Competition between stacked and hydrogen bonded structures of cytosine aggregates

Jorge González [a], Imanol Usabiaga [a], Pedro F. Arnaiz [a], Iker León [a], Rodrigo Martínez [b], Judith Millán [b], José A. Fernández [a]

a. Department of Physical Chemistry, Faculty of Science and Technology, University of the Basque Country (UPV/EHU), Barrio Sarriena s/n, Leioa, Spain

b. Department of Chemistry, Faculty of Science and Technology, University of La Rioja, Madre de Dios, 53, Logroño, Spain

Abstract

The four bases of DNA constitute what is known as the ‘‘alphabet of life’’. Their combination of protondonor and acceptor groups and aromatic rings allows them to form stacking structures and at the same time establish hydrogen bonds with their counterparts, resulting in the formation of the well-known double-helix structure of DNA. Here we explore the aggregation preferences of cytosine in supersonicexpansions, using a combination of laser spectroscopic techniques and computations. The data obtained from the experiments carried out in the cold and isolated environment of the expansión allowed us to establish which are the leading interactions behind aggregation of cytosine molecules. The results obtained demonstrated that ribbon-like structures held together by hydrogen bonds are the preferred conformations in the small clusters, but once the tetramer was reached, the stacking structures became enthalpically more stable. Stacking is further favoured when cytosine is replaced by its 10-methylated version, as demonstrated by quantum-mechanical calculations performed using the same level that reproduced the experimental results obtained for cytosine aggregates. A discussion on the biological implications that such observations may have is also offered.

Introduction

DNA is formed by a combination of sugar phosphates attached to four bases, or nucleotides: cytosine, guanine, adenine and thymine (CGAT). Usually represented by their initial letter, they constitute what has been called ‘‘the alphabet of life’’. Why these four molecules were chosen to build the DNA instead of others that are already present in the cell is still a matter of debate.1,2 One of the most commonly accepted ideas is that the atmosphere of the primitive earth was not able to filter the UV as efficiently as it does nowadays and therefore those molecules that were not stable under strong UV irradiation suffered from photodissociation or photoinduced degradation.3–7 In agreement with this hypothesis, the four bases present a remarkably short electronic excited state lifetime, and it even decreases with the attachment of a sugar unit (or other substituents).7–11 However, there may be additional reasons behind the selection of CGAT to construct one of the most important molecules of life. To build DNA, the bases must be able to form stable pairs in biological environments and stack them. For the former, the DNA bases present a number of chemical groups, {NH, NH2, {CQO, able to aggregate the bases in pairs by formation of strong hydrogen bonds, whereas the aromaticity of the bases results in p–p interactions of moderate strength that combined with their hydrophobicity result in a remarkably stable structure, reinforced by a sugar phosphate skeleton. However, the canonical dimeric C-G/A-T structures are not the only ones found in DNA. Conversely, formation of trimers and quadruplexes is not unusual. For example, telomeres are a kind of ‘‘end-of-the-road’’ signal found at the end of the chromosomes, formed by repetition of a TTAGGG sequence in vertebrates.12,13 At those places, the DNA strand is bent in such a way that allows the guanine units to form quadruplexes, usually stabilized by the presence of a cation.12,14 Cytosine has also been found forming quadruplexes in the so-called i-motifs.15 Likewise, thymine and adenine are also able to form other non-canonical structures.16,17 The existence of such structures seems to be possible for a given sequence of bases and it is not clear if it is the sugar skeleton that forces the bases to form them or it is the strong propensity of the bases to establish hydrogen bonds that are able to bend the strand and to form quadruplexes only when certain sequences occur.2 These observations raise the question of which are the real aggregation preferences of the DNA bases and what is the role that such preferences played in the appearance of the first DNA strands. To answer them, we used a combination of laser ablation and supersonic expansions to create the required conditions to form cytosine aggregates. The collisions

in the first instants of the expansion cool the molecules to B100 K and stabilize the aggregates, which will afterwards travel in the expansion without any external perturbation, allowing us to obtain data on their structure, using a combination of mass-resolved spectroscopic techniques, mainly REMPI (resonance-enhanced multiphoton ionization) and IR/UV double resonance. The IR spectra recorded with the latter carry important structural information, but require accurate calculations for their interpretation, as will be shown below.

This combined experimental/computational approach is not new and has been used before to tackle the spectroscopy of numerous systems, including DNA bases.7,18–24 However, it was so far limited to the exploration of dimers. Thanks to the finely tuned experimental conditions we extend here such studies up to the tetramer. Most of the information available on the spectroscopy and aggregation preferences of DNA bases has been obtained by the de Vries group, who explored the spectroscopy of the isolated bases and the formation of dimers in supersonic expansions.20,25 The spectroscopy of the derivatives of the bases has also been explored24 in an attempt to understand how different substituents affect the structure of the dimers. Several pure computational studies have also been published, mainly aiming at understanding the formation of dimers.26–30 Recently, an exhaustive exploration of the aggregation preferences of keto-cytosine up to the hexamer appeared, using a force field fitted to reproduce the ab initio computed structure of the dimer.31 The extension of the experimental studies to larger clusters has been hampered by the short electronic excited state lifetime of cytosine, and by the dynamics of its excited states that significantly reduces the signal rendered by the aggregates in experiments of excitation spectroscopy. Thanks to the finely tuned experimental conditions, we were able to tackle the spectroscopy of cytosine aggregates up to the tetramer. Interpretation of those experimental results also required a deep exploration of the conformational landscape and a large number of calculations using density functional theory (DFT) and bulky basis sets. To complement the results on cytosine, a computational experiment was carried out on 1-methylcytosine aggregates to block the nitrogen where deoxyribose attaches to cytosine. The results obtained and their interpretation may help understand the origin of the nature of the DNA bases, although the species generated are in principle only stable under jet conditions.

