

SUPPORTING INFORMATION

PREDICTION OF SELF-ASSEMBLY OF ADENOSINE ANALOGUES IN SOLUTION: A COMPUTATIONAL APPROACH VALIDATED BY ISOTHERMAL TITRATION CALORIMETRY

Luca Redivo,^{‡a} Rozalia-Maria Anastasiadi,^{a‡} Marco Pividori,^b Federico Berti,^c Maria Peressi,^d Devis Di Tommaso,^{a*} Marina Resmini^{a*}

^aSchool of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London, United Kingdom, E1 4NS

^bDipartimento di Fisica, Università degli Studi di Trieste, via A. Valerio 2, Trieste, Italy, 34149

^cDipartimento di Scienze Chimiche e Farmaceutiche, Università degli Studi di Trieste, via L. Giorgieri 1, Trieste, Italy, 34149

^dIstituto Officina dei Materiali, Consiglio Nazionale delle Ricerche, S.S. 14 km 163.5, Trieste, Italy, 34149

TABLE OF CONTENTS

S.1 Assessment of the methods in terms of the ability to locate minima on the potential energy surface of caffeine dimer	2
S.2 Accuracy in the evaluation of association free energy in solution.....	7
S.3 Dimer, Trimer and Tetramer structures of caffeine.....	9
S.4 Dimers of paraxanthine	12
S.5 Experimental methods	16
S.6 Supplementary references.....	18

S.1 Assessment of the methods in terms of the ability to locate minima on the potential energy surface of caffeine dimer

A critical step to conduct DFT calculations of caffeine aggregates is the definition of the level of theory (density functional and basis set) used to describe the interaction between caffeine molecules. From the crystallographic data caffeine two interaction patterns were identified: a “face-to-back” (hereafter known as parallel) and “face-to-face” configuration (hereafter known as anti-parallel), respectively. The structures of are reported in Figure S.1 and S.2.^{1,2} In both synthons the aromatic rings distant 3.4 Å and the methyl groups are rotated to allow the formation of weak $\text{CH}_3 \cdots \text{X}$ ($\text{X} = \text{O}$ and N) hydrogen (H) bonds between two caffeine molecules. These two dimeric motifs were used as the initial configurations to conduct geometry optimization using a selected set of generalized-gradient-corrected GGA, hybrid DFT (HDFT), hybrid-meta DFT (HMDFT) and long-range corrected HDFT (LC-HDFT) exchange-correlation functionals (all methods are listed in Table S.1 and S.2).

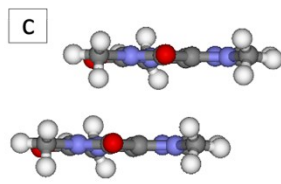
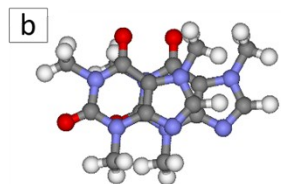
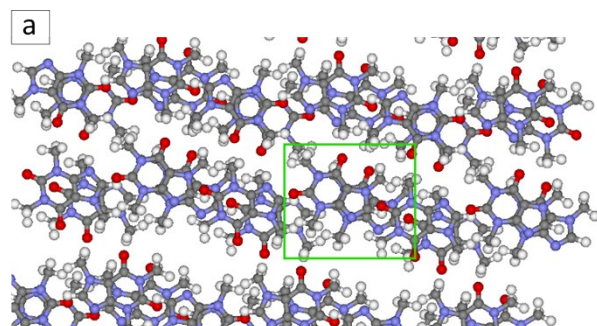


Figure S.1 The crystallographic structure of caffeine.³ Panel a: dimer geometry taken as a model highlighted in green. Panel b: top view of the structure of the dimer showing the overlap between the two aromatic rings. Panel c: side view of the structure of the dimer showing the orientation of the methyl groups towards the O.

