairs as a function of mutual arrangements of monomers. The obvious advantage of such a model is the ability to apply precise IIE computations via quantum chemistry methodology to structures of biochemical relevance. The relative IIE values computed with respect to canonical G-C of given conformation were used for inspection of the energetic consequences of C5 cytosine substitution. The results of performed IIE computations for canonical G-C pair are presented in Fig.7 (see supporting materials). The first important conclusion is that the amount of the total available interaction energy used for the stabilization of systems in biological conditions is significantly lower compared to the ground state of G-C pair. One potential source of this fact is the relaxation of monomer geometries upon pair formation. However, in the case of guanine and cytosine these correction are almost identical and are equal to 1.5 kcal/mol. Thus, due to steric restrictions imposed by B-DNA conformations the formations of the most stable conformation between nucleic acid bases is restricted and helix stabilization is to be achieved by only a portion of the available attraction energy. This

fraction can still be quite significant and in the case of G-C pairs reaches 91% of the ground state IIE values. It is also worth mentioning that the range of both angular and displacement parameters characterizes the highest possible attraction range of G-C pair found in analysed d(GpC) dinucleotides in B-DNA conformation. The sensibility of IIE to parameter changes varies for different parameters and changes of the following parameters *Stretch*, *Opening* and *Shear* are the most influential on the stabilization energy. On the contrary the values of *Stagger*, *Buckle* and *Propeller* can be changed within a broad range without significant energetic penalties. It is interesting to see if modifications of cytosine at C5 position will result in changes of bonding patterns in pairs of model d(GpC*) dinucleotides. The performed energy scans were provided in supporting materials as Fig. 8 and 9. The most important conclusion coming from the obtained plots is that the energy differences observed for global minima remain important also for a broad range of displacement and angular parameters defining pairs conformations in model d(GpC) dinucleotides. However, due to the sensitivities of intermolecular interactions to mutual arrangements of monomers the modification of cytosine at C5 site can significantly alter the actual energy profiles. Indeed, the least stable pairs comprising C5ac and C5for remain least favourable compared to canonical pairs also in the whole ranges of all displacement and angular parameters. However, the lowering of *Shear* values and the rise of *Stretch* values significantly reduces the observed ΔE_{HB} difference. This suggests that the conformations of G-C5ac and G-C5for pairs will be significantly different in B-DNA double helix compared to canonical G-C pair. Similar effects are to be expected also in the case of C-5hm, for which the increase of *Stagger* and decrease of *Propeller* with respect to G-C pair are to be expected. Although these remarks of a qualitative nature do not allow for prediction of actual energies of modified pairs in B-DNA they provide at least a clue about expected trends in conformation-dependent intermolecular interactions between guanine and epigenetically modified cytosine. The detailed study on actual conformations adopted by modified dinucleotide in full B-DNA double helix environment are worth undertaking.

Conclusions

The modification of cytosine at C5 position is important from the biochemical perspective. The detailed inspection of chemical consequences on hydrogen bonding patterns of cytosine analogues led to conclusions allowing for a deeper understanding of observed phenomena in the biological context. First of all, the substituent effects documented by linear relationships of several structural, electronic and energetic properties with variation of Hammett constant or cSAR index were observed as for example the stabilization of hydrogen bonded pairs, the electrostatic properties on sites immediately involved in HB formation as well as individual bond strength. The resulting trends provide a clear and systematic overview of chemical modifications in the context of the pairing abilities of cytosine analogues.

