SUPPLEMENTARY INFORMATION

Computational investigation of uracil dimers in water and the role of classical potentials

Tea Ostojić¹, Juraj Ovčar², Ali Hassanali³, and Luca Grisanti^{4*}

^{1,2,4}Ruđer Bošković Institute, Divison of Theoretical Physics, Bijenička cesta 54, 10 000 Zagreb, Croatia

³Condensed Matter and Statistical Physics, The Abdus Salam International Centre for Theoretical Physics, strada Costiera 11, 34151 Trieste, Italy

^{2,4}CNR-IOM, Consiglio Nazionale delle ricerche - Istituto Officina dei Materiali, c/o SISSA, via

Bonomea 265, 34136 Trieste, Italy

*e-mail: grisanti@iom.cnr.it

Complementary dataset for this publication is available at: https://doi.org/10.5281/zenodo.15262465.

Note: citation numbers refer to the specific bibliography of this document.

Contents

S1 wt-MetaMD and force field validation	3
S2 Additional analysis and properties	7

List of Figures

S 1	Chen&Garcia Gibbs Energy dependence on time for HT minima	3
S2	Chen&Garcia Gibbs Energy dependence on time for 2H minima	3
S 3	Chen&Garcia Gibbs Energy dependence on time for ST minima	4
S 4	ff99 Gibbs Energy dependence on time for HT minima	4
S 5	ff99 Gibbs Energy dependence on time for 2H minima	4
S6	ff99 Gibbs Energy dependence on time for ST minima	4
S 7	The Gibbs free energy surface: 2 uracil molecules with water (Chen&Garcia with walls)	6
S 8	Intermolecular angular heatmaps	7
S 9	TD-DFT absorption spectra	8
S10	CO bond lengths and absorption spectral shifts	8

List of Tables

S 1	Chen&Garcia Minima	5
S2	ff99 Minima	5
S 3	Energy dependence on number of water molecules	5
S 4	Force field comparison	5
S 5	2H populations according to HB contacts	7
S 6	ST populations based on angle analysis.	7

S1 wt-MetaMD and force field validation

To estimate the uncertainty of the energies, we analyzed the energy variation over time for the final \sim 300*ns* of the simulation, with a 1 ns time step, resulting in a total of \sim 300 energy points, G(t), for each minimum basin. For each dataset, we performed a linear regression to estimate the energy at the final time point of the simulation, G(t = 1000ns) for Chen&Garcia force field and G(t = 907ns) for ff99. To assess the energy uncertainties, we calculated the energy at each time point using the linear regression fit equation and subtracted it from the original energy value. The uncertainty was then determined as the standard deviation of the resulting data set for each minimum basin. Both procedures were carried out using the reparameterized AMBER99 force field with TIP3P water and the original AMBER99 force field with OPC water.





Figure S1: Chen&Garcia Gibbs Energy dependence on time for HT minima. Results observed during the wT-metaD, plotted against a linear regression.

Figure S2: Chen&Garcia Gibbs Energy dependence on time for 2H minima. Results observed during the wT-metaD, plotted against a linear regression.





Figure S3: Chen&Garcia Gibbs Energy dependence on time for ST minima. Results observed during the wT-metaD, plotted against a linear regression.

Figure S4: **ff99 Gibbs Energy dependence on time for HT minima.** Results observed during the wT-metaD, plotted against a linear regression.





Figure S5: **ff99 Gibbs Energy dependence on time for 2H minima.** Results observed during the wT-metaD, plotted against a linear regression.

Figure S6: **ff99 Gibbs Energy dependence on time for ST minima.** Results observed during the wT-metaD, plotted against a linear regression.

Minimum	G_r (1000 ns, kJ/mol)	linear regression equation $G(t)$	$\sigma_{E_{corr}}$	$G_r \; (kJ/mol)$
НТ	-33.80	-0.015t - 19.06	0.07	2.26 ± 0.07
2H	-36.06	-0.016t - 19.58	0.11	0.00 ± 0.11
ST	-27.54	-0.015t - 12.07	0.05	8.52 ± 0.05

Table S1: **Chen&Garcia Minima.** Linear regressions coefficients to estimate the average free energies and their uncertainties.

Minimum	G_r (907 ns, kJ/mol)	linear regression equation $G(t)$	$\sigma_{E_{corr}}$	$G_r \; (kJ/mol)$
НТ	-31.85	-0.018t - 15.01	0.09	0.00 ± 0.09
2H	-30.01	-0.016t - 15.31	0.08	1.84 ± 0.08
ST	-29.44	-0.017t - 13.70	0.06	2.41 ± 0.06

Table S2: ff99 Minima. Linear regressions coefficients to estimate the average free energies and their uncertainties.

