PCCP

The nucleobase cytosine and the cytosine dimer investigated by double resonance laser spectroscopy and *ab initio* calculations

E. Nir,[a] I. Hünig,[b] K. Kleinermanns[b] and M. S. de Vries[c]

[a] *Department of Chemistry, The Hebrew University, Jerusalem 91904, Israel*
[b] *Institut für Physikalische Chemie und Elektrochemie I, Heinrich Heine Universität Düsseldorf, 40225, Düsseldorf, Germany*
[c] *Dept. of Chemistry and Biochemistry, Santa Barbara, California 93106, USA*

Received 27th August 2003, Accepted 29th August 2003
First published as an Advance Article on the web 24th September 2003

The vibronic spectrum of laser desorbed and jet cooled cytosine consists of bands from two major tautomers (keto and enol) as revealed by UV-UV and IR-UV double resonance spectroscopy and methyl blocking experiments. Only one isomer each was observed for the cytosine dimer and for the cytosine - 1-methylcytosine mixed dimer. These isomers form C=O···HNH/NH···N hydrogen bonds. Cytosine - 5-methylcytosine exhibits three isomers: one again with C=O···HNH/NH···N connectivity, the second with C=O···HNH/NH···N interaction but one cytosine in the enol form and the third with symmetrical C=O···NH/NH···O=C bonds. These are also the most stable clusters according to molecular dynamics/quenching and *ab initio* quantum chemical calculations. The experimental IR spectra of these isomers agree well with the calculated normal mode vibrational spectra. The vibronic spectra of the clusters are blue shifted relative to the monomer spectra by more than 1000 cm^{-1} indicating a considerable reduction of dimer stability upon electronic excitation.

I. Introduction

The relative stability of tautomers of the nucleobases is important for the structure and functioning of DNA. The occurence of certain tautomers has been suggested as a possible mechanism of spontaneous mutation.[1] Numerous calculations have been reported of the lowest energy tautomers of cytosine[2–5] and its methyl derivatives,[6] both as isolated molecules and interacting with water molecules. Matrix isolation studies have been reported as well.[7,8] Recently we reported resonant two-photon ionization spectra of laser desorbed, jet cooled cytosine and several of its methylated derivatives.[9] Unlike other pyrimidine bases, cytosine exhibits vibronic spectra with sharp features. In this work we show by methyl blocking and IR-UV double resonance experiments that the spectra arise from one keto tautomer and one enol tautomer. This agrees with microwave spectroscopy investigations of cytosine[10] where two major tautomers, the keto and the enol *cis* form, could be identified by comparison of the mesured rotational constants and dipole moment components with results from *ab initio* calculations. A third imino tautomer showed approximately one quarter of the abundance of the other two tautomers.[10]

The different cytosine tautomers and their methylated derivatives can form dimers in the jet, for which we have reported first spectroscopic data.[11] These dimers are investigated in further detail in this work. We compare our experimental results with molecular dynamics/quenching calculations and correlated *ab initio* quantum chemical[12] and normal mode vibrational calculations to identify the structures of these dimers and to explore their potential energy surface.

II. Experimental and theoretical methods

The measurements with laser desorbed nucleobases were performed with an apparatus described in detail elsewhere.[13] In short, material is laser desorbed from a graphite sample in front of a pulsed nozzle. A typical fluence of the Nd:YAG desorption laser, operated at 1064 nm (where graphite absorbs but cytosine does not) is about 1 mJ cm^{-2} or less, which is considerably lower than the fluences normally used for ablation. The laser is focused to a spot of about 0.5 mm diameter within 2 mm in front of the nozzle. We used a pulsed valve (General Valve; Iota One) with a nozzle diameter of 1 mm at a backing pressure of 5 atm argon drive gas. The skimmed molecular beam crosses the ionization laser at right angle inside the source region of a reflectron time-of-flight (TOF) mass spectrometer. By monitoring the parent mass peaks while varying the two photon, two color ionization wavelength (resonant two photon ionization: R2PI), we obtained mass selected excitation spectra. The first photon came from a frequency doubled dye laser and the second photon from an ArF excimer laser at 193 nm.