The provided diversities of expected intermolecular interactions in model dinucleotide allow to distinguish three classes of pyrimidine analogues. To the first one belong C5for and C5ac, derivatives obtained after enrichment of cytosine with strong electro-accepting character. These two derivatives if involved in pair formation with guanine lead to weakening hydrogen bonded pattern with respect of G-C, both in global minimum and in conformations found in d(GpC) dinucleotides of B-DNA conformations. The second type represented by C5m is characterized by an increase of hydrogen bonding with guanine as the electro-donation is the most significant in the case of the methyl group. The rest of the studied cytosine analogues are almost indistinguishable from the energetic point of view, although many other properties are affected by C5 modification. Our results may be seen in the context of the recognition of epigenetic marks by different proteins. They may, at least partially, explain why TDG can recognize and excise from DNA C5for and C5ca while not C, C5m and C5hm when paired with guanine[53]. The results may also be helpful in explaining the finding that the presence of C5for and C5ca on the template DNA strand resulted in a great reduction, of those observed for C or C5hm, in the rate of guanine incorporation in the complementary position of transcribed RNA[54].

Acknowledgements

This research was supported in part by PL-Grid Infrastructure. The allocation of computational facilities of Academic Computer Centre "Cyfronet" AGH / Krakow / POLAND is also acknowledged.

Notes and references

1 W. Reik, W. Dean and J. Walter, Epigenetic reprogramming in mammalian development, *Science*, 2001, 293, 1089.
2 S. Guibert and M. Weber, Functions of DNA methylation and hydroxymethylation in mammalian development, *Curr. Top. Dev. Biol.*, 2013, **104**, 47.
3 W. Reik and W. Dean, DNA methylation and mammalian epigenetics, *Electrophoresis*, 2001, **22**, 2838.
4 S. Feng, S. E. Jacobsen and W. Reik, Epigenetic reprogramming in plant and animal development, *Science*, 2010, **330**, 622.
5 N. Bhutani, D. M. Burns and H. M. Blau, DNA demethylation dynamics, *Cell*, 2011, **146**, 866.
6 C. S. Nabel, S. A. Manning, and R. M. Kohli, The curious chemical biology of Cytosine: deamination, methylation,and oxidation as modulators of genomic potential, *ACS Chem. Biol.*, 2012, **7**, 20.
7 C. G. Spruijt et al, Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives, *Cell*, 2013, **152**, 1146.
8 H. W. Gabel, B. Kinde, H. Stroud, C. S. Gilbert, D. A. Harmin, N. R. Kastan, M. Hemberg, D. H. Ebert and M. E. Greenberg, Disruption of DNA-methylation-dependent long gene repression in Rett syndrome, *Nature*, 2015, **522**, 89.
9 C. Luo, J. R. Ecker, Exceptional epigenetics in the brain, *Science*, 2015, **348**, 1094.
10 L. Chen, K. Chen, L. A. Lavery, S. A. Baker, C. A. Shaw, W. Li, and H. Y. Zoghbi, MeCP2 binds to non-CG methylated DNA as

