Uracil Dimer: Potential Energy and Free Energy Surfaces. Ab Initio beyond Hartree−Fock and Empirical Potential Studies

Martin Kratochvíl, Ola Engkvist, Jiří Šponer, Pavel Jungwirth, and Pavel Hobza*

J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, 182 23 Prague, Czech Republic

Received: March 26, 1998; In Final Form: June 15, 1998

The first complete theoretical analysis of the gas-phase formation of a nucleic acid base pair (uracil dimer) has been performed. The study is based on a combination of AMBER 4.1 empirical potential, correlated ab initio quantum chemical methods, computer simulations, and statistical thermodynamical methods. In total, 11 low-energy minima structures were located on the potential energy surface of the uracil dimer: seven of them are H-bonded, one is T-shaped, and three correspond to various stacked arrangements. The most stable structure is a H-bonded dimer with two $N_1-H\cdots O_2$ H-bonds, designated as HB4; it has an energy minimum of −15.9 kcal/mol at the MP2/6-31G*(0.25)//HF/6-31G** level of theory. T-shaped structure and stacked structures are less stable than H-bonded ones. Thermodynamic characteristics were obtained using the rigid rotor−harmonic oscillator−ideal gas (RR-HO-IG) approximation adopting the AMBER 4.1 and ab initio characteristics. Furthermore, the population of various structures was determined by computer simulations in the NVT canonical and NVE microcanonical ensembles. Results obtained from the RR-HO-IG approximation and the NVT ensemble are very similar and differ from the result of the NVE ensemble. The present analysis demonstrates that different gas-phase experimental techniques can be used for investigating different regions of the conformational space for nucleic acid base pairs. The fact that entropy is always significant and differs for H-bonded and stacked structures is of importance.

1. Introduction

Two different types of complexes with nucleic acid bases exist: hydrogen-bonded (H-bonded) structures that are approximately planar and stacked structures. H-bonding is usually stronger than base stacking. H-bonding is mainly stabilized by electrostatic forces while the base stacking is mainly due to dispersion forces.[1,2] As a consequence, H-bonding of bases is much more specific than base stacking. However, both interactions are of equal importance in nucleic acids.

Structures of base pairs (and of any molecular complexes) are determined, not only by the interaction energy, the entropy should also be taken into account. It means the change of Gibbs free energy upon complexation should be considered. The role of entropy becomes especially important for floppy systems. The potential energy and free energy surfaces may be very different for floppy systems. (See our calculations on the benzene$\cdots Ar_8$ complex.[3]) Entropy terms for H-bonded and for stacked structures are expected to be different: Therefore, the ultimate evaluation of the relative stability of H-bonded and stacked base pairs must be based on the free energy values.

Recently, we carried out an exhaustive characterization of the potential energy surfaces of nucleic acid base pairs.[1b,2b−e] The aim of the present paper is to study the potential energy surface and free energy surface of the uracil (U) dimer. It is expected that due to the small dipole moment of uracil the stacked structures may be closer to the H-bonded ones. Thus, we propose that there may be a competition between stacked and H-bonded structures of the uracil dimer in gas phase. This problem is of importance for any experimental treatment of nucleobase complexes. Selection of this dimer was also prompted by the ongoing gas-phase experiments on the U$\cdots$U complex by Saykally and co-workers.[4] Theoretical calculations on structure, energetics, and vibrational frequencies can be of crucial importance in the assignment of experimental data. Furthermore, both H-bonding and stacking complexes of uracil, including the uracil dimer, are important in nucleic acids. Noncanonical U$\cdots$U base pairs occur in several RNAs.[5] They include not only the usual structures bonded through two N−H$\cdots$O hydrogen bonds, but the r(UUCGCG) A-RNA hexamer also contains the planar uracil dimer with one N−H$\cdots$O and one C−H$\cdots$O H-bond.[5d] The latter H-bond is expected to be energetically considerably weaker than the former one. Evidently, uracil dimer can form a number of H-bonded structures of biological importance. Thus, it is of primary interest to investigate the role of interaction energy and entropy contributions in these structures.

Competition between H-bonding and stacking has been recently proposed for the pyrimidine dimer. This hypothesis has been investigated by McCarthy et al.[6] using ab initio calculations and infrared (IR) spectroscopy in an Ar matrix. The analysis of experimental data indicated that the dominant contribution is from the planar structure. However, the interaction with the Ar atoms of the matrix is expected to preferentially stabilize the H-bonded pair with respect to the more compact antiparallel stacked dimer.[6] In contrast to the present paper, McCarthy et al.[6] do not investigate the thermodynamics of the dimer formation. Also, the gradient optimization of the stacked

* To whom correspondence should be addressed. E-mail: hobza@indy.jh-inst.cas.cz.