Methods

Experimental

The set up used in this work has been previously described in detail,32 and therefore only a brief description will be offered here. Cytosine (Z99% purity, Sigma-Aldrich) was mixed with carbon nanotubes (Multi-Walled Carbon Nanotubes, purity490%, 10–30 mm diameter, Sun Nanotech Co. Ltd) and deposited on the surface of a cylindrical sample holder (4.5 4.5 10 mm3). Then, the sample was introduced in the vacuum chamber of a linear time-of-flight (TOF) mass spectrometer. Desorption of cytosine was accomplished using a tightly focused Nd/YAG laser (Quantel Ultra, 20 mJ per pulse at 1064 nm, usually operated at B1 mJ per pulse, 8 ns pulse duration). Desorption was synchronized with the aperture of a pulsed valve that created a supersonic expansion of Ar (10 bar typical pressure) that picked the ablated material and cooled it, resulting in the formation of the desired molecular aggregates. The expansion carrying the aggregates entered the ionization region of the TOF through a 2 mm skimmer that selected the colder portion of the beam and was interrogated using a combination of UV and IR pulsed lasers (ScanMate pumped by Brilliant B 2W, Fine Adjustment pumped by Brilliant B 3W, Quantel TDL90 pumped by Quantel YG980 2W and LaserVision OPO/OPA pumped by Continuum Surelite) that excited and ionized the molecules. The ions created using 1-color REMPI were sent to the MCP (micro-channel plate) detector of the TOF using an electric field (voltages in plates: 4000, 3700 and 0 V) and the electric currentgenerated at the MCP was recorded using a digital oscilloscope.

Computational

The computational procedure was already tested in systems of similar complexity,32–34 and it consists of three stages. In the first stage, the conformational landscape for the interaction of the molecules was explored using molecular mechanics (MMFFs35,36) and Schro¨dinger's suite. This stage was required due to the complexity of the systems, which can interact in multiple different orientations, and present several tautomers. All possible combinations of keto and enol tautomers wereaken into account. The exploration usually resulted in a large number of conformers (typically thousands) that were grouped into families attending to their similarity. Then, in the second stage, representative members of each family in a reasonable stability window (30 kJ mol1) were subjected to full

optimization at the M06-2X/6-311++G(d,p) calculation level, using the Gaussian package.37 The resulting structures were tested as true minima using a normal mode analysis that was also used to apply the zeropoint energy (ZPE) correction to the energy of the system. The nomenclature used to name the structures is as follows: k or e was used to denote keto-/enol-cytosine tautomers in each species, followed by a number that denotes its relative stability, according to the (ZPE-corrected) DH value. Thus, kke1 is the global mínimum of a cytosine trimer containing two keto cytosines and one enol cytosine; kkkk5 is the fifth most stable tetramer and contains keto-cytosines exclusively. Finally, the DG values for all the structures calculated in stage 2 were computed using the procedure described in ref. 32, starting from 0 K and up to 700 K, the temperature at which organic matter decomposes. The DG values presented in this work were also corrected for the BSSE (basis set superposition error) using the counterpoise method.38 All the structures calculated in this work, together with representation of the variation of their DG with the temperature, can be found in the ESI.

Calculation of DG assumes that the system is in thermodynamicequilibrium to allow one to compare with biological systems. However, the expansion is not at equilibrium. A similar calculation can be found in the ESI for the systems assuming a rotational temperature of 4 K, which is approximately the rotational temperatura of the molecules in the beam. Such a calculation demonstrates that the vibrational entropic term is the main source of differences between the entropy of the different isomers for a given stoichiometry.

Results

REMPI spectroscopy

Cytosine can exist in multiple tautomeric forms, each of them presenting UV absorptions in different parts of the spectrum. Under the conditions of our expansion, the two most abundant species are the keto and enolic forms (Scheme 1), although the latter presents also two different isomers, based on the orientation of the hydroxylic hydrogen with respect to the rest of the molecule. Such a small structural difference does not introducenoticeable differences in the spectroscopy of the molecule, though, and

therefore it is very likely that both spectra overlap at the spectroscopic resolution of our experiments.

Conversely, the tautomeric conformation strongly modula es the absorption spectrum of cytosine: the keto form000 transition23,24 lies around 31 826 cm1 and presents a well-resolved spectrum, while the absorption spectrum of the enol tautomer is very noisy, complicating the identification of the origin band, but it lies around 36 225 cm1. This large energy difference in the S1 ' S0 transition also results in very different D+0 'S1 values.5,6 Previous studies reported the spectrum of both isomers, using 1+10 REMPI and an excimer laser emitting at 193 nm to obtain energeticenough radiation to drive the molecule to ionization.23,24 A similar laser is not available in our laboratory and therefore we were not able to record the spectrum of the keto tautomer of cytosine. The lower D+0 'S1 transition of the enol conformer, on the other hand, allowed us to record the noisy, congested trace in Fig. 1 for the cytosine monomer. Formation of the dimer usually results in a red shift both on the S1'S0 and D+0 'S1 transitions, allowing us to explore the spectroscopy of the complex built on the keto tautomer. Thus, the spectrum of the cytosine dimer collected in Fig. 1 was recorded in the vicinity of the 000 transition of the cytosine's keto tautomer. The first band in the spectrum appears at 33 499 cm1, which is in good agreementwith previous publications (33483 cm1, ref. 23). That cytosine dimer's spectrum appears in this región also means that at least one of the twomolecules of cytosine is in the keto conformation. Recording of the corresponding spectrum in the vicinity of enol-cytosine was not possible due to either a loss of signal intensity or to the absence of enol-cytosine as part of the dimer. Certainly, as will be shown below, formation of the dimer may result in a direct isomerization to the keto tautomer.The electronic spectra of the trimer and tetramer were recorded also in the 33 300–34 200 cm1 region, pointing once more to the presence of at least one keto-cytosine in the complex. The traces (Fig. 1) are broad absorptions, probably due to excited state dynamics or to the large number of low-frequency vibrations of theaggregates, or to a combination of both. In any case, both traces(cytosine trimer and tetramer) present a small red shift comparedto the cytosine dimer. It is worth noting that this is the first timethat the gas phase spectra of these species are reported. In the following we will present the structural information obtained from the IR/UV double resonance spectroscopy and we will interpret it in light of the MM/DFT calculations.

IR/UV spectra

Fig. 2 shows the mass-resolved IR/UV spectra of the four aggregates tudied in this work. The simulated spectra of the species to which they have been assigned are also shown for comparison. Scheme 1 Keto and enol tautomers of cytosine, together with their atom numbering.

Previous studies reported the UV absorption spectrum of both keto and enol forms of cytosine, together with the UV–UV hole burning and the mass-resolved IR/UV spectrum of ketocytosine. 23–25 Dong and Miller39 reported the IR spectrum of adenine and cytosine recorded in helium droplets. Using a large dc field to orient the molecules, the authors were able to isolate the vibrations from the single keto and the two enol tautomers of cytosine. However, to the best of our knowledge, no previous report on the mass-resolved IR/UV spectroscopy of enol-cytosine existed.