neurons mature, influencing transcription and the timing of onset for Rett syndrome, *PNAS*, 2015, **112**, 5509.
11 P. Lonkar, P. C. Dedon, Reactive species and DNA damage in chronic inflammation: Reconciling chemical mechanisms and biological fates, *Int. J. Cancer*, 2011, **128**, 1999
12 Kang JI. and L. C. Sowers, Examination of hypochlorous acid-induced damage to cytosine residues in a CpG dinucleotide in DNA, *Chem. Res. Toxicol.*, 2008, **21**, 1211.
13 V. V. Lao, A. Darwanto, L. C. Sowers, Impact of base analogs within a CpG dinucleotide on the binding of DNA by the methyl-binding domain of MeCP2 and methylation by DNMT1. *Biochemistry*, 2010, **49**, 10228–10236.
14 A. H. Vermes, J. Guchelaar, J. Dankert, Flucytosine: a review of its pharmacology, clinical indications, pharmacokinetics, toxicity and drug interactions, *J. Antimicrob. Chem.*, 2000, **46**, 171.
15 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. P. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez and J. A. Pople, Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.
16 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, A consistent and accurate ab initio parameterization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, *J. Chem. Phys.*, 2010, **132**, 154104.
17 N. Sadlej-Sosnowska, Evolution of Quantum Similarity Measures: How They Perform in Modeling Hammett Constants, *Pol. J. Chem.*, 2009, **83**, 2215.
18 N. Sadlej-Sosnowska, Substituent active region - a gate for communication of substituent charge with the rest of molecule: Monosubsituted benzenes, *Chem. Phys. Lett.*, 2007, **447**, 192.
19 S. F. Boys and F. Bernardi, Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors, *Mol. Phys.*, 1970, **19**, 553.
20 S. Simon, M. Duran and J. J. Dannenberg, How does basis set superposition error change the potential surfaces for hydrogen bonded dimers?, *J. Chem. Phys.,* 1996, **105**, 11024.
21 The Quantum Theory of Atoms in Molecules: From Solid State to DNA and Drug Design, Edited by C.F. Matta and R.J. Boyd, WILEY-VCH, Weinham, 2007.
22 AIMAll (Version 14.11.23), A. K. Todd, TK Gristmill Software, Overland Park KS, USA, 2014 (aim.tkgristmill.com).
23 R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell and M. Head-Gordon, Unravelling the origin of intermolecular interactions using absolutely localized molecular orbital, *J. Phys. Chem.* A, 2007, **111**, 8753.
24 Y. Shao, L. Fusti-Molnar, Y. Jung, J. Kussmann, Ch. Ochsenfeld, S. T. Brown, A. T.B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill, R. A. DiStasio Jr., R. C. Lochan, T. Wang, G. J.O. Beran, N. A. Besley, J. M. Herbert, Ch. Y. Lin, T. Van Voorhis, S. H. Chien, A. Sodt, R. P. Steele, V. A. Rassolov, P. E. Maslen, P. P. Korambath, R. D. Adamson, B. Austin, J. Baker, E. F. C. Byrd, H. Daschel, R. J. Doerksen, A. Dreuw, B. D. Dunietz, A. D. Dutoi, T. R. Furlani, S. R. Gwaltney, A. Heyden, S. Hirata, Ch. Hsu, G. Kedziora, R. Z. Khaliullin, P. Klunzinger, A. M. Lee, M. S. Lee, W. Z. Liang, I. Lotan, N. Nair, B. Peters, E. I. Proynov, P. A. Pieniazek, Y. M. Rhee, J. Ritchie, E. Rosta, C. D. Sherrill, A. C. Simmonett, J. E. Subotnik, H. L. Woodcock III, W. Zhang, A. T. Bell, A. K. Chakraborty, D. M. Chipman, F. J. Keil, A. Warshel, W. J. Hehre, H. F. Schaefer III, J. Kong, A. I. Krylov, P. M.W. Gill and M. Head-Gordon., Advances in quantum chemical methods and algorithms in the Q-Chem 3.0 program package, Phys. Chem. Chem. Phys., 2006, 8, 3172.
25 A.I. Krylov and P.M.W. Gill, Q-Chem: An engine for innovation, WIREs Comput. Mol. Sci., 2013, 3, 317.
26 B. C. Narayanan, J. Westbrook, S. Ghosh, A. I. Petrov, B. Sweeney, C. L. Zirbel, N. B. Leontis and H. M. Berman, The Nucleic Acid Database: new features and capabilities, *Nucleic Acids Res.,* 2013, **42**, D114.
27 W. K Olson, M. Bansal, S. K Burley, R. E Dickerson, M. Gerstein, S. C Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, Ch. S. Tung, E. Westhof, C. Wolberger and H. M Berman, A standard reference frame for the description of nucleic acid base-pair geometry, *J. Mol. Biol.*, 2001, **313**, 229.
28 X. J. Lu and W. K. Olson, 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures, Nucleic Acids Res, 2003, 31, 5108.
29 P. Cysewski and B. Szefler. Environment influences on the aromatic character of nucleobases and amino acids, *J. Mol. Model.,* 2010, **16**, 1709.
30 P. Cysewski, A post-SCF complete basis set study on the recognition patterns of uracil and cytosine by aromatic and p–aromatic stacking interactions with amino acid residues, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2636.
31 P. Cysewski, Intra-strand stacking interactions in B-DNA crystals characterized by post-SCF quantum chemistry computations, *New J. Chem.*, 2009, **33**, 1909.
32 P. Cysewski, Ż. Czyżnikowska, R. Zaleśny and P. Czeleń, The post-SCF quantum chemistry characteristics of the guanine–guanine stacking B-DNA, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2665.
33 P. Cysewski, Non-additive interactions of nucleobases in model dinucleotide steps occurring in B-DNA crystals, *J. Mol. Model.,* 2010, **16**, 1721.
34 P. Cysewski and P. Czeleń, Structural and energetic consequences of oxidation of d(ApGpGpGpTpT) telomere repeat unit in complex with TRF1 protein, *J. Mol. Model.*, 2010, **16**, 17970.
35 P. Czeleń and P. Cysewski, Structural and energetic properties of canonical and oxidized telomeric complexes studied by molecular dynamics simulations, *J. Mol. Model.*, 2013, **19**, 3339.
36 C. Hansch, A. Leo, R. W. Taft, A Survey of Hammett Substituent Constants and Resonance and Field Parameters, *Chem. Rev*. 1991, **97**, 165.
37 J. Sponer, P Jurecka, Pavel Hobza, Accurate Interaction Energies of Hydrogen-Bonded Nucleic Acid Base Pairs, *J. Am. Chem. Soc.* 2004, **126**, 10142.
38 J. Sponer, J. Leszczynski, P. Hobza, Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases, *Biopolymers,* 2001-2002, **61**, 3.
39 J. F. Beck and Y. Mo, How resonance assists hydrogen bonding interactions: An energy decomposition analysis, *J. Comput. Chem.,* 2007, **28**, 455.
40 H. Karabıyık, Resul Sevinçek and H. Karabıyık, Pi-Cooperativity effect on the base stacking interactions in DNA: is there a novel stabilization factor coupled with base pairing H-bonds? *Phys.Chem.Chem.Phys.*, 2014, **16**, 15527.