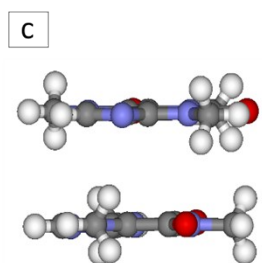
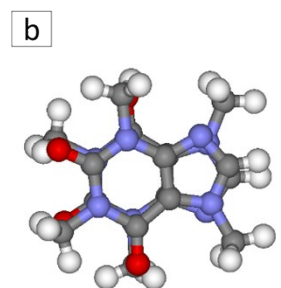
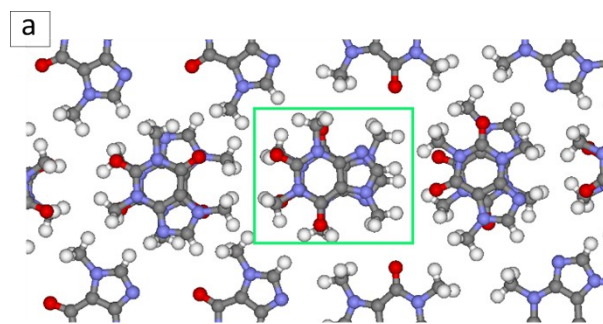


Figure S.2 The crystallographic structure of anhydrous caffeine.² Panel a: dimer geometry taken as a model highlighted in green. Panel b: top view of the structure of the dimer showing the overlap between the two aromatic rings. Panel c: side view of the structure of the dimer showing the orientation of the methyl groups towards the N and O.

A systematic study of DFT methods by Sherrill and co-workers⁴ concluded that the long-range corrected hybrid ω B97X-D functional gives an excellent description of non-covalent (hydrogen-bonded and dispersion-bound) complexes. The geometry of the caffeine dimer optimized using the ω B97X-D functional (Figure S.3(g)) is consistent with the face-to-face crystalline dimeric motif and the binding energy at this level of theory is -14.1 kcal mol⁻¹.

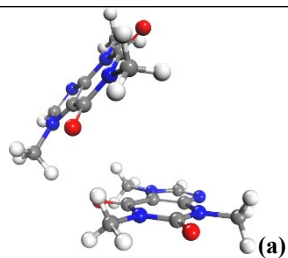
Pure GGA (PBE and BLYP) and HDFT (B3LYP and PBE1PBE) functionals are unable to locate a local minimum corresponding to the parallel crystalline dimeric motif on the potential energy surface (PES) of (CAF)₂, (Table S.1). These methods neglect dispersion and cannot accurately describe the combination of π - π and CH₃...X (X = O and N) interactions (Figure S.3). The caffeine molecules in the dimer optimised at the PBE, BLYP, B3LYP and PBE1PBE methods are arranged in a way to maximize the CH₃...O H-bonding intermolecular interaction and are not the π - π stacking arrangement [Figure S.3(d)]. This shows that misleading results could be obtained by the application of these functionals. The dispersion corrected PBE function (PBE-D3) locates the crystal dimer motif on the PES of (CAF)₂ [Figure S.3(g)], but the binding energies (ΔE) computed using diffuse Gaussian functions [6-31+G(d,p) and 6-311+G(d,p)] and plane wave basis sets are approximately -11 kcal mol⁻¹. This value is four kcal mol⁻¹ lower than the binding energy obtained using the ω B97X-D method. On the other hand, Grimme's -D3 corrected BLYP and B3LYP methods as well as the M06-2X functional, which was specifically developed for non-covalent interactions,⁵ are unable to locate the parallel dimeric configuration (Figure S.1). The GGA dispersion corrected B97-D2 and B97-D3 functionals perform very well both in terms of ability to reproduce the structural properties and the energetics of interaction in the dimer. The energy of formation and ability of DFT methods to locate the anti-parallel crystal dimeric motif are reported in Table S.2. With the exception of BLYP/6-31+G(d,p), the structures optimized with all other methods correspond to the dimer motif found in the crystal. However, pure GGA (BLYP and PBE) and hybrid DFT (B3LYP and PBE1PBE) functionals give values for the gas-phase dimerization energy that are 10 to 15 kcal.mol⁻¹ lower than the binding energies computed with the ω B97D functional. The B97-D2 functional together with the 6-31+G(d,p) provides a good compromise between accuracy and computational cost⁶ and therefore this method was used to compute the structure and energetics in the gas phase associated with the formation of caffeine oligomers.