Also, previous studies on the cytosine dimer24 reported its IR/UV spectrum in the 3400–3700 cm1 region, where four bands were found. We extended here the scanned región towards the red, where a broad absorption was found. To the best of our knowledge there are no previous reports on the IR/UV spectra of the cytosine trimer and tetramer. The spectra in Fig. 2 show a clear progression from the bare molecule to the tetramer. The spectrum of the cytosine monomer presents three bands, corresponding to the symmetricandantisymmetric stretches of the amino group (at 3462 and 3575 cm1, respectively, see Fig. 3) and to the stretching of the hydroxyl group (at 3613 cm1). In good agreement, the computed spectra present three bands, although with small shifts, mainly due to the anharmonic nature of the real vibrations.

Formation of aggregates produces two effects in the spectrum: a broad band appears in the 2700–3300 cm1 region of the spectrum, and a change in the number of bands due to free NH stretches. Both changes are well reproduced in the spectra predicted for the most stable structures computed in this work. The assignments presented in Fig. 2 are based on the direct comparison between simulated and experimental spectra and on the relative stability of the computed structures (Fig. 4). The complete set of computed structures and their relative stability can be found in the ESI.

The increase in the number of bands due to free NH stretches is a direct consequence of the increase in the number of molecules in the aggregate. However, the broad absorption is a somehow unexpected effect. Formation of hydrogen bonds usually induces a shift to the red in the stretch of the protondonor group, and at the same time the vibrational band's width increases due to anharmonicity. However, the shifts observed in Fig. 2B–D are very large and the bands are very broad, pointing to extraordinarily strong hydrogen bonds. An additional broad absorption appears in the spectrum of the tetramer, centred at 3340 cm1 (shaded in yellow in Fig. 2D). Assignment of such a band requires the introduction of additional conformers with a completely different geometry, as will be explained below.

Structure of the aggregates

Comparison between the experimental and calculated IRspectra of the cytosine monomer in Fig. 2A demonstratesthe sensitivity of the technique. While the orientation of thehydroxyl group introduces a subtle difference in the positionof the N–H stretches, the spectrum predicted for the ketoconformer is significantly different and does not match withthe experimental trace, recorded probing the enol-cytosineS1 ' S0 transition.

From an energetic point of view, the two isomers of enolcytosineand the keto-cytosine tautomer lie in a narrow stabilitywindow of less than 5 kJ mol1 and therefore we expected tofind all three tautomers in the expansion. Thus, the spectrumin Fig. 2A very likely contains the contribution from bothisomers of the enol-cytosine tautomer.The DDG data in Fig. 4A also show that, as temperatureincreases, entropy favours the keto tautomer, becoming isoenergeticwith one of the enol isomers around 300 K, while theimino tautomers are too high in energy to present a noticeableconcentration at the temperature of the expansion (100–200 K).Assignment of the experimental spectrum of the cytosinedimer points to the existence of two isomers, or put in anotherway, the molecules can interact in two different orientations,leading to structures of almost equal stability, according to thedata in Fig. 4B. Both isomers are formed by two keto-cytosines.According to the calculations, those dimers containing enoltautomers are too high in energy to be detected in the beam.Furthermore, the interaction between the

molecules facilitatesthe keto–enol tautomerism, explaining the absence of enol-baseddimers, despite the fact that the enol monomer was detected inthe expansion.

The broad absorption observed in the spectrum is, accordingto the assignment, due to the N–H bonds taking part in the twointermolecular hydrogen bonds and points to a delocalizationof the protons between the two molecules. Previous studies onprotonated cytosine demonstrated the delocalization of a protonbetween the two molecules.40 The results presented here seem toindicate that the presence of an additional proton may not berequired for the two interacting cytosines to share their protons,even in the ground electronic state. This would be in line withprevious theoretical studies on the ability of DNA bases to sharetheir protons, jumping between tautomers in a fs time scale.If such is the case, this dynamics could be in part responsible forthe broad absorption observed in the spectrum.30,41,42Assignment of the experimental spectrum of the trimershows that it contains the most stable structure of the dimeras a core and the new cytosine molecule binds in the twoorientations already observed in the dimer.Addition of a fourth cytosine molecule produces two noticeableeffects in the mass resolved IR spectrum (Fig. 2D): areuction in the number of resolved bands, accompanied bythe appearance of a second broad absorption (shaded in yellow)between the broad band already present in the trimer and thediscrete transitions. Also, all the bands in the 3400–3600 cm1region present shoulders that seem to indicate the presenceof several isomers. The computational analysis offered someinteresting results. Stacking structures replaced the ribbon-likestructures as the most stable ones when (ZPE-corrected) DHalone was taken into account. However, when (BSSE-corrected)DG was used, the energy difference between stacked and linearstructures disappeared and the latter became significantlymore stable (Fig. 4D) even at very low temperatures. Thus, inthe temperature interval of our molecular beam (slightly above100 K), linear structures are expected to be the most abundantones, although one cannot rule out the presence of stackedstructures, from a stability point of view.

Comparison of the experimental spectrum in Fig. 2D withthose predicted for the most stable linear structures showsthat the whole spectrum can be explained by isomer kkkk6(the entropically most stable one at 100 K), apart from thebroad absorption centred at B3331 cm1. No planar structurein a reasonable range of stability can explain such absorptionand only stacked structures are predicted to present bands onthat region. Thus,

the IR spectrum in Fig. 2D may be pointingto the coexistence on the beam of planar and stacked structuresof the cytosine tetramer.

Discussion

The experimental results demonstrated that the main interactionmechanism of isolated cytosine aggregates proceeds throughformation of hydrogen bonds, resulting in ribbon-like structuressimilar to those found when cytosine is deposited over goldsurfaces.31 Such structures are also similar to those reported incrystals (see the ESI), although the intermolecular distanceschange probably due to the interaction with the neighbouringribbons or to the packing in the crystal.

Also, only keto-based structures were found in the aggregates,despite the fact that the isolated enol tautomer is more stable.The loss of stability due to tautomerization into the keto form islargely compensated by the increase in interaction energy betweenthe two molecules. This is clearly demonstrated both by thecomputational results and by the detection of aggregates containingexclusively keto tautomers.

There are two dominant orientations for the interactionof cytosine molecules: N1–HN30//HNH0OQC2 or the symmetricN1–HOQC20//N1–H0OQC2. Both are isoenergeticin the dimer, but it seems that the former is favoured ‘‘in thelong run’’, as aggregates larger than the dimer prefer the formerorientation.

