PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

First principles potential for the cytosine dimer

Artür Manukyan and Adem Tekin*

Informatics Institute, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey

E-mail: adem.tekin@be.itu.edu.tr

Abstract

We developed a new first principles potential for the cytosine dimer. The *ab initio* calculations were performed with DFT-SAPT combination of symmetry-adapted perturbation method and density functional theory, and fitted to a model site-site functional form. The model potential was used to predict cluster structures up to cytosine hexamers. The global cluster structure optimizations showed that the new potential is able to reproduce the some of the 2-D filament structures. Moreover many new non-planar cytosine cluster structures were also discovered. Interaction energies of these clusters were compared with B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 and AMBER. It has been shown that the model agrees well with all *ab initio* methods especially for cytosine hexamer. The model potential outperforms AMBER force field and therefore it can be exploited to study more challenging larger systems.

I. Introduction

Deoxyribonucleic acid (DNA) contains required genetic information of all known living beings and some viruses. Thus, DNA is considered to be the most important biological molecule. Non-covalent interactions, electrostatic (hydrogen bonding between O-H and N-H) and π-stacking, ensures the stabilization of DNA and RNA bases. Besides Watson-Crick base pairing, it is known

*To whom correspondence should be addressed

that DNA bases form triplexes, quadruplexes and many other complex structures.[1–12] Existence of H-bonded quadruplexes was confirmed by scanning tunneling microscopy (STM) and Atomic force microscopy (AFM).[13,14] Even mostly the quanine quartet was experimentally detected, Patel et al.[6] found a cytosine tetrad which is sandwiched between two quanine quartets in the sequence of d-TGGGCGGT. This planar structure is formed by H-bonds between O of cytosine and NH_2 of the other cytosine. Besides this pure c-tetrad, a mixed one (GCGC), sandwiched between two quanine quartets, having a Watson-Crick G-C pair interaction was also experimentally determined.[15]

Nowadays, single-molecule methods[16] are quite advanced, and this allows to embed individual DNA bases on metal surfaces. Resulting metal-DNA complexes are used in the fabrication of biochip sensors, organic semi-conducters, and organic photovoltaic tools.[17–19] To investigate the self-assembly formation mechanism of DNA base molecules, cytosine,[20–23] guanine,[21,24] adenine,[21,25–28] and thymine[21,29,30] were deposited on various metal surfaces. It has been observed via STM that all DNA bases forms one-dimensional (1D) or 2D (periodic or disordered) supramolecular planar netwoks which show distinct patterning based on the type of DNA base, metal surface and experimental conditions. In these networks, DNA bases are stabilized with hydrogen bonds and the metal surface acts as a platform to sustain the planarity of the network.[20] Computational approaches can be used to identify the existing structural motifs in STM images[31] e.g., Otero et al.[20] found that cytosine indicates three elementary patterns (zigzag filaments and five- and six-membered rings) on a gold(111) surface using density functional theory (DFT) calculations. In addition to DFT, molecular dynamics (MD) simulations can also be utilized to reveal the formation of structural motifs in comparison with STM images.[32] In such studies, force fields are required to compute the interactions between base-base and base-surface. More specifically, Maleki et al.[32] employed a combination of Lennard-Jones potential (van der Waals (vdW) parameters are obtained from AMBER force field) and a Coulomb term to account the interactions between base pairs. It is certain that improving the force fields for the base-base interactions might enhance the accuracy of the MD simulations.

Even though the selected dimers,[33–38] trimers[37] and quartets[37,39,40] of cytosine have been the-

oretically investigated, as far as we know, no attempt has been made to generate a force field to describe the interactions in cytosine oligomers. In this study, we implemented a new force field for cytosine based on the density functional theory-symmetry-adapted perturbation theory (DFT-SAPT) developed by Hasselmann and Jansen.[41–44] The model potential was then used to search the potential energy surfaces (PES) of cytosine clusters up to hexamer case. The results obtained so far offer indicate that the model potential might also be invoked in MD simulations.

The outline of this paper is as follows. Section II presents details on the quantum chemical calculations performed with the DFT-SAPT combination of symmetry-adapted perturbation theory and density functional theory and comparison to other high-accurate quantum chemical methods and AMBER force field. This section also reports the details of the fitting strategy. In Section III, the efficiency of all methods were compared and cytosine cluster structures generated by the model potential were discussed. Finally, we draw some conclusions in Sec. IV.

II. Method

A. Computational details

Selection of the theoretical method to compute the PES of any dimer has a critical importance for the assessment of the predictive power of the force field. Cytosine is not a symmetrical molecule and this certainly increases the number of unique orientations in the PES. Therefore, the preferred method of choice must be accurate and fast enough. Single and double excitation coupled cluster theory including perturbative triple excitations (CCSD(T)) is known as the most accurate method for the calculation of the intermolecular interactions. However, it is extremely time demanding e.g., for the computation of the PES of cytosine dimer. Much cheaper methods compared to CCSD(T) e.g., the second order Møller-Plesset perturbation theory (MP2), spin-component scaled MP2 (SCS-MP2),[45] SCS-MI-MP2[46] and DFT-D are more frequently employed supermolecular approaches for the tretament of the interaction energies. Here, B3LYP-D[47,48] was chosen as a DFT-D approach and it has been observed that its performance depends on the considered system

similar to PBE-D and meta-generalized gradient approximation functionals M05-2X and M06-2X, e.g., even it produces interaction energies which are similar to CCSD(T) for pyrazine[49–51] and paralle-displaced benzene[52] it fails for stacked benzene and H_2S-benzene dimer interactions.[52] As an alternative to supermolecular approach, one can also use DFT-SAPT to obtain the interaction energy as a sum of physically distinct components such as electrostatics, dispersion, induction and their exchange counterparts. It has been shown that DFT-SAPT with its superior scaling features with system and basis set size is in quite agreement with CCSD(T).[51,53,54]

In this study, we have compared the performance of DFT-SAPT and counterpoise (CP)-corrected supermolecular MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D and CCSD(T) for the stacked and hydrogen-bonded cytosine dimer orientations taken from reference[33] as shown in Figure 1. All the interaction energy computations were performed using MOLPRO program package[55] and density fitting implementation of DFT-SAPT (DF-DFT-SAPT)[56] has been used. The DF approximation was also employed in the supermolecular Hartree-Fock (HF), MP2 and SCS-MP2 calculations. In all these computations, Dunning's aug-cc-pVXZ (aVXZ) basis sets with X = D, T and in some cases Q were employed. Weigends's[57] cc-pV(X+1)Z JK-fitting basis conjunction with an aug-cc-pVXZ basis set was used in DF-HF and DF-DFT-SAPT. The corresponding aug-cc-pVXZ MP2-fitting basis[58] set was employed in DF-MP2 and DF-SCS-MP2. Asymptotically corrected PBE0AC[41] and LPBE0AC[56] exchange-correlation (xc) potentials in combination with the adiabatic local density approximation (ALDA) xc-kernel were utilized in DF-DFT-SAPT computations. Asymptotic correction required for the xc-potentials was calculated as the sum of experimental ionization potential of cytosine (8.80 eV)[59] and the highest occupied molecular orbital energy (-6.83 eV).

The developed force field for cytosine has been utilized to predict the structures of cytosine oligomers. Some of the resulting structures were further relaxed at PBE, SCS-MP2 and CP-SCS-MP2 levels of theory employing aug-cc-pVDZ basis set. For the larger cluster sizes, geometry optimizations were carried out only at SCS-MP2/aug-cc-pVDZ level. All these computations were performed using TURBOMOLE V6.1 program package.[60]

In addition to the comparisons performed with *ab initio* methods, AMBER non-bonded empirical potential was also utilized to calculate the interaction energies of the cytosine clusters. In AMBER force field,[61] as described in Morgado et al.[62] only vdW and Coulombic terms were considered. The atom-centered point charges were obtained by an electrostatic potential (ESP) fitting at PBE1PBE/aug-cc-pVDZ level of theory using the Merz-Kollman[63,64] scheme as implemented in Gaussian 09[65] software package.

B. PES and fitting details

A fixed geometry of cytosine, which was optimized at CC2/aug-cc-pVTZ, was used in the dimer configurations of the six-dimensional PES. The first dimension, R, defines the center of mass (cms) distance between the monomers and it was generated from the following set of distances R=3.0, 4.0, 5.0, 6.0, 7.0, 9.0 and 12.0 Å. To vary the position of the cms distance vector two polar angles with [0, 45 and 90°] and [28.13, 45.00, 92.52, 154.25, 208.25, 260.38, 285.38 and 318.82°] intervals were used. The reason of the employment of an unregular grid for the second polar angle is to allow the second cytosine to settle around the hydrogens, oxygen and nitrogen of the first cytosine. To determine the orientation of the second cytosine three polar angles with [0, 45, 90 and 270°] and [0, 45, 90, 135, 180, 225, 270 and 315°] intervals were employed. This selection of polar angles allows us to generate the hydrogen-bonded and stacked cytosine dimers. After elimination of symmetry-redundant and close-contact orientations, this grid resulted in 6477 dimer geometries.

Recently, one of the current author successfully developed a first principles potential for the acetylene dimer[66] using DFT-SAPT(PBE0AC) and resulting interaction energies were fitted to a site-site model potential. The same potential function shown below was used to fit the cytosine dimer:

$$V = \sum_{i\in A}^{sites}\sum_{j\in B}^{sites}\left\{\alpha_{ij}exp(-\beta_{ij}r_{ij}) + \frac{C_{ij}}{(r_{ij}^6 + c_{ij}^6)} + f_0(\delta_0^{ij}, r_{ij})\frac{q_i q_j}{r_{ij}}\right\}, \quad (1)$$

where r_{ij} denotes the distance between two sites i and j on the monomers A and B and α,

β, C are the fitting parameters. Each element in cytosine was assumed to be a different site and this gives a total of 10 pair interaction and 30 fit parameters. In equation 1, q_i and q_j are the ESP fitted partial charges (which were obtained as described in the previous subsection) and f_0 is the Tang-Toennies damping function[67] with $\delta = n = 1$

$$f_n(\delta, r) = 1 - exp(-\delta r) \sum_{m=0}^{n} \frac{(\delta r)^m}{m!}, \tag{2}$$

which eliminates the divergence of the Coulomb interaction term for $r_{ij} \rightarrow 0$. Dispersion interaction term is also damped using manually adjusted c_{ij} parameters to avoid unphysically large values for them for $r_{ij} \rightarrow 0$. Fit parameters were determined employing a Levenberg-Marquardt nonlinear weighted least squares method by minimizing

$$\chi^2 = \sum_{i=1}^{N} \sigma_i (y_0(x_i) - y(x_i; \alpha_{ij}, \beta_{ij}, C_{ij}))^2, \tag{3}$$

where y_0 and y refer to the ab-initio and model eenrgy at dimer geometry x_i, respectively. Individual weight terms, σ_i, were determined to each dimer geometry based on its interaction energy (E_{int}): σ_i was set to $1/y_0$ for $E_{int} > 1$ mH (2.6 kJ/mol) and to exp((1-y_0)/5) for $E_{int} \leq 1$ mH.

III. Results and discussion

A. Performance of Methods

Figures 2 and 3 compare the CP-corrected supermolecular (MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D and CCSD(T)) and DFT-SAPT interaction energies employing aug-cc-pVDZ (aVDZ) for the stacked and H-bonded dimer geometries of cytosine shown in Figure 1 as a function of the intermolecular distance R. All considered methods gave the H-bonded dimer to be more stabilized than the stacked one, e.g., H-bonded isomer is lower in energy than the stacked one by 41.63 kJ/mol at the CCSD(T) level of theory. As seen in Figure 3, MP2 generally overestimates the interaction energy compared to CCSD(T) for the stacked orientation. However, it is in very good

agreement with CCSD(T) for the H-bonded isomer. Such a performance of MP2 have already been reported for acetylene-furan[68] and pyrazine and triazine dimers.[51] Even though SCS-MP2 corrects MP2, it produces higher interaction energies than CCSD(T). SCS-MI-MP2 performs better than SCS-MP2 and it is very good agreement with CCSD(T). Among the methods, B3LYP-D gives the most overestimated energies similar to the behaviour of MP2 observed for the stacked dimer. In particular, B3LYP-D and MP2 produce the same results for the stacked dimer at very short intermolecular separations. However, this changes at the larger R where MP2 becomes lower in energy than B3LYP-D. DFT-SAPT(PBE0) performs similar to SCS-MP2 for the H-bonded geometry while it slightly underestimates the energies than SCS-MP2 for the stacked cytosine dimer. When LPBE0AC xc potential is used instead of PBE0AC, DFT-SAPT(LPBE0AC) almost reproduces CCSD(T) energies for both dimers.

Table 1 indicates the interpolated energies and the corresponding R for the lowest energy geometries of the potential energy curves (PEC) shown in Figures 2 and 3. As clear from this Table, the method that agrees most with CCSD(T) is DFT-SAPT(LPBE0AC) with a negligible differences in R (0.01 and 0.02 Å) and interaction energy (0.42 and 0.07 kJ/mol) for H-bonded and stacked dimers respectively. DFT-SAPT(LPBE0AC) was followed by SCS-MI-MP2 which shows the best performance among MP2 methods.