The double resonance experiments were performed with the same set up. We performed spectral hole burning (SHB) by using two counterpropagating dye laser pulses with a delay of about 150 ns. This results in two peaks in the TOF spectrum–the first from the "burn" laser and the second from the "probe" laser. When both lasers are tuned to the resonance of the same tautomer the burn laser causes a decrease in the signal of the probe laser. Generally, we scan the burn laser while the probe laser frequency is fixed to an intense band of one tautomer. If a significant band of the R2PI spectrum is missing in the burn spectrum it belongs to another tautomer or to a hot band. In the next step we probe at this frequency while scanning the pump laser to reveal the spectrum of the next tautomer.

We performed IR-UV SHB with the same method but using a difference frequency IR laser as the burn laser. The radiation from an infrared dye (a mixture of Styryl 8 and Styryl 9) was aligned collinearly with the perpendicularly polarized Nd-YAG fundamental (1064 nm) and directed through

 DOI: 10.1039/b310396j

a MgO-doped $LiNbO_3$ crystal to generate 3300–4000 cm^{-1} tunable IR light. Suitable dielectric mirrors separate the Nd-YAG fundamental and the dye laser beam behind the crystal. We typically use 50 mJ of the YAG fundamental and 10 mJ of the dye laser to obtain around 1 mJ $pulse^{-1}$ IR radiation between 3400 and 4000 cm^{-1} with a bandwidth of <0.1 cm^{-1}. The IR laser was calibrated by recording a water vapor spectrum. Color centers in the $LiNbO_3$ crystal lead to a decrease of the IR intensity from 3515 to 3550 cm^{-1}. In that spectral range we used another $LiNbO_3$ crystal with an intensity gap in a different region. Cytosine and the methylated cytosine derivatives were obtained from Sigma-Aldrich and used without further purification.

The calculations were carried out using the Gaussian 98 program package.[14] We performed calculations at the HF 6-31 G(d,p) level for the clusters. All structures were fully optimized with 10^{-8} E_h as the SCF convergence criterion and 1.5×10^{-5} $E_h\ a_0^{-1}$ and $E_h\ (°)^{-1}$, respectively, as the convergence criteria for the gradient optimization. The vibrational frequencies were obtained by performing a normal mode analysis on the optimized geometries using analytical gradients of the energy.

The stabilization energies D_e were corrected for zero point energy (ZPE) using the harmonic frequencies. The dissociation energies D_0 of the clusters were corrected for basis set superposition error (BSSE).

III. Results and discussion

Fig. 1 shows the structures of the most stable cytosine–cytosine (CC) isomers calculated at the HF/6-31G(d,p) level and the dissociation energies (including BSSE and ZPE correction). K and E denote a keto or enol tautomer of cytosine (C), *cis*/*trans* denotes the conformation of the OH group with respect to the N_1–C_2 bond and –1, –2… associates clusters with the same H bond arrangement and orders them according to their stability. Cluster structures with symmetrical C=O···NH/NH···O=C bonding (CK–CK–1 and for the enol tautomer CE*cis*–CE*cis*–1), C=O···HNH/NH···N interaction (CK–CK–2 and CK–CE*cis*–2) and HNH···N/N···HNH bonding (CK–CK–3, CK–CE*cis*–3 and CE*cis*–CE*cis*–3) are especially stable: see also Table 1 and Table 2.

Unlike most cytosine cluster calculations in the literature we have calculated dimer structures resulting from both the keto and the enol tautomer. The question arises whether both cytosine tautomers are abundant in the jet. Vibronic spectroscopy and UV–UV SHB can answer this question. Cytosine, unlike other pyrimidine bases, exhibits vibronic spectra with sharp features in two spectral regions, separated by about 4000 cm^{-1}, *cf.* Fig. 2 and ref. 9. 5-Methylcytosine (C5m) exhibits spectra in both these spectral regions while 1-methylcytosine (C1m), for which the enol form is blocked, absorbs only in