The hydrogen bonds between cytosine monomers are extraordinarily strong, as reflected by their appearance in the spectrum as very broad absorptions, shifted to the red, to a region usually occupied by the CH stretches. Such stronginteractions may compensate the stress in the DNA strandsintroduced by the torsions required to put in close contact thefour molecules of the quadruplex. However, further experimentswill be required to understand why those structures are formedonly in DNA sections with a certain sequence of nucleotides.As the aggregate grows, alternative interaction structuresappear, such as formation of cycles (for exampl, kkk14) orstacking structures. The latter type is dominant in the tetramer,from an enthalpic point of view. However, when the BSSE correctionis taken into account and the temperature is introduced inthe equation, ribbon-like structures

become more stable. It maywell be that the special conditions of the expansion favourdetection of stacking structures although they are higher inenergy. Certainly, stacked dimers may be formed by collisionof dimers if the energy released by the stacking interaction isnot large enough to dissociate the pre-existing structures orto surmount the barriers connecting them to the ribbon-likeisomers. Tetramers may also be formed by collision of a trimerwith a monomer. Although the most abundant species is themonomer, dimeric aggregates are significantly more abundantthan trimers. More difficult to evaluate is the collision crosssection for the aggregation of a monomer + trimer compared todimer + dimer. In any case, such mechanisms cannot be ruledout as responsible for the detection of the stacking structures,despite the fact that they are approx. 5 kJ mol1 higher in energyat 100 K.

Extrapolation to a biological environment

A deoxyribofuranose is attached to N1 of cytosine in DNA, blocking one of the preferred sites of the molecule to form intermolecular hydrogen bonds. This situation can be simulated by adding a methyl group to cytosine in position 1, to make 1-methylcytosine (MeCyt). Computations of the aggregation preferences of MeCyt, Fig. 5, demonstrate that the substitution significantly increases the propensity of the base towards the formation of stacked structures already in the trimer and they become both enthalpically and entropically more stable for thetetramer.

The preference for the stacking conformation will be further favoured in the nucleoside due to the ribose in the N1 position. Previous studies using NMR on the interaction between polysaccharides43 and on supersonic expansions using similar techniques to those employed in this work44 show that sugars tend to form stacking structures, especially in solution, because in that way the hydrophobic surface in contact with water is reduced. In good agreement, in an exploration of the aggregation preferences of adenosine dimers in jets, Asami et al. Found almost exclusive formation of stacked dimers.45 Unfortunately, the authors did not explore larger aggregates. Probably the small S/N ratio rendered by such a difficult system precluded the recording of the spectroscopy of higher-order clusters. We have also explored the aggregation of sugar units, observing the same trend towards formation of stacked aggregates in dimers.32,44

It is also difficult to extrapolate such observations to solution. On the one hand, in solution and in the temperature range of life, entropy certainly plays an important role, and therefore one would expect to find ribbon-like structures only. However, water may overcompensate the entropic effect. Previous computational studies on the effect of water molecules on the formation of cytosine–guanine46 and adenine–thymine dimers47 clearly demonstrated that two water molecules are enough to favor stacking structures over hydrogen bonded ones. In those studies they also observed a preference for the interaction between keto tautomers, in agreement with our own results. Thus, the studies seem to point to the existence of stacking dimers in solution. All these pieces of evidence point to a reinforcement of the stacking interactions in cytidine compared with cytosine that combined with the effect of solvation may well tip the balance towards the formation of stacked structures of dimers. From such structures it would be relatively easy to evolve to the well-known DNA structure, just by adding a link between sugar units, as we know it nowadays. Thus, the appearance of the first DNA (or RNA) structure based on CGAT (CGAU) may in part be a consequence of the intrinsic aggregation preferences of the nucleobases. We are currently running experiments on the rest of the nucleobases to further explore this hypothesis.

Conclusions

We present here the first study on the structure of cytosine aggregates up to the tetramer in supersonic expansion. Species formed by keto tautomers were preferentially detected. Planar aggregates are initially preferred, but as the size of the aggregates increases, formation of stacked dimers becomes enthalpically favoured, although entropy still favours planar structures. When a methyl group is added in position 1, the tendency towards the formation of stacking structures increases. These structures are stable only under jet conditions. Thus, extrapolation to nucleosides in solution needs to balance the effect of entropy, which favours planar structures with the effect of the solvent, which usually favours stacking interactions, and of the sugar substituent, which also favours staking. Thus, our results point to a natural tendency of cytosine (and probably reinforced in cytidine) to form stacking dimers in solution. Such a tendency may have facilitated the formation of the first DNA strands in the primitive Earth.

Acknowledgements

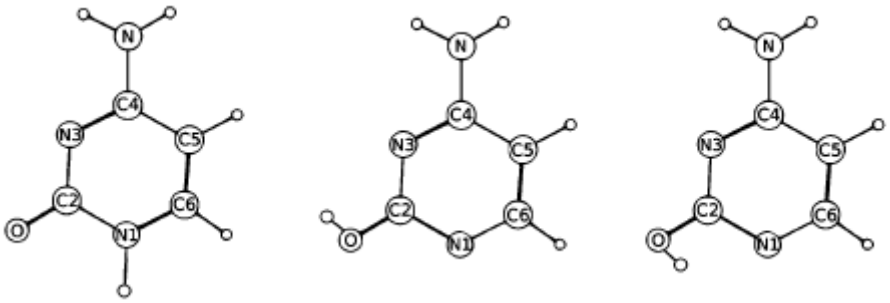
The research leading to these results has received funding from the Spanish MINECO (CTQ-2015-68148) FEDER EU. The experimental measurements were carried out at the SGIker Laser Service of the UPV/EHU. Computational resources from the SGI/IZO-SGIker network were used for this work. J. G. thanks the University of the Basque Country for a predoctoral fellowship. I. U. thanks the Basque Government for a predoctoral fellowship. I. L. would like to thank the MINECO for a Juan de la Cierva postdoctoral fellowship.