Table 2 compares the dependence of interaction energies to the basis set for the selected minimum energy orientations of H-bonded (R=5.31 Å) and stacked (R=3.50 Å) dimers. Moreover, this table also includes the extrapolated interaction energies to the complete basis set (cbs) limit using aVTZ and aVQZ energies in the two-point formula of Bak et al.[69] Unfortunately, we could not carry out CCSD(T) computations employing either aVTZ or aVQZ basis set due to the requirement of extreme computer resources. Therefore, we applied the strategy reported by Janowski et al.[70] to obtain the extrapolated energies of CCSD(T). Here, the difference between the CCSD(T) and SCS-MP2 interaction energies as obtained at the aVDZ level was assumed to be constant and the cbs limit for CCSD(T) was determined by the addition of this difference to the extrapolated SCS-MP2 energies. We found that employment of MP2 energies instead of SCS-MP2 has a negligible effect

on the extrapolated CCSD(T) energies. This strategy can also be verified using the interaction energies computed for the stacked benzene dimer in the study of Miliordos et al.[71] In particular, they obtained an energy difference between CCSD(T) and MP2 to be (1.168 - 2.253 kcal/mol) and (2.017 - 2.279 kcal/mol) for plain (cc-pVnZ) and augmented (aug-cc-pVnZ) basis sets, respectively, by using the CP-corrected interaction energies. These ranges were changed to (1.989 - 2.387 kcal/mol) and (2.144 - 2.375) upon the employment of the non CP-corrected energies. On the other hand, their results indicated that the difference between CCSD(T) and SCS-MP2 is more fluctuating compared to CCSD(T) and MP2. In the case of DFT-SAPT energies, only second-order $E_{disp}^{(2)}$ and its exchange counterpart, $E_{exch-disp}^{(2)}$, were used in the extrapolation formula and the remaining terms were taken from aVQZ level. The electron correlation (E_{corr}) and total intereaction energy have been extrapolated for MP2 (and SCS-MP2) and B3LYP-D respectively. As indicated in Table 2, for the H-bonded dimer MP2 energy was lowered by 1.95 and 1.36 kJ/mol from aVTZ to aVQZ and from aVQZ to cbs limit, respectively. These values were obtained to be 0.95 and 0.66 kJ/mol for the stacked geometry. A similar amount of decrease in the energies were calculated for SCS-MP2 and DFT-SAPT. It is also apparent that employment a larger basis set does not lead to a significant decrease in the interaction energy of B3LYP-D. Overall, DFT-SAPT(LPBE0AC) is in a very good agreement with the extrapolated CCSD(T) energies. All these findings motivated us to employ DFT-SAPT(LPBE0AC) to calculate the PES of cytosine dimer.

One of the most important feature of DFT-SAPT against to supermolecular approach is its ability to decompose the total interaction energy into the relevant energy components such as electrostatics ($E_{el}^{(1)}$),induction ($E_{ind}^{(2)}$), dispersion ($E_{disp}^{(2)}$), their exchange counterparts ($E_{exch}^{(1)}$, $E_{exch-ind}^{(2)}$ and $E_{exch-disp}^{(2)}$) and an error term ($\delta(HF)$). This offers an easy way to highlight the which forces are the most stabilizing factor in the dimer. Figures 4 and 5 indicate the energy components obtained with DFT-SAPT(LPBE0AC) using aVDZ as a function of R. Among them, $E_{disp}^{(2)}$, $E_{ind}^{(2)}$ and $\delta(HF)$ are always negative and their exchane counterparts are positive. $E_{el}^{(1)}$ is mostly negative but it can be positive based on R and the orientation. For H-bonded and stacked cytosine dimer, $E_{el}^{(1)}$ was determined to be negative. The magnitude of the repulsive contributions changes according to

the following ordering: $E^{(2)}_{exch-disp} < E^{(2)}_{exch-ind} < E^{(1)}_{exch}$. $E^{(1)}_{exch}$ collects almost more than half of the exchange energy. $E^{(2)}_{ind}$ contributes to total energy especially at short R while $E^{(2)}_{disp}$ remains significant even at longer R. $E^{(1)}_{el}$ and $E^{(2)}_{disp}$ were found to be the most dominant component throughout PEC of H-bonded and stacked dimers, respectively. Employment of LPBE0AC instead of PBE0AC leads to lowers the total energy and this constitutes the harmony between DFT-SAPT(LPBE0AC) and CCSD(T). The main reason of this decrease in DFT-SAPT(LPBE0AC) is the contribution of dispersion term. In particular, $E^{(2)}_{disp}$ was obtained to be lower energy than PBE0AC by 9.19 and 5.00 kJ/mol for H-bonded and stacked dimers, respectively.

Figure 6 shows basis set dependence of DFT-SAPT(LPBE0AC) for H-bonded (R=5.31 Å) and stacked (R=3.50 Å) dimers. It is clear that except $E^{(2)}_{disp}$ and $E^{(2)}_{exch-disp}$ all the remaining contributions were already converged at aVTZ. Therefore, in the extrapolation of DFT-SAPT energies to the cbs limit only these two components were considered.

B. Fitting

Interaction energies of 6477 dimer geometries in the computational grid were calculated with DFT-SAPT(LPBE0AC)/aVDZ and then fitted to equation 1 using Levenberg-Marquardt nonlinear weighted least squares method. The resulting parameters were listed in Table 3. When all grid points were considered, the standard deviation of the fit was calculated to be 13.81 kJ/mol. Taking only the subset of 4710 geometries with $E_{int} \leq 2.6$ kJ/mol into account, the standard deviation becomes 1.69 kJ/mol, while for the remaining structures with $E_{int} > 2.6$ kJ/mol it is 24.46 kJ/mol. Figures 7 and 8 compares the interaction energies resulting from the model potential to the original DFT-SAPT(LPBE0AC)/aVDZ values for the set of 6477 dimer structures. Our fitting is quite promising especially for the low-energy orientations. In particular, for structures with E_{int} < 30 kJ/mol, model was able to reproduce the DFT-SAPT values with an error less than 5 kJ/mol. However, the fit loses its prediction power for structures with $E_{int} > 30$ kJ/mol. The model potential estimates the most important stacked and hydrogen bonded orientations accurately as shown in Figure 9. Amongst the considered dimer structures, the stacked one at the top of Figure 9 has a

positive interaction energy at DFT-SAPT(LPE0AC) and this was successfully reproduced with the model. Other two orientations shown in Figure 9 are similar to ones depicted in Figure 1 and there is also a very good agreement between the model and DFT-SAPT energies for them. Here, model is only compared with the dimer structures which were used in the fitting. In the next section, the PES of model will unbiasedly be searched by a global optimization method and this is certainly a very though test for the judgement of predictive power of the force field.

C. Cytosine clusters

Dimer

The resulting 6-dimensional PES of DFT-SAPT model potential has been searched through a simulated annealing approach (SA).[72] In the SA implementation, the approval of a movement on the potential energy surface is determined via the Metropolis criterion[73] and the annealing is controlled by taking half of the temperature of the previous step. In this way, six H-bonded minima shown in Figure 10 were found and almost all of them were previously located by Kelly at. al.[35] Amongst them, C-2(1) and C-2(3) contain two H-bonds between O-H and N-H, respectively. C-2(2) is the only one with three H-bonds (two O-H and one N-H). The remaining ones consist of a single H-bond between N-H and O-H. We performed geometry relaxations at PBE, SCS-MP2 and CP-SCS-MP2 with aVDZ starting from SA structures to verify the quality of the predicted orientations. Table 4 compares H-bond and cms distances of model and ab-initio geometries. Mostly, initial structures obtained from SA were slightly changed upon the ab-initio relaxations. However, C-2(5) was transformed into a pseudo-stacked orientation after the CP-SCS-MP2/aVDZ optimization. There was also a notable difference in C-2(4) after SCS-MP2 and CP-SCS-MP2 relaxations, but the planarity of the resulting geometry is preserved. It seems that all PBE/aVDZ optimized geometries are in an agreement with the model and the structures reported in Ref.[35] In general, the shortest distances were obtained from PBE and longer ones from SCS-MP2 and CP-SCS-MP2. For C-2(1), C-2(2) and C-2(3), distances in model geometries were in excellent agreement with the ab-initio ones. In particular, for C-2(1), H-bond and cms distances of model were different from

CP-SCS-MP2 by only 0.02 and 0.05 Å, respectively. This agreement was partially deteriorated for the high energy orientations (C-2(4), C-2(5) and C-2(6)). Table 5 lists the interaction energies obtained from ab-initio methods (B3LYP-D, MP2, SCS-MP2, , SCS-MI-MP2, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ basis set), AMBER force field and model employing both the SA and CP-SCS-MP2 relaxed geometries. It seems that B3LYP-D overestimates the interaction energies and there is a similarity between MP2 - DFT-SAPT(LPBE0AC) and SCS-MP2 - DFT-SAPT(PBE0AC). Almost always, the magnitude of the interaction energy was obtained in the following order: B3LYP-D < MP2 $\approx$ SCS-MI-MP2 $\approx$ DFT-SAPT(LPBE0AC) < SCS-MP2 $\approx$ DFT-SAPT(PBE0AC). The model predicted C-2(1) as the lowest energy orientation and it was lower in energy than C-2(2) and C-2(3) by 1.88 and 5.12 kJ/mol, respectively. However, all considered ab-initio methods estimated that C-2(2) is lower in energy than C-2(1) by at most 3.95 kJ/mol. There is a clear consensus about the ranking of C-2(3): all methods and model predicted C-2(3) as the third most stable orientation. The model gives similar interaction energies for the remaining dimer structures and only C-2(4) was predicted different (the model prefers it almost two times more stable) from the ab-initio methods. In order to clarify the disagreement in energy ordering obtained from model and ab-initio methods, we further calculated the energies of dimer geometries which were optimized at CP-SCS-MP2/aVDZ. As already mentioned above, there is small differences between model and CP-SCS-MP2 structures and this may lead to a change in the energy ordering. As shown in Table 5, there is a significant reduction up to nearly 20 kJ/mol especially in the ab-initio energies. These computations reveal that C-2(1) is indeed the lowest energy orientation and it was lower in energy than C-2(2) by at most 4.01 kJ/mol. However, this finding was not reproduced by the model and C-2(2) was obtained lower in energy than C-2(1) by 1.24 kJ/mol. Actually, such an error is acceptable when the standard deviation of the fitting is considered. Interaction energies obtained employing CP-SCS-MP2 geometries are in accordance with the reported results of Kelly et al.[35] In particular, they calculated the energies of the three lowest energy orientaions to be 95.52, 92.63 and 83.94 kJ/mol, respectively and these values are similar to the ones obtained from SCS-MP2 and DFT-SAPT(PBE0AC). There is only a contradiction for

C-2(5) for which Kelly et al. computed an interaction energy of -41.49 kJ/mol and this value were obtained around 20 kJ/mol lower with SCS-MP2 and DFT-SAPT(PBE0AC). As a summary, the model potential successfully located the most important three (C-2(1), C-2(2) and C-2(3)) cytosine homodimers and their corresponding energy ordering was also in agreement to the one determined upon the employment of CP-SCS-MP2 geometries. This offers this model potential can be further exploited for the determination of larger cytosine homo-oligomers. Even AMBER force field fails for C-2(4), it locates C-2(2) as the lowest energy isomer similar to the model potential when CP-SCS-MP2 geometries were considered. However, in contrast to all other methods, AMBER prefers C-2(3) as the second stable isomer.

Trimer

For cytosine trimer and beyond, in the SA approach, the total interaction energy of the cluster was determined by summing all possible dimer interactions in the system. Thus, the many-body effects were neglected. By using this strategy, eight cytosine trimers shown in Figure 11 have been located. Amongst them, C-3(1), C-3(3) and C-3(6) are a combination of the most stable cytosine dimers of C-2(1)-C-2(2), C-2(1)-C-2(3) and C-2(2)-C-2(3), respectively. C-3(2) includes two C-2(2) interactions. The remaining ones contain at least one of the low-energy dimer configuration. All cytosine trimers are planar except C-3(8) which also contains stacking interaction. These trimers were further relaxed at PBE, SCS-MP2 and CP-SCS-MP2 levels employing aVDZ basis set and no significant change has been observed. CP-corrected interaction energies of trimers generated by the model were calculated at B3LYP-D, MP2 and SCS-MP2 levels with aVDZ basis set and listed in Table 6. C-3(1) was found as the lowest energy orientation by the model. Actually, the model predicted that except C-3(8) all the other structures are isoenergetic. In particular, C-3(1) is lower in energy than C-3(2) and C-3(3) by only 0.87 and 2.01 kJ/mol, respectively. In contrast to model, ab-initio interaction energy computations showed that C-3(2) is slightly lower in energy than C-3(1) and C-3(6) is the third most stable isomer. Both C-3(1) and C-3(6) were already reported by Kelly et al.[23] as the most common trimers connecting two cytosine filaments.

These filaments also contain C-3(2) and C-3(3) trimers as the main repeating units. Employment of CP-SCS-MP2/aVDZ optimized structures in the interaction energy computations changed the energy ordering: C-3(1) was found to be the lowest energy structure in all ab-initio methods and it was followed by C-3(2), C-3(3) and C-3(6).

