PCCP

Accepted Manuscript

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/pccp

Journal Name

ARTICLE

Quantum chemical MP2 results on some hydrates of cytosine: Binding sites, energies and the first hydration shell

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

DOI: 10.1039/x0xx00000x

www.rsc.org/

Géza Fogarasi[a],† and Péter G. Szalay[a]

A detailed quantum chemical investigation was undertaken to obtain the structure and energetics of cytosine hydrates Cyt.nH_2O, with n = 1 to 7. The MP2(fc)/aug-cc-pVDZ level was used as standard, with some DFT (B3LYP) and coupled cluster calculations, as well as calculations with the aug-cc-pVTZ basis set added for comparison. In a systematic search for microhydrated forms of cytosine, we have found several structures not yet reported in the literature. The energies of the different isomers, as well as the binding energies are compared. When predicting the stability of a complex, we suggest using a scheme where the water molecules are extracted from a finite model of bulk water. Finally, based on energetic data, we suggest a rational definition of the first hydration shell; with this definition, it contains just six water molecules.

Introduction

The fundamental significance of DNA bases in molecular biology justifies a close scrutiny into every aspect of their structures. We have been studying the structure and *tautomerism* of cytosine for some time, currently investigating the electronic excited states of its "canonical" amino-oxo form (Cyt),[1] and its hydrates.[2] As part of the latter study we obtained the geometries of several complexes, and it seemed now reasonable to extend the search for more isomers, investigating their structures in detail.

There have been numerous quantum chemical (QC) computations on the interaction between cytosine and water. One approach is to use a continuum model as, for example, in refs 3 and 4. Most studies, and specifically those using higher QC levels, treat the interactions in a model of 'microhydration', attaching individual water molecules to cytosine. The number of water molecules may be just one,[5–9] or up to five waters that can bind directly to cytosine.[10–13] Some specific structures with as many as 14 water molecules have also been constructed and their geometries optimized.[14–16] In the latter structures, however, several water molecules bind to each other and not to cytosine. The present study investigates Cyt.nH_2O hydrates in a systematic way for n = 1 to 7, with the objective of identifying as many isomers as possible. The question of the first hydration shell will also be addressed.

The notation of individual complexes is based on the scheme introduced in our first study, one on the monohydrates.[9] Moieties (neighbourhoods) around Cyt are defined as follows. *A*: N1(ring)–H with C=O(carbonyl), *B*: N8(amino)–H_{syn} with N3(ring), *C*: C=O(carbonyl) with N3(ring), *D*: N8(amino)–H_{anti} with C5–H (the position of amino hydrogens given relative to N3). These are shown in Fig. 1, along with the numbering of atoms. Note that the moieties *A*, *B* and *C* may give place to two H–bonds (on *C*, however, see later below), while at *D* water builds one H–bond only. Throughout the paper, moieties will be given in italics, while the molecules themselves in normal fonts. For the larger complexes then, for example AB refers to a dihydrate with waters at neighbourhoods *A* and *B*, AABC a tetrahydrate with two waters in the moiety *A*, one at *B* and one at *C*, and so on.

Our primary objective is to get an overview on the possibilities for Cyt to bind water molecules, to establish the energetic differences between a variety of isomers and to determine individual binding energies. In relation to the latter, general comments will be given on the uncertainty of BSSE (basis set superposition error) corrections.

Computational details

Electronic structure calculations have been performed at MP2 and coupled cluster singles, doubles with approximate triples CCSD(T) levels, as well as by the DFT method using the B3LYP functional. Double and triple zeta correlation consistent basis sets with diffuse functions (aug-cc-pVDZ, aug-cc-pVTZ) were used. Core electrons were frozen in all post-Hartree-Fock calculations. The standard level of theory in this study was MP2(fc)/aug-cc-pVDZ. For comparison, numerous DFT(B3LYP)/aug-cc-pVDZ calculations were also performed. Vibrational frequencies based on analytical second derivatives were routinely included in the DFT calculations; MP2 frequencies were obtained from numerical differentiation of first derivatives. Single point CCSD(T) calculations of energies

---

[a] *Institute of Chemistry, Eötvös University, H-1518 Budapest, P. O. Box 32, Hungary.*
†Corresponding Author. Phone: +36 1 372 2540. Fax: +36 1 372 2592. E-mail: fg@chem.elte.hu.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

were carried out at some of the stationary points. MP2 and DFT calculations were performed using PQS,[17] while coupled cluster energies were calculated by CFOUR.[18] Pure water clusters have also been optimized by MP2(fc)/aug-cc-pVDZ, making use of excellent starting geometries from The Cambridge Cluster Database.[19]

Results and discussion

Testing the level of QC computations

For a systematic study, one first has to check the stability of the results with respect to computational level. The MP2(fc)/aug-cc-pVDZ level used as standard in several of our studies including the present work, is compared to higher levels of theory in Table 1. As to the basis set, going from aug-cc-pVDZ to the corresponding augmented triple-zeta basis (aug-cc-pVTZ), the maximum change in the *relative* energies is insignificant, 0.13 kcal/mol. Concerning the role of electron correlation, the CCSD(T) data confirm the MP2 results within 0.1 kcal/mol for the dihydrates, with a maximum deviation of 0.18 kcal/mol for the trihydrates (ABC). The tests thus prove that the MP2(fc)/aug-cc-pVDZ level is well adequate for the present purpose. (In fact, we stressed already in an earlier study[9] that for distinguishing between different *tautomers* of the nucleotide bases, a high level method of electron correlation is obligatory, while for a given tautomer, as in the present case, hydration can be well described by MP2 or DFT methods.)

Besides checking MP2 by coupled cluster calculations, we have included DFT results in Table 1. Overall, the latter perform also well. In fact, it is reassuring to see that in 10 of the 11 sets of calculations, the order of energies is the same: AA < AB < BB <BC, and AAB < ABB < ABC < AAA, for the di– and trihydrates, respectively; the only discrepancy is found among the trihydrates, where B3LYP places AAA below ABC in energy, with both basis sets. In the quantitative details, however, the DFT results for the isomers' relative energies differ from the wave function results by up to a factor of 2. The latter will be discussed in relation to literature results further below.

Hydrates identified and their relative energies

Table 2 gives an overview of the final results, showing the relative energies of individual isomers within each group defined by the number of water molecules, n. The table includes also the zero point vibrational energies. Looking at the numbers in parentheses in the first column of the table, an observation seems worth mentioning: as the number of water molecules increases, the change in ZPE is fairly constant, 15-16 kcal/mol per one additional water. This may indicate that, in the most stable isomers, the additional hydrogen bonds, whose number increases by two with each additional water, are quite similar in strength and nature.

As to the (quite small) relative energies between isomers (in a given row of Table 2), the ZPEs are not negligible but, fortunately, do not change the order of relative energies, except for one case: among the tetrahydrates, AABC and BBCC are equal in energy without ZPE, while the addition of ZPE puts BBCC more than 1 kcal/mol above AABC.

Two sets of literature results are included in Table 2 for comparison: traditional B3LYP/DZP++ calculations by Kim and Schaefer[13] and an intriguing recent study by Thicoipe et al.[20] The latter authors used a global search algorithm (GSAM) to find the microhydrated forms, with up to five waters (and for all the five nucleic acid bases). The QC method they used was DFT (B3LYP and B3LYP-D, with the 6-311+G(d,p) basis set). It is somewhat surprising that, as will be seen below, the original B3LYP functional performs in this case better than its variant B3LYP-D, augmented with an empirical dispersion term.