Table S.1 Gas phase energy of formation ($\Delta E_{e,\text{gas}}$) of caffeine dimer as computed by the DFT methods. The initial configuration of the two caffeine molecules configuration corresponded to the dimer motif (a) found in the crystal structure of caffeine.

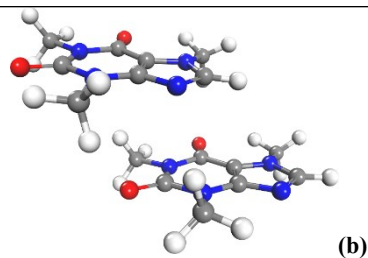
	DFT Method	Basis set	ΔE (kcal mol ⁻¹)	Ability to locate motif (a)	
GGA	BLYP	6-31G(d)	-5.89	×	
		6-31+G(d,p)	-2.32	×	
	PBE	6-31G(d)	-5.50	×	
		6-31+G(d,p)	-5.61	×	
		Plane Waves	-0.68	√	
GGA + Grimme	BLYP-D3	6-31G(d)	-24.97	×	
		6-31+G(d,p)	-18.72	×	
	PBE-D3	6-31G(d)	-14.60	√	
		6-31+G(d,p)	-11.51	√	
			6-311+G(d,p)	-11.28	√
			Plane Waves	-10.68	√
	B97-D2	6-31G(d)	-16.33	√	
		6-31+G(d,p)	-13.00	√	
		6-311+G(d,p)	-13.60	√	
			6-311+G(2df,2p)	-13.16	√
	B97-D3	6-31G(d)	-15.61	√	
		6-31+G(d,p)	-12.57	√	
		6-311+G(d,p)	-13.12	√	
	HDFT	B3LYP	6-31G(d)	-6.52	×
6-31+G(d,p)			-2.91	×	
PBE1PBE		6-31G(d)	-4.14	×	
		6-31+G(d,p)	-2.88	×	
HDFT + Grimme	B3LYP-D3	6-31G(d)	-16.06	√	
		6-31+G(d,p)	-17.91	×	
	PBE1PBE-D3	6-31G(d)	-19.99	×	
		6-31+G(d,p)	-11.76	√	
HMDFT	M06-2X	6-31G(d)	-19.27	×	
		6-31+G(d,p)	-17.84	×	
LC-HDFT	ω B97X-D	6-31+G(d,p)	-14.08	√	

Table S.2 Gas phase energy of formation ($\Delta E_{e, \text{gas}}$) of caffeine dimer as computed by the DFT methods. The initial configuration of the two caffeine molecules configuration corresponded to the dimer motif (b) found in the crystal structure of caffeine.

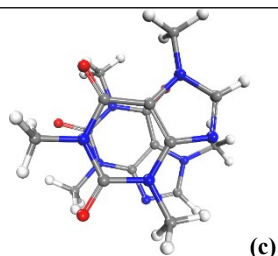
	DFT Method	Basis set	ΔE (kcal/mol)	Ability to locate motif (b)
GGA	BLYP	6-31G(d)	-5.82	√
		6-31+G(d,p)	-1.52	×
	PBE	6-31G(d)	-9.49	√
		6-31+G(d,p)	-2.88	√
GGA + D3	BLYP-D3	6-31G(d)	-25.15	√
		6-31+G(d,p)	-18.80	√
	PBE-D3	6-31G(d)	-21.16	√
		6-31+G(d,p)	-16.54	√
	B97-D2	6-31G(d)	-22.10	√
		6-31+G(d,p)	-17.66	√
	B97-D3	6-31G(d)	-21.26	√
		6-31+G(d,p)	-16.92	√
HDFT	B3LYP	6-31G(d)	-6.41	√
		6-31+G(d,p)	-2.45	√
	PBE1PBE	6-31G(d)	-8.16	√
		6-31+G(d,p)	-5.28	√
HDFT + Grimme	B3LYP-D3	6-31G(d)	-22.83	√
		6-31+G(d,p)	-17.90	√
	PBE1PBE-D3	6-31G(d)	-19.99	√
		6-31+G(d,p)	-16.75	√
HMDFT	M06-2X	6-31G(d)	-19.89	√
		6-31+G(d,p)	-18.51	√
LC-HDFT	ω B97X-D	6-31+G(d,p)	-18.51	√