As we declared above, the many-body interactions were omitted in the SA approach for the calculation of the total interaction energy. However, it is interesting to show the mangitude of these effects. Therefore, we computed the individual two- and three-body interactions of eight cytosine trimers shown in Figure 11. The corresponding energies were listed in Table 7. Here, the CP-corrected total interaction energy can be calculated using the following formula:

$$\Delta E_{int}^{CP} = E_{ABC}^{ABC} - E_{A}^{ABC} - E_{B}^{ABC} - E_{C}^{ABC} \tag{4}$$

These total trimer interaction energy may be broken down into two- and three-body components as follows:

$$\Delta^2 E_{int,AB}^{CP} = E_{AB}^{ABC} - E_{A}^{ABC} - E_{B}^{ABC} \tag{5}$$

$$\Delta^2 E_{int,AC}^{CP} = E_{AC}^{ABC} - E_{A}^{ABC} - E_{C}^{ABC} \tag{6}$$

$$\Delta^2 E_{int,BC}^{CP} = E_{BC}^{ABC} - E_{B}^{ABC} - E_{C}^{ABC} \tag{7}$$

The total interaction energy is then written as a sum of these two-body interaction energies plus a three-body interaction energy,

$$\Delta E_{int}^{CP} = \Delta^2 E_{int,AB}^{CP} + \Delta^2 E_{int,AC}^{CP} + \Delta^2 E_{int,BC}^{CP} + \Delta^3 E_{int,ABC}^{CP} \tag{8}$$

where ΔE_{int}^{CP} is the total trimer interaction energy, $\Delta^2 E_{int,AB}^{CP}$, $\Delta^2 E_{int,AC}^{CP}$ and $\Delta^2 E_{int,BC}^{CP}$ are the two-body components and $\Delta^3 E_{int,ABC}^{CP}$ is the three-body term. In the right hand side of these equations,

the superscripts denote the basis used and the subscripts denote the geometry. It is clear from Table 7 that $\Delta^3 E^{CP}_{int,ABC}$ is only negative if three cytosine monomers are close to each other as in the case of C-3(4), C-3(5) and C-3(7). In particular, the lowest three-body contribution was obtained for C-3(4) and it comprises 5 - 8 % of ΔE^{CP}_{int} depending on the method. These findings support the neglection of many-body interaction energy components during the computation of total interaction energy of trimers and larger clusters.

Tetramer

Investigations on different DNA sequences showed the existence of DNA base tetrads which are related to aging and cancers. Patel et.al[6] observed a novel cytosine tetrad in the DNA sequence of d-TGGGCGGT. This planar tetrad is formed via H-bonds between O and NH_2 of the neighboring cytosine. In the NMR structure, C-tetrad was actually found as stacked in between different guanine tetrads. This clearly suggests that the stability of the C-tetrad does not solely depend on the stability of the tetrad itself and the stacking interactions with the G-tetrads are also important. Gu et al.[40] optimized the coordinates of this C-tetrad using B3LYP/6-31G(d,p) method and their vibrational analysis showed that it is a minimum of the PES. This C-tetrad and seventeen other stacked, one-dimensional chain (filament) and planar cytosine tetramers shown in Figure 12 were found by the SA method and their corresponding interaction energies were included in Table 8. All the resulting tetramer structures were further relaxed at the SCS-MP2/aVDZ level and no significant change in the orientations were observed. Except C-4(14) which is not planarly aligned, these tetramers either contain a trimer or dimer motif shown in Figures 11 and 10. Amongst the filaments, C-4(4) and C-4(9) were already found by Kelly et al.[23] and the model was able to generate new filaments, stacked and a mixture of H-bonded and stacked cytosine tetramer orientations. The model potential preferred C-4(1) containing H-bonding and stacking interactions over the filament strutures as the lowest energy isomer. The second stable isomer in the model, C-4(2), is also a non-planar orientation and it can be derived from C-3(3). These two structures are isoenergetic with only a difference of 1.68 kJ/mol. In contrast, a filament structure, C-4(9), was found to be

the lowest energy tetramer at B3-LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 levels. Similarly, another filament, C-4(5), was obtained as the second stable isomer and it was higher in energy than C-4(9) by only 5.64, 4.10, 3.90 and 2.03 kJ/mol respectively at B3-LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 levels. These two filaments were actually higher in energy than C-4(1) by 3.64 and 4.38 kJ/mol, respectively, in the model. The C-tetrad with only four H-bonds was found to be the least stable tetramer by both the model and SCS-MP2/aVDZ level. Further relaxation of the model geometries with SCS-MP2/aVDZ altered the energy ordering both at model and other considered ab-initio levels. In particular, B3-LYP-D reversed ordering between the filaments, C-4(5) and C-4(9). MP2 and model favoured C-4(6), contains both H-bonds and stacking interactions, as the lowest energy isomer and this was followed by C-4(5) and C-4(9). Similar to B3-LYP-D, SCS-MP2 and SCS-MI-MP2 located the filaments C-4(5) and C-4(9) as the most stable tetramers. In contrast to all methods, AMBER prefers the stacked C-4(13) as the most stable isomer.

Pentamer

Cytosine pentamers have been recently found on the Au(111) surface via STM in clean ultrahigh vacuum conditions.[23] In the study of Kelly et al.,[23] they observed disordered structures which include filaments and five- and six-fold rings instead of an ordered arrangement of cytosine filaments. These filament and ring structures as well as the non-planar orientations were found by the SA method using the cytosine intermolecular potential energy function. Figure 13 shows these nineteen cytosine pentamer structures and Table 9 lists the corresponding interaction energies. For the selected cases (two non-planar (C-5(1) and C-5(2)), two ring (C-5(3) and C-5(17)) and one filament (C-5(6))), further geometry optimization of structures found from SA method were performed at the SCS-MP2/aVDZ level of theory and the corresponding interaction energies were calculated and they were also reported in Table 9. The SA method was able to reproduce five of the ring structures (C-5(17), C-5(15), C-5(16), C-5(19) and C-5(14)) already found by Kelly et al.[23] Moreover, another ring structure, C-5(13), was also found and which is quite similar to C-5(9) of Kelly et al.[23] Filament structures located by the SA method are mostly based on the tetramers

shown in Figure 12. The non-planar pentamer structures, C-5(1) and C-5(2), can be assemblied by adding a cytosine to C-4(11) and C-4(2), respectively. Similar to the tetramer case, interaction energy orderings were differed in each computational level of theory. In particular, when the interaction energies calculated for the SA structures are considered, B3-LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 favor the filament C-5(6) structure whereas the model gives the non-planar C-5(1) as the lowest energy orientation which is 24.45 kJ/mol lower in energy than C-5(6). For B3-LYP-D, MP2, SCS-MP2 and SCS-MI-MP2, C-5(6) was found to be lower in energy than C-5(1) by 34.14, 7.40, 23.22 and 18.99 kJ/mol, respectively. In all methods, interaction energies of ring structures were calculated to be less stable compared to the filament and non-planar structures. When the SA structures were further relaxed at the SCS-MP2/aVDZ level, even though no significant structural change has been observed, the corresponding energy values were lowered around 100 kJ/mol. C-5(6) is still the lowest energy structure for B3-LYP-D with an enegy difference of 1.64 kJ/mol as compared to C-5(1). On the other hand, energy ordering in MP2 and SCS-MI-MP2 was completely changed by favoring C-5(1) as the lowest energy orientation. This was followed by C-5(6) which is 16.74 and 3.34 kJ/mol higher in energy than C-5(1), respectively. In the case of SCS-MP2, C-5(6) was still the lowest energy structure, however, the energy difference with C-5(1) was reduced to only 7.32 kJ/mol. Similar to MP2 and SCS-MI-MP2, the model preferred C-5(1) against to C-5(6) with an energy difference of 30.07 kJ/mol. AMBER force field also locates non-planar C-5(1) as the most stable isomer and this was followed by C-5(6).

Hexamer

For cytosine hexamers, the SA method produced ring and many different filament and non-planar orientations. The ring structure (C-6(7)) found by the SA approach was also considered in the study of Kelly et al.[23] In addition to these structures, the structures C-6(13) and C-6(14) appeared in Kelly et al.[23] were also relaxed at the SCS-MP2/aVDZ level of theory. The resulting optimized geometries and the corresponding interaction energies were shown in Figure 14 and Table 10. Among these structures the first six ones are the 1D-filament orientations. C-6(1) includes C-4(10)

and C-2(2) motifs. C-6(2) is the structure resembling to C-1D(2) in Ref[23] and this structure can also be build from C-5(11) by adding an extra cytosine molecule. Similar to C-6(2), C-6(3) was already introduced by Kelly et al.[23] as C-1D(1) and this filament can be assemblied by adding a cytosine monomer to C-5(6). The other filament structures (C-6(4), C-6(5) and C-6(6)) include the following patterns observed in (C-4(10), C-2(2)), (C-5(10),C-2(2)) and (C-4(15),C-2(1)), respectively. In addition to the ring and filament structures, SA approach also yielded many different non-planar orientations (C-6(8) - C-6(12)) which can be constructed by adding one or two cytosine molecules to C-4(2), C-5(2), C-5(1), C-4(1) and C-5(1), respectively. The model potential and MP2 prefers non-planar structures over the filemants, when the model structures are considered. Just in contrast, filaments were found to be the lowest energy structures with B3LYP-D, SCS-MP2, SCS-MI-MP2 and AMBER. It is also worth to mention that AMBER force field fails for several structures such as C-6(7), C-6(8), C-6(11) and C-6(12). These preferences were dramatically changed upon the SCS-MP2/aVDZ relaxations. This means that the non-planar structures are the lowest energy orientations in all methods. In particular, B3LYP-D gives C-6(10) as the lowest energy structure and it is lower in energy than the lowest filament (C-6(3)) by 9.84 kJ/mol. In addition, the ring structures were obtained to be the least stable structures. MP2 gives the same picture and in this case the energy difference between C-6(10) and C-6(3) was reached to 36.86 kJ/mol. Even SCS-MP2 and SCS-MI-MP2 agrees with MP2, the energy difference between C-6(10) and C-6(3) was obtained only to be 2.60 and 14.33 kJ/mol, respectively. AMBER force field also favoured a non-planar structure as the lowest energy structure, however, it was C-6(9) instead of C-6(10) and this structure was lower in energy than C-6(3), which is the lowest enegy filament structure, by 45.32 kJ/mol. Our model potential also gave the similar results and the energy difeerence between the lowest non-planar (C-6(10)) and filament (C-6(3)) were found to be 48.70 kJ/mol. Similar to B3LYP-D, the ring structures (C-6(7), C-6(13) and C-6(14)) were also the least stable isomers in the other considered methods including the model potential.

IV. Conclusions

A new first principle potential for cytosine has been developed by fitting DFT-SAPT interaction energies which agrees very well with CCSD(T) to a Buckingham type site site model. The fitting was very successful especially for the structures with interaction energies lower than 30 kJ/mol. The performance of the model potential was tested by searching the cytosine clusters (from dimer to hexamer) with a global optimization method, simulated annealing. The model potential was able to reproduce the known cytosine cluster structures as well as new orientations. The cluster structures generated by the model were further relaxed at CP-SCS-MP2/aVDZ (for dimer and trimer) and SCS-MP2/aVDZ (for the rest) leves of theory. It has been noticed that there is a slight change in the model and the corresponding ab-initio structures. Global cluster structure optimizations yielded non-planar structures in addition to the planar ones especially starting from cytosine trimer. In general, MP2 favors non-planar structures over the planar ones similar to the model potential, while B3LYP-D and SCS-MP2 prefer planar structures. However, with the increasing cluster size, the non-planar structures were obtained to be the lowest energy orientations by all ab-initio methods and model potential. Even AMBER force field mostly agrees with ab-initio methods and the model potential, in some cases it fails to produce the correct interaction energies. The model potential performs better compared to AMBER for each cluster size. Therefore, it can also be applied in simulations of larger cytosine clusters.

Acknowledgments

This work was financially supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK-110T805). Computing resources are provided by the National Center for High Performance Computing of Turkey (UHEM), under the Grant Number 1002132012, TÜBİTAK ULAKBİM, High Performance and Grid Computing Center (TRUBA resources) and Informatics Institute of İstanbul Technical University.

References

(1) Lacroix, L.; Mergny, J.-L.; Leroy, J.-L.; Hélène, C. *Biochemistry* **1996**, *35*, 8715–8722.

(2) Borbone, N.; Amato, J.; Oliviero, G.; D'Atri, V.; Gabelica, V.; Pauw, E. D.; Piccialli, G.; Mayol, L. *Nucleic Acids Research* **2011**, *39*, 7848–7857.

(3) Pagba, C. V.; Lane, S. M.; Wachsmann-Hogiu, S. *BIOMEDICAL OPTICS EXPRESS* **2011**, *2*, 207.

(4) Guéron, M.; Leroy, J.-L. *Current Opinion in Structural Biology* **2000**, *10*, 326–331.

(5) Egli, M. *Current Opinion in Chemical Biology* **2004**, *8*, 580–591.

(6) Patel, P.; Bhavesh, N. S.; Hosur, R. *Biochemical and Biophysical Research Communications* **2000**, *270*, 967–971.

(7) Patel, P.; Bhavesh, N. S.; Hosur, R. *Biochemical and Biophysical Research Communications* **2000**, *278*, 833–838.