S1089-5639(98)01641-7 CCC: $15.00
Published on Web 08/06/1998

structure has not been fully carried out with the inclusion of electron correlation. On the other hand, they presented an extensive vibrational analysis to make the calculations comparable with the IR experiments.

The present study represents the first attempt to completely characterize a complexation of a nucleic acid base pair in the gas phase, by combining ab initio, empirical potential, computer simulations, and statistical thermodynamical techniques. This is the only way that permits the theoretical predictions to be compared with experimental data.

2. Strategy of Calculations

The potential energy surface (PES) of the uracil dimer (as well as of any other base pair) is too complex, so that one cannot rely on the chemical intuition or experience. It is necessary to use some objective method for determining the geometry of the stationary points: These methods are usually based on computer simulations. Localization of all stationary points is beyond the capability of ab initio techniques and must be done using an empirical potential. Ab initio calculations are then necessary to verify the quality of the force field. As in our previous paper,[3] the quenching/molecular dynamics/AMBER 4.1 method was used. Let us now discuss why the AMBER 4.1 force field has been selected. The AMBER 4.1 potential[7] was shown to reproduce best (among various empirical potentials) the ab initio stabilization energies of H-bonded and stacked base pairs.[8]

After localization of stationary points, the calculated (AMBER) stabilization energies were compared with stabilization energies obtained from correlated ab initio calculations. We have fully optimized, using gradient optimization, structures of all stationary points found at the AMBER PES. Thermodynamic characteristics were evaluated using the rigid rotor−harmonic oscillator−ideal gas (RR-HO-IG) approximation on the basis of AMBER and ab initio characteristics. Furthermore, the populations of various structures were determined in the NVT canonical and NVE microcanonical ensembles. Finally, the second virial coefficient (SVC) was evaluated with the AMBER 4.1 potential. Also, contributions to the SVC from different conformations were calculated.

3. Calculations

3.1. Quantum-Chemical Calculations. Geometries of H-bonded structures were optimized at the Hartree−Fock (HF) level with the 6-31G** basis set using the gradient optimization method. The optimization of H-bonded structures at the HF level is justified because these structures are stabilized by electrostatic (dipole−dipole) interactions that are included at this level. This is even more true in the present case of uracil dimer, because the dipole moment of uracil is smaller than that of other bases, and the change when correlation is taken into account is therefore small. However, the HF method does not include the dispersion attraction and is thus not sufficient for the evaluation of accurate interaction energies. The dispersion energy is due to electron correlation, and its inclusion requires beyond-HF techniques. The stabilization energies were calculated using the second-order Møller−Plesset perturbation method (MP2) utilizing basis sets with more diffuse polarization functions (6-31G*(0.25)).[1b,2b] The flat polarization functions improve the description of the dispersion energy. The interaction energies were corrected for the basis set superposition error, and HF deformation energies of the monomers were also considered. The deformation energy of monomers is a repulsive contribution which is associated with the deformation of monomers upon formation of the complex. Gradient optimization of the stacked structure is a much more difficult task. Stacked structures should be optimized at the correlated level, because these structures are mainly stabilized by dispersion forces. In our previous studies, the conformational space of stacked nucleobase dimers was investigated by a set of single-point MP2/6-31G*(0.25) calculations with rigid monomers; i.e., no gradient optimization has been applied. Unfortunately, the 6-31G*(0.25) basis set is not suitable for a gradient optimization for two reasons: (i) the intramolecular geometries and flexibility of the monomers could be spoiled by the use of diffuse functions; (ii) the currently available gradient optimization techniques are not corrected for the basis set superposition error. The sum of HF and MP2 basis set superposition errors (6-31G*-(0.25) basis set) would be, near the vdW minimum, about 5−6 kcal/mol. This is roughly as much as the stabilization of the dimer itself! This problem could be overcome by working with basis sets containing two sets of polarization functions: MP2 gradient optimization of a base pair with such a basis set is, however, impractical. Hence, the only way to make a gradient optimization of stacked nucleobase dimers is to use a basis set that underestimates the dispersion attraction. This partially counterbalances, by cancellation of errors, the deterioration of the PES due to the basis set superposition error. Therefore, the stacked structures were optimized at the MP2/6-31G* level, using the standard exponents of polarization functions. The interaction energy was subsequently evaluated at the MP2/6-31G*(0.25)//MP2/6-31G* level and corrected for the basis set superposition error. The deformation energies of monomers were standardly included at the MP2/6-31G* level.