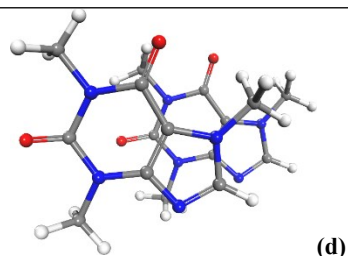
BLYP/6-31+G(d,p)



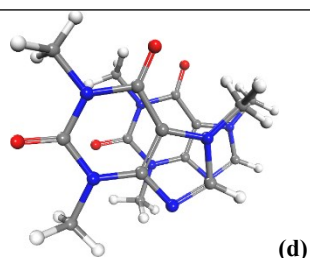
PBE/6-31+G(d,p)



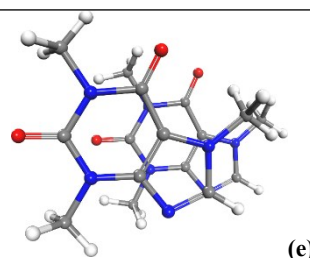
BLYP-D3/6-31+G(d,p)



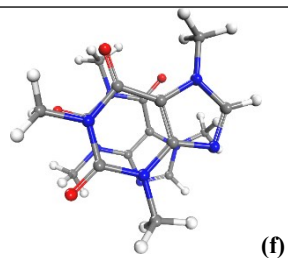
PBE-D3/6-31+G(d,p)



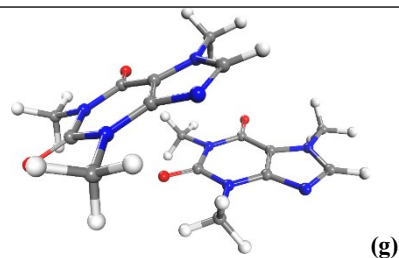
B97-D2/6-31+G(d,p)



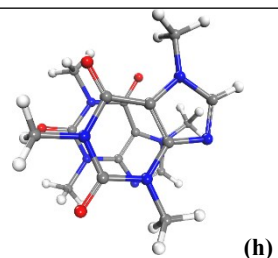
B97-D3/6-31+G(d,p)



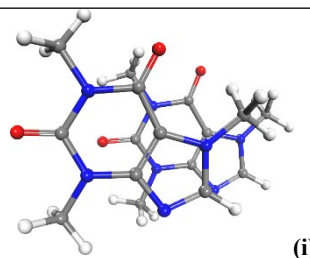
B3LYP/6-31+G(d,p)



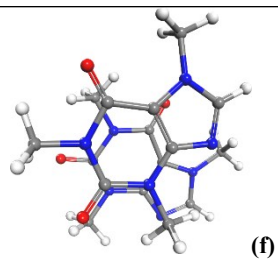
PBE1PBE/6-31+G(d,p)



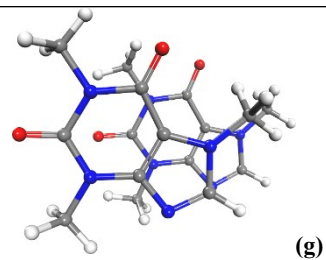
B3LYP-D3/6-31+G(d,p)



PBE1PBE-D3/6-31+G(d,p)



M06-2X/6-31+G(d,p)



wB97X-D/6-31+G(d,p)

Figure S.3 Optimized structures of the caffeine dimer as computed by the DFT methods. The initial configuration for the geometry optimization was the to the dimeric configuration (a) found in the crystal structure of caffeine