(8) Kondo, J.; Adachi, W.; ichi Umeda, S.; Sunami, T.; Takénaka, A. *Nucleic Acids Research* **2000**, *32*, 2541–2549.

(9) Mergny, J.-L.; Lacroix, L.; Han, X.; Leroy, J.-L.; Hélène, C. *J. Am. Chem. Soc.* **1995**, *117*, 8887–8898.

(10) Li, T.; Ackermann, D.; Hall, A. M.; Famulok, M. *J. Am. Chem. Soc.* **2012**, *134*, 3508–3516.

(11) Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčová, M. *Nuclear Acids Research* **2009**, *37*, 1713–1725.

(12) Pan, B.; Shi, K.; Sundaralingam, M. *PNAS* **2006**, *103*, 3130–3134.

(13) Vesenka, J.; Marsh, T.; Henderson, E.; Vellandi, C. *Scanning Microscopy* **1998**, *12*, 329–342.

(14) Mendez, M. A.; Szalai, V. A. *Nanoscale Research Letters* **2013**, *8*, 210.

(15) Kettani, A.; Bouaziz, S.; Gorin, A.; Zhao, H.; Jones, R. A.; Patel, D. J. *Journal of molecular biology* **1998**, *282*, 619–636.

(16) Erdmann, M.; David, R.; Fornof, A. R.; Gaub, H. E. *Nature chemistry* **2010**, *2*, 745–749.

(17) Boon, E. M.; Ceres, D. M.; Drummond, T. G.; Hill, M. G.; Barton, J. K. *Nature biotechnology* **2000**, *18*, 1096–1100.

(18) Wang, J. *Nucleic Acids Research* **2000**, *28*, 3011–3016.

(19) McKendry, R.; Zhang, J.; Arntz, Y.; Strunz, T.; Hegner, M.; Lang, H. P.; Baller, M. K.; Certa, U.; Meyer, E.; Güntherodt, H.-J.; Gerber, C. *Proceedings of the National Academy of Sciences* **2002**, *99*, 9783–9788.

(20) Otero, R.; Lukas, M.; Kelly, R. E. A.; Xu, W.; Lægsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Science* **2008**, *319*, 312–315.

(21) Furukawa, M.; Tanaka, H.; Kawai, T. *The Journal of Chemical Physics* **2001**, *115*, 3419.

(22) Frankel, D. J.; Chen, Q.; Richardson, N. V. *The Journal of chemical physics* **2006**, *124*, 204704.

(23) Kelly, R. E. A.; Lukas, M.; Kantorovich, L. N.; Otero, R.; Xu, W.; Mura, M.; Lægsgaard, E.; Stensgaard, I.; Besenbacher, F. *The Journal of chemical physics* **2008**, *129*, 184707.

(24) Otero, R.; Schöck, M.; Molina, L. M.; Lægsgaard, E.; Stensgaard, I.; Hammer, B.; Besenbacher, F. *Angewandte Chemie International Edition* **2005**, *44*, 2270–2275.

(25) Kelly, R. E. A.; Xu, W.; Lukas, M.; Otero, R.; Mura, M.; Lee, Y.-J.; Lægsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Small* **2008**, *4*, 1494–1500.

(26) Lukas, M.; Kelly, R. E. A.; Kantorovich, L. N.; Otero, R.; Xu, W.; Lægsgaard, E.; Stensgaard, I.; Besenbacher, F. *The Journal of chemical physics* **2009**, *130*, 024705.

(27) Perdigão, L. M. A.; Staniec, P. A.; Champness, N. R.; Kelly, R. E. A.; Kantorovich, L. N.; Beton, P. H. *Physical Review B* **2006**, *73*, 195423.

(28) Preuss, M.; Bechstedt, F. *Surface Science* **2008**, *602*, 1643–1649.

(29) Xu, W.; Kelly, R. E. A.; Otero, R.; Schöck, M.; Lægsgaard, E.; Stensgaard, I.; Kantorovich, L. N.; Besenbacher, F. *Small* **2007**, *3*, 2011–2014.

(30) Krull, C.; Valencia, S.; Pascual, J. I.; Theis, W. *Applied Physics A* **2009**, *95*, 297–301.

(31) Kelly, R. E. A.; Kantorovich, L. N. *Journal of Materials Chemistry* **2006**, *16*, 1894–1905.

(32) Maleki, A.; Alavi, S.; Najafi, B. *The Journal of Physical Chemistry C* **2011**, *115*, 22484–22494.

(33) JurecÌŇka, P.; Šponer, J.; Hobza, P. *The Journal of Physical Chemistry B* **2004**, *108*, 5466–5471.

(34) Czyżnikowska, Ż.; Zaleśny, R. *Biophysical Chemistry* **2009**, *139*, 137–143.

(35) Kelly, R. E. A.; Lee, Y. J.; Kantorovich, L. N. *The Journal of Physical Chemistry B* **2005**, *109*, 22045–22052, PMID: 16853862.

(36) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlík, P.; Vondrášek, J. *Journal of Computational Chemistry* **1997**, *18*, 1136–1150.

(37) Williams, N. G.; Williams, L. D.; Shaw, B. R. *J. Am. Chem. Soc.* **1989**, *111*, 7205–7209.

(38) Šponer, J.; Riley, K. E.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2595.

(39) Gu, J.; Leszczynski, J. *The Journal of Physical Chemistry A* **2000**, *104*, 7353–7358.

(40) Gu, J.; Leszczynski, J. *Chemical Physics Letters* **2002**, *351*, 403–409.

(41) Jansen, G.; Haßelmann, A. *J. Phys. Chem. A* **2001**, *105*, 11156.

(42) Haßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464.

(43) Haßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319.

(44) Haßelmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778.

(45) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.

(46) Distasio, R. A.; Head-Gordon, M. *Molecular Physics* **2007**, *105*, 1073.

(47) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.

(48) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787.

(49) Busker, M.; Svartsov, Y. N.; Häber, T.; Kleinermanns, K. *Chem. Phys. Lett.* **2009**, *467*, 255.

(50) Mishra, B. K.; Arey, J. S.; Sathyamurthy, N. *J. Phys. Chem. A* **2010**, *114*, 9606.

(51) Sütay, B.; Tekin, A.; Yurtsever, M. *Theoretical Chemistry Accounts* **2012**, *131*, 1–13.

(52) Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146.

(53) Haßelmann, A.; Jansen, G. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5010.

(54) Tekin, A.; Jansen, G. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1680–1687.

(55) Werner, H. J. et al. MOLPRO, version 2009.1, a package of ab initio program. `http://www.molpro.net`.

(56) Haßelmann, A.; Jansen, G.; Schütz, M. *J. Chem. Phys.* **2005**, *122*, 014103.

(57) Weigend, F. *Phys. Chem. Chem. Phys.* **2002**, *4*, 4285.

(58) Weigend, F.; Köhn, A.; Hättig, C. *J. Chem. Phys.* **2002**, *110*, 3175.

(59) Close, D. M. *J. Phys. Chem. A* **2004**, *108*, 10376–10379.

(60) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165.

(61) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(62) Morgado, C. A.; Jurečka, P.; Svozil, D.; Hobza, P.; Šponer, J. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3522.

(63) Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1984**, *5*, 129.

(64) Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comp. Chem.* **1990**, *11*, 431.

(65) Frisch, M. J. et al. Gaussian 09, Revision A.02, Gaussian, Inc., Wallingford CT, 2009. `http://http://www.gaussian.com`.

(66) Leforestier, C.; Tekin, A.; Jansen, G.; Herman, M. *Journal of Chemical Physics* **2011**, *135*, 234306.

(67) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1984**, *80*, 3726.

(68) Sanchez-Garcia, E.; Mardyukov, A.; Tekin, A.; Crespo-Otero, R.; Montero, L. A.; Sander, W.; Jansen, G. *Chemical Physics* **2008**, *343*, 168–185.

(69) Bak, K. L.; Jorgensen, P.; Olsen, J.; Helgaker, T.; Klopper, W. *J. Chem. Phys.* **2000**, *112*, 9229.

(70) Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27.

(71) Miliordos, E.; Aprà, E.; Xantheas, S. S. *J. Phys. Chem. A* **2014**, *118*, 7568.

(72) Corana, A.; Marchesi, M.; Martini, C.; Ridella, S. *Assoc. Comput. Mach. Trans. Math. Software* **1987**, *13*, 262–280.

(73) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

Table 1: Cytosine A ve B isomers for aug-cc-pvdz basis set with MP2, SCS-MP2, SCS-MI-MP2, B3LYP-D, DFT-SAPT(PBE0AC), DFT-SAPT(LPBE0AC) and CCSD(T) methods: Minimum energies and Center of mass distances via spline interpolation.

Dimer	Method	C.m.s. Å	kJ/mol
A	MP2	5.40	-87.43
	SCS-MP2	5.44	-80.98
	SCS-MI-MP2	5.41	-87.14
	B3LYP-D	5.30	-97.68
	PBE0AC	5.42	-79.45
	LPBE0AC	5.37	-85.27
	CCSD(T)	5.38	-85.69
B	MP2	3.38	-50.50
	SCS-MP2	3.44	-39.74
	SCS-MI-MP2	3.42	-45.75
	B3LYP-D	3.29	-49.00
	PBE0AC	3.43	-41.04
	LPBE0AC	3.42	-44.13
	CCSD(T)	3.44	-44.06

Table 2: Interaction Energies of Chosen Dimer A (C.m.s. 5.31 Å) and B (C.m.s. 3.50 Å) with basis sets aug-cc-pVXZ(X=D, T and Q). aug-cc-pVTZ and aug-cc-pVQZ were used for the extrapolation of cbs

Isomer	X	E_{int}						
		MP2	SCS-MP2	SCS-MI-MP2	B3LYP-D	DFT-SAPT (PBE0AC)	DFT-SAPT (LPBE0AC)	CCSD(T)
A	D	-86.92	-79.23	-86.54	-97.65	-78.60	-85.15	-85.86
	T	-92.38	-84.47	-90.39	-97.97	-83.21	-90.11	
	Q	-94.33	-86.52	-91.48	-98.12	-84.33	-91.29	
	cbs*	-95.69	-87.96	-92.22	-98.23	-85.09	-91.29	-94.59
B	D	-49.11	-39.30	-45.01	-46.46	-40.50	-43.41	-43.61
	T	-51.44	-41.32	-46.63	-46.77	-42.16	-45.24	
	Q	-52.39	-42.25	-47.24	-46.76	-42.89	-45.99	
	cbs*	-53.05	-42.90	-47.65	-46.75	-43.36	-46.50	-47.21

Table 3: Fitted parameters for Cytosine potential energy function. Here, parameters have been given in "bohr" and shown as "b". "H" is used for "Hartree". "i" and "j" represent the site of first and second cytosine monomer, respectively.

i-j	$\alpha[H]$	$\beta[b^{-1}]$	$C[Hb^6]$	$c[b]$
C-C	-0.434	0.781	188.630	1.00
C-H	13.126	2.062	-0.931	1.70
C-O	23.988	1.746	-15.215	1.80
C-N	91.986	2.247	37.734	1.30
O-O	1.279	1.510	-12.530	1.25
O-N	43.538	1.776	-40.943	0.75
O-H	79.879	1.707	-180.876	1.50
N-N	6.179	1.342	-135.350	1.50
H-N	33.274	1.484	-313.225	1.50
H-H	1.284	1.510	-10.336	1.50

Table 4: Comparison of model and ab-initio geometries obtained at PBE, SCS-MP2 and CP-SCS-MP2 using aVDZ basis set.

Dimer	Bond	Model	PBE	SCS-MP2	CP-SCS-MP2
C-2(1)	O-H	1.74	1.66	1.72	1.72
	O-H	1.74	1.66	1.72	1.72
	CMS	6.08	5.98	6.02	6.03
C-2(2)	O-H	1.76	1.72	1.80	1.80
	N-H	1.78	1.77	1.81	1.82
	O-H	3.01	2.94	2.91	2.92
	CMS	5.36	5.35	5.38	5.39
C-2(3)	N-H	1.88	1.84	1.91	1.92
	N-H	1.88	1.84	1.91	1.92
	CMS	5.38	5.38	5.39	5.39
C-2(4)	O-H	1.80	2.34	2.60	2.70
	N-H	2.01	2.13	2.60	2.68
	CMS	5.96	6.27	6.65	6.72
C2-(5)	O-H	1.75	1.90	1.94	
	N-H	2.75	2.64	2.49	
	CMS	6.28	6.20	6.14	
C2-(6)	O-H	1.77	1.90	2.02	2.00
	CMS	6.81	6.83	6.65	6.72

Table 5: Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2, AMBER, DFT-SAPT(PBE0AC) and DFT-SAPT(LPBE0AC) using aVDZ) of cytosine dimers shown in Figure 10.