41 C. F. Guerra, T. van der Wijst, F. M. Bickelhaupt, Substituent Effects on Hydrogen Bonding in Watson–Crick Base Pairs. A Theoretical Study, *Struct. Chem.*, 2005, **16**, 211.
42 C. F. Guerra, F. M. Bickelhaupt, Charge Transfer and Environment Effects Responsible for Characteristics of DNA Base Pairing, *Angewandte Chemie Int. Ed.*, 1999; **38**, 2942.
43 Ż. Czyżnikowska, R. W. Góra, R. Zaleśny, P. Lipkowski, K. N. Jarzembska, P. M. Dominiak and J. Leszczyński, Structural variability and the nature of intermolecular interactions in Watson-Crick B-DNA base pairs, *J. Phys. Chem. B*, 2010, **114**), 9629.
44 Ż. Czyżnikowska, On the importance of electrostatics in stabilization of stacked guanine–adenine complexes appearing in B-DNA crystals, *J. Mol. Structure: THEOCHEM*, 2009, **895**, 161.
45 Ż. Czyżnikowska, How does modification of adenine by hydroxyl radical influence the stability and the nature of stacking interactions in adenine-cytosine complex?, *J. Mol. Model.*, 2009, **15**, 615.
46 M. Swart, C. F. Guerra, F. M. Bickelhaupt, Hydrogen Bonds of RNA Are Stronger than Those of DNA, but NMR Monitors Only Presence of Methyl Substituent in Uracil/Thymine, *J. Am. Chem. Soc.* 2004, **126**, 16718.
47 J Poater, M Solà, FM Bickelhaupt, Hydrogen–Hydrogen Bonding in Planar Biphenyl, Predicted by Atoms-In-Molecules Theory, Does Not Exist, *Chemistry-A European J.*, 2006, 12, 2889.
48 Grabowski, Hydrogen Bonding - New Insights, Springer, 2006
49 R. Parthasarathi, R. Amutha, V. Subramanian, B. U. Nair, and T. Ramasami, Bader's and reactivity descriptors' analysis of DNA base pairs, *J. Phys. Chem. A,* 2004, **108**, 3817.
50 R. Parthasarathi, V. Subramanian, and N. Sathyamurthy, Hydrogen Bonding without Borders: An Atoms-in-Molecules Perspective. *J. Phys. Chem.* A, 2006, **110**, 3349.
51 R. Z. Khaliullin, A. T. Bell and M. Head-Gordon, Analysis of charge transfer effects in molecular complexes based on absolutely localized molecular orbitals, *J. Chem. Phys.*, 2008, **128**, 184112.
52 P. Cysewski, The post-SCF quantum chemistry characteristics of inter- and intra-strand stacking interactions in d(CpG) and d(GpC) steps found in B-DNA, A-DNA and Z-DNA crystals, *J. Mol. Model.*, 2009, **15**, 597.
53 Y. F. He, B. Z. Li et al, Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA, *Science*, 2011, , 1300.
54 M. W. Kellinger, C. X. Song, J. Chong, X. Y. Lu, C. He and D. Wang, 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription, *Nat. Struct. Mol. Biol.*, 2012, **19**, 831.