References

1. E. T. Kool, J. C. Morales and K. M. Guckian, Angew. Chem., Int. Ed., 2000, 39, 990–1009, DOI: 10.1002/(SICI)1521-3773(20000317)39:6o990::AID-ANIE99043.0.CO;2-0.
2. S. A. Benner, ACS Cent. Sci., 2016, 2, 882–884.
3. S. Lobsiger, M. Etinski, S. Blaser, H. Frey, C. Marian and S. Leutwyler, J. Chem. Phys., 2015, 143, 234301, DOI: 10.1063/ 1.4937375.
4. C. G. Triandafillou and S. Matsika, J. Phys. Chem. A, 2013, 117, 12165–12174, DOI: 10.1021/jp407758w.
5. D. Mishra and S. Pal, J. Mol. Struct.: THEOCHEM, 2009, 902, 96–102, DOI: 10.1016/j.theochem.2009.02.018.
6. O. Kostko, K. Bravaya, A. Krylov and M. Ahmed, Phys. Chem. Chem. Phys., 2010, 12, 2860–2872, DOI: 10.1039/B921498D.
7. A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. de Vries, Proc. Natl. Acad. Sci. U. S. A., 2005, 102, 20–23, DOI: 10.1073/pnas.0408574102.
8. K. Kleinermanns, D. Nachtigallová and M. S. de Vries, Int. Rev. Phys. Chem., 2013, 32, 308–342, DOI: 10.1080/ 0144235X.2012.760884.
9. C. T. Middleton, d. L. Harpe, C. Su, Y. K. Law, C. E. CrespoHernández and B. Kohler, Annu. Rev. Phys. Chem., 2009, 60, 217–239, DOI: 10.1146/annurev.physchem.59.032607.093719.
10. E. Mburu and S. Matsika, J. Phys. Chem. A, 2008, 112, 12485–12491, DOI: 10.1021/jp807145c.
11. Z. Gengeliczki, M. P. Callahan, N. Svadlenak, C. I. Pongor, B. Sztaray, L. Meerts, D. Nachtigallova, P. Hobza, M. Barbatti, H. Lischka and M. S. de Vries, Phys. Chem. Chem. Phys., 2010, 12, 5375–5388, DOI: 10.1039/B917852J.

12. Z. Kan, Y. Lin, F. Wang, X. Zhuang, Y. Zhao, D. Pang, Y. Hao and Z. Tan, Nucleic Acids Res., 2007, 35, 3646–3653, DOI: 10.1093/nar/gkm203.
13. Q. Wang, J. Liu, Z. Chen, K. Zheng, C. Chen, Y. Hao and Z. Tan, Nucleic Acids Res., 2011, 39, 6229–6237.
14. G. N. Parkinson, M. P. H. Lee and S. Neidle, Nature, 2002, 417, 876–880.
15. H. A. Day, P. Pavlou and Z. A. E. Waller, Bioorg. Med. Chem., 2014, 22, 4407–4418, DOI: 10.1016/j.bmc.2014.05.047.
16. L. Cai, L. Chen, S. Raghavan, A. Rich, R. Ratliff and R. Moyzis, Nucleic Acids Res., 1998, 26, 4696–4705, DOI: 10.1093/nar/26.20.4696.
17. P. K. Patel and R. V. Hosur, Nucleic Acids Res., 1999, 27, 2457–2464, DOI: 10.1093/nar/27.12.2457.
18. E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, J. Chem. Phys., 2001, 115, 4604–4611.
19. C. Plutzer, E. Nir, M. de Vries and K. Kleinermanns, Phys. Chem. Chem. Phys., 2001, 3, 5466–5469.
20. E. Nir, K. Kleinermanns and M. S. de Vries, Nature, 2000, 408, 949–951.
21. C. Plutzer, I. Hunig, K. Kleinermanns, E. Nir and M. de Vries, ChemPhysChem, 2003, 4, 838–842.
22. E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, Phys. Chem. Chem. Phys., 2002, 4, 732–739.
23. E. Nir, M. Mu¨ller, L. I. Grace and M. S. de Vries, Chem. Phys. Lett., 2002, 355, 59–64, DOI: 10.1016/S0009-2614(02)00180-X.
24. E. Nir, I. Hunig, K. Kleinermanns and M. S. de Vries, Phys. Chem. Chem. Phys., 2003, 5, 4780–4785, DOI: 10.1039/B310396J.
25. M. S. de Vries, in Gas-Phase IR Spectroscopy and Structure of Biological Molecules, ed. M. A. Rijs and J. Oomens, Springer International Publishing, Cham, 2015, pp. 271–297.
26. P. Mignon, S. Loverix, J. Steyaert and P. Geerlings, Nucleic Acids Res., 2005, 33, 1779–1789, DOI: 10.1093/nar/gki317.
27. J. Florián, J. Leszczynski and S. Scheiner, Mol. Phys., 1995, 84, 469–480, DOI: 10.1080/00268979500100321.
28. P. Jurecka and P. Hobza, J. Am. Chem. Soc., 2003, 125, 15608–15613, DOI: 10.1021/ja036611j.

29. J. Sponer, J. Leszczynski and P. Hobza, J. Phys. Chem., 1996, 100, 1965–1974, DOI: 10.1021/jp952760f.
30. G. Villani, ChemPhysChem, 2013, 14, 1256–1263, DOI: 10.1002/cphc.201200971.
31. A. Manukyan and A. Tekin, Phys. Chem. Chem. Phys., 2015, 17, 14685–14701, DOI: 10.1039/C5CP00553A.
32. I. Usabiaga, J. Gonzalez, P. F. Arnaiz, I. Leon, E. J. Cocinero and J. A. Fernandez, Phys. Chem. Chem. Phys., 2016, 18, 12457–12465.
33. I. León, J. Millán, E. J. Cocinero, A. Lesarri and J. A. Fernández, Angew. Chem., Int. Ed., 2013, 52, 7772–7775.
34. I. León, J. Millán, E. J. Cocinero, A. Lesarri and J. A. Fernández, Angew. Chem., Int. Ed., 2014, 53, 12480–12483, DOI: 10.1002/anie.201405652.
35. T. A. Halgren, J. Comput. Chem., 1996, 17, 616–641.
36. T. A. Halgren, J. Comput. Chem., 1999, 20, 730–748.
37. M. Frisch, Gaussian 09, Rev. A02, Gaussian Inc., Wallingford CT, 2009.
38. S. F. Boys and F. Bernardi, Mol. Phys., 1970, 19, 553–566.
39. F. Dong and R. E. Miller, Science, 2002, 298, 1227–1230.
40. Y. Valadbeigi, M. Soleiman-Beigi and R. Sahraei, Chem. Phys. Lett., 2015, 629, 1–7, DOI: 10.1016/j.cplett.2015.03.007.
41. G. Villani, Chem. Phys., 2005, 316, 1–8, DOI: 10.1016/ j.chemphys.2005.04.030.
42. G. Villani, Chem. Phys., 2006, 325, 389–396, DOI: 10.1016/ j.chemphys.2006.01.015.
43. A. G. Santana, E. Jiménez-Moreno, A. M. Gómez, F. Corzana, C. González, G. Jiménez-Oses, J. Jiménez-Barbero and J. L. Asensio, J. Am. Chem. Soc., 2013, 135, 3347–3350, DOI: 10.1021/ja3120218.
44. I. Usabiaga, J. González, I. León, P. F. Arnaiz, E. J. Cocinero and J. A. Fernández, J. Phys. Chem. Lett., 2017, 8, 1147–1151.
45. H. Asami, K. Yagi, M. Ohba, S. Urashima and H. Saigusa, Chem. Phys., 2013, 419, 84–89, DOI: 10.1016/j.chemphys. 2013.01.038.
46. T. Zeleny, P. Hobza and M. Kabelac, Phys. Chem. Chem. Phys., 2009, 11, 3430–3435, DOI: 10.1039/B819350A.
47. M. Kabelac, F. Ryjacek and P. Hobza, Phys. Chem. Chem. Phys., 2000, 2, 4906–4909, DOI: 10.1039/B007167F.