Dimer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	PBE0AC	LPBE0AC	AMBER	Model
C-2(1)[a]	-87.94	-78.72	-70.37	-84.07	-71.53	-77.10	-71.00	-78.41
C-2(2)[a]	-91.89	-81.39	-72.96	-84.14	-74.17	-80.49	-74.28	-76.53
C-2(3)[a]	-84.33	-74.11	-66.73	-73.46	-66.94	-73.08	-69.94	-73.29
C-2(4)[a]	-28.39	-22.39	-15.26	-22.83	-18.27	-22.64	109.54	-45.85
C-2(5)[a]	-40.19	-36.78	-32.14	-37.00	-33.22	-36.95	-42.82	-44.61
C-2(6)[a]	-41.66	-37.52	-33.51	-38.59	-34.44	-37.96	-40.91	-44.43
C-2(1)[b]	-114.13	-101.78	-92.43	-108.50	-92.49	-99.58	-75.18	-81.16
C-2(2)[b]	-110.12	-98.74	-90.24	-102.28	-90.48	-97.25	-80.94	-82.40
C-2(3)[b]	-97.73	-88.44	-81.51	-88.18	-80.37	-86.37	-76.56	-78.36
C-2(4)[b]	-48.41	-45.41	-41.83	-46.17	-42.20	-45.20	-42.78	-43.04
C-2(5)[b]	-71.11	-67.47	-57.56	-64.78	-59.64	-64.73	-64.12	-68.59
C-2(6)[b]	-48.16	-45.54	-41.97	-46.31	-42.35	-45.35	-42.80	-43.06

[a] Using model geometries.
[b] Using CP-SCS-MP2/aVDZ geometries.

Table 6: Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 and AMBER using aVDZ) of cytosine trimers shown in Figure 11.

Trimer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model	Motifs involved
C-3(1)[a]	-173.80	-154.72	-137.87	-161.64	-140.76	-150.16	C-2(1),C-2(2)
C-3(2)[a]	-178.91	-158.35	-141.32	-163.15	-144.77	-149.29	C-2(2)
C-3(3)[a]	-166.67	-147.56	-131.79	-151.64	-137.19	-148.15	C-2(1),C-2(3)
C-3(4)[a]	-153.33	-137.09	-121.72	-139.02	-133.99	-147.49	C-2(3),C-2(6)
C-3(5)[a]	-144.02	-124.98	-106.76	-124.74	8.15	-147.16	C-2(3),C-2(4)
C-3(6)[a]	-171.00	-150.58	-134.70	-152.25	-140.89	-146.70	C-2(2),C-2(3)
C-3(7)[a]	-145.01	-127.13	-110.13	-137.19	33.24	-146.62	C-2(1)
C-3(8)[a]	-143.44	-136.03	-113.82	-131.66	-124.93	-129.36	C-2(2)
C-3(1)[b]	-220.88	-198.08	-180.30	-207.19	-152.62	-159.48	
C-3(2)[b]	-217.93	-195.80	-178.74	-202.25	-158.52	-161.24	
C-3(3)[b]	-208.85	-187.89	-171.73	-193.68	-148.79	-156.53	
C-3(4)[b]	-192.07	-175.91	-161.87	-179.13	-147.75	-154.91	
C-3(5)[b]	-189.66	-174.47	-160.57	-175.00	-143.83	-154.16	
C-3(6)[b]	-204.74	-184.64	-169.30	-187.47	-154.57	-157.99	
C-3(7)[b]	-192.54	-175.40	-159.65	-186.47	-121.31	-146.25	
C-3(8)[b]	-188.03	-178.08	-151.52	-174.31	-147.45	-148.84	

[a] Using model geometries.
[b] Using CP-SCS-MP2/aVDZ geometries.

Table 7: Total (ΔE_{int}^{CP}), two-body ($\Delta^2 E_{int,AB}^{CP}$, $\Delta^2 E_{int,AC}^{CP}$ and $\Delta^2 E_{int,BC}^{CP}$) and three-body ($\Delta^3 E_{int,ABC}^{CP}$) interaction energies (calculated at B3LYP-D, MP2, SCS-MP2 and SCS-MI-MP2 using aVDZ) of cytosine trimers shown in Figure 11.

Method	Interaction	C-3(1)[a]	C-3(2)[a]	C-3(3)[a]	C-3(4)[a]	C-3(5)[a]	C-3(6)[a]	C-3(7)[a]	C-3(8)[a]
B3LYP-D	ΔE_{int}^{CP}	-173.76	-178.88	-166.66	-153.33	-144.02	-170.96	-145.00	-143.44
	$\Delta^2 E_{int,AB}^{CP}$	-92.05	3.84	-84.19	-30.93	-27.84	-91.90	-87.97	-37.02
	$\Delta^2 E_{int,AC}^{CP}$	4.79	-91.98	-87.89	-31.68	-83.51	3.47	-12.38	-15.97
	$\Delta^2 E_{int,BC}^{CP}$	-87.97	-92.07	3.92	-83.63	-24.53	-84.22	-36.71	-90.76
	$\Delta^3 E_{int,ABC}^{CP}$	1.46	1.33	1.50	-7.09	-8.14	1.68	-7.94	0.31
MP2	ΔE_{int}^{CP}	-154.72	-158.35	-147.56	-137.09	-124.98	-150.58	-127.14	-136.03
	$\Delta^2 E_{int,AB}^{CP}$	-81.55	3.52	-73.93	-27.13	-21.65	-81.50	-79.09	-39.09
	$\Delta^2 E_{int,AC}^{CP}$	4.34	-81.55	-78.75	-29.60	-72.80	3.28	-6.30	-16.76
	$\Delta^2 E_{int,BC}^{CP}$	-78.84	-81.53	3.73	-73.04	-22.33	-73.96	-33.81	-80.25
	$\Delta^3 E_{int,ABC}^{CP}$	1.32	1.21	1.40	-7.33	-8.19	1.59	-7.93	0.07
SCS-MP2	ΔE_{int}^{CP}	-137.87	-141.32	-131.79	-121.72	-106.76	-134.70	-110.13	-113.83
	$\Delta^2 E_{int,AB}^{CP}$	-73.16	3.69	-66.56	-23.39	-14.29	-73.10	-70.93	-29.75
	$\Delta^2 E_{int,AC}^{CP}$	4.52	-73.12	-70.44	-25.47	-65.16	3.39	-2.32	-12.35
	$\Delta^2 E_{int,BC}^{CP}$	-70.53	-73.10	3.83	-65.58	-19.11	-66.58	-28.87	-71.56
	$\Delta^3 E_{int,ABC}^{CP}$	1.31	1.21	1.38	-7.26	-8.20	1.59	-8.01	-0.17
SCS-MI-MP2	ΔE_{int}^{CP}	-161.64	-163.15	-151.64	-139.02	-124.74	-152.25	-137.19	-131.66
	$\Delta^2 E_{int,AB}^{CP}$	-84.28	4.00	-73.28	-28.54	-22.02	-84.23	-84.32	-35.31
	$\Delta^2 E_{int,AC}^{CP}$	5.24	-84.27	-84.07	-31.13	-72.12	3.54	-9.61	-13.55
	$\Delta^2 E_{int,BC}^{CP}$	-84.15	-84.26	4.14	-72.39	-22.83	-73.30	-35.27	-82.88
	$\Delta^3 E_{int,ABC}^{CP}$	1.55	1.38	1.57	-6.96	-7.76	1.74	-7.98	0.08
		C-3(1)[b]	C-3(2)[b]	C-3(3)[b]	C-3(4)[b]	C-3(5)[b]	C-3(6)[b]	C-3(7)[b]	C-3(8)[b]
B3LYP-D	ΔE_{int}^{CP}	-220.85	-217.89	-208.81	-192.05	-189.64	-204.70	-192.53	-188.03
	$\Delta^2 E_{int,AB}^{CP}$	-112.02	4.01	-99.41	-38.36	-46.71	-111.15	-113.75	-44.21
	$\Delta^2 E_{int,AC}^{CP}$	5.27	-111.11	-115.59	-45.64	-98.28	3.86	-26.88	-34.91
	$\Delta^2 E_{int,BC}^{CP}$	-115.57	-111.59	4.54	-99.58	-34.10	-99.07	-43.13	-106.86
	$\Delta^3 E_{int,ABC}^{CP}$	1.46	0.81	1.65	-8.47	-10.56	1.65	-8.77	-2.05
MP2	ΔE_{int}^{CP}	-198.08	-195.80	-187.89	-175.91	-174.47	-184.64	-175.40	-178.08
	$\Delta^2 E_{int,AB}^{CP}$	-100.69	3.68	-90.27	-33.83	-42.71	-99.95	-102.44	-45.18
	$\Delta^2 E_{int,AC}^{CP}$	4.78	-99.94	-103.38	-43.25	-89.08	3.67	-23.57	-34.41
	$\Delta^2 E_{int,BC}^{CP}$	-103.42	-100.21	4.31	-90.13	-32.00	-89.88	-40.56	-95.99
	$\Delta^3 E_{int,ABC}^{CP}$	1.25	0.67	1.45	-8.70	-10.66	1.51	-8.84	-2.50
SCS-MP2	ΔE_{int}^{CP}	-180.30	-178.74	-171.73	-161.87	-160.57	-169.30	-159.65	-151.52
	$\Delta^2 E_{int,AB}^{CP}$	-92.28	3.85	-83.44	-30.26	-38.13	-91.56	-93.49	-32.96
	$\Delta^2 E_{int,AC}^{CP}$	4.95	-91.53	-94.13	-39.63	-82.34	3.79	-21.53	-28.39
	$\Delta^2 E_{int,BC}^{CP}$	-94.20	-91.74	4.42	-83.32	-29.39	-83.04	-35.71	-87.37
	$\Delta^3 E_{int,ABC}^{CP}$	1.22	0.69	1.42	-8.65	-10.70	1.51	-8.93	-2.79
SCS-MI-MP2	ΔE_{int}^{CP}	-207.19	-202.25	-193.68	-179.13	-175.00	-187.47	-186.47	-174.31
	$\Delta^2 E_{int,AB}^{CP}$	-104.36	4.18	-90.06	-35.28	-43.08	-103.46	-109.20	-40.51
	$\Delta^2 E_{int,AC}^{CP}$	5.75	-103.44	-110.10	-45.60	-88.90	3.95	-26.32	-32.08
	$\Delta^2 E_{int,BC}^{CP}$	-110.13	-103.86	4.78	-89.94	-32.85	-89.66	-42.13	-99.33
	$\Delta^3 E_{int,ABC}^{CP}$	1.54	0.86	1.69	-8.30	-10.17	1.71	-8.82	-2.38

[a] Using model geometries.
[b] Using CP-SCS-MP2/aVDZ geometries.

Table 8: Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 and AMBER using aVDZ) of cytosine tetramers shown in Figure 12.

Tetramer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model	Motifs involved
C-4(1)[a]	-232.94	-210.07	-178.64	-217.24	-86.74	-227.54	C-3(1)
C-4(2)[a]	-233.68	-210.14	-183.12	-214.24	-205.33	-225.86	C-3(3)
C-4(3)[a]	-247.04	-220.79	-195.97	-222.83	-208.47	-224.27	C-2(3),C-2(2)
C-4(4)[a]	-250.00	-221.88	-197.70	-230.86	-204.32	-223.93	C-2(1),C-2(3)
C-4(5)[a]	-261.70	-232.42	-206.94	-241.55	-205.32	-223.90	C-3(3),C-2(1)
C-4(6)[a]	-238.09	-225.53	-187.26	-222.64	-212.16	-223.42	C-3(1),C-2(2)
C-4(7)[a]	-253.63	-226.15	-200.89	-231.57	-209.03	-223.43	C-3(6),C-2(1)
C-4(8)[a]	-232.61	-210.35	-186.88	-215.01	-210.91	-223.19	C-3(1),C-2(2)
C-4(9)[a]	-267.34	-236.52	-210.84	-243.58	-199.85	-223.16	C-2(3)
C-4(10)[a]	-254.28	-224.88	-200.59	-231.42	-212.58	-222.48	C-3(6),C-2(2)
C-4(11)[a]	-221.74	-214.31	-180.44	-214.97	-195.65	-222.19	C-2(3)
C-4(12)[a]	-259.74	-230.87	-205.58	-239.33	-209.36	-222.16	C-2(2)
C-4(13)[a]	-247.94	-231.27	-195.75	-227.06	-216.27	-219.23	C-3(2),C-2(2)
C-4(14)[a]	-258.22	-227.55	-203.15	-231.59	-194.56	-220.92	
C-4(15)[a]	-246.28	-217.24	-194.05	-220.14	-210.79	-218.58	C-2(2)
C-4(16)[a]	-214.30	-186.97	-161.10	-202.30	138.48	-217.95	C-2(1)
C-4(17)[a]	-225.01	-202.13	-179.69	-200.39	-204.07	-211.26	C-3(3),C-2(3)
C-4(18)[a]	-214.84	-194.60	-177.31	-197.00	-168.12	-185.48	C-2(6)
C-4(1)[b]	-304.34	-282.60	-246.83	-287.42	-215.61	-235.51	
C-4(2)[b]	-298.98	-273.10	-246.53	-279.29	-222.36	-234.53	
C-4(3)[b]	-301.25	-275.53	-251.91	-279.74	-226.36	-237.50	
C-4(4)[b]	-317.49	-284.17	-258.50	-296.02	-228.57	-234.34	
C-4(5)[b]	-326.45	-292.18	-265.49	-304.33	-220.47	-238.27	
C-4(6)[b]	-321.29	-299.77	-251.86	-298.45	-229.19	-246.49	
C-4(7)[b]	-306.09	-277.91	-231.65	-287.65	-226.45	-235.22	
C-4(8)[b]	-297.97	-273.70	-251.40	-280.96	-228.82	-236.84	
C-4(9)[b]	-325.10	-291.38	-265.26	-301.03	-224.11	-240.61	
C-4(10)[b]	-314.43	-281.99	-257.03	-290.94	-232.48	-236.38	
C-4(11)[b]	-288.49	-277.91	-231.65	-274.88	-205.67	-227.00	
C-4(12)[b]	-324.58	-290.89	-264.43	-302.33	-223.74	-237.25	
C-4(13)[b]	-313.95	-290.66	-247.38	-286.52	-235.11	-245.29	
C-4(14)[b]	-310.73	-279.18	-255.00	-285.16	-205.73	-238.35	
C-4(15)[b]	-303.22	-272.88	-249.47	-277.83	-233.36	-232.27	
C-4(16)[b]	-275.66	-251.85	-228.70	-267.50	-165.20	-219.69	
C-4(17)[b]	-276.72	-255.64	-234.37	-255.60	-222.39	-221.75	
C-4(18)[b]	-255.59	-234.23	-217.44	-237.49	-195.03	-202.76	

[a] Using model geometries.
[b] Using SCS-MP2/aVDZ geometries.