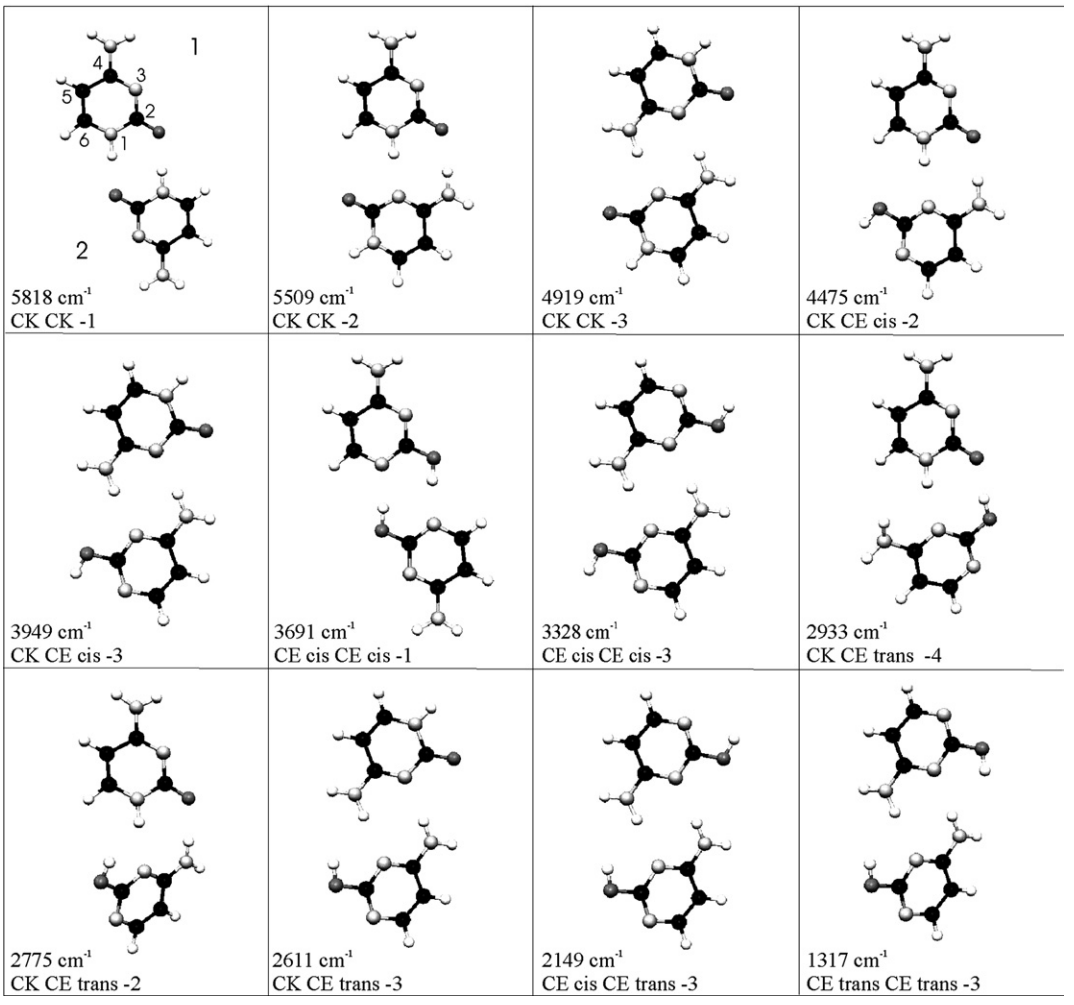

Fig. 1 Structures and stabilities of the most stable cytosine–cytosine (CC) isomers calculated at the HF/6-31G(d,p) level. K and E denote a keto or enol tautomer of cytosine (C), *cis*/*trans* denotes the conformation of the OH group with respect to the N_1–C_2 bond and the indices –1,–2… associate clusters with the same H bond arrangement and orders these cluster families according to their stability. The numbers 1 and 2 designate the upper and lower monomer in the dimer figures to simplify discussion of dimer structures in the text. For example CKCK–1 labels the keto–keto dimer with C=O–NH/NH–O=C bonding which is the most stable cluster in this family of H bond arrangements (and the most stable one overall) at the HF/6-31G(d,p) level. The numbers in the figures are the cluster dissociation energies (in cm^{-1}) including BSSE correction and ZPE.

Table 1 Dissociation energies D_0 (in cm^{-1}) of the most stable cytosine dimer structures calculated at the HF/6-31G(d,p) level. D_0 is obtained by correcting the electronic dissociation energy D_e (with BSSE correction) for the zero point vibrational energy (ZPE): D_e–ZPE=D_0. The labels correspond to those in the figures

Structure	D_e	ZPE	D_0
CK–CK–1	6389	571	5818
CK–CK–2	6307	797	5510
CK–CK–3	5575	656	4919
CK–CE*cis*–2	5080	605	4475
CK–CE*cis*–3	4639	690	3949
CE*cis*–CE*cis*–1	4243	552	3691
CE*cis*–CE*cis*–3	3851	523	3328
CK–CE*trans*–4	3512	579	2933
CK–CE*trans*–2	3291	516	2775
CK–CE*trans*–3	3202	591	2611
CE*cis*–CE*trans*–3	2574	425	2149
CE*trans*–CE*trans*–3	1686	369	1317