S.2 Accuracy in the evaluation of association free energy in solution

The methodology adopted in this work to compute the free energy of association in solution (ΔG_{ass}^*) has an error $\varepsilon(\Delta G_{ass}^*)$ which is given by the sum of the error in the evaluation of the gas phase free energy of reaction, $\varepsilon(\Delta G_{gas}^\circ)$, and the error in the evaluation of the solvation free energies of the reactants and products, $\varepsilon(\Delta\Delta G_{solv}^*)$:

$$\varepsilon(\Delta G_{ass}^*) = \varepsilon(\Delta G_{gas}^\circ) + \varepsilon(\Delta\Delta G_{solv}^*) \quad (1)$$

In Eq. (1), $\varepsilon(\Delta\Delta G_{solv}^*)$ is the difference between the solvation free energy of the products and of the reactants. For the process of dimerization, the term $\varepsilon(\Delta\Delta G_{solv}^*)$ is given by:

$$\varepsilon(\Delta\Delta G_{solv}^*) = \varepsilon[\Delta G_{solv}^*(D)] - 2\varepsilon[\Delta G_{solv}^*(M)] \quad (2)$$

where $\varepsilon[\Delta G_{solv}^*(D)]$ and $\varepsilon[\Delta G_{solv}^*(M)]$ are the errors in the computed values of the hydration free energies of caffeine monomers and dimers. Table S.3 compares the hydration free energies of caffeine predicted using the SMD solvation model with the experimental value.⁷

At the B97-D2 level of theory, the effect of the basis set on the hydration free energy is significant and the error $\varepsilon[\Delta G_{solv}^*]$ is in the range of 0.9–2.7 kcal.mol⁻¹. On the other hand, the M06-2X is less sensitive to the basis set employed and the solvation free energy computed using the SMD/M06-2X/6-31+G(d,p) method differs by only 0.1 kcal.mol⁻¹.

In this work, the solvation free energies of the monomer, dimer, trimer and tetramer forms were computed using the SMD/B97-D2/6-31+G(d,p) and SMD/M06-2X/6-31+G(d,p) levels of theory. Based on the analysis reported in Table S.3, the error $\varepsilon(\Delta G_{solv}^*)$ is approximately 1 kcal.mol⁻¹ with the SMD/B97-D2/6-31+G(d,p) level of theory, and 0.1 kcal.mol⁻¹ with the SMD/M06-2X/6-31+G(d,p) method.

Table S.3 Hydration free energy of caffeine molecule. Solvation calculations conducted using the gas phase geometry of caffeine optimised at the B97-D2/6-31+G(d,p) level of theory. Values of kcal.mol⁻¹.			
DFT functional	Basis set	ΔG_{solv}^*	$\varepsilon[\Delta G_{solv}^*]$
B97-D2	6-31+G(d,p)	-11.76	0.88
	6-311+G(d,p)	-11.42	1.22
	6-311+G(2df,2p)	-9.95	2.69
M06-2X	6-31+G(d,p)	-12.78	-0.14
	6-311+G(d,p)	-12.86	-0.22
	6-311+G(2df,2p)	-12.09	0.55
Expt.⁷		-12.64	

Table S.4. reports the energetics of formation in the gas phase of dimeric structure evaluated using hierarchical basis sets. As experimental values of the binding energy and free energy of caffeine are not known, the distribution of the values of ΔG_{gas}° have been used to estimate the error $\varepsilon(\Delta G_{gas}^{\circ})$ in Eq. (1). The difference between the minimum and maximum value of ΔG_{gas}° is about 0.5 kcal.mol⁻¹, which can therefore be used as an estimate for $\varepsilon(\Delta G_{gas}^{\circ})$.