To justify the use of the reparameterized force field by Chen and Garcia, we performed a comparison with a more commonly used force field and water model. The procedure is schematically illustrated in the main text (Figure 1.), and additional details are provided here. The configurations used for the force field comparison differed not only in the type of intermolecular interactions between uracils but also in the number of water molecules in the clusters. The clusters were optimized using the force field corresponding to the simulation from which the configuration originated. Single point energy calculations were then performed using the Orca 5.0.2. software package at the PBE0/def2-TZVP D3BJ level of theory. However, the energies could not be directly compared due to differences in the number of atoms, arising from the varying numbers of water molecules. To enable a direct comparison, we transformed the energies by accounting for the water contribution. Specifically, we performed a linear regression of the energy dependence on the number of water molecules.

method	linear regression equation $E(N(H_2O))$	R^2
Chen&Garcia	$-35.9N(H_2O) - 756.2$	0.85
ff99	$-38.5N(H_2O) - 760.5$	0.87
DFT(Chen&Garcia, Ha)*	$-76.4N(H_2O) - 829.1$	1.00
DFT(ff99, Ha)*	$-76.4N(H_2O) - 829.1$	1.00

Table S3: Energy dependence on number of water molecules. Energy dependence on $N(H_2O)$ for each set of data optimized with force fields rAMBER99/TIP3P and AMBER99/OPC. *DFT calculations are done separately on each set for consistency reasons.

From the linear regression equation, we calculated the energy contribution of water molecules for each configuration and subtracted this value from the initial energy. By applying this transformation, the energies became directly comparable on a linear scale, and the results of this comparison are presented in Figure 4 of the main text.

method	bd linear regression equation $E'_{DFT}(E'_{f.f.})$			
Chen&Garcia	$1.12E'_{f.f.} + 0.00$	0.94		
ff99	$0.85E'_{f.f.} + 2.71$	0.86		

Table S4: Force field comparison. Linear regression of transformed DFT energies and classical energies



Figure S7: The Gibbs free energy surface: 2 uracil molecules with water (Chen&Garcia with walls). To additionally validate the usage of wT-MetaD we have run the molecular dynamics with walls defined at s = 0.3 and h = 0.3. The walls act as high barriers that prevent the system from visiting the configurations characterized by s and h lower than 0.3. In this way, we eliminated the area that is out of our interest from the exploration. The relative minima free Gibbs energies at the end of the simulation are 2H: 0.00 kJ/mol, HT: 1.31 kJ/mol and ST: 8.53 kJ/mol. The resulting potential energy surface has a similar shape to the one obtained with wt-MetaD and the trends in minima are the same. See main text for further details.

S2 Additional analysis and properties

labels	contacts		population (%)
HB1	O2.N3'	N1.O2'	19.5
HB2	N3.O2'	O2.N3'	6.7
HB3	O4.N3'	N3.O2'	17.8
HB4	O2.N1'	N1.O2'	17.7
HB5	N3.O4'	O4.N3'	11.3
HB6	N3.O2'	O4.N1'	18.9
HB7	O4.N1'	C5.N3'	_

Table S5: **2H populations according to HB contacts.** Results for the 2H basin obtained in this work. Labelling and basin definition is consistent with the structures in vacuum previously reported [1].

labels	bin size $lpha$	bin size $ au$	population (%)
ST1	0-45	0-60	16.0
ST2	0-45	120-180	21.6
tot. parallel (F2B)	0-45	any	53.7
ST3	135-180	60-120	19.3
tot. antiparallel (F2F)	135-180	any	46.0

Table S6: **ST populations based on angle analysis..** Results for ST basin obtained in this work. F2F and F2B refer to previously reported labelling schemes [2]. See main text.



Figure S8: **Intermolecular angular heatmaps.** For a) HT and b) 2H, as equivalent to what is presented in the main text for Fig. 5. See text for definition of geometrical parameters.



Figure S9: **TD-DFT absorption spectra.** Results obtained starting from configurations (clusters of uracil molecule(s) and water) sampled from MD and then optimized according to the protocol presented in the methods. TDDFT results are presented for two functionals.



Figure S10: **CO bond lengths and absorption spectral shifts.** Correlation between the sum of C4O4 bond lengths and a quantitative estimation of the redshift as computed from optical spectra reported in Fig S9 as $\int_0^{\omega_{\text{max}}} S_{2U}(\omega) - 2S_{\text{monomer}}(\omega)d\omega$, where $2S_{\text{monomer}}(\omega)$ is the monomer absorption spectra in eV units peaking at energy ω_{max} , and $S_{2U}(\omega)$ is the spectra for the dimer in the same eV units. Coloring is associated to the COM-to-COM distance between the two molecular units. Left panel: PBE0; right panel: ω B97X. We also found that for **2H** the configurations showing exclusively a redshift correspond to states labeled as HB4, HB5, HB6.

References

- [1] Martin Kratochvíl, Ola Engkvist, Jiří Šponer, Pavel Jungwirth, and Pavel Hobza. Uracil dimer: Potential energy and free energy surfaces. ab initio beyond hartree-fock and empirical potential studies. *The Journal of Physical Chemistry A*, 102:6921–6926, 08 1998.
- [2] Branislav Milovanović, Jurica Novak, Mihajlo Etinski, Wolfgang Domcke, and Nada Došlić. On the propensity of formation of cyclobutane dimers in face-to-face and face-to-back uracil stacks in solution. *Physical Chemistry Chemical Physics*, 24:14836–14845, 2022.