Table 2 Dissociation energies D_0 (in cm^{-1}) of the most stable C–C5m structures calculated at the HF/6-31G(d,p) level. D_0 is obtained by correcting the electronic dissociation energy D_e (BSSE correction included) for the zero point vibrational energy (ZPE): D_e–ZPE=D_0. The labels correspond to those in the figures

Structure	D_e	ZPE	D_0
CK–C5mK–1	6367	508	5859
CK–C5mK–2.1	6384	724	5660
CK–C5mK–2.2	6286	727	5559
CK–C5mK–3	5625	772	4853
CK–C5mE*cis*–2	5111	590	4521
CE*cis*–C5mK–2	5042	540	4502
CE*cis*–C5mK–3	4677	618	4059
CK–C5mE*cis*–3	4650	674	3976
CE*cis*–C5mE*cis*–1	4234	381	3853
CE*cis*–CE*cis*–3	3862	506	3356
CE*trans*–C5mK–2	3314	444	2870
CK–C5mE*trans*–2	3304	504	2800
CE*trans*–C5mK–3	3241	516	2725
CK–C5mE*trans*–3	3202	582	2620

the region around 32 000 cm^{-1}. Therefore we ascribe the vibronic spectrum around 32 000 cm^{-1} to the keto tautomer of C and the spectrum around 36 000 cm^{-1} to the enol tautomer of C. This view is supported by the IR–UV spectrum of the C tautomer absorbing around 32 000 cm^{-1} which is displayed in Fig. 4. The spectrum shows the antisymmetric and symmetric NH_2 and the N1H stretching vibrations at their typical frequencies but does not contain the OH vibration. The C5m and C1m tautomers absorbing around 32 000 cm^{-1} exhibit a very similar IR spectrum as C(keto), merely the N1H vibration is missing in C1m. Hence both the keto and the enol tautomer are present in the jet and we may expect dimers from both tautomers.

The cytosine dimers C–C and C–C1m both absorb around 33 500 cm^{-1} and their UV–UV SHB spectra each exhibit only one conformer in this spectral region, as shown in Fig. 3. In contrast the vibronic spectra of C–C5m and C5m–C5m, absorbing around 32 500 cm^{-1}, show contributions from three isomers, based on the UV–UV SHB measurements in Fig. 3. We designate these as "red", "middle" and "blue", according to the spectral region where they absorb within the REMPI spectrum. As we will see later most of these dimers consist of keto moieties and hence absorb at least 1000 cm^{-1} blueshifted relative to their monomers, see Table 3. This implies that these dimers are considerably less stable in the electronically excited state than in the ground state.

The IR spectra of the different dimers are shown in Fig. 4. The inset displays the three most stable dimer structures according to molecular dynamics/quenching calculations with an empirical force field and correlated *ab initio* quantum chemical calculations from the Hobza group[12] in order of decreasing population.[12] Cluster structures with C=O–HNH/NH–N, C=O–NH/NH–O=C, and HNH–N/N–HNH bonding are the most stable clusters according to our calculations as well, *cf.* Fig. 1. With these three main types of interaction we can easily assign the cluster structures based on their H bond connectivities.

The CC IR spectrum shows the same frequencies as the C(keto) spectrum with an additional band at 3525 cm^{-1}. This band lies in the typical frequency range of an antisymmetric NH_2 vibration involved in a hydrogen bond with a C=O group.[15] Hence the CC IR spectrum corresponds to one free and one bound NH_2(a) vibration which only cluster 1 in the inset of Fig. 4 exhibits. The 1-methylcytosine–cytosine dimer shows the same spectrum as CC, except that the free N1H vibration is missing. There is only one free N1H group in cluster 1 which is blocked by methylation in C–C1m. The hydrogen bound N1H and the bound symmetric NH_2 stretching vibration are presumably shifted to lower frequencies outside of the investigated spectral range.