Table S.4 Energetics of formation of caffeine dimer (a) found in the crystal structure of caffeine. Values of kcal.mol⁻¹.				
DFT functional	Basis set	$\Delta E_{e,gas}$	$\Delta\delta G_{VRT,gas}^{\circ}$	ΔG_{gas}°
B97-D2	6-31+G(d,p)	-13.00	15.95	2.95
	6-311+G(d,p)	-13.60	16.41	2.81
	6-311+G(2df,2p)	-13.16	15.15	1.99

The analysis conducted in this section indicates that free energy of dimerization of caffeine in water computed at the B97-D2/6-31+G(d,p) method together with the SMD/B97-D2/6-31+G(d,p) solvation model should have an error of $\varepsilon(\Delta G_{dim}^*) = \varepsilon(\Delta G_{gas}^{\circ}) + \varepsilon(\Delta\Delta G_{solv}^*) = 1 \text{ kcal mol}^{-1} + 1 \text{ kcal mol}^{-1} = 2 \text{ kcal mol}^{-1}$. The error associated with the B97-D2/6-31+G(d,p) method together with the SMD/M06-2X/6-31+G(d,p) solvation model is $\varepsilon(\Delta G_{dim}^*) = \varepsilon(\Delta G_{gas}^{\circ}) + \varepsilon(\Delta\Delta G_{solv}^*) = 1 \text{ kcal mol}^{-1} + 0.1 \text{ kcal mol}^{-1} = 1.1 \text{ kcal mol}^{-1}$.