The monohydrates have been studied many times by computational chemistry still some discussion seems to be in order. There are, in principle, four possible positions for water to attach to Cyt, as shown in Fig. 2. Most QC calculations,[5–9] including our coupled cluster results,[9] agree that A is the strongest binding moiety, favoured over B; the MP2 energy difference in Table 2 is ΔE(B-A) = 0.72 kcal/mol (0.55 with ZPE), the CCSD(T) result of ref. 9 was 0.82 kcal/mol (no ZPE). DFT calculations give a larger difference, roughly between 1.0 –1.5 kcal/mol.[8,13,20] As noted above, the new B3LYP-D fails here, giving a virtually zero difference between B and A (0.1 kcal/mol, Table 2, taken from ref. 20). This is too low in the light of all other results, while the original B3LYP with Δ(B–A) = 0.7 kcal/mol is well acceptable. Another remark concerns the work by Sivanesan et al.[10]: these authors reported moiety B as the lower-energy, i.e. stronger, binding site which is definitely wrong in the light of all more recent calculations.

Moiety C deserves a special comment. Here, there is a possibility that water binds with both of its protons as donors, and such an energy minimum was indeed reported on the basis of MP2 calculations.[4] By contrast, it was emphasized both by DFT calculations[8] and by MP2 calculations[9] that this is not the case. More recently, Hunter and Wetmore[21] investigated the question in detail using DFT calculations. Their finding was that an energy minimum does exist here, but this is a „ring-opened" H–bond structure. The authors emphasize the role of diffuse functions: without them, the H–bond ring is allowed. We have also investigated this question in some detail now, running both B3LYP and MP2 optimizations with several basis sets, checking also the nature of the stationary points from the corresponding vibrational frequencies. The result is in line with ref. 21: independent of the electron correlation method (DFT or MP2), several basis sets do indeed give double H–donor structures as true minima. (The hydrogen bonds are weak, however, as indicated by bond lengths above 2 Å; with MP2/cc-pVDZ, r(N3…H) = 2.12, r(C=O…H) = 2.32 Å). And the important point is the role of diffuse functions: as soon as one goes from, say, 6-311G(2d,2p) to 6-311++G(2d,2p), or from cc-pVTZ to aug-cc-pVTZ basis set, only one H–bond survives. As a tentative idea, one can perhaps understand this strange phenomenon from the fact that two H–bonds involve a strained ring in which the water's own H–O–H angle is forced to decrease. (For example, its MP2/cc-pVDZ value is 98.2^{o}.) The latter is „less acceptable" by the augmented basis set which gives, for isolated water itself, a larger angle than the basis without diffuse functions (103.9^o vs. 101.9^o; the former is also closer to the experimental value of 104–105^o). It is also interesting that in the single H–bond which does form, water donates its hydrogen to the carbonyl oxygen, rather than to N3 (Fig. 2, C). We tried several optimizations to bind the water–hydrogen to N3, but in all of these attempts the optimization stubbornly led to complex B.

Finally, a water molecule can bind at position D to the amino-hydrogen *anti* to N3 (Fig. 2, D). This bond is isolated, so it is often omitted in the studies. It is significantly weaker than other H-bonds in the hydrates, as can be inferred from the long H(amino)..O(water) distance; e.g., our MP2(fc)/aug-cc-pVDZ calculation has given a distance of 2.03 Å, while the other H–bonds in the monohydrates are in the range of 1.82–1.98 Å. (The last value refers to the other amino–water bond in complex B.) As expected, isomer D with its energy of 6.1 kcal/mol (5.1 kcal/mol with ZPE) is the less stable form among the monohydrates. We will use this isolated site *D* for accepting water only in the larger systems (with 6 or 7 waters), where it may be the only possibility to attach the next hydrogen directly to Cyt.

For the dihydrates, we have identified four isomers as shown in Fig. 3 and listed in Table 2. It is interesting that structure AA, with both waters in one ring, has the lowest energy. Mourik et al. reported on these structures some time ago[11] and commented that, as a general trend, „the second water appears to preferably bind to the first one". The present results show, however, that the trend is not general: the other, alternative one-ring system appears in BB, but the second most stable structure is AB, with a relative energy of 0.4 - 0.5 kcal/mol only, while BB with ΔE = 1.16 kcal/mol is significantly higher. Our findings agree roughly with the DFT results of Kim and Schaefer,[13] in that the latter work lists (in our notation) AA, AB and BB as the main species. At the same time they report, apparently unaware of both refs 11 and 9, an AC-type structure. This fifth isomer does not exist as an energy minimum in the present MP2 results: our attempts to localize AC have led to AB, in analogy with the result above on the C-type monohydrate. In fact, already ref. 11 has noted the uncertainty about AC (23 in their notation). In addition to the dihydrates in Fig. 3 and Table 2, ref. 13 reports two structures where position D is occupied; as expected, their relative energies are high, 6–7 kcal/mol. As noted above we have searched for structures using site D in the larger hydrates only. We have, however, found isomer BC (Fig. 3), with moderate energy of 1.7 kcal/mol (Table 2), which was not considered in ref. 13. Thicoipe et al.[20] report a "(BC)2" structure with an energy of 1.9 kcal/mol, their notation indicating that waters are 'hovering' between B and C. In all probability this is our MP2-calculated BC dihydrate. (As a check, we have also found BC by DFT, the latter giving relative energy of 2.6 kcal/mol by B3LYP/aug-cc-pVDZ).

For the trihydrates the seven structures shown in Fig. 4 comprise all possible variations, if site D is omitted (see above) and only one water is allowed at neighbourhood C (as will be seen below, the latter restriction may be questioned). For comparison, ref. 13 reported 4+4 isomers (by '+4' we indicate that each structure uses one additional water molecule at site D.) They did not consider 3-water rings, so the present AAA and BBB do not appear in ref. 13. AAA does appear in ref. 20; in fact, with the B3LYP potential they have found two complexes of this type (A3 and A'3 in their notation), one at 0.8 kcal/mol, the other at 1.3 kcal/mol, significantly lower than our result of 2.5 kcal/mol. We were also trying to find a second AAA-type structure but no other energy minimum could be identified. On the other hand, BBB, which is roughly at the same energy as AAA, is missing in ref. 20. It should also be noted that ABC, lying at 1.1 kcal/mol (0.8 kcal/mol with ZPE) in the present MP2 result, is at much higher energy of about 2.5 kcal/mol in both of the DFT results.[13,20] The DFT calculations of the present work show the same (Table 1). In both cases, ABC and AAA, the CCSD(T) data in Table 1 support the MP2 results. It is also interesting to see, however, that –while DFT gives larger ΔE values than the wave function methods in most cases – an opposite trend is found for AAA. (We have checked that AAA in MP2 and DFT is indeed the same type of structure.)

For the tetrahydrates six structures are shown in Fig. 5, with energies compiled in Table 2. As may be expected from the above, the lowest energy complex is AABB. Isomer AABC, in which one B–type water moved closer to the carbonyl group, thus becoming C (Fig. 5), lies only 0.8 kcal/mol higher in our result; the DFT calculation[13] put it more than twice as high, 2.1 kcal/mol, while the global search[20] did not find it. Both studies,[13,20] as well as the present one, have found an isomer shown as ABBC in Fig. 5, but our MP2 result of $\Delta E \cong 1.3$ kcal/mol lies again significantly lower than predicted by DFT, 2.2 and 1.8 kcal/mol, respectively. Note that a more recent study, one on the theoretical vibrational frequencies by Carbonniere et al.[22] considers only two isomers, AABB and AAAB, as the most stable tetrahydrates. While all results agree that AABB is the most stable isomer, Table 2 suggests that well below AAAB in energy there are three to four structures. Both DFT studies[13,20] list some further, high energy structures by putting a water to site D; as noted above, these are not considered in the present study. At the same time, it is quite remarkable that we have found a new isomer, BBCC; what is more, it is at a quite low energy of ~0.7 kcal/mol (with ZPE it lies significantly higher, at 1.65 kcal/mol). As shown in Fig. 5, it has an interesting three-dimensional arrangement of the H–bonds – we will see this structural motif again in the larger systems below.