Scheme 1 Keto and enol tautomers of cytosine, together with their atom numbering.

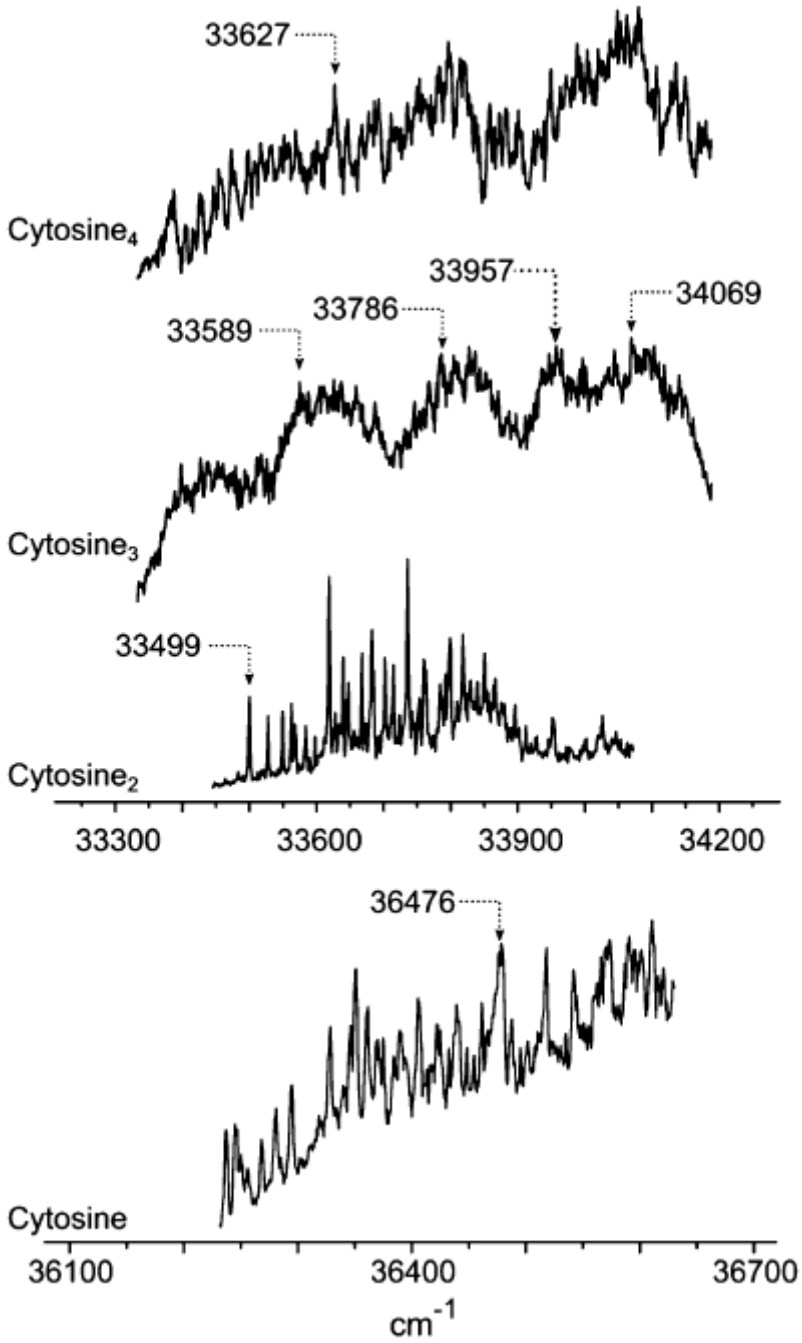


Fig. 1 REMPI spectra of enol cytosine (bottom panel) and keto cytosine aggregates (top panel). The arrows indicate the wavenumbers probed to record the IR/UV spectra.