Table 9: Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 and AMBER using aVDZ) of cytosine pentamers shown in Figure 13.

Pentamer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model	Motifs involved
C-5(1)[a]	-321.44	-307.18	-257.03	-304.85	-286.80	-321.08	C-4(11)
C-5(2)[a]	-310.73	-278.71	-234.40	-282.27	-159.01	-315.26	C-4(2)
C-5(3)[a]	-323.51	-291.38	-258.43	-295.92	-273.93	-298.97	C-4(3)
C-5(4)[a]	-349.78	-310.29	-276.17	-321.58	-283.22	-297.26	C-4(9),C-2(1)
C-5(5)[a]	-336.85	-298.80	-266.06	-309.58	-276.06	-296.71	C-4(10),C-2(1)
C-5(6)[a]	-355.58	-314.58	-280.25	-323.84	-287.43	-296.63	C-4(9)
C-5(7)[a]	-342.13	-302.56	-269.63	-311.21	-280.09	-295.88	C-3(2),C-2(1),C-2(3)
C-5(8)[a]	-340.93	-301.57	-268.70	-309.89	-279.70	-295.24	C-4(14),C-2(1)
C-5(9)[a]	-340.80	-301.57	-268.78	-309.75	-279.18	-294.58	C-4(10),C-2(2)
C-5(10)[a]	-345.95	-305.21	-272.21	-311.36	-283.77	-294.32	C-3(2),C-2(2),C-2(3)
C-5(11)[a]	-329.32	-291.26	-259.65	-299.01	-271.83	-294.00	C-4(4),C-2(2)
C-5(12)[a]	-333.03	-294.00	-262.25	-298.76	-274.73	-291.30	C-4(15),C-2(1)
C-5(13)[a]	-296.82	-259.33	-224.20	-274.13	54.81	-287.78	
C-5(14)[a]	-289.89	-255.05	-218.30	-265.36	107.91	-286.72	C-5(7) in Ref[23]
C-5(15)[a]	-313.50	-271.25	-232.77	-275.88	-107.23	-281.06	C-5(2) in Ref[23]
C-5(16)[a]	-306.78	-274.01	-240.01	-281.87	-86.47	-279.61	C-5(3) in Ref[23]
C-5(17)[a]	-317.67	-280.46	-243.90	-288.25	-192.42	-279.25	C-5(1) in Ref[23]
C-5(18)[a]	-305.55	-268.75	-235.81	-275.48	-228.57	-269.43	C-3(2),C-2(5)
C-5(19)[a]	-284.79	-246.07	-210.77	-251.99	-32.46	-264.47	C-5(6) in Ref[23]
C-5(1)[b]	-432.37	-410.65	-346.73	-405.21	-318.38	-349.87	
C-5(2)[b]	-395.42	-370.14	-317.93	-367.51	-306.79	-327.66	
C-5(3)[b]	-405.40	-372.00	-340.30	-379.94	-301.40	-316.66	
C-5(4)[b]							
C-5(5)[b]							
C-5(6)[b]	-434.01	-393.91	-354.05	-401.87	-312.11	-319.80	
C-5(7)[b]							
C-5(8)[b]							
C-5(9)[b]							
C-5(10)[b]							
C-5(11)[b]							
C-5(12)[b]							
C-5(13)[b]							
C-5(14)[b]							
C-5(15)[b]							
C-5(16)[b]							
C-5(17)[b]	-400.82	-366.09	-333.13	-377.56	-274.30	-297.72	
C-5(18)[b]							
C-5(19)[b]							

[a] Using model geometries.
[b] Using SCS-MP2/aVDZ geometries.

Table 10: Interaction energy calculations (at B3LYP-D, MP2, SCS-MP2, SCS-MI-MP2 and AMBER using aVDZ) of cytosine hexamers shown in Figure 14.

Pentamer	B3LYP-D	MP2	SCS-MP2	SCS-MI-MP2	AMBER	Model	Motifs involved
C-6(1)[a]	-434.59	-385.47	-343.01	-398.18	-353.08	-368.83	C-4(10),C-2(2)
C-6(2)[a]	-412.44	-365.39	-325.37	-378.02	-339.80	-369.62	C-5(11),C-1D(2) in Ref[23]
C-6(3)[a]	-444.05	-392.83	-349.86	-404.35	-358.79	-370.33	C-5(6),C-1D(1) in Ref[23]
C-6(4)[a]	-422.39	-374.63	-333.47	-386.79	-345.40	-367.88	C-4(10),C-2(2)
C-6(5)[a]	-433.84	-382.93	-341.30	-391.21	-355.06	-367.96	C-5(10),C-2(2)
C-6(6)[a]	-419.75	-370.82	-330.57	-377.37	-345.64	-364.10	C-4(15),C-2(1)
C-6(7)[a]	-419.75	-370.82	-330.57	-377.37	357.98	-303.45	C-6(5) in Ref[23]
C-6(8)[a]	-399.08	-371.16	-304.38	-358.07	-186.19	-417.35	C-4(2)
C-6(9)[a]	-416.74	-384.62	-317.21	-370.17	-301.69	-416.38	C-5(2)
C-6(10)[a]	-417.82	-397.87	-331.32	-391.07	-375.13	-414.93	C-5(1)
C-6(11)[a]	-409.55	-378.89	-309.06	-374.59	-80.10	-414.43	C-4(1)
C-6(12)[a]	-395.91	-379.03	-306.46	-367.29	-183.29	-405.40	C-5(1)
C-6(13)[a]							C-6(1) in Ref[23]
C-6(14)[a]							C-6(2) in Ref[23]
C-6(1)[b]	-539.39	-483.67	-439.83	-501.07	-384.02	-395.87	
C-6(2)[b]	-523.12	-469.08	-427.05	-486.21	-366.57	-388.49	
C-6(3)[b]	-543.51	-487.16	-443.27	-503.26	-389.05	-399.16	
C-6(4)[b]	-530.67	-476.14	-433.27	-492.92	-374.86	-391.30	
C-6(5)[b]	-525.80	-472.27	-430.65	-484.22	-388.20	-397.52	
C-6(6)[b]	-515.91	-464.02	-423.43	-474.35	-377.59	-390.46	
C-6(7)[b]	-415.18	-388.91	-361.07	-395.06	-281.37	-303.21	
C-6(8)[b]	-542.47	-502.99	-430.91	-493.44	-417.05	-430.32	
C-6(9)[b]	-552.96	-513.95	-440.55	-500.83	-434.37	-447.79	
C-6(10)[b]	-553.35	-524.02	-445.87	-517.59	-418.87	-447.86	
C-6(11)[b]	-537.38	-511.53	-431.44	-502.82	-391.92	-437.09	
C-6(12)[b]	-529.88	-505.72	-420.84	-492.38	-399.35	-441.26	
C-6(13)[b]	-513.81	-471.09	-427.07	-481.83	-340.38	-372.12	
C-6(14)[b]	-506.69	-461.61	-416.07	-477.98	-346.15	-360.48	

[a] Using model geometries.
[b] Using SCS-MP2/aVDZ geometries.

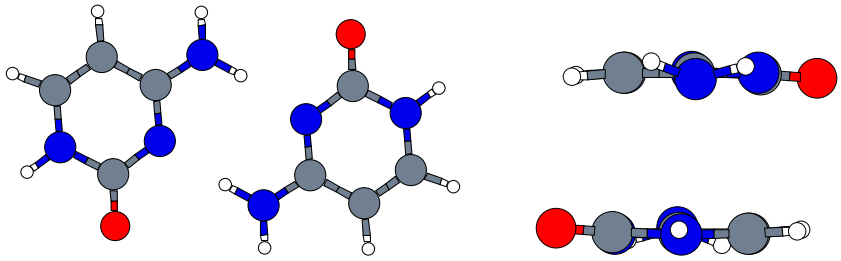

Figure 1: Cytosine Dimers used in interaction energy computations: a) H-bonded (left) and b) Stacked (right).

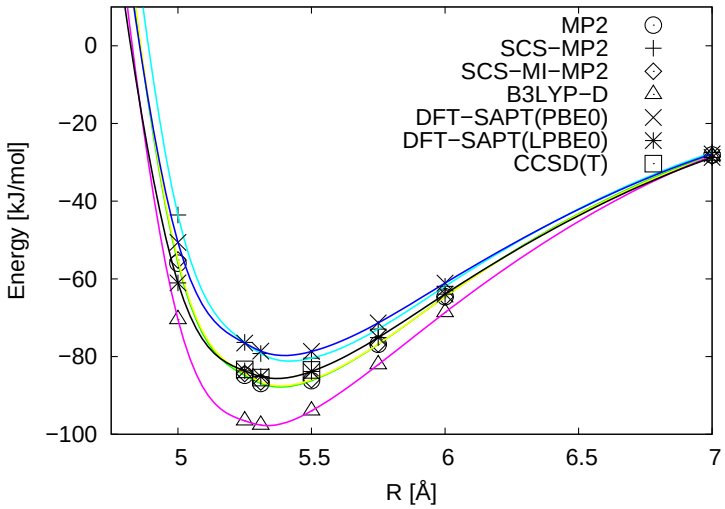

Figure 2: Potential energy curve of H-bonded cytosine dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black

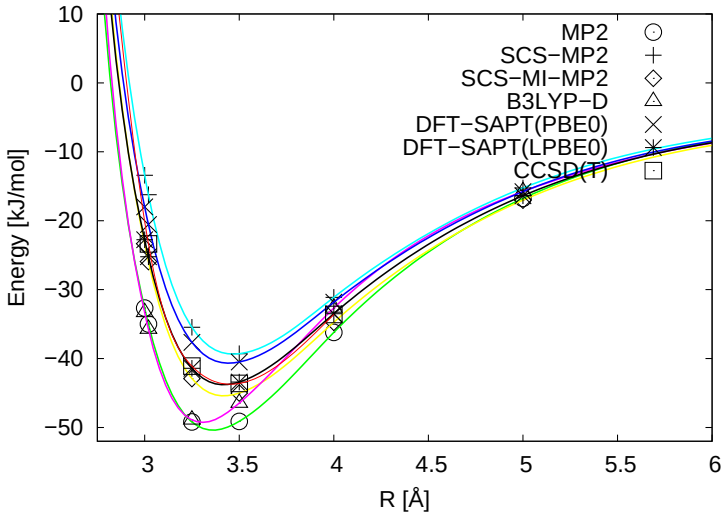

Figure 3: Potential energy curve of stacked cytosine dimer. Representative colours: MP2, green; SCS-MP2, cyan; SCS-MI-MP2, yellow; B3LYP-D, purple; DFT-SAPT(PBE0AC), blue; DFT-SAPT(LPBE0AC), black; CCSD(T), red.

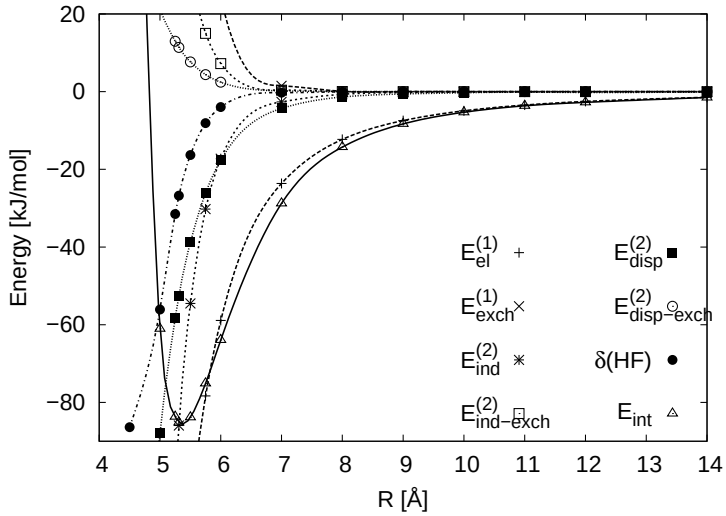

Figure 4: DFT-SAPT(LPBE0AC) energy contributions for the H-bonded orientation. Long-dashed lines represent $E_{\mathrm{el}}^{(1)}$ and $E_{\mathrm{exch}}^{(1)}$, short-dashed $E_{\mathrm{ind}}^{(2)}$ and $E_{\mathrm{exch-ind}}^{(2)}$, dotted $E_{\mathrm{disp}}^{(2)}$ and $E_{\mathrm{exch-disp}}^{(2)}$, dotted long-dashed $\delta(\mathrm{HF})$, and the solid line E_{int}.