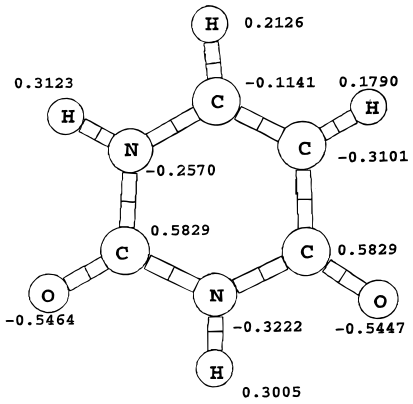

Figure 1. Structure of uracil with RESP/HF/6-31G* atomic charges.

Stabilization energies were corrected for the basis set superposition error with the counterpoise method; all orbitals of the ghost subsystem were included. The deformation energy of the bases upon complexation was determined as the difference of the energies of optimized bases and energies of bases having the same geometry as within the base pair.

3.2. Empirical Potential. The Cornell et al. force field in the original parametrization[7] was used. The atomic charges of uracil were determined consistently with the AMBER 4.1 force field, i.e., using the restrained electrostatic potential fitting procedure (RESP) at the HF/6-31G* level. The charges are summarized in Figure 1. It should be noted that the HF/6-31G* approximation overestimates the dipole moments of bases. Fortunately, this is not so critical for the uracil due to its small dipole moment, which in addition does not change significantly when passing from the HF to MP2 level.

3.3. Statistical Thermodynamical Treatment. The process leading to the formation of the uracil dimer (U···U) from the monomers will be studied:

$$U + U \rightleftharpoons U\cdots U \quad (1)$$

The equilibrium constant (K) of the process at temperature T is related to the change in the Gibbs free energy $\Delta G°$ by the following equation:

$$\Delta G° = -RT \ln K_T \quad (2)$$

The $\Delta G°$ term can be determined from the enthalpy and entropy of the uracil pair formation, $\Delta H°$ and $\Delta S°$, using the equation

$$\Delta G° = \Delta H° - T\Delta S° \quad (3)$$

To evaluate the thermodynamic functions of process 1, it is necessary to know the interaction energy, equilibrium geometry, and frequencies of the normal vibration modes of both U and U···U. These characteristics were evaluated using the AMBER 4.1 potential and compared, wherever possible, with ab initio calculations. Partition functions are evaluated within the (RR-HO-IG) approximation[9] at a temperature of 298 K and a pressure of 1 atm.

3.4. NVT Canonical Ensemble and Second Virial Coefficient Analysis. In the NVT canonical ensemble, the complex is in a temperature equilibrium with the surroundings, and accordingly, the NVT ensemble gives information of the behavior of the complex when it is interacting with the surroundings. However, since the number of considered (intermolecular) degrees of freedom is small, it is convenient to perform a direct numerical integration of the configuration integral, instead of evaluating the configuration integral in a Monte Carlo simulation. In the NVT ensemble the energy can be always expressed as a sum of kinetic and potential contributions; i.e., the partition function factorizes into a product of kinetic and potential parts.[9] The kinetic part will be equal for all conformations, and only the configuration integral (Z) has to be considered. The configuration integral is given by

$$Z_{NVT} = \int d\mathbf{r} \exp(-V(\mathbf{r})/k_B T) \quad (4)$$

where V is the potential energy, k_B is the Boltzmann constant, and T is the absolute temperature. The coordinates of the center of mass (r, θ, ϕ) and the Euler angles (α, β, γ) are chosen as integration variables. Therefore, the configurational integral is given by

$$Z_{NVT} = \int_0^\infty \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} \exp(-V/k_B T) r^2 \sin\theta \sin\beta \, dr \, d\theta \, d\phi \, d\alpha \, d\beta \, d\gamma \quad (5)$$

Due to the planarity of the uracil molecule, only one-half of the configuration integral has to be calculated. The configurational integral is formally not convergent. However, it is made convergent by discarding all points that have potential energy above a certain threshold. The energy threshold must be smaller than the dissociation energy. It is equivalent to say that all configurations with an energy above this threshold are considered as nonbonded and could be excluded when a comparison between the contribution to the configuration integral from different conformations is made. In the calculations presented here a threshold of -2.0 kcal/mol was chosen. Calculations performed with a threshold of -1.0 kcal/mol yield very similar results. So in practice for each point (r, θ, ϕ, α, β, γ) in the grid, gradient optimization is performed to determine to which minimum the point belongs.