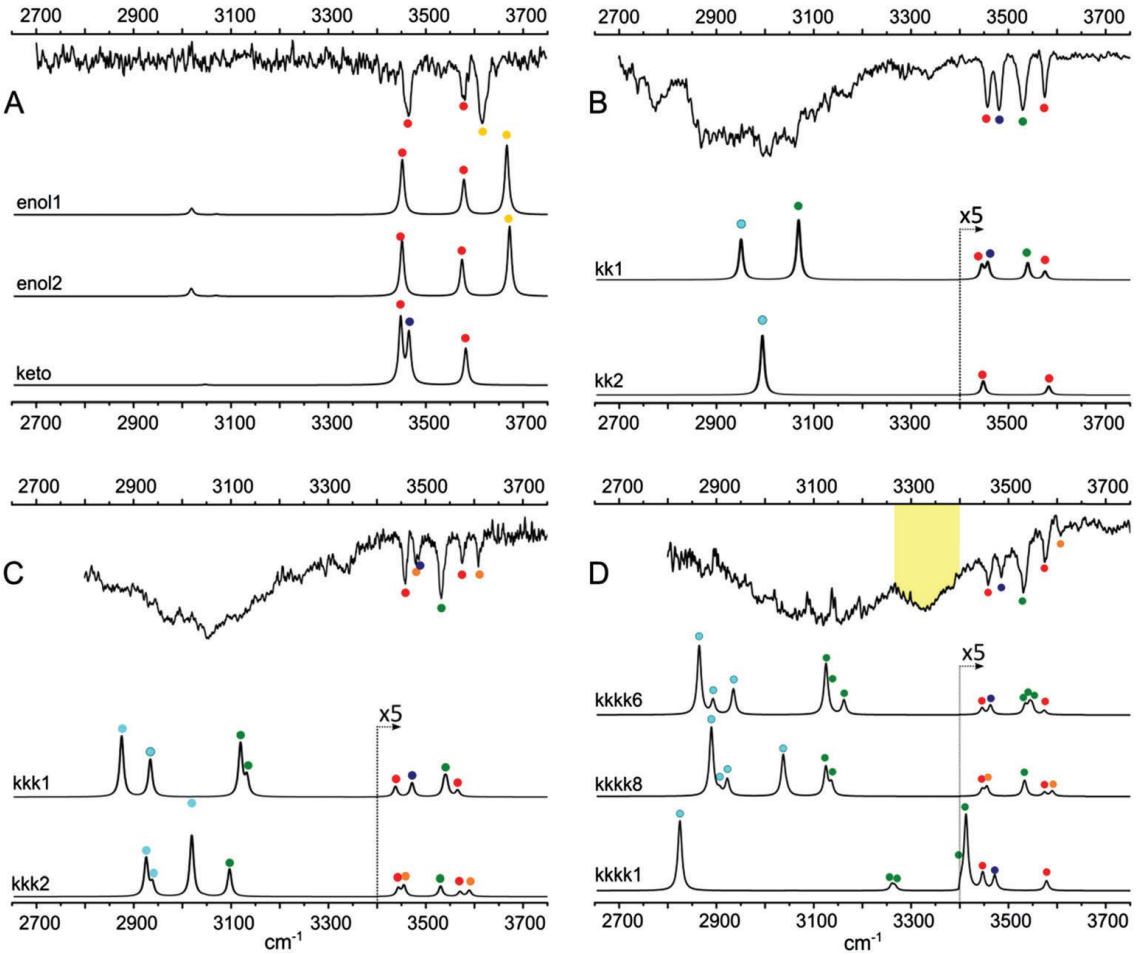


Fig. 2 Comparison between the mass-resolved IR/UV spectrum and the simulated spectra of some selected conformers of (A) enol-cytosine, (B) the cytosine dimer, (C) the cytosine trimer and (D) the cytosine tetramer. Comparison with the rest of the calculated species can be found in the ESI. A correction factor of 0.9483 was used to account for anharmonicity. The colour dots indicate the assignment of each transition and correlate with the shaded groups of the ball & stick models in Fig. 3. Comparison with the simulated spectra of additional computed structures can be found in the ESI.

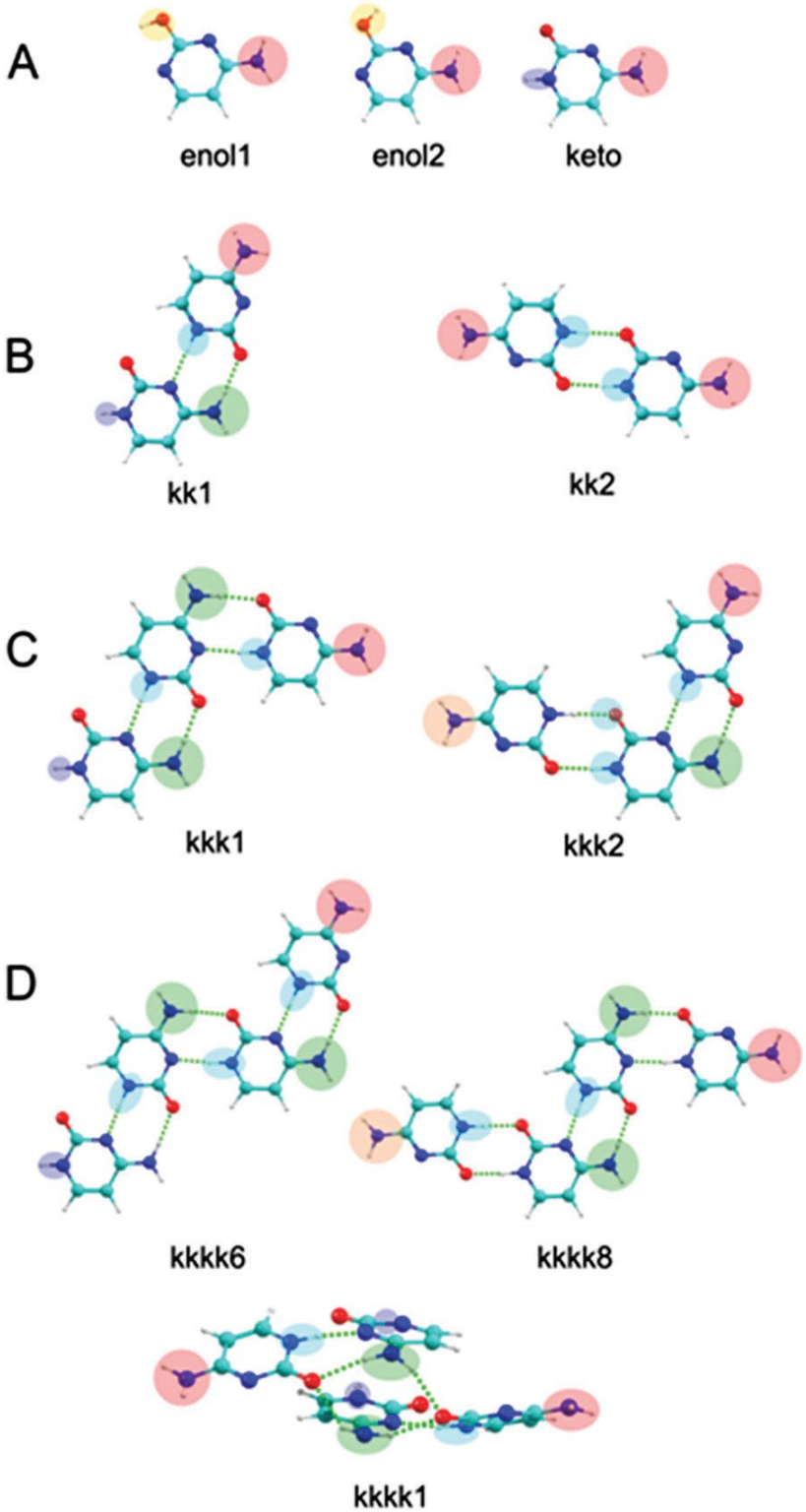


Fig. 3 Structures of the isomers whose simulated spectra are presented in Fig. 2: (A) the three most stable tautomers of cytosine; (B) the two assigned isomers of the cytosine dimer; (C) the two assigned isomers of the cytosine trimer; and (D) the assigned isomers of the cytosine tetramer. The rest of the calculated structures can be found in the ESI. The colours of the shaded groups match those used in the assignment of the transitions on the experimental spectra in Fig. 3.