Among the pentahydrates (Fig. 6), AABBC had been the most stable structure (Table 2) for some time in our search. This seemed logical, supported by both earlier studies.[13,20] In fact, except for some high energy structures using site D, ref. 13 lists this structure only. Further search for energy minima in the present study revealed, however, that there exists the isomer ABBCC, with its BBCC part showing the 3D structure (Fig. 6, two views) seen already above in the tetrahydrate BBCC. What is more, this is the lowest energy complex. In the light of the stability of BBCC discussed above, the stability of ABBCC is, in fact, not surprising. We have included in the present study two structures with three–water rings, AAABB and AABBB, respectively (Fig. 6). In line with the smaller hydrates above, the complex with an AAA-ring lies at relatively high energy of 2.2 kcal/mol. Interestingly, the complex with a BBB-ring is lying lower by about 1.0 kcal/mol. The global search[20] for the pentahydrates has found six structures in the low range of 1 kcal/mol, plus the AAABB isomer at 2.0 kcal/mol. According to Table 2 we are in agreement with ref. 20, except for the fact that in our calculations ABBCC is the lowest energy isomer. It is confirmed that three structures, including AABBC and AABBB, are within a range of about 1 kcal/mol only. However, there is a surprising discrepancy: ref. 20 reports two structures of AABBB-type, A2B3 and A2B’3 in their notation. We checked their Cartesian coordinates from the paper’s ESI supplement, by calculating from them sensitive quantities like the nuclear repulsion and the moments of inertia. Specifically: in the principal axis system, the coordinates agree to 10^{-5} - 10^{-6} Å, with a maximum deviation of 1×10^{-4} Å, *except for the signs:* these agree along the *a* and *c* axis but are opposite along *b,* indicating that the two structures are just mirror images, i.e. *enantiomers*. This would change of course the percentage compositions they presented, too (Fig. 5 in ref.20). We searched also for a further isomer, ABBDD, inspired by Alemán’s early study[14] which reported this structure on the basis of Hartree-Fock calculations. The existence of this isomer is now confirmed at the MP2/aug-cc-pVDZ level, with an energy of 2.80 kcal/mol relative to ABBCC. Note, however, that in this structure the second D–water binds to waters only, with no bond to Cyt (Fig. 6). (See also remarks below on the hydration shells.)

Two hexahydrates and one heptahydrate have been optimized as complexes with the highest number of water molecules in the present study (Fig. 7). As explained above, with these larger systems we have routinely included site D in the calculations. In fact, when searching for hexahydrates, we instinctively tried to attach the sixth water at this site, to the yet “free” amino-proton (*anti* to N3). One finds here indeed a true minimum but, as it turned out, putting the sixth water in the “crowded” region at moieties B and C is energetically favourable. As was the case with the tetra– and pentahydrates above, the four water molecules form a 3D arrangement (Fig. 7) and AABBCC becomes much more stable than the structure occupying site D: as Table 2 shows, AABBCD lies higher in energy by 4.2 kcal/mol (3.1 kcal/mol with ZPE). This reminds one of a recent study by Mourik et al.[16]: they performed B3LYP, LMP2 and (semiempirical) PM6 computations on cytosine.14H_2O and adenine.16H_2O comparing in each system two optimized structures, „one with water molecules distributed around the central base and one with a clustering of water molecules. The clustered structures were found to be energetically favoured”, by all three methods. As we see now, this “clustering” appears already in the BBCC moiety. A similar structure, now with one water also at site D, was found for the heptahydrate (Fig. 7).

The search for the hexahydrates and heptahydrates of cytosine was certainly not exhaustive in the present work. More structures could easily be obtained by inserting further water molecules into the H–bond rings. This way, even Cyt.14H_2O has been constructed long ago.[15] However, the distance from Cyt is getting larger and larger then, and the mid-ring waters do not bind to Cyt, so that the term “hydration” may be questioned. For example, the AA and BB rings could be increased to AAA and BBB as above in the smaller systems. In the AAA ring, the mid-ring (“apex”) water is at a distance of about 4.5 Å from the closest atom in Cyt, as compared to the two “terminal” waters with corresponding distances of 2.7–2.8 Å. In the BBB case, the relevant distances are 5.0 and 2.8 Å, respectively. In fact, when discussing the binding energies below, we will bring forward arguments that the first hydration shell may end with six or seven water molecules.

Hydration energies

The calculated binding energies of the hydrates Cyt.nH_2O, n=1, 2, … 7 are given in Table 3. For each hydrate, only the lowest-energy isomer (cf. Table 2) is listed here. For the discussion below we will need also the corresponding water clusters. Optimized at the same level as the cytosine complexes, MP2(fc)/aug-cc-pVDZ, these are included in column 3 of Table 3. The calculation of binding energies from these data deserves some comments. Most studies consider simply the reaction

$$\text{Cyt} + n\text{H}_2\text{O} \rightarrow \text{Cyt.}n\text{H}_2\text{O} \qquad (1)$$

so that the calculation of the reaction energy is straightforward from the individual energies. This describes the hypothetical situation of Cyt combining with *n* isolated H_2O molecules, all in the gas phase. It is perhaps more “to the point”, however, to consider a situation where the waters on the left side are not free molecules; rather, they exist already in an $(H_2O)_n$ cluster. This way one gets information about the competition between cytosine–water and water–water bindings.

Thus, instead of Eq. 1, we suggest the use of the following reaction:

$$\text{Cyt} + (\text{H}_2\text{O})_n \rightarrow \text{Cyt.}n\text{H}_2\text{O} \qquad (2)$$

Eq. 2 describes the reaction between Cyt and a water cluster, both in the gas state. One can also try to go one step further: imagine that the $(H_2O)_n$ cluster comes from a huge cluster $(H_2O)_N$, an approximation for bulk water. The reaction in this case is:

$$\text{Cyt} + \text{W}_N \rightarrow \text{Cyt.}n\text{H}_2\text{O} + \text{W}_{N\text{-}n} \qquad (3)$$

where the notation W distinguishes the fictitious water from a true H_2O.

In Eq. 3 we need the energy of *n* molecules leaving the cluster (to join cytosine). If the cluster is large, all molecules are “equal” on average; therefore, for the (leaving) waters we take the average energy of $E_W = E(W_N)/N$. To this purpose, we optimized water clusters up to N=14, at the same level used for the complexes, MP2(fc)/aug-cc-pVDZ. The graph in Fig. 8 indicates that for the average energy we may be not too far from convergence; so an estimate of the average energy of $E_W = -76.2785\ E_h$ will be used below. This is equivalent to a relative energy of −11.04 kcal/mol for bound water W, with respect to free water (Table 3).

As a next question, whichever scheme we apply from above, the energies of the left hand side reactants are subject to basis set superposition error, BSSE. The usual correction by the counterpoise method (CPC) of Boys and Bernardi[23] improves (lowers) the reactants' energies by applying „ghost" basis functions. In case of a reaction of two components, the scheme requires four additional energy calculations, a rather large extra effort if one does it for each complex separately. Thus, for a given number of waters in the complex, *n*, these calculations were done for the most stable isomer only (Table 3, CPC data given explicitly in the footnote).