The following grid points are used in the gradient optimization. The step length in the numerical integration was 0.125−0.25 Å and 5−10°, depending on where on the PES the point was situated. Test calculations with smaller grids verified convergence of the obtained values.

The SVC and its temperature dependence are important properties of a dimer and can be measured experimentally. Comparison of experimental and calculated data represents an important test of quality of theoretical intermolecular potential. However, the SVC only reflects the average interaction over all conformations. It has been shown that for the benzene dimer[10] different intermolecular potentials can lead to the same SVC, although they describe the PES very differently.

The SVC is closely related to the configuration integral, where the SVC is given by

$$B(T) = -(N_a/16\pi^2) \int_0^\infty \int_0^\pi \int_0^{2\pi} \int_0^{2\pi} \int_0^\pi \int_0^{2\pi} (\exp(-V/k_B T) - 1) r^2 \sin\theta \sin\beta \, dr \, d\theta \, d\phi \, d\alpha \, d\beta \, d\gamma \quad (6)$$

The same integration procedure as described for the configuration integral above was used.

3.5. NVE Microcanonical Ensemble. In the NVE microcanonical ensemble all the systems have the same energy: each system is individually isolated. Populations of various structures were obtained by long runs of molecular dynamics (MD). Constant-energy MD simulations were performed assuming rigid uracil monomers (quaternion formalism). The respective code[11] uses a fifth-order predictor−corrector formalism. A 1 fs integration time step was used. The total energy of the dimer was conserved within 1.4×10^{-2} kcal/mol during the MD run, and this fluctuation originates from the numerical method used.

In the course of the MD simulation, after a specific number of steps, the simulation is stopped and a minimization using the AMBER 4.1 potential is performed: Intermolecular as well as intramolecular coordinates were optimized. We compared three different optimization techniques: steepest descent, conjugate gradient, and Newton−Raphson. After different numbers of steps all methods provided equivalent minima. We finally applied the conjugate gradient method because of its fast convergency. The minimal energy and coordinates are stored, and the MD starts again from the point where it was stopped. Convergence was checked by dividing the MD simulation into five parts where the population distribution for each part was calculated separately. A further verification of the convergency was achieved by satisfying the requirement that the population was equal for each isomer of the conformations. The calculation of relative populations from quenching is possible in a rather narrow energy interval. The energy should be sufficiently high to allow high frequency of interconversions among different isomers. The constant energy MD does not, however, allow us to control fully the temperature selection. In the present study, quenches were made after 10 ps, and we made about 10^9 time steps (1000 ns).

The following characteristics describing the dimer are utilized: (i) Relative population (RP)—the relative abundance of one conformation with respect to other conformations. The RP is evaluated as the ratio between the time the dimer spends in the particular conformation and the total simulation time. The total simulation time should be long enough to yield converged populations; i.e., populations should not change with increasing simulation time. (ii) Number of interconversions—the number of transitions from one structure to another one. The error in determining the relative populations depends mainly on the number of interconversions.

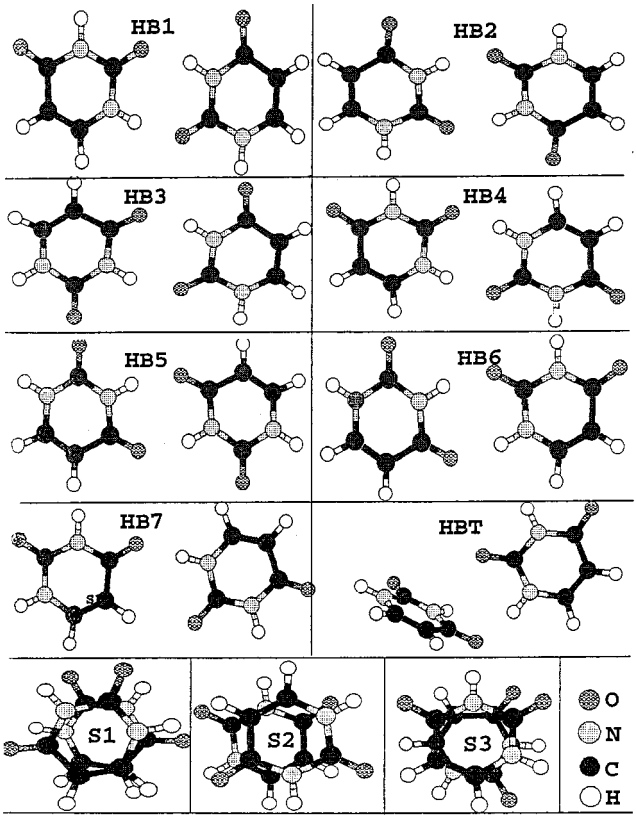

Figure 2. Structures of the uracil dimer; HB, T, and S mean H-bonded, T-shaped, and stacked structures.