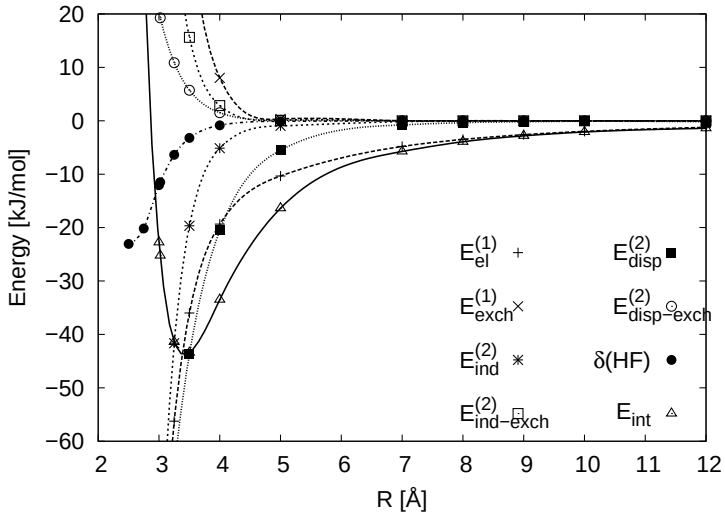

Figure 5: DFT-SAPT(LPBE0AC) energy contributions for the stacked orientation. Long-dashed lines represent $E_{\mathrm{el}}^{(1)}$ and $E_{\mathrm{exch}}^{(1)}$, short-dashed $E_{\mathrm{ind}}^{(2)}$ and $E_{\mathrm{exch-ind}}^{(2)}$, dotted $E_{\mathrm{disp}}^{(2)}$ and $E_{\mathrm{exch-disp}}^{(2)}$, dotted long-dashed $\delta(\mathrm{HF})$, and the solid line E_{int}.

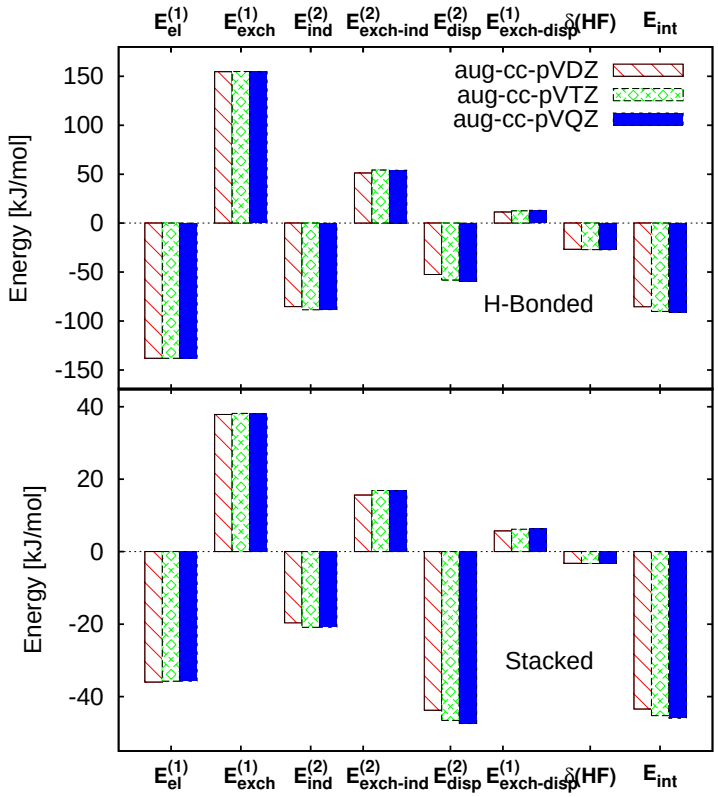

Figure 6: Basis set dependence of DFT-SAPT(LPBE0AC) energy components.

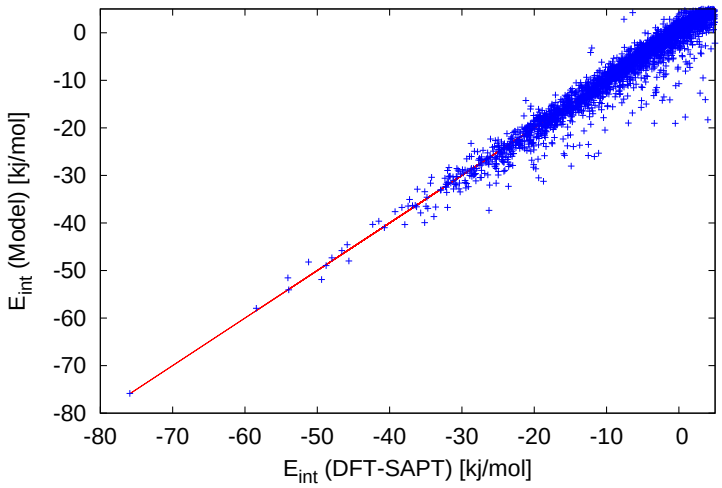

Figure 7: Comparison of the model and the DFT-SAPT(LPBE0AC)/aVDZ interaction energies.

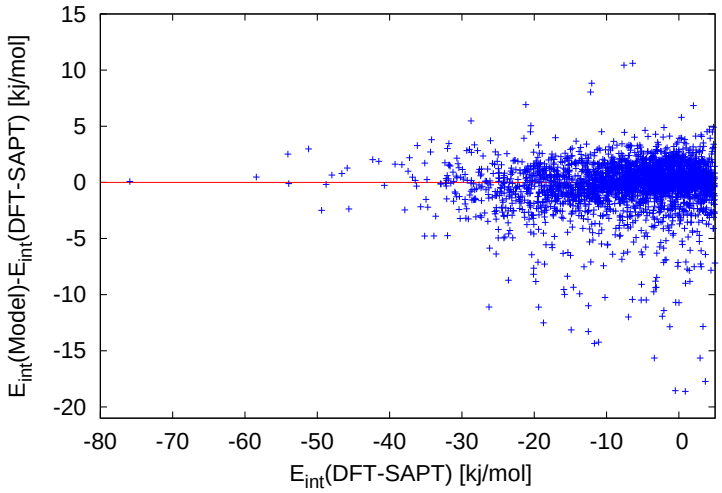

Figure 8: Deviations of the model potential from the DFT-SAPT(LPBE0AC)/aVDZ.

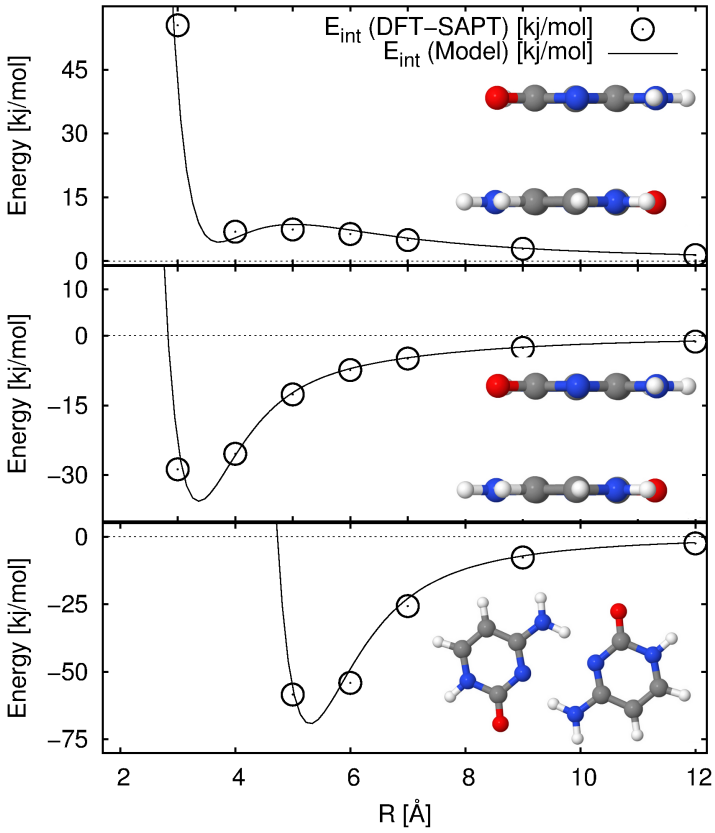

Figure 9: Comparison of the model and the DFT-SAPT(LPBE0AC)/aVDZ interaction energies for the stacked (top), anti-stacked (middle) and H-bonded (bottom).

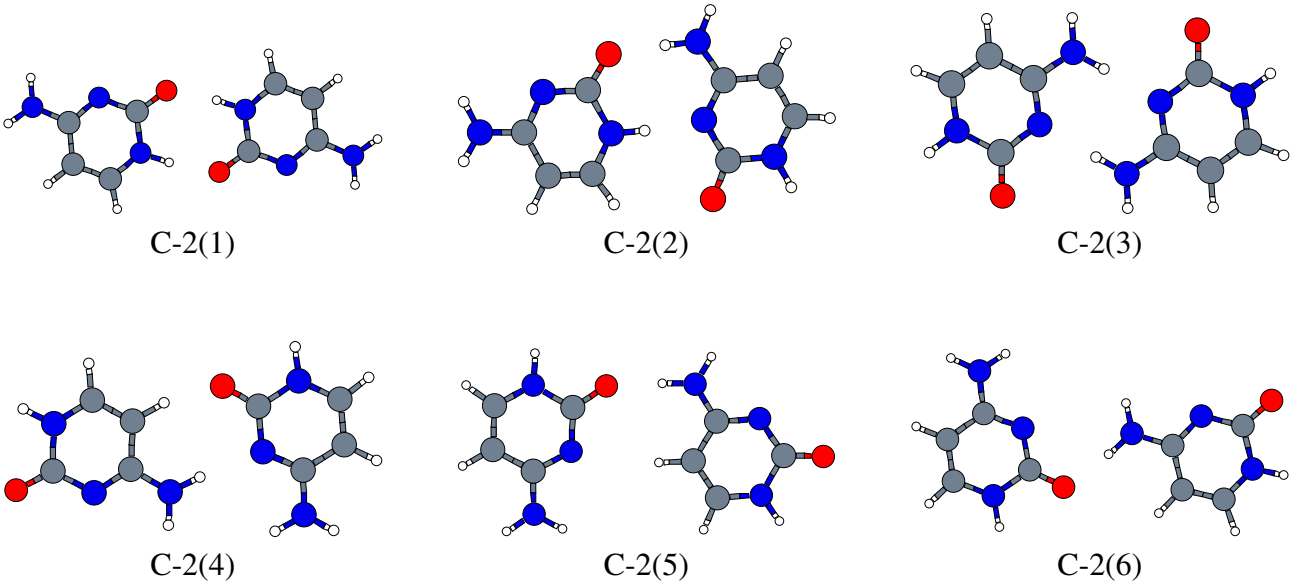

Figure 10: Six cytosine dimer structures found by the SA approach.

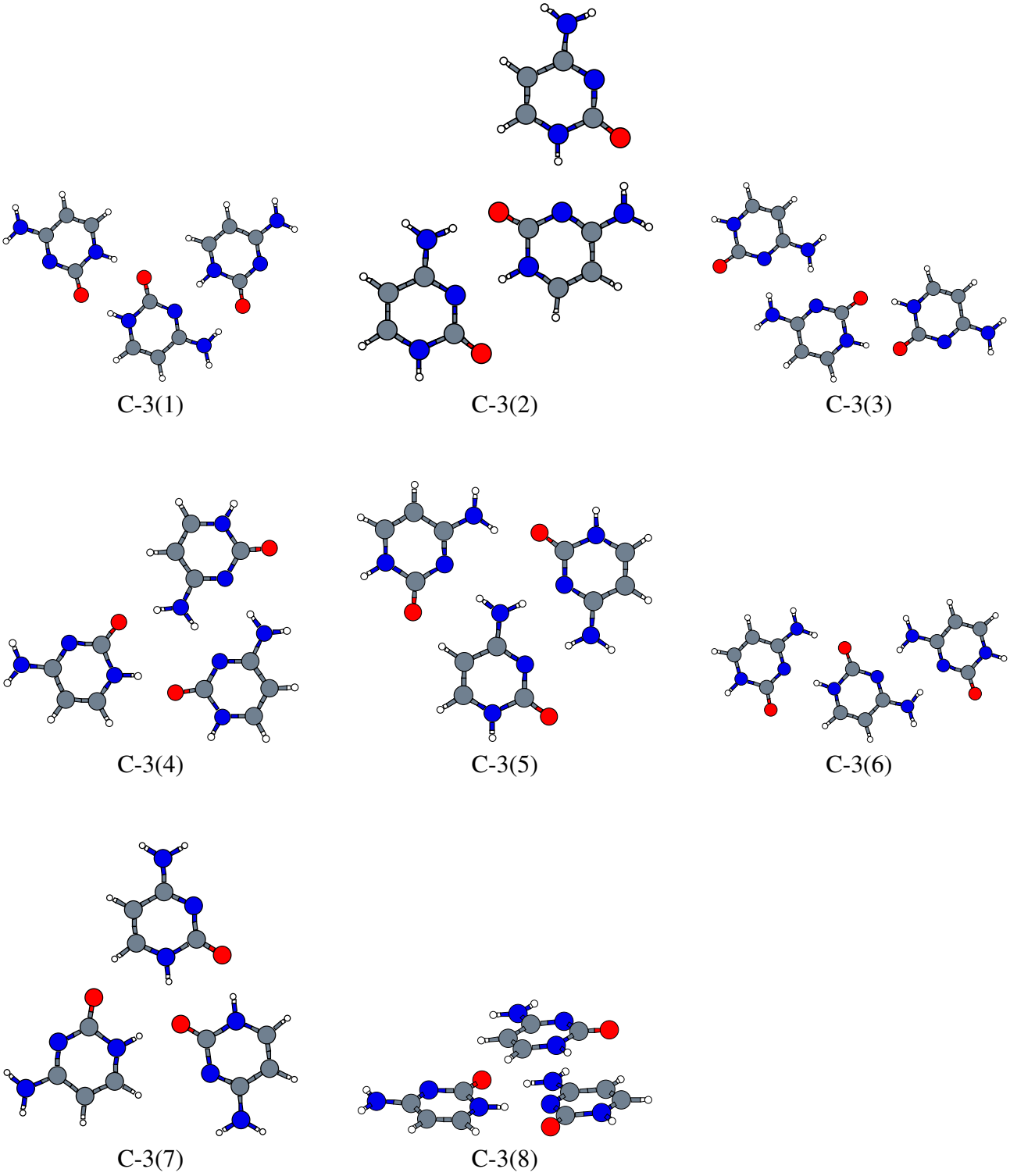

Figure 11: Eight cytosine trimer structures found by the SA approach.