C–C5m "red" shows the same IR spectrum as CC except that the N1H vibration is missing while there is an additional free OH vibration, as indicated by the arrows in Fig. 4. This agrees with structure 1 with one C in the enol form (the one

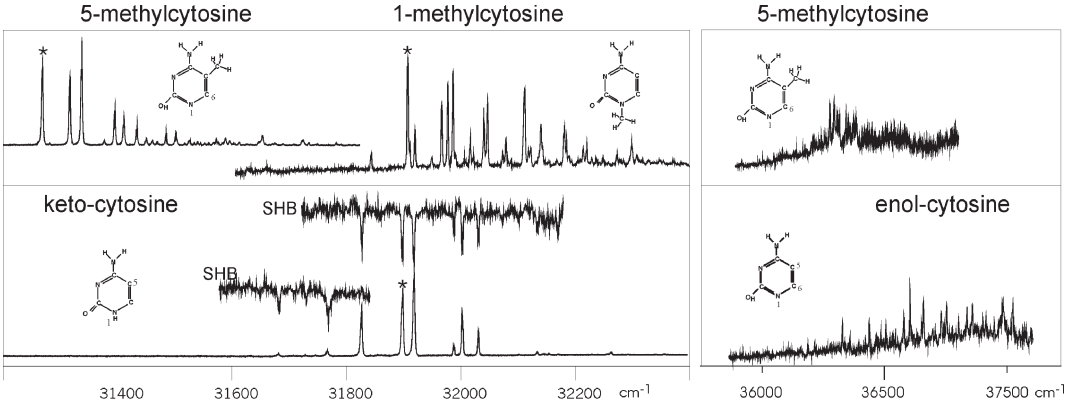

Fig. 2 R2PI and UV–UV hole burning (SHB) spectra of the keto and enol tautomer of cytosine (C). 5-methylcytosine (C5m) exhibits spectra in both spectral regions while 1-methylcytosine (C1m) for which the enol form is blocked, absorbs only in the region around 32 000 cm^{-1} (keto tautomer). The SHB spectra of the keto tautomer show that the weak peaks at <31 800 cm^{-1} belong to a different ground state population (hot band, isomeric form or dissociating cluster). The enol tautomer can have a *cis* or *trans* conformation of the OH group.

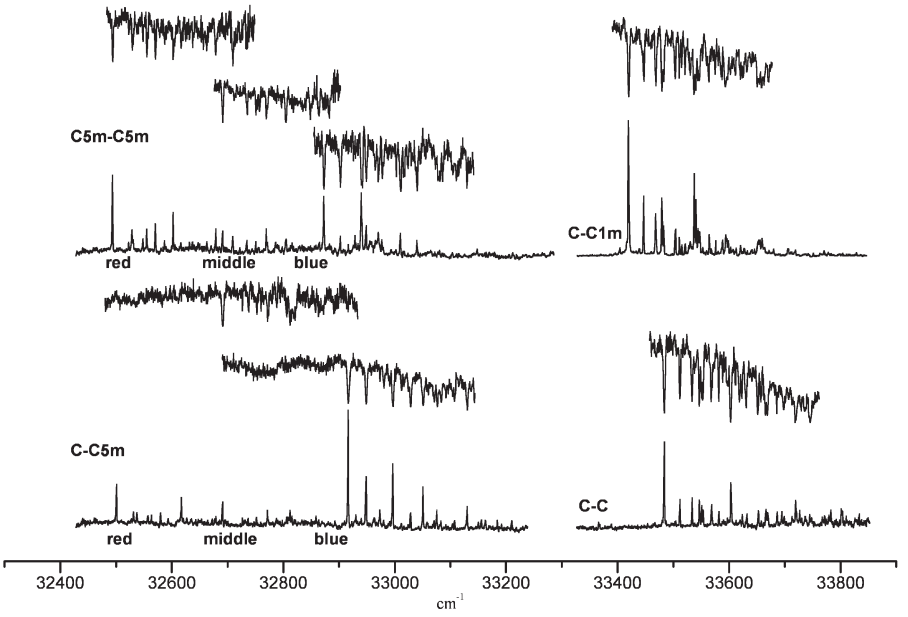

Fig. 3 R2PI and UV–UV hole burning (SHB) spectra of the cytosine dimer C–C and the methylated dimers C–C1m, C–C5m and C5m–C5m. SHB shows that the C–C and C–C1m spectra can be traced back to 1 isomer each while C–C5m and C5m–C5m exhibit 3 conformers in the investigated spectral range which absorb in the ''red'', ''middle'' and ''blue'' part of each R2PI spectrum.