TABLE 1: Interaction Energies (ΔE, in kcal/mol) and Changes of Gibbs Energies ($\Delta G°$, in kcal/mol) Calculated by AMBER 4.1 and ab Initio Method for Various Structures of the Uracil Dimer

		ΔE		ΔG^o		
struct[a]	o[b]	AMBER 4.1	ab initio	AMBER 4.1[c]	ab initio[c]	NVT[d]
HB1	2	−13.0	−12.7[e]	−1.3	−0.9	−1.5
HB2	1	−11.1	−10.8[e]	0.9	1.5	0.5
HB3	2	−11.1	−10.4[e]	0.3	1.1	0.3
HB4	1	−15.9	−15.9[e]	−3.5	−3.0	−3.5
HB5	1	−11.0	−10.5[e]	0.9	1.5	0.8
HB6	2	−13.4	−12.4[e]	−1.7	−1.0	−1.7
HB7	2	−10.7	−10.5[e]	−0.7	0.2	−0.9
T	4	−10.1	−7.8[e]	−0.5	1.5	0.1
S1	2	−9.4	−7.7[f]	1.5		1.7
S2	2	−8.7	−7.4[f]	2.0		2.2
S3	4	−8.1		1.7		2.1

[a] Cf. Figure 1; HB, T, and S mean H-bonded (planar), T-shaped, and stacked structure, respectively. [b] Symmetry number, i.e., the number of indistinguishable orientations of the particular structure. [c] Rigid rotor−harmonic oscillator−ideal gas approximation; $T = 298$ K, $p = 1$ atm. [d] Relative values from the NVT analysis using AMBER 4.1 potential; for comparison with RR-HO-IG/AMBER 4.1 analysis the $\Delta G°$ of HB4 was set to −3.5 kcal/mol. [e] MP2/6-31G*(0.25)//HF/6-31G**. [f] MP2/6-31G*(0.25)//MP2/6-31G*.

4. Results and Discussion

4.1. Potential Energy Surface. The MD/quenching/AMBER 4.1 investigation of the potential energy surface resulted in 11 minima structures.[12] Seven of them are planar H-bonded ones, one is T-shaped, and three are stacked. Structures of these stationary points are visualized in Figure 2. Their stabilization energies are given in Table 1. The most stable structure is the H-bonded structure 4, followed by H-bonded structures 6, 1 and 2, 3, 5. All of these structures have two H-bonds of the C═O···H−N type. H-bonded structure 7 is slightly less stable than other H-bonded structures which is due to the fact that one C═O···H−N H-bond is replaced by the weaker C═O··· H−C bond. Rotation around the C═O···H−N H-bond in the HB7 leads to T-shaped structure which is slightly less stable than all H-bonded structures. Following expectations, stacked structures are the least stable among all structures. However, the energy difference is not too large. Among the three stacked structures, structure S1 is the most stable. Structure S2, having antiparallel orientation of dipole moments, is less stable. This could be explained by the fact that dipole moment of uracil is rather small, and the dipole−dipole electrostatic interaction in the stacked structure of the dimer is not as dominant as for more polar bases.[2b,8]

Five transition structures were located at the AMBER 4.1 PES. The highest energy barrier (5 kcal/mol) was found for the transition HB1 → HB3 and the lowest one (0.7 kcal/mol) for the transition S2 → T. The energy of the transition structure separating structures HB7 (having C−H···O H-bond) and T is localized 2.9 kcal/mol above the energy of HB7.