Supporting materials

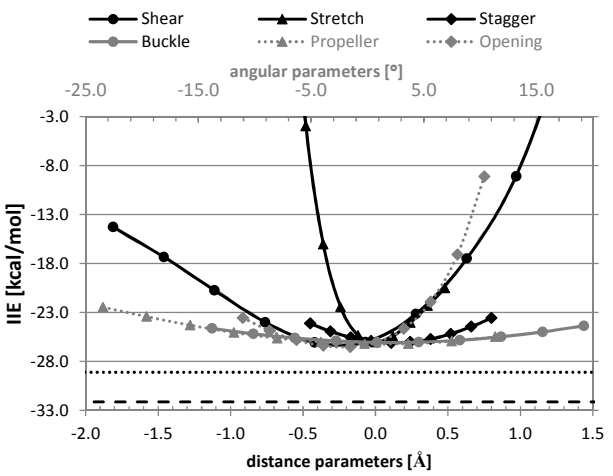


Fig. 7. The diversity of intermolecular interaction energy of canonical G-C pair in conformations characteristics to B-DNA double helix. Black lines represent variations of IIE as a function of displacement parameter, while grey curves stand for angular parameters. The corresponding energy level is denoted by dashed line. Since the energy scans do not include this contribution the correction for geometry relaxation were included and presented dotted line in Fig.7.