A remark on this point seems to be of general interest. BSSE may be, in fact, different for different isomers within a group of the same *n*. We have tested this question on the two isomers AA and AB of Cyt.2H_2O. The *direct* calculated energies give AB just a little *over* AA, the most stable form: Δ = E(AB) − E(AA) = 0.45 kcal/mol (see Table 2). Suppose now that we follow an indirect approach, through the formation (complexation) energies ΔE of AA and AB, respectively, from their (same) *fragments*, Cyt + $(H_2O)_2$, using eq. 2. In principle, E(AB) − E(AA) = ΔE(AB) −ΔE(AA) should hold. In the actual calculations, however, the equality stands only if the CPCs in the two complexation reactions are the same for AA and AB. We have calculated the CPCs, obtaining 2.80 kcal/mol for AA and 3.78 kcal/mol for AB. Thus, the *difference* in the correction is 1 kcal/mol, twice as large as Δ itself (and is in the opposite direction). Using CPC thus *would change the order of energies*, putting AB below AA. In the case of comparing energies just for isomers, the direct calculation of energies for each isomer is the accepted practice (when doing this, one tacitly assumes that the BSSE −relative to an infinite basis set − is the same for two isomers). The present discrepancy, namely the finding that the direct way and the way through formation energies give different results indicates clearly that the use of CPC is questionable.

While there have been polemic discussions on BSSE quite early,[24,25] an especially critical study has come out recently by Mentel and Baerends.[26] Their intriguing conclusion was that the widely used counterpoise correction (CPC)[23] is in fact not justified. This view was accepted, for example, in a recent study by Cremer and co-workers in their computations on the formic acid dimer.[27] At the same time, a most recent paper on the same system by Miliordos and Xantheas[28] argues again for the use of CPC. In our opinion, the investigation of ref. [26] was so thorough and detailed that one has to accept their case against CPC, except for one serious limitation: they studied a system of two atoms only, the beryllium dimer. Our present experience with a much more complex system, as noted above, supports, however, the renegade idea that the universally accepted handling of BSSE can be seriously challenged. Nevertheless, this important subject would require further detailed investigations. In the present study, CPC data are included if appropriate but in the discussion we will focus on the uncorrected data.

With all these uncertainties, the data in Table 3 are instructive. First, it is obvious that formation energies of complexes, like those in column 5 (based on eq. 1), may be misleading. From such data a conclusion is often drawn as, for example: "Each successive water molecule is bound by 7−10 kcal/mol to the relevant cytosine complex".[13] While the statement is of course true, as soon as we take into consideration the *ab ovo* formation of water clusters, the binding energies get drastically reduced, as seen in column 6, based on eq. 2. (Just to indicate the possible effect of BSSE, the CPC−corrected values are included in parentheses.) Finally, the data in the last column refer to the situation when the water molecules are coming from W_N, the latter representing a "discrete" model of bulk water as described above in connection with eq. 3. In this case, the energy gain by complexation is just a minimum. More importantly, the data in the last column of Table 3 show that the successive steps of adding water molecules are exothermic only up to n = 4, with Cyt.5H_2O being already less exothermic than Cyt.4H_2O. Finally, by adding the seventh water molecule −attached to the amino group at site D− the total reaction has become endothermic. (Note that temperature effects are not included in these considerations.) We suggest that this energy criterion may be used for a definition of the first hydration shell around cytosine. The latter then contains just six water molecules, and does not use site D.

We realize the uncertainties in this statement. Additional water molecules can be built into the H-bond rings, forming larger rings. The latter, however, would be just part of an H-bonded chain of water molecules, no more bound directly to the base molecule. Also, as noted above, the distance from cytosine is getting larger for these extra waters. The notion of separated hydration shells being itself uncertain, the present suggested picture on the first shell seems acceptable.

Structural changes in the cytosine skeleton upon hydration

From chemical intuition one would not necessarily expect significant structural changes in the cytosine ring upon hydration. Nevertheless, on the basis of DFT calculations it has been pointed out some time ago by Leszczynski et al.[15] that microhydration does make significant changes even in bond lengths. What is more, ref. 15 suggested that the lengthening of the C=O bond indicates a shift from the canonical oxo-amino form to the hydroxo-imino tautomer, in its *zwitterionic* form.

The same group performed also dynamics studies on the subject.[29,30] The latter impressive work applied Car−Parrinello calculations (for all four nucleotide bases), using DFT with the BLYP functional and a plane wave basis for the electronic structure. In the case of cytosine, the system was put in a cubic cell (with periodic boundary conditions) and contained 57 water molecules with one cytosine. It may be of interest to compare our results on individual hydrates with the average structural properties given by dynamics. Among the basic results, the dynamics study has given average coordination numbers. Specifically for cytosine, this number was 6.12 water molecules around Cyt. In the light of this, our suggestion above, that the first hydration shell ends with the AABBCC structure containing 6 water molecules, seems justified.

In the structural details, ref. 29 emphasizes that in the neighbourhood BC (in our notation) the water molecules form a three-dimensional structure (rather than essentially planar rings). The present results for individual structures with 4 –7 waters show this explicitly. As done for the pentahydrate above, two views are shown now in Fig. 7 for both AABBCC and AABBCCD, demonstrating clearly the 3D structure.

In the second part of the dynamics study[30] individual geometric changes of the nucleotide bases themselves upon hydration are given. Inspired by this, we have compiled some of the relevant bond lengths in Table 4. The dynamical averages shown in Fig. 2 of ref. 30 are cited as "Cyt in bulk water" in our Table 4. The "bulk" data (in parentheses) may be compared with the largest complex, the heptahydrate: Δ(C=O7) = 0.028 (0.036); Δ(N1–C2) = –0.031 (–0.053); Δ(N3–C4) = 0.028 (0.028); Δ(C4–N8) = –0.031 (–0.030) (all figures in Å). The agreement is quite amazing: the only significant difference is in the N1–C2 bond shortening. To check the role of the electronic structure method, we have made our own DFT calculations, given in the last two rows of Table 4. The result proves clearly that the discrepancy is between static vs. dynamic treatment and not between DFT and MP2.

As noted above, the most intriguing suggestion in ref. 15 was that the elongated carbonyl C=O bond is a sign of the structure moving towards an enolic form. "*Its elongation and shift to enolic-like character results in the formation of three lone pairs on oxygen. Thus, oxygen atom of carbonyl group can form three hydrogen bonds (for certain period of time).*" This trivalent oxygen is really intriguing; by contrast, in the present results the carbonyl group participates in at most two H–bonds. As seen in Table 4, the C=O bond length changes from 1.231 Å in isolated cytosine by 0.015 Å in A and AA, and the bond is slightly lengthening further in the larger hydrates (n = 5 – 7), to a maximum of 1.263 Å in the hexahydrate AABBCD (1.259 Å in the heptahydrate). We think that a lengthening of 0.03 Å is significant but a bond distance of 1.26 Å is still far from a typical C–O single bond of 1.35-1.45 Å. An important argument in ref. 29 is the appearance of out-of-plane (o-o-p) H-bonds characterized by H(water)–O7–C–C torsional angles, suggesting "*a tetrahedral type of coordination and sp3-hybridization of oxygen.*" In the present results for the heptahydrate the two H–bonds of the carbonyl group are o-o-p indeed, but only by 13–14^{o} (N1–C2=O7–H19 =–12.8^{o}, N3–C2=O7–H24 = 13.9^{o}.) For comparison, the two H–bonds at the ring *nitrogen* N3 are much more o-o-p: O–C2–N3–H31 = 34.4^{o} , O–C2–N3–H25=–56.0^{o}.