The quality of the AMBER 4.1 predictions was verified by performing correlated ab initio calculations on all the 11 energy minima found. Total energies, geometries, rotational constants, and dipole moments of all 11 structures optimized in this study at HF/6-31G** (H-bonded structures) and MP2/6-31G* (stacked structures) are available upon request from the authors. Stabilization energies determined consistently at the MP2/6-31G*-(0.25) level are summarized in Table 1. From the table it is evident that the H-bonded structure HB4 corresponds to the global minimum. This structure is stabilized by two H-bonds of the C═O···H−N type which are almost linear (both N−H−O angles equal to 174°) with rather short O···H distances of 1.916 Å. Other structures also having these two H-bonds are less stable which is due to fact that in these cases H-bonds are less linear and intermolecular O···H distances are larger. For example, in the case of the first local minimum (structure HB1) with the stabilization energy of 12.7 kcal/mol, the N−H−O angles are 167° and 166°, and the O···H distances are equal to 1.985 and 1.993 Å, respectively. As demonstrated in Table 1, we found an excellent agreement between AMBER 4.1 and corelated ab initio results. It must be mentioned that this is partly due to the fact that our ab initio calculations were performed with the same basis sets as used in the AMBER/RESP procedure. Best agreement between AMBER and ab initio was found for H-bonded structures where the largest absolute error is about 1 kcal/mol. The AMBER 4.1 stabilization energy for the T-shaped structure and stacked structures is overestimated—in the former case by about 2 kcal/mol while in the latter cases by about 1.5 kcal/mol. This is, in fact, remarkable success for the AMBER potential since evaluation of geometries and stabilization energies of stacked base pairs is extremely difficult and requires the use of sophisticated beyond-Hartree−Fock techniques. Correlated ab initio calculations on base stacking based on gradient optimization is the subject of our forthcoming paper.[13]

4.2. Free Energy Surface. *Rigid Rotor−Harmonic Oscillator−Ideal Gas Approximation.* The AMBER 4.1 free energy values are summarized in Table 1. The entropy term is important and compensates for the interaction energy (enthalpy) term. A similar type of compensation has also been found in the case of DNA base pairs.[14] The H-bonded structure 4 remains the most stable, and also HB6 and HB1 structures remain as the second and third most stable ones. However, the following order of stability is changed. The H-bonded

TABLE 2: Second Virial Coefficient Calculated with the AMBER 4.1 Potential as a Function of the Absolute Temperature (*T*)

T, K	HB1[a]	HB2[a]	HB3[a]	HB4[a]	HB5[a]	HB6[a]	HB7[a]	T[a]	S1[a]	S2[a]	S3[a]	Σ[b]
300	3	0	0	90	0	5	1	0	0	0	0	-28×10^6
350	5	0	1	79	0	9	3	1	0	0	0	-20×10^5
400	7	1	2	67	1	13	5	3	1	0	0	-20×10^4
450	9	1	2	55	1	15	8	4	1	0	1	-97×10^3
500	10	2	3	46	2	17	10	5	2	1	1	-41×10^3
550	11	2	4	40	2	18	11	5	2	1	1	-22×10^3
600	11	2	4	37	3	18	12	5	3	1	2	-13×10^3
650	11	2	4	34	3	18	13	6	4	1	2	-93×10^2
700	11	2	4	33	3	19	13	6	4	1	2	-69×10^2
750	11	2	4	30	3	19	14	6	4	1	2	-54×10^2
800	11	2	4	30	3	19	14	6	4	1	2	-45×10^2
850	11	2	4	30	3	19	14	6	4	1	2	-37×10^2

[a] Relative contributions in %. [b] Total second virial coefficient in cm^3/mol.

structure 7 and the T-shaped structure are surprisingly more stable than H-bonded structures 2, 3, and 5. Analyzing various contributions to entropy, we found that it is the vibration entropy which favors structures HB7 and T over the HB2, HB3, and HB5. Investigating the ab initio free energy values (cf. Table 1), we again found very good agreement with the AMBER 4.1 characteristics. This also concerns the preference of H-bonded structure 7 over structures HB2, HB3, and HB5. The ab initio $\Delta G°$ values could not be predicted for stacked structures because evaluation of MP2 vibration frequencies for systems as large as uracil dimer is still impractical.

NVT Ensemble. Analysis in the NVT ensemble (with consideration of AMBER 4.1 potential) is given in Table 1, and it is seen that the conformation distribution is very similar to the RR-HO-IG analysis. This rather surprising result indicates that the nonharmonicity neglected within the framework of the RR-HO-IG analysis is not of major importance for the conformation distribution of the present system. Further, a nice agreement between both results gives support for further application of easily performable RR-HO-IG/ AMBER 4.1 analysis. Let us recall that both the RR-HO-IG method and NVT ensemble analysis (with some different approximations involved) correspond to a situation when the dimer is in thermal equilibrium with the surroundings.