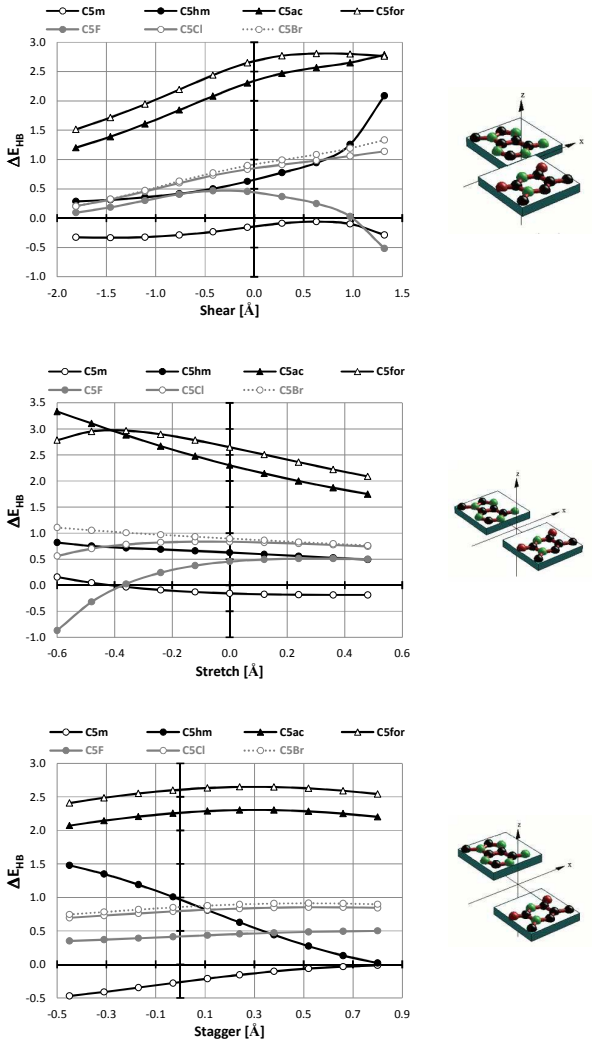


Fig.8. Scan of the relative intermolecular interaction energies of canonical guanine with 5-modified cytosine derivatives as a function of hydrogen bonding separation parameters defining. The reference points corresponds to values characterizing G-C pair documented in Fig.7. The positive values indicate lower pairs stability compared to analogical conformations of canonical G-C pair.

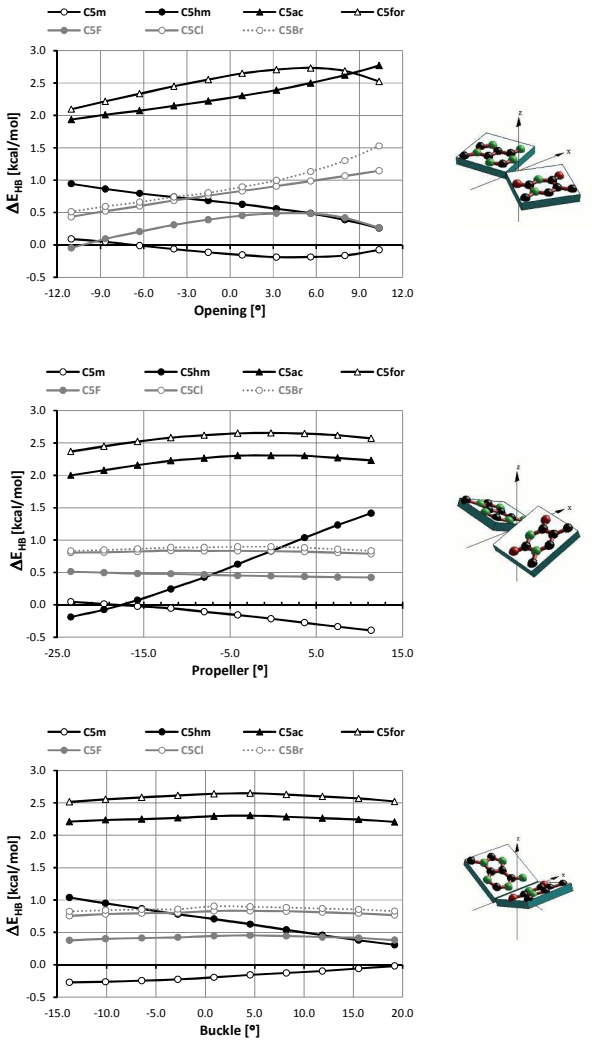


Fig.9. Scan of the relative intermolecular interaction energies of canonical guanine with 5-modified cytosine derivatives as a function of hydrogen bonding angle parameters defining. The reference points corresponds to values characterizing G-C pair documented in Fig.7. The positive values indicate lower pairs stability compared to analogical conformations of canonical G-C pair.