As a further inquiry about the C=O bond lengthening and the related enolic structure, we have performed a quick test on the heptahydrate AABBCCD *and* its zwitter-ionic tautomer (moving the N1-H hydrogen to the –N8H$_2$ amino group): optimization of both forms at the B3LYP/6-31++G(d,p) level has given energies of –930.12553 vs. –930.09441 Eh. Thus, even in the presence of 7 water molecules, the ionic form is about 0.85 eV higher in energy than the neutral tautomer, suggesting that the enolic structure is not preferred. However, for a reliable description of this charge separation one should use a multiconfiguration method.

Conclusions

In a systematic search for microhydrated forms of cytosine, up to a heptahydrate, we have found several structures not yet reported in the literature. On the basis of comparative CCSD(T) calculations on the smaller complexes, the relative energies, as obtained from MP2(fc)/aug-cc-pVTZ calculations, are expected to be more accurate than in previous DFT results, although the general picture given by the two methods is quite similar. When applying counterpoint corrections (CPC) on BSSE, we called attention to serious uncertainties in the latter, with the conclusion that it may be better not to use CPC at all.

We also suggest that when studying binding energies, one should take into consideration the competition between cytosine-water and water-water clusters, which dramatically changes the picture on the energetics of hydration. If n water molecules are extracted from a finite model of bulk water, W$_N$, the net hydration energy is increasingly exothermic up to n = 4 only, then is falling off and with the 7th fictitious water, W, the complexation is already endothermic.

For a definition of the first hydration shell we propose that beyond considering whether a water molecule is bound directly to the central base (as proposed, for example, in a recent study on glycine[32]), the relevant energy change upon attaching the next water molecule should also be checked. In the latter, the competition between cytosine – water and water – water interactions is important. As noted above, the reaction Cyt + 7W $\rightarrow$ Cyt.7H$_2$O is already endothermic. With the present definition then, the first hydration shell ends with the sixth water molecule.

Acknowledgements

This work has been supported by *Országos Tudományos Kutatási Alap* (National Research Fund of Hungary): OTKA; Grant No. 104672.

Notes and references

§ The authors declare no competing financial interest

1 P. G. Szalay, T. Watson, A. Perera, V. Lotrich and R. J. Bartlett, Benchmark Studies on the Building Blocks of DNA. 1 Superiority of Coupled Cluster Methods in Describing the Excited States of Nucleobases in the Franck-Condon Region. *J. Phys. Chem. A,* 2012, **116**, 6702–6710.
2 P. G. Szalay, T. Watson, A. Perera, V. Lotrich, G. Fogarasi, and R. J. Bartlett, Benchmark Studies on the Building Blocks of DNA. 2. Effect of Biological Environment on the Electronic Excitation Spectrum of Nucleobases. *J. Phys. Chem. A* 2012, **116**, 8851–8860.
3 I. R. Gould, D. V. S. Green, P. Young and I. H. Hillier, A theoretical study using ab initio methods of tautomerism in

cytosine in the gas phase and in water. *J. Org. Chem.* 1992, **57**, 4434–4437.

4 A. Broo, and A. Holmén, A. Calculations and Characterization of the Electronic Spectra of DNA Bases Based on ab Initio MP2 Geometries of Different Tautomeric Forms. *J. Phys. Chem. A* 1997, **101**, 3589–3600.

5 A. L. Sobolewski and L. Adamowicz, Theoretical investigations of proton transfer reactions in a hydrogen bonded complex of cytosine with water. *J. Chem. Phys.* 1995, **102,** 5708–5718.

6 L. Gorb and J. Leszczynski, Intramolecular Proton Transfer in Monohydrated Tautomers of Cytosine: An Ab Initio PostHartree-Fock Study. *Int. J. Quant. Chem.* 1998, **70**, 855–862.

7 C. Alemán, Solvation of Cytosine and Thymine Using a Combined Discrete/SCRF Model. *Chem. Phys. Lett.* 1999, **302**, 461–470.

8 A. K. Chandra, M. T. Nguyen and T. Zeegers-Huyskens, Theoretical study of the protonation and deprotonation of cytosine. Implications for the interaction of cytosine with water. *J. Mol. Struct.* 2000, **519**, 1–11.

9 G. Fogarasi and P. G. Szalay, The Interaction Between Cytosine Tautomers and Water: an MP2 and Coupled Cluster Electron Correlation Study. *Chem. Phys. Lett.* 2002, **356**, 383–390.

10 D. Sivanesan, K. Babu, S. R. Gadre, V.Subramanian, and T. Ramasami, Does a Stacked DNA Base Pair Hydrate Better than a Hydrogen-Bonded One? An ab Initio Study. *J. Phys. Chem. A* 2000, 104, 10887–10894.

11 T. van Mourik, D. M. Benoit, S. L. Price and D. C. Clary, Ab initio and diffusion Monte Carlo study of uracil-water, thymine-water, cytosine-water, and cytosine-(water)2. *Phys. Chem. Chem. Phys.* 2000, **2** (6), 1281–1290.

12 S. A. Trygubenko, T. V. Bogdan, M. Rueda, M. Orozco, F. J. Luque, J. Sponer, P. Slavicek and P. Hobza, Correlated ab Initio Study of Nucleic Acid Bases and Their Tautomers in the Gas Phase, in a Microhydrated Environment and in Aqueous Solution. *Phys. Chem. Chem. Phys.* 2002, **4**, 4192–4203.

13 S. Kim, S. and H. F. Schaefer, H. F. Microhydration of Cytosine and its Radical Anion: Cytosine .(H_2O)n (n=1–5). *J. Chem. Phys.* 2007, **126,** 064301.

14 C. Alemán, Hydration of cytosine using combined discrete/SCRF models: influence of the number of discrete solvent molecules. *Chem. Phys.* 1999, **244**, 151–162.

15 O. V. Shishkin, L. Gorb and J. Leszczynski, Does the Hydrated Cytosine Molecule Retain the Canonical Structure? A DFT Study. *J. Phys. Chem. B* 2000, **104**, 5357–5361.

16 T. van Mourik, V. I. Danilov, E. Gonzalez, A. Deriabina and V. I. Poltev, Indication of Water Droplet Formation in Hydrated Clusters of Cytosine and Adenine: An Electronic Structure Study Using B3LYP, LMP2 and PM6. *Chem. Phys. Lett.* 2007, **445**, 303–308.

17 PQS Ab Initio Program Package, version 2.4, Parallel Quantum Solutions, 2013 Green Acres, Suite A, Fayetteville, Arkansas 72703, USA

18 CFOUR, a quantum chemical program package written by J. F. Stanton, J. Gauss, M. E. Harding, P. G. Szalay with contributions from A. A. Auer, R. J. Bartlett, U. Benedikt, C. Berger, D. E. Bernholdt, O. Christiansen, M. Heckert, O.Heun, C. Huber, D. Jonsson,J Juselius,K. Klein, W. J. Lauderdale, D. Matthews, T. Metzroth, D. P. O'Neill, D. R. Price, E. Prochnow, K. Ruud, F. Schiffmann, S. Stopkowicz, A. Tajti, M. E. Varner, J. Vazquez, F. Wang, J. D. Watts and the integral packages MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, and J. Olsen), and ECP routines by A. V. Mitin and C. van Wüllen. For the current version, see http://www.cfour.de (accessed Spring 2012).

19 The Cambridge Cluster Database (http://www-wales.ch.cam.ac.uk/~wales/CCD/TIP4P-water.html), based on: D. J. Wales and M. P. Hodges, *Chem. Phys. Lett.* 1998, **286**, 65–72.