Table 2 presents the SVC and the contribution to the SVC from different conformations. The absolute values of the SVC are large mainly due to the depth of the global energy minimum. As discussed above, the difference between the NVT ensemble analysis and the SVC calculations is rather small. Therefore, not surprisingly, the same contributions are dominant within both methods. The H-bonded structure HB4 contributes mostly, followed by the H-bonded structure HB6. At low temperatures, the HB1 structure follows while at high temperatures it is the structure HB7. This reflects the fact that HB1 has a deeper energy minimum, and HB7 is entropically favored. At higher temperatures the contribution from stacked structures to the SVC is about 7%. Results mentioned are in a full agreement with the RR-HO-IG results. At low temperatures structure HB4 is completely dominant, and the other structures increase their relative weight with increasing temperature.

NVE Ensemble. MD calculations yielded 15 energy minima, of them, four were populated quite insignificantly. Relative populations of the remaining 11 minima are shown in Figure 3. The clearly dominant peak corresponds to H-bonded structure 4, which was also dominant at the PES. Qualitatively, a new feature at the free energy surface (with respect to the PES) is the population of stacked structures in general, and population of the stacked structure 3 in particular: The population of the latter structure is in fact the second highest. Population of stacked structures 1 and 2 is also quite high and is comparable

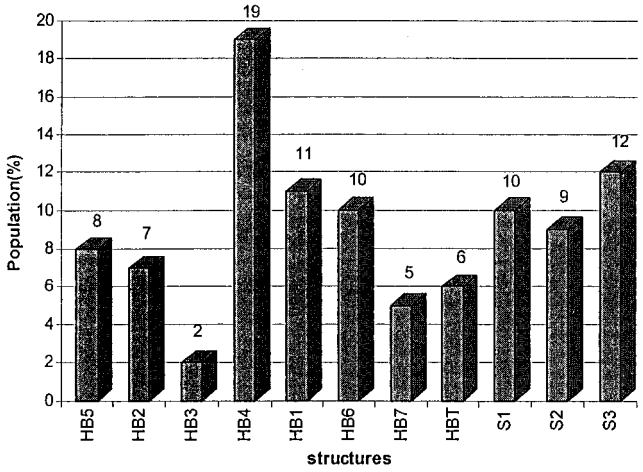

Figure 3. Relative populations of 11 energy minima obtained from MD/quenching/AMBER 4.1 simulations. The maximal convergence error, obtained from averaging over five trajectories whose length increased from 4000 to 10 000 quench steps, is about 6%. The populations in the figure correspond to the largest trajectories (10 000 quench steps). The number of interconversions was 78 008 (from 97 175 quench steps).

to the populations of H-bonded structures 1 and 6. H-bonded structures 2 and 5 are populated slightly more than structures HB7 and T. Population of H-bonded structure 3 is the lowest. Let us recall that the NVE ensemble (contrary to the previously discussed NVT ensemble) gives a property of a dimer that does not interact with the surroundings.

Comparing RR-HO-IG and NVT ensemble results with those for the NVE ensemble, we found full agreement in the case of the global minimum HB4. However, the situation is very different for the stacked structures: Their population is quite high in the case of NVE ensemble and negligible in the case of RR-HO-IG and NVT ensemble results. Other important differences concern low populations of H-bonded structure HB7 and structure T from the NVE ensemble, in comparison with quite high population of these structures from RR-HO-IG and NVT ensemble calculations. Clearly, the NVE and NVT ensemble results must coincide in the limit of infinitely large systems.[15,16] However, the uracil dimer with only six degrees of freedom explicitly considered is all but an infinite size system. This has important consequences for the probability $P(E) = \Omega(E) \exp(-E/kT)$ (where $\Omega(E)$ is the density of states with energy E) of finding the NVT ensemble at energy E. While for large systems with high density of states this canonical probability approaches a δ-function peaked at the energy of a corresponding microcanonical ensemble,[16] small systems rather approach one degree of freedom limit $P(E) = \exp(-E/kT)$. In

a complementary way, the smallness of the investigated system is reflected in the large temperature fluctuations present in the microcanonical MD simulations. While the average temperature was 298 K, fluctuations in the range 10−1000 K were observed.

In both ensembles we see that the entropy works for the energetically higher stacked structures compared to the H-bonded ones. As already mentioned, it follows from our calculations that the dominant contribution to the configurational entropy comes from the six intermolecular vibrations. Not surprisingly then, the more weakly bound stacked structures are entropically favored. In light of the above findings, it is also understandable why the NVE MD simulations relatively favor the stacked structures much more than the NVT ensemble calculations. Namely, in a few-dimensional NVT system low-energy configurations are extensively sampled due to the Boltzmann $\exp(-E/k_BT)$ factor. Of course, the lower the total energy, the less the energetically higher stacked structures are populated. Also, at lower energies the system samples more extensively the bottoms of the potential wells, which are more harmonic. As a consequence, the relevant parts of the potential surface are more similar to each other for different configurations than in high-energy situations when strongly anharmonic regions are sampled (especially for the stacked geometries), which further diminishes the entropic advantage of stacked structures. Finally, the more "harmonic character" of the NVT ensemble compared to the NVE simulations explains the surprisingly good agreement between the two NVT methods used: the (harmonic) RR-HO-IG approach and the (beyond-harmonic) NVT configuration integral methods.