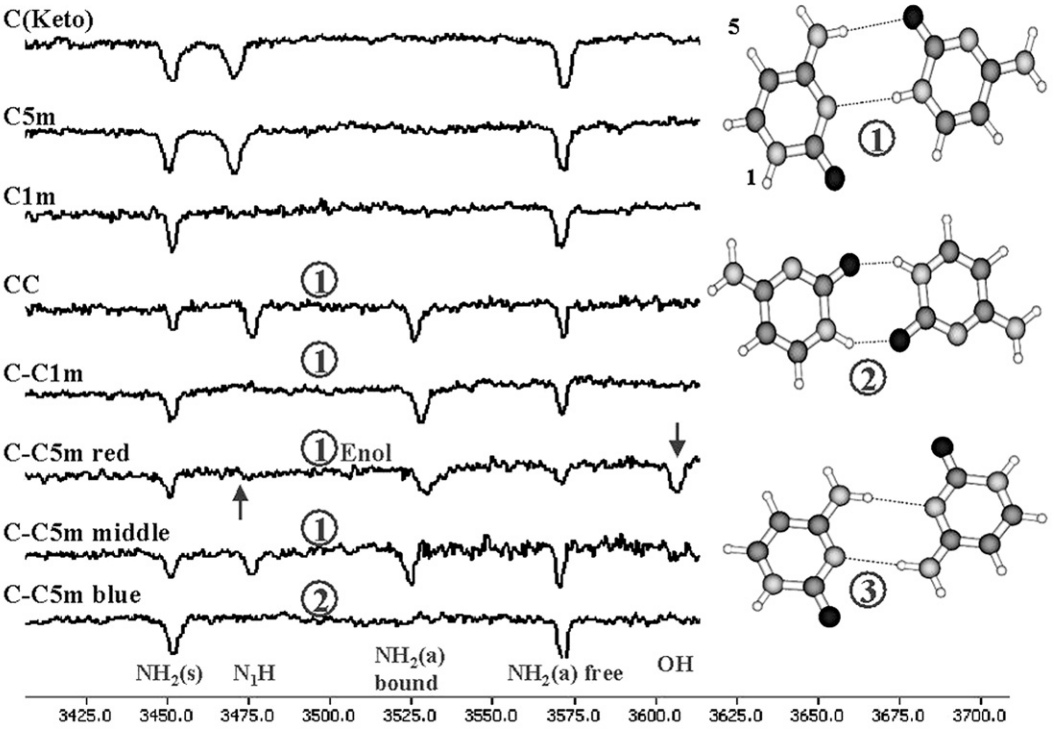

Fig. 4 IR-UV spectra of cytosine (keto tautomer; see text), 5-methylcytosine (C5m), 1-methylcytosine (C1m) and their dimers C–C, C–C1m and C–C5m. The inset displays the 3 most stable dimer structures according to molecular dynamics/quenching calculations.[12] 1,2,3 indicate the assignment of the IR spectra to specific cluster structures, see text. NH_2(s), N1H, NH_2(a) bound, NH_2(a) free and OH denote the symmetric NH_2, N1–H, bound (hydrogen bound) antisymmetric NH_2, free (not hydrogen bound) antisymmetric NH_2 and O–H stretching vibrations.

with N1H not involved in a hydrogen bond). C–C5m ''middle'' shows exactly the same IR spectrum as CC. Hence this isomer can be assigned to structure 1 with one of the C methylated in position 5. C–C5m ''blue'' shows no free OH, free N1H and bound antisymmetric NH_2 vibrations in the investigated IR range in agreement with the symmetrical structure 2. Here either the free NH_2 vibrations are coupled and only one component of them is sufficiently infrared active, or these are local vibrations for which the frequencies nearly coincide so that only one NH2(a) and NH2(s) band is observed. The N1H vibrations are involved in strong symmetrical hydrogen bonds and probably shifted below 3400 cm^{-1}. Hence the two most stable dimers are observed in excellent agreement with the calculations.[12]

Fig. 5 supports these considerations in a more quantitative sense. Shown are the IR-UV spectrum of the cytosine dimer C–C and the vibrational spectra and dissociation energies D_0 of the most stable C–C conformers calculated at the HF

Table 3 Electronic origins for the redmost bands of the cytosine and methyl substituted cytosine tautomers and their dimers. The spectral blue shifts upon dimerization are also given

	Origin bands/cm^{-1}
Cytosine	31 826
1-Methylcytosine	31 908
5-Methylcytosine	31 269
C–C	33 483
C–1MC	33 419
C–5MC	32 500, 32 691, 32 916
5MC–5MC	32 493, 32 691, 32 872

6-31G(d,p) level. Experimental and theoretical spectra agree only for isomer CK–CK–2 which is isomer 1 in Fig. 4. This is also the most populated cluster according to the molecular dynamics/quenching calculations.[12] None of the other spectra match nearly as well.