20 S. Thicoipe, P. Carbonniere and C. Pouchan, The Use of the GSAM Approach for the Structural Investigation of Low-Lying Isomers of Molecular Clusters from Density-Functional-Theory-Based Potential Energy Surfaces: The Structures of Microhydrated Nucleic Acid Bases. *J. Phys. Chem. A* 2013, **117**, 7236–7245.

21 K. C. Hunter and S. D. Wetmore, Hydrogen-Bonding Between Cytosine and Water: Computational Evidence for a Ring-Opened Complex. *Chem. Phys. Lett.* 2006, **422**, 500–506.

22 P. Carbonniere, S. Thicoipe, T. Very, and X. Assfeld, Vibrational analysis beyond the harmonicity from Ab initio molecular dynamics: Case of cytosine in its anhydrous and aqueous forms. *Int. J. Quant. Chem.* 2012, **112**, 2221–2230.

23 S. F. Boys and F. Bernardi, The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* 1970, **19**, 553–566.

24 D. W. Schwenke and D. G. Truhlar, Systematic study of basis set superposition errors in the calculated interaction energy of 2 HF molecules. *J. Chem. Phys.* 1985, **82**, 2418–2426.

25 M. J. Frisch, J. E. DelBene, J. S. Binkley and H. F. Schaefer, Extensive theoretical-studies of the hydrogen-bonded complexes (H2O)2, (H2O)2H+, (HF)2, (HF)2H+, F2H-, and (NH3)2. *J. Chem. Phys.* 1986, **84**, 2279-2289.

26 L. M. Mentel and E. J. Baerends, Can the Counterpoise Correction for Basis Set Superposition Effect Be Justified? *J. Chem. Theor. Comp.* 2014, **10**, 252–267.

27 R. Kalescky, E. Kraka and D. Cremer, Accurate determination of the binding energy of the formic acid dimer: The importance of geometry relaxation. *J. Chem. Phys.* 2014, **140**, 084315.

28 E. Miliordos, and S. S. Xantheas, On the validity of the basis set superposition error and complete basis set limit extrapolations for the binding energy of the formic acid dimer. *J. Chem. Phys.* 2015, **142**, 094311.

29 A. Furmanchuk, O. Isayev, O. V. Shishkin, L. Gorb, and J. Leszczynski, Hydration of Nucleic Acid Bases: A Car–Parrinello Molecular Dynamics Approach. *Phys. Chem. Chem. Phys.* 2010, **12**, 3363–3375.

30 A. Furmanchuk, O. V. Shishkin, O. Isayev, L. Gorb, and J. Leszczynski, New Insight on Structural Properties of Hydrated Nucleic Acid Bases From ab Initio Molecular Dynamics. *Phys. Chem. Chem. Phys.* 2010, **12**, 9945–9954.

31 R. Car, M. Parrinello, Unified Approach for Molecular Dynamics and Density-Functional Theory. *Phys. Rev. Lett.* 1985, **55**, 2471–2474.

32 Wei, Z.; Chen, D.; Zhao, H.; Li, Y.; Zhu, J.; Liu, B**.** Ab initio investigation of the first hydration shell of protonated glycine *J. Chem. Phys.* **2014**, *140*, 085103**.**

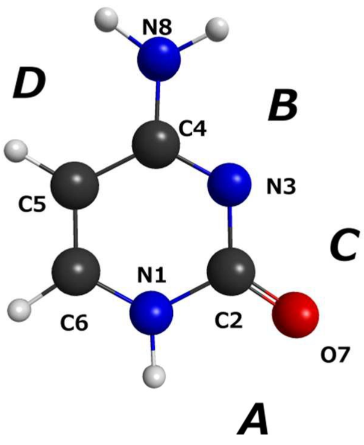

Fig. 1 The canonical amino-oxo form of cytosine (Cyt), with the moieties suitable for hydration indicated as *A, B, C, D.*