S.3 Dimer, Trimer and Tetramer structures of caffeine

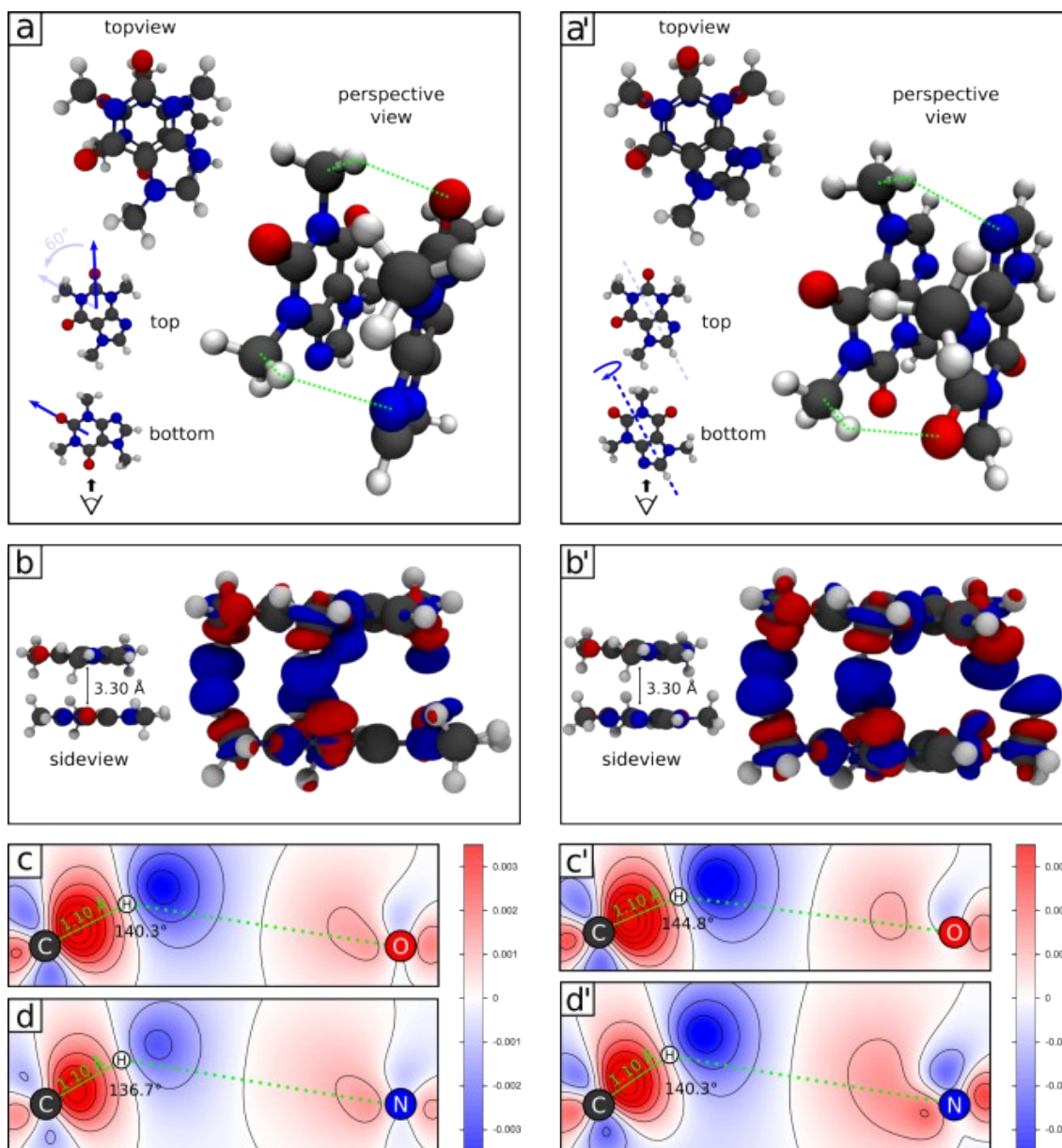


Figure S.4 Structural and electronic properties of dimer (D3) and (D7) computed at the PBE-D3/PW level basis set: parallel (a-d), anti-parallel configuration (a'-d'). (a / a') Ball-and-stick model for the two configurations: top and perspective view. The two molecules (top and bottom) are shown in order to emphasize their relative positions. The perspective views highlight the bonds shown in the panels (c - d) / (c' - d'). (b / b') Charge rearrangement in the dimer configurations with respect to the isolated molecules, corresponding to isosurfaces of $\pm 0.0006 e/\text{\AA}^{-3}$. Blue/red indicates depletion/excess of electron density. The ball-and-stick models on the left help understanding the geometry of the minimum energy configuration. (c - d) / (c' - d') Charge density plots on the bond planes for (c - d) / (c' - d') the C-H...O and (d / d') the C-H...N hydrogen bonds depicted in the panels (a / a').