Fig. 6 further strengthens our arguments. The IR spectrum of C–C5m "red" agrees only with the calculated spectra of CK–C5mE*cis*–2 and C5mK–CE*cis*–2 which correspond to conformer 1 in Fig. 4 with one of the cytosines in the enol form. It is hard to decide from the spectra where the methyl group is located but the spectrum with CH_3 in position 5 of the keto form (C5mK–CE *cis*–2) agrees somewhat better. The IR spectrum of C–C5m "middle" agrees quite well with the spectrum from CK–C5mK–2 and sligthly better with the spectrum from C5mK–CK–2. Again the other spectra do not agree nearly as well. The IR spectrum of C–C5m "blue" looks similar to the spectrum of CK–C5mK–1. No other calculated

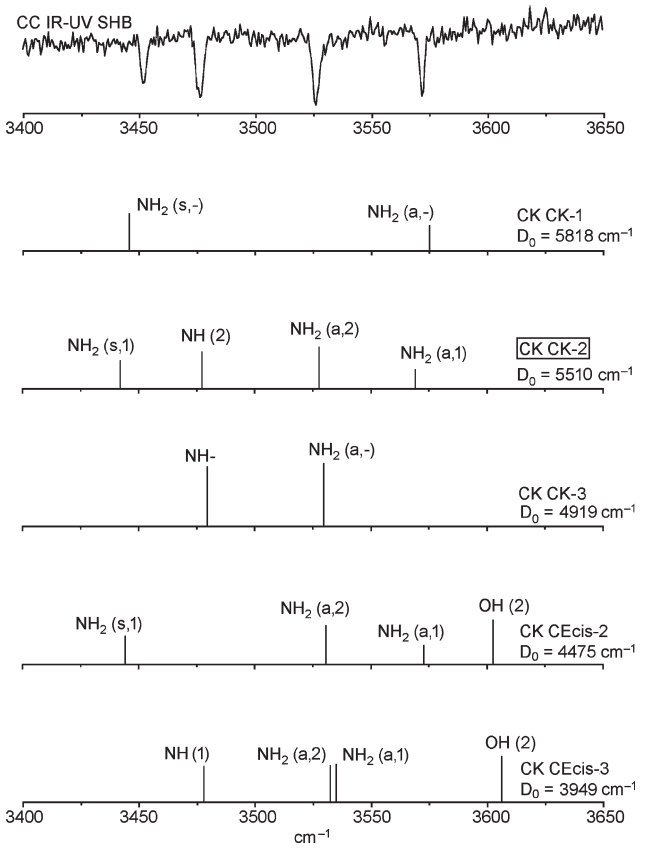

Fig. 5 IR–UV spectrum of the cytosine dimer C–C and vibrational spectra and dissociation energies D_0 of the most stable C–C isomers calculated at the HF 6-31G(d,p) level. The labeling scheme is the same as in Fig. 1. The cluster vibrations are designated as follows: s = symmetric; a = antisymmetric; 1,2 = moiety involved in the local vibration, moiety 1 (2) is the upper (lower) molecule in Fig. 1; – marks coupled vibrations of opposite phase.