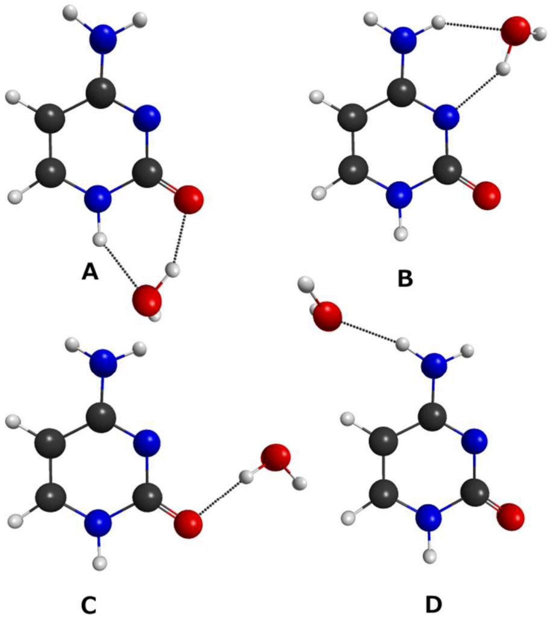

Fig. 2 The four monohydrates of cytosine, Cyt.H_2O

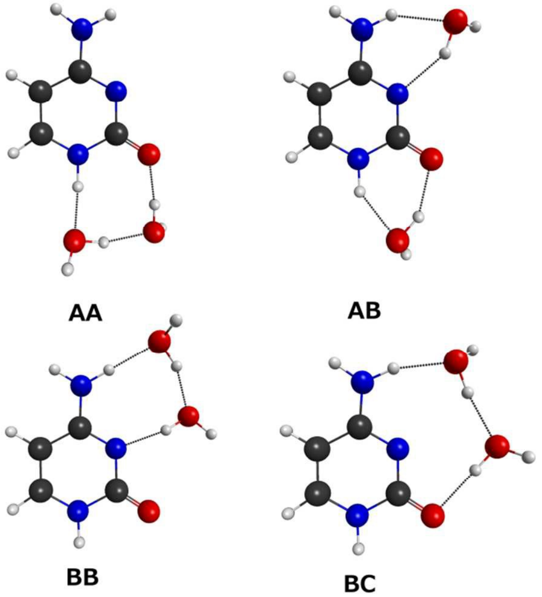

Fig. 3 Dihydrates of cytosine, Cyt.$2H_2O$

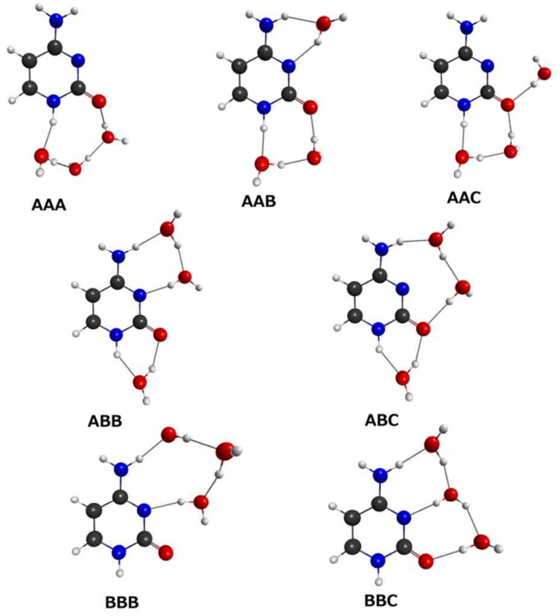

Fig. 4 Trihydrates of cytosine, Cyt.$3H_2O$

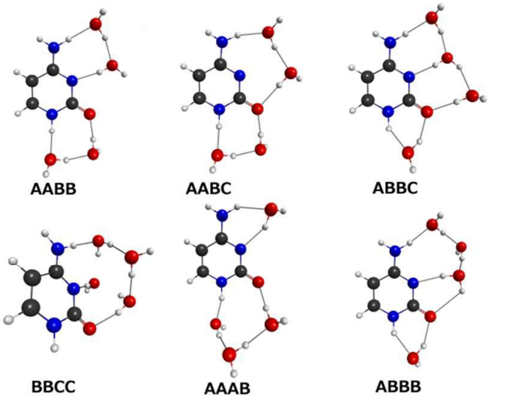

Fig. 5. Tetrahydrates of cytosine, Cyt.4H_2O

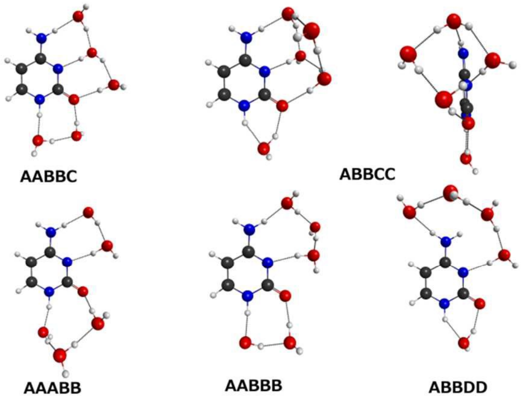

Fig. 6. Pentahydrates of cytosine, Cyt.5H_2O. For ABBCC, two views are given to show the 3D arrangement of water molecules.

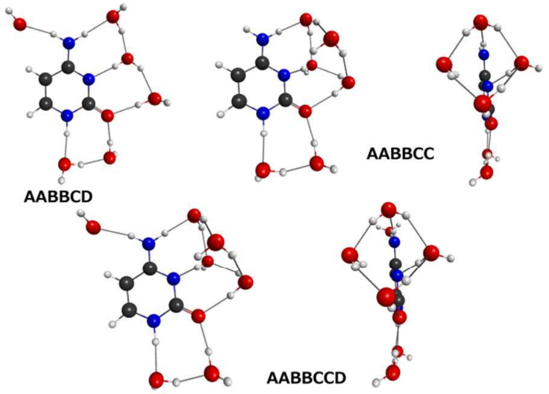

Fig. 7. Two hexahydrates and one heptahydrate of cytosine, Cyt.6H_2O and Cyt.7H_2O. For AABBCC and AABBCCD two views are given to show the 3D arrangement of water molecules.

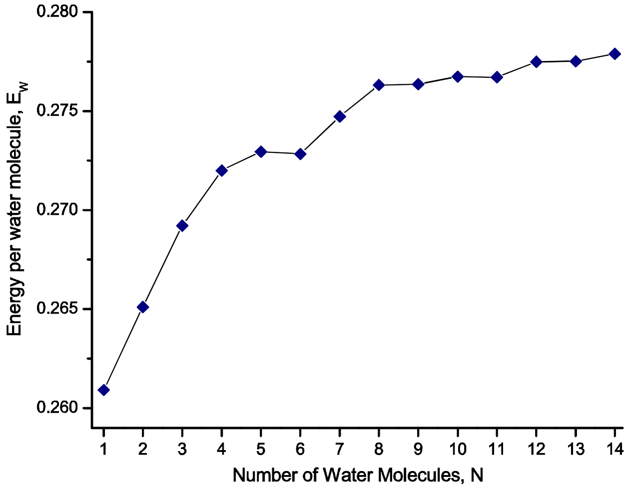

Fig. 8. Average energy, $E_W = E(W_N)/N$, of water molecules in a W_N cluster, as a function of N. MP2(fc)/aug-cc-pVDZ results. For a better overview, the graph shows $-(E_W + 76)/E_h$.

TOC

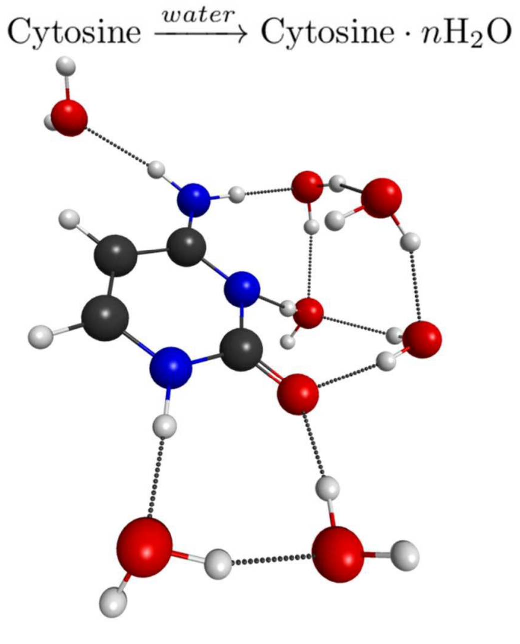

Table 1. Checking the Level of Theory: Energies of Some Cytosine Hydrates[a]. E – Total Energy in E_h[b], Δ – Energy Relative to the Most Stable Isomer in that Row in kcal.mol^{-1}, the Latter Given in Bold Letters.

Selected Dihydrates	Energies	AA	AB	BB	BC
Standard method of this study MP2(fc)/aug-cc-pVDZ	E + 546 Δ	–0.47293 **0**	-0.47221 **0.45**	-0.47108 **1.16**	-0.47023 **1.69**
Larger basis set MP2(fc)/aug-cc-pVTZ	E + 546 Δ	–0.9418 **0**	–0.9413 **0.32**	–0.9399 **1.19**	–0.9389 **1.82**
Coupled Cluster (in standard geom.) CCSD/aug-cc-pVDZ	E+ 546 Δ	–0.54038 **0**	-0.53955 **0.52**	-0.53867 **1.07**	-0.53813 **1.41**
Coupled Cluster (in standard geom.) CCSD(T)/aug-cc-pVDZ	E+ 546 Δ	–0.61129 **0**	-0.61046 **0.52**	-0.60941 **1.18**	-0.60875 **1.59**
DFT B3LYP/ aug-cc-pVDZ	E+ 547 Δ	-0.92150 **0**	-0.91966 **1.15**	-0.91913 **1.49**	-0.91740 **2.57**
DFT B3LYP/ aug-cc-pVTZ	E+ 548 Δ	-0.05872 **0**	-0.05681 **1.20**	-0.05638 **1.47**	-0.05473 **2.50**
Selected Trihydrates	Energies	AAA	AAB	ABB	ABC
Standard method of this study MP2(fc)/aug-cc-pVDZ	E+ 622 Δ	-0.74897 **2.49**	-0.75294 **0**	-0.75198 **0.60**	-0.75115 **1.12**
Coupled Cluster (in standard geom.) CCSD/aug-cc-pVDZ	E+ 622 Δ	-0.82562 **2.28**	-0.82926 **0**	-0.82834 **0.58**	-0.82780 **0.92**
Coupled Cluster (in standard geom.) CCSD(T)/aug-cc-pVDZ	E+ 622 Δ	-0.90262 **2.51**	-0.90662 **0**	-0.90574 **0.55**	-0.90512 **0.94**
DFT B3LYP/aug-cc-pVDZ	E+ 624 Δ	-0.38013 **1.49**	-0.38251 **0**	-0.38116 **0.85**	-0.37924 **2.05**
DFT B3LYP/aug-cc-pVTZ	E+ 624 Δ	-0.53860 **1.26**	-0.54061 **0**	-0.53917 **0.90**	-0.53738 **2.03**

[a] For notation of the complexes see text (Introduction). [b] E_h (hartree)/molecule ≅ 627.5 kcal/mol.