5. Conclusions

There exist 11 different low-energy minima on the potential energy surface of the uracil dimer: seven of them are H-bonded, one is T-shaped, and three are stacked. The global minimum is the H-bonded structure number 4 (HB4) with a stabilization energy of 15.9 kcal/mol. Entropy was shown to be always important. Statistical thermodynamical analysis with the rigid rotor−harmonic oscillator−ideal gas approximations showed that H-bonded structures are more populated than stacked and T-shaped structures, and the HB4 structure is the dominant conformation. The same conclusion was drawn from the study in the NVT ensemble. A constant energy molecular dynamics simulation in the NVE ensemble agreed as to the dominant conformation; other structures such as the stacked ones were, however, also significantly populated. These results demonstrate that in this case (no interaction of the dimer with surroundings) entropy differs considerably for H-bonded and stacked structures. Consequently, the order of stability of various dimer structures at PES and free energy surface (NVE ensemble) differs. These findings are very important and indicate that various experimental techniques can yield different results. Experimental techniques where the dimer is in thermal equilibrium with the surroundings should give results similar to our rigid rotor−harmonic oscillator−ideal gas calculation and our analysis in the NVT ensemble. On the other hand, experimental techniques studying the isolated dimer should give similar results as obtained in the constant energy molecular dynamics simulation.

Acknowledgment. This work has been supported by a grant 203/97/0029 from the Grant Agency of the Czech Republic. O.E. gratefully acknowledges grants from the Swedish Institute and Chemical Centre, University of Lund, Sweden.

References and Notes

(1) (a) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302. (b) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.
(2) (a) Aida, M.; Nagata, C. *Int. J. Quantum Chem.* **1986**, *29*, 1253. (b) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 5590. (c) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Comput. Chem.* **1996**, *17*, 841. (d) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117. (e) Šponer, J.; Gabb, H. A.; Leszczynski, J.; Hobza, P. *Biophys. J.* **1997**, *73*, 76.
(3) Vacek, J.; Hobza, P. *J. Phys. Chem.* **1995**, *99*, 17088.
(4) Saykally, R., personal communication, 1996.
(5) (a) Cech, T. R.; Damberger, S. H.; Gutell, R. R. *Nat. Struct. Biol.* **1994**, *1*, 273. (b) Grüne, M.; Fürste, J. P.; Klussmann, S.; Erdmann, V. A.; Brown, L. R. *Nucleic Acid Res.* **1996**, *24*, 2592. (c) Lietzke, S. E.; Barnes, C. L.; Berglund, J. A.; Kundrot, C. E. *Structure* **1996**, *4*, 917. (d) Wahl, M.; Rao, S. T.; Sundaralingam, M. *Nat. Struct. Biol.* **1996**, *3*, 24.
(6) McCarthy, W.; Smets, J.; Adamowicz, L.; Plokhotnichenko, A. M.; Radchenko, E. D.; Sheina, G. G.; Stepanian, S. G. *Mol. Phys.* **1997**, *91*, 513.
(7) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. E.; Kollman, P. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(8) Hobza, P.; Kabeláč, M.; Šponer, J.; Mejzlík, P.; Vondrášek, J. *J. Comput. Chem.* **1997**, *18*, 1136.
(9) Atkins, P. W. *Physical Chemistry*; Oxford University Press: Oxford, 1994; Chapter 20.
(10) Smith, G. D.; Jaffe, R. L. *J. Phys. Chem.* **1996**, *100*, 9624.
(11) Heindenreich, A.; Jortner, J. Package of MD programs for molecular clusters, 1992.
(12) Besides 11 low-energy structures, there exist several H-bonded structures with higher energy, all with one or two C−H···O═C H-bonds. However, in our MD and NVT ensemble analysis these structures were insignificantly populated.
(13) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1998**, *288*, 7.
(14) Hobza, P.; Šponer, J. *Chem. Phys. Lett.* **1996**, *261*, 379.
(15) Allen, M. P.; Tildesey, D. J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, 1987.
(16) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: Oxford, 1987.