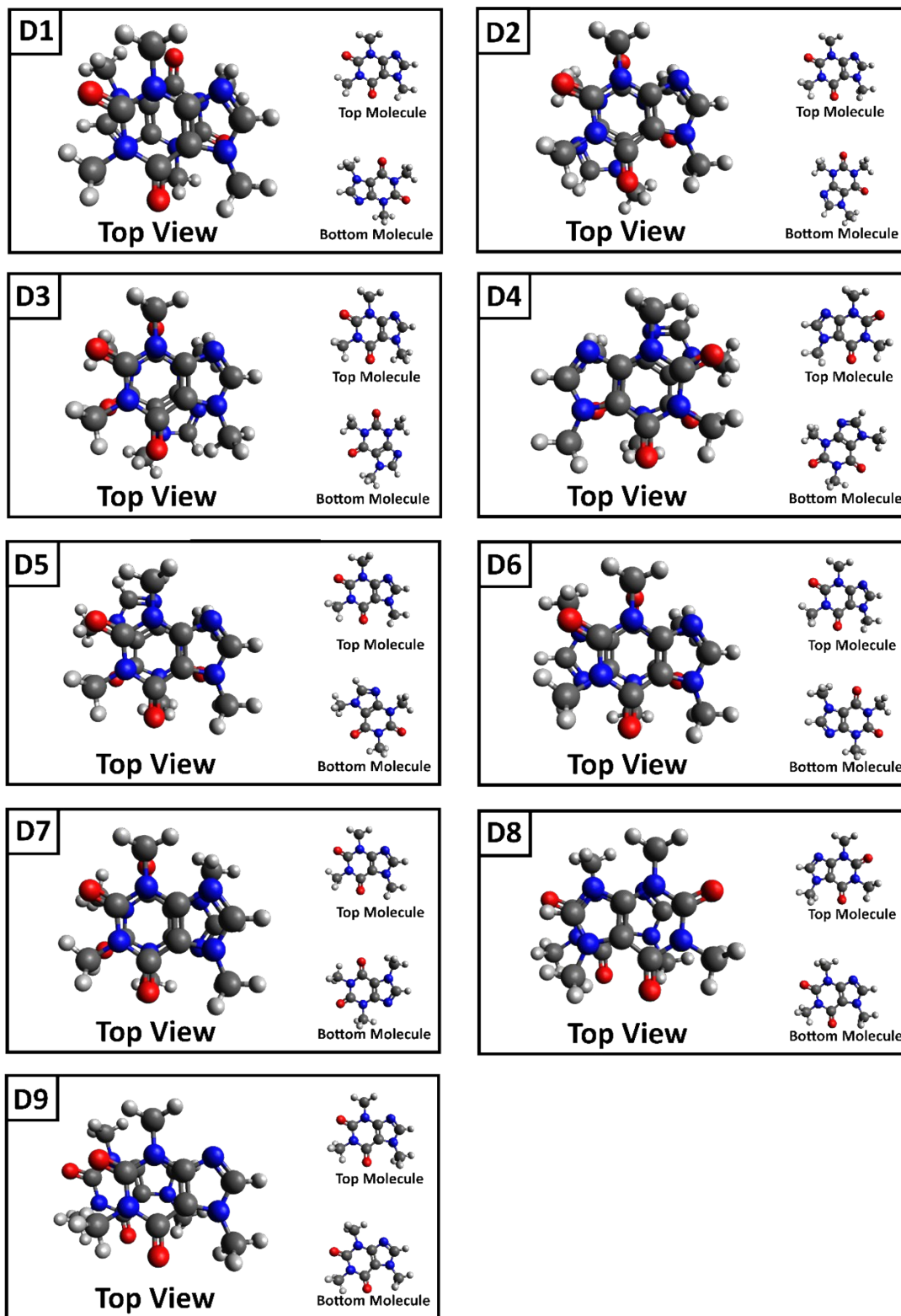


Figure S.5 The nine low-energy minima were located on the PES of $(CAF)_2$, following the proposed computational approach. For clarity the dimer configurations are depicted from a top view and aside the isolated top and bottom molecules are reported to show the reciprocal orientation.

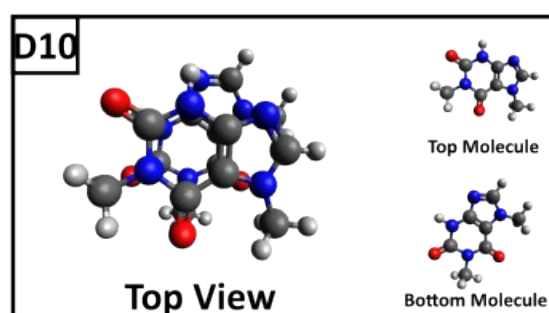
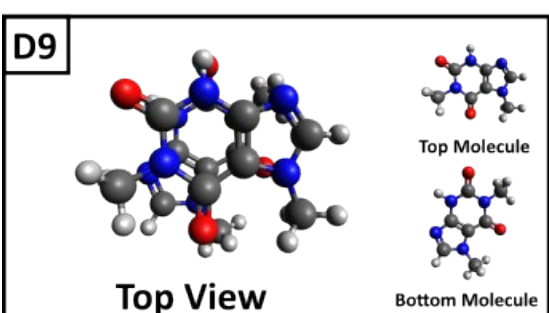
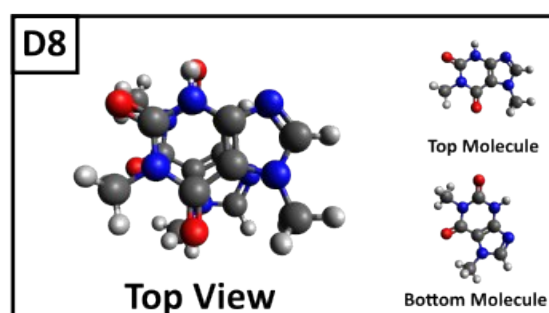