Table 2. Relative Energies of Cytosine Hydrates, MP2(fc)/aug-cc-pVDZ results.[a]
For comparison, some DFT results are listed in brackets, first: B3LYP/DZ++ [b], second: B3LYP-D/6-311+G(d,p)[c]

Cyt.H_2O -470.19228 (77.42)	**A** 0 (0)	B 0.72 (0.55) [0.6 (0.5); (0.1)]	C:1[d] 5.12 (4.30) [x; (1.0)]	D:1[d] 6.09 (5.07) [6.1(5.1); x]			
Cyt.$(H_2O)_2$ -546.47293 (93.27)	**AA** 0 (0)	AB 0.45 (0.18) [1.5 (1.2); (0.6)]	AC[e] x [2.8 (2.4); x]	BB 1.16 (0.98) [1.4 (1.2); (0.5)]	BC 1.69 (1.15) [x; (1.9)]		
Cyt.$(H_2O)_3$ -622.75294 (108.85)	AAA 2.49 (2.39) [x, (0.8)]	**AAB** 0 (0)	AAC 4.09 (3.46) [x; x]	ABB 0.60 (0.55) [1.0 (0.9); (0.4)]	ABC 1.12 (0.76) [2.4 (2.1); (2.7)]	BBB 2.65 (2.92) [x, x]	BBC 1.61 (2.07) [1.1 (1.7); 1.2]
Cyt.$(H_2O)_4$ -699. 03272 (124.69)	AAAB 2.52 (2.26) [x; 3.0]	**AABB** 0 (0)	AABC 0.78 (0.38) [2.1 (1.6); x]	ABBB 2.09 (2.12) [x, x]	ABBC 1.32 (1.59) [2.2 (2.5); (1.8)]	BBCC 0.73 (1.65) [x, x]	
Cyt.$[H_2O)_5$ -775.31209 (140.88)	AABBC 0.31 (0.26) [0 (0); (0)]	AAABB 2.20 (1.58) [x; (2.0)]	AABBB 1.31(1.07) [x; (0.9)[f,g]]	**ABBCC** 0 (0) [x; (0.6)[g]]	ABBDD 2.80 (2.18) [x; x]		
Cyt.$(H_2O)_6$ -851. 59058 (156.97)	AABBCD 4.18 (3.08) [--]	**AABBCC** 0 (0) [--]					
Cyt.$(H_2O)_7$ -927.86368 (172.08)	AABBCCD 0 (0) [--]						

[a]The total energy of the most stable species in a row, denoted by bold letters, is given in the first column in E_h; below it in parentheses the zero point vibrational energy is quoted in kcal/mol. In the rest of the table each cell refers to an individual isomer indicated by its symbol. Their relative energies are given in kcal/mol with respect to the most stable species. In parentheses the same data are listed after ZPE correction. In the literature values – third row of the cells – ref. 13 has quoted both uncorrected and ZPE-corrected figures, while ref. 20 listed only the latter. At this same level, the energies of isolated cytosine and water are: –393.91102 E_h and –76.26091 E_h, respectively.
[b]ref. 13. [c]ref. 20; *x* indicates that the complex was not reported. [d] C:1 and D:1 indicate that water binds with *one* H-bond only.
[e]All attempts have led to structure AB. [f]ref. 20 reports *two* structures here, see text. [g]These literature values refer to *their* lowest-energy complex, AABBC.

Table 3. Derivation of Hydration Energies for Cytosine Hydrates, MP2(fc)/aug-cc-pVDZ Results[a].

n; Bold: symbol of isomer[b]	Cyt.nH_2O[b]	$(H_2O)_n$[c]	nH_2O → $(H_2O)_n$[d]	Cyt + nH_2O → Cyt.nH_2O[e]	Cyt + $(H_2O)_n$ → Cyt.nH_2O[f]	Cyt + nW_b → Cyt.nH_2O[g]
0	-393.91102					
1 **A**	-470.19228	-76.26091	n/a	–12.77	–12.77 (–10.85)	–0.86
2 **AA**	-546.47293	-152.53021	–5.26	–25.16	–19.89 (–17.09)	–1.79
3 **AAB**	-622.75294	-228.80763	–15.63	–37.14	–21.52 (–16.86)	–1.88
4 **AABB**	-699.03272	-305.08800	–27.84	–48.98	–21.15 (–15.55)	–2.22
5 **ABBCC**	-775.31209	-381.36473	–37.76	–60.57	–22.50 (–16.01)	–2.21
6 **AABBCC**	-851.59058	-457.64109	–47.46	–71.60	–24.14 (–16.11)	–1.46
7 **AABBCCD**	-927.86368	-533.92343	–60.91	–79.23	–18.33 (–8.83)	+2.42

[a] Energies in E_h (hartree)/molecule ≡ 627.5 kcal/mol, energy differences in kcal/mol.

[b] The lowest energy complex from Table 2.

[c] Optimized water cluster.

[d] Stabilization energy of the optimized water cluster relative to *n* independent water molecules; no correction for BSSE at this point.

[e] Energy of the optimized cytosine-water complex relative to free cytosine plus *n* water molecules, no BSSE correction at this point.

[f] Energy of the optimized cytosine-water complex relative to free cytosine and the relevant water *cluster*; in parentheses: data after correction for BSSE. The latter CPC values for the relevant *reaction*, as obtained by the standard Boys-Bernardi method, are 1.92, 2.80, 4.66, 5.60, 6.49 8.03 and 9.50 kcal/mol, for *n* = 1 to 7, respectively.

[g] W_b denotes a fictitious water molecule bound in a huge water cluster; its energy relative to an isolated water molecule was estimated as 11.04 kcal/mol, see text. BSSE applies here for Cyt only; the CPC values that *reduce* the energy of Cyt in the equation, are 0.87, 1.29, 2.15, 2.61, 3.17, 3.91 and 4.39 kcal/mol for n = 1 to 7, respectively.

Table 4. Characteristic Bond Lengths (Å) of Cytosine and their Changes upon Hydration. MP2(fc)/aug-cc-pVDZ Results, with Some Comparative DFT Data[a]

Species[b]	N1-C2	C2=O7	N3-C4	C4-N8
Cyt	1.422	1.231	1.327	1.372
A	–0.011	+0.014	+0.003	–0.001
B	–0.004	+0.002	+0.009	–0.015
C	–0.009	+0.008	+0.002	–0.007
D	–0.000	+0.002	+0.004	–0.011
AA	–0.018	+0.016	+0.004	–0.000
AAB	–0.020	+0.017	+0.013	–0.015
AABB	–0.024	+0.020	+0.015	–0.021
AABBC	–0.030	+0.030	+0.019	–0.024
AABBCC	–0.030	+0.026	+0.025	–0.027
AABBCCD	–0.031	+0.028	+0.028	–0.031
Cyt in Bulk Water, DFT[c]	–0.053	+0.036	+0.028	–0.030
Cyt, DFT[d]	1.426	1.223	1.323	1.360
AABBCCD DFT[d]	–0.033	+0.028	+0.029	–0.027

[a] Bond lengths given for Cyt, and their changes listed for the other species. [b]Cyt denotes cytosine (amino-keto form), A, B, C and D and their combinations refer to various hydrates, see Fig. 1 and text. [c]Average from Car-Parrinello dynamics calculations, with B3LYP functional and a plane wave basis, Fig. 2. of ref. 28. [d]This work, B3LYP/aug-cc-pVDZ.