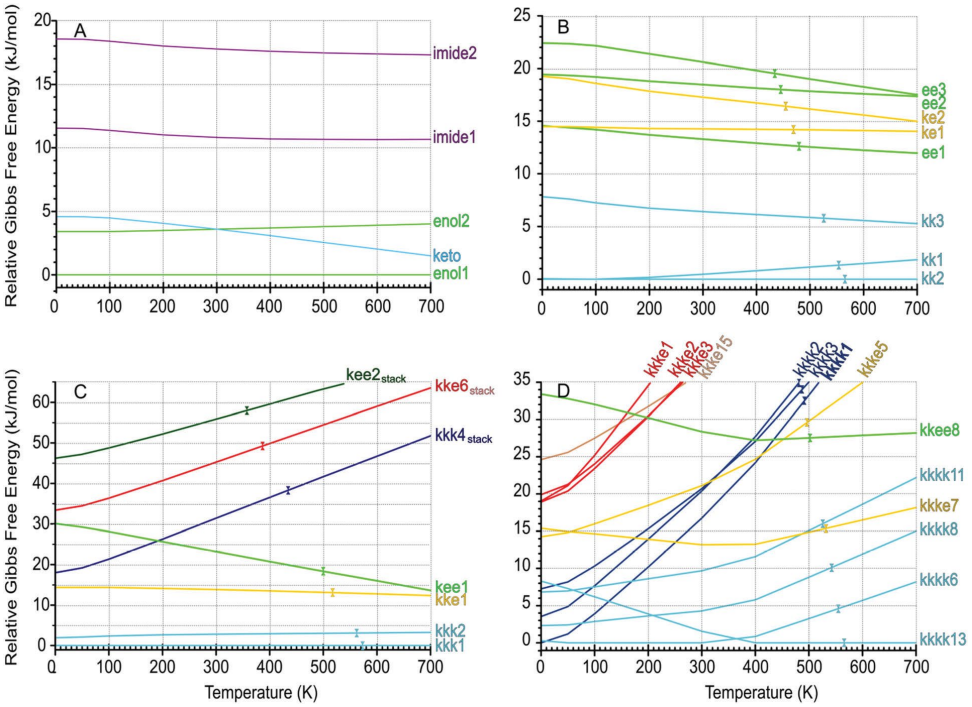


Fig. 4 Relative Gibbs free energy of some selected calculated species of cytosine and cytosine aggregates in the 0–700 K interval. Clearly, the dimers formed exclusively by keto tautomers are significantly more stable. The triangles mark the DG = 0 temperature for each species: (A) monomer; (B) dimer; (C) trimer; and (D) tetramer. The computed DG for all the calculated structures can be found in the ESI. A similar figure but assuming a non-equilibrium system can be found in the ESI.

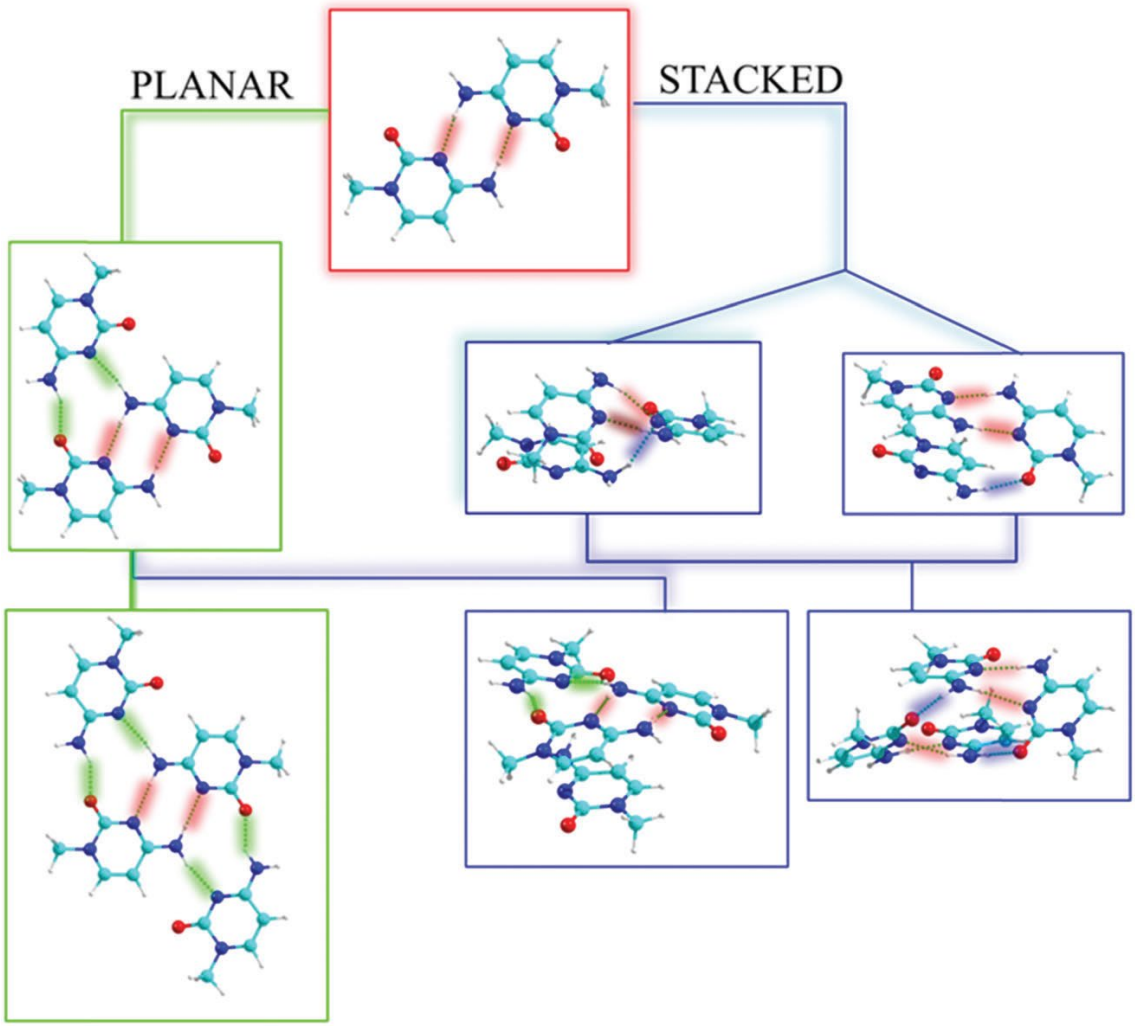


Fig. 5 Most stable structures of 1-methylcytosine aggregates computed at the M06-2X/6-311++G(d,p) level. The complete set of structures together with their relative stability according to DH and DG can be found in the ESI.