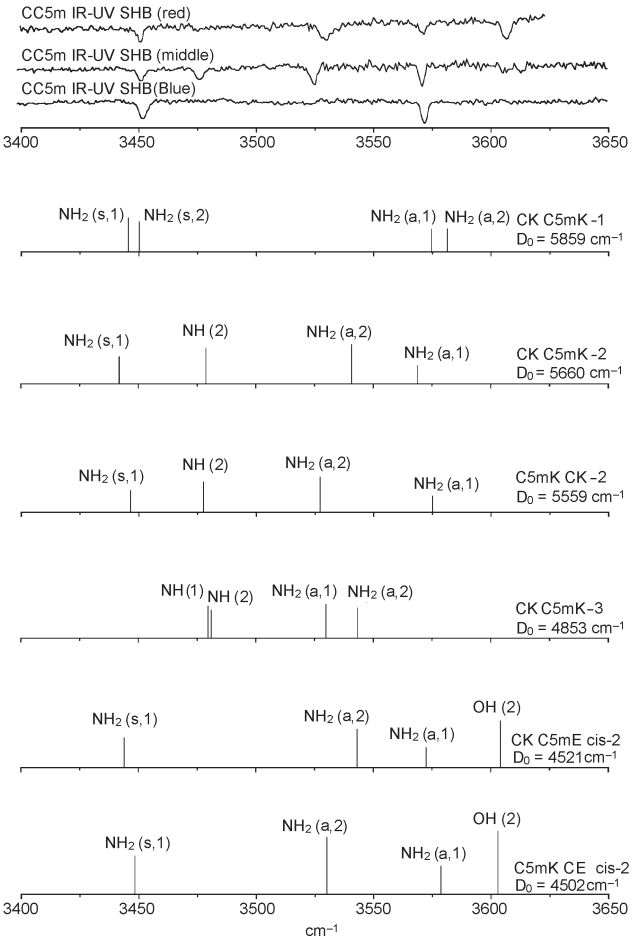

Fig. 6 IR–UV spectra of the three observed C–C5m conformers and vibrational spectra and dissociation energies D_0 of the most stable C–C5m conformers calculated at the HF 6-31 G(d,p) level. The labeling scheme is the same as in Fig. 1. The cluster vibrations are described as follows: s = symmetric; a = antisymmetric and 1,2 = upper, lower cluster moiety involved in the local vibration.

spectrum agrees with the experimental spectrum. It seems however that the calculations underestimate the coupling of the NH_2 vibratons in this mixed dimer or their similarity in frequency. The slight splitting of the calculated NH_2 vibrations is not reflected in the experimental pattern which resembles the calculated spectrum of CK–CK–1 in Fig. 5 more closely.

Altogether the agreement between experimental and calculated IR spectra is excellent and allows an unambiguous assignment of cluster connectivities. C5m–C5m shows a vibrational pattern very similar to C–C5m and the spectra are therefore not shown here. Ostensibly the same cluster structures are formed in C–C5m and C5m–C5m. We do not observe the spectrum of the symmetrical structure even though that is the most stable one, according to the calculations. This is analogous to our observations for GG.[15] A possible explanation is that this structure produces a hydrogen bond shift and an exciton splitting of such magnitude that the allowed component absorbs outside the investigated wavelength range. Another explanation is a photoinduced process which is especially fast for the symmetrical dimer. These explanations are consistent with the fact that we do observe this structure for 5-methylated dimers, because in that case the symmetry is broken.

Acknowledgements

This work is supported by the Deutsche Forschungsgemeinschaft Kl 531/20-2 and the United States-Israel Binational Science Foundation.

References

1 B. Pullmann and A. Pullmann, *Adv. Heterocycl. Chem.*, 1971, **13**, 77.
2 M. Hanus, F. Ryjacek, M. Kabelac, T. Kubar, T. V. Bogdan, S. A. Trygubenko and P. Hobza, *J. Am. Chem. Soc.*, 2003, **125**, 7678.
3 S. A. Trygubenko, T. V. Bogdan, M. Rueda, M. Orozco, F. J. Luque, Jiri Sponer, P. Slavíek and P. Hobza, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4192.
4 C. Colominas, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 1996, **118**, 6811.
5 J. R. Sambrano, A. R. de Souza, J. J. Queralt and J. Andres, *Chem. Phys. Lett.*, 2000, **317**, 437.
6 J. R. Sambrano, A. R. de Souza, J. J. Queralt, M. Oliva and J. Andres, *Chem. Phys.*, 2001, **264**, 333.
7 J. Smets, A. Destexhe, L. Adamowicz and G. Maes, *J. Phys. Chem. B*, 1997, **101**, 6583.
8 A. Les and L. Adamowicz, *J. Mol. Struct.*, 1990, **221**, 209.
9 E. Nir, M. Muller, L. I. Grace and M. S. de Vries, *Chem. Phys. Lett.*, 2002, **355**, 59.
10 R. D. Brown, P. D. Godfrey, D. McNaughton and A. Pierlot, *J. Am. Chem. Soc.*, 1989, **111**, 2308.
11 E. Nir, C. Pluetzer, K. Kleinermanns and M. de Vries, *Eur. Phys. J. D*, 2002, **20**, 317.
12 M. Kabelac and P. Hobza, *J. Phys. Chem. B*, 2001, **105**, 5804.
13 G. Meijer, M. S. de Vries, H. E. Hunziker and H. R. Wendt, *Appl. Phys. B*, 1990, **51**, 395.
14 M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez and J. A. Pople, *Gaussian 94, Revision A.4*, Gaussian, Inc., Pittsburgh PA, 1995.
15 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2002, **4**, 740.