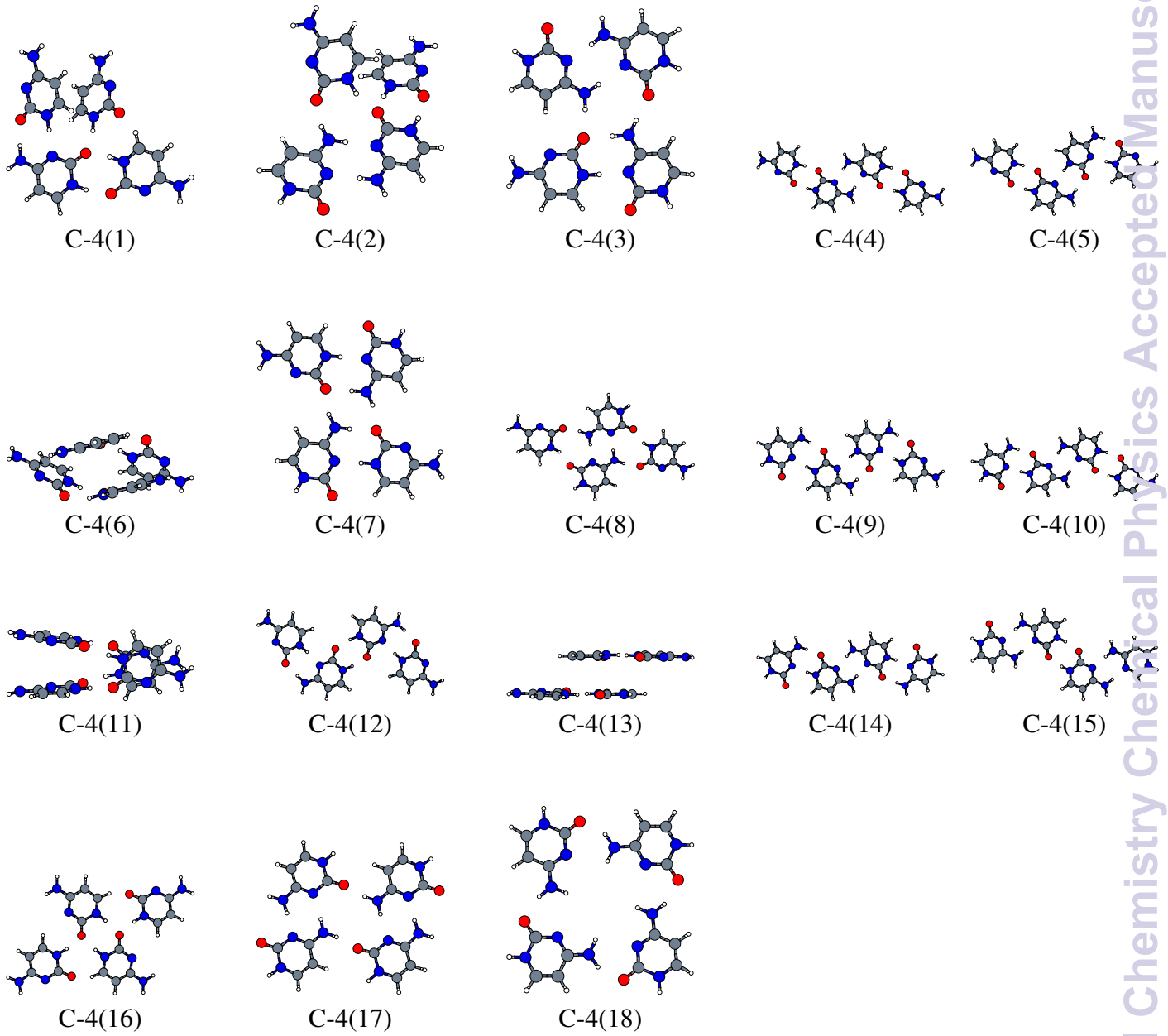

Figure 12: Cytosine tetramer structures found by the SA approach.

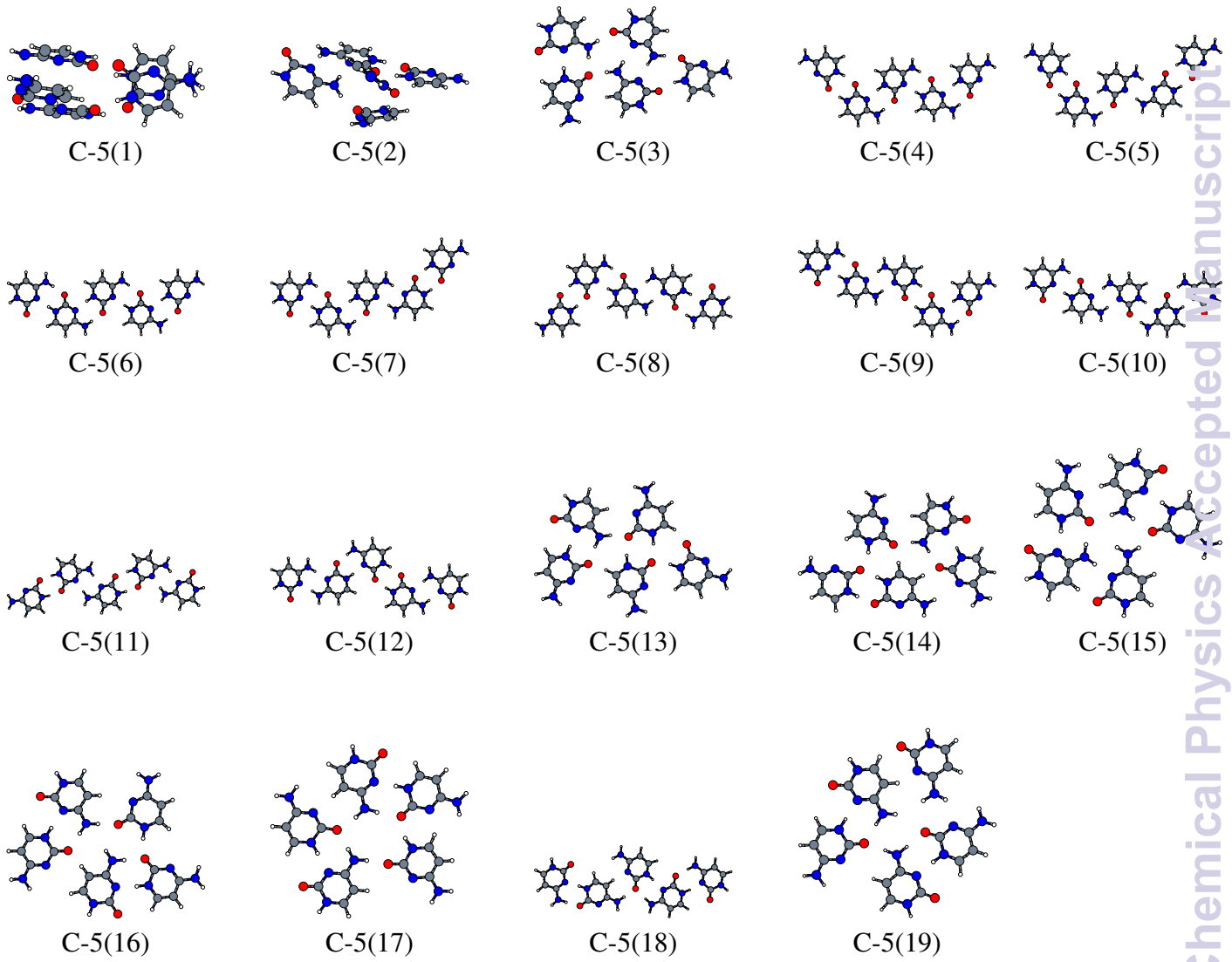

Figure 13: Cytosine pentamer structures found by the SA approach.

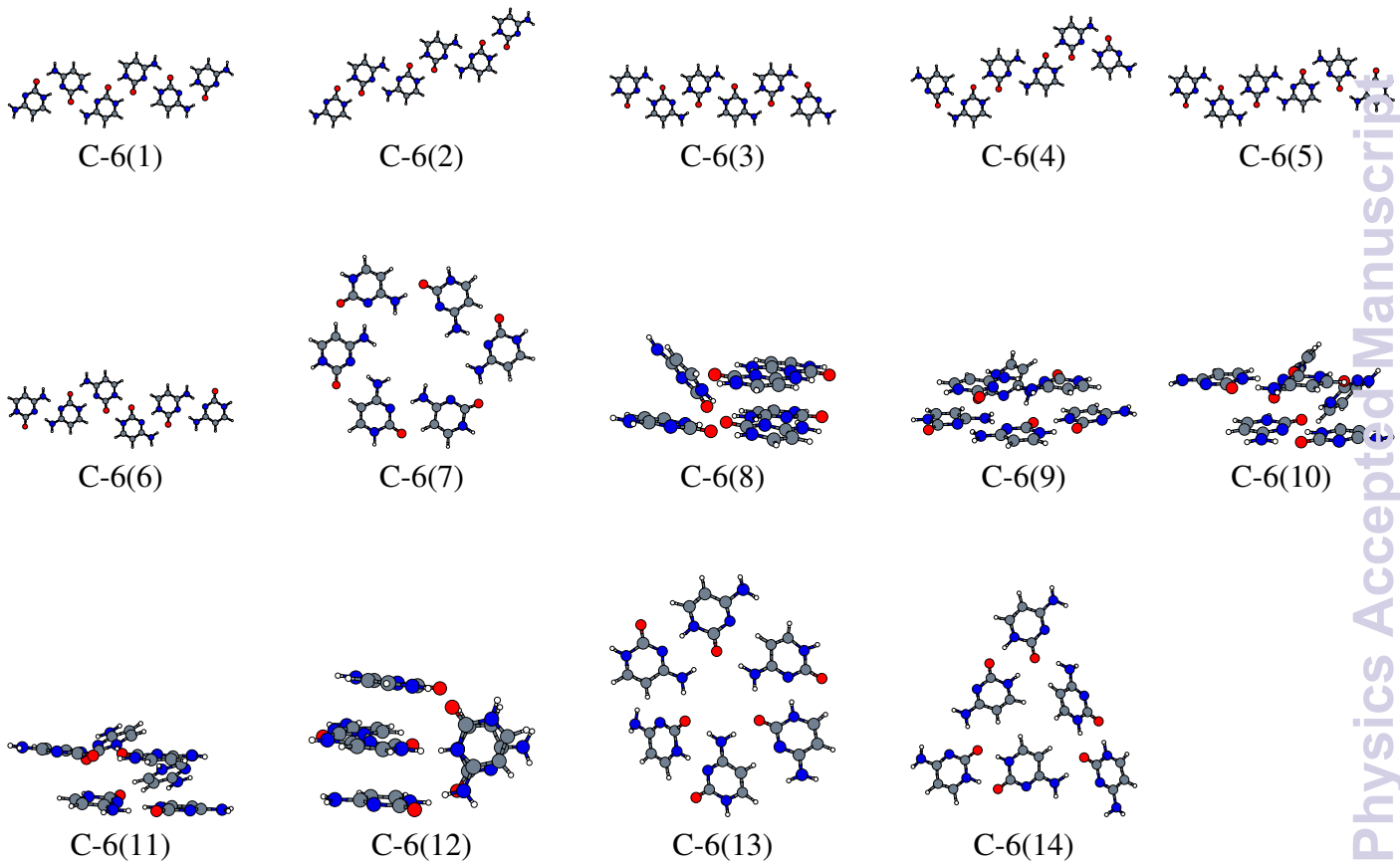

Figure 14: Cytosine hexamer structures found by the SA approach (C-6(1)-C-6(12)) and taken from Ref.[23] (C-6(13)-C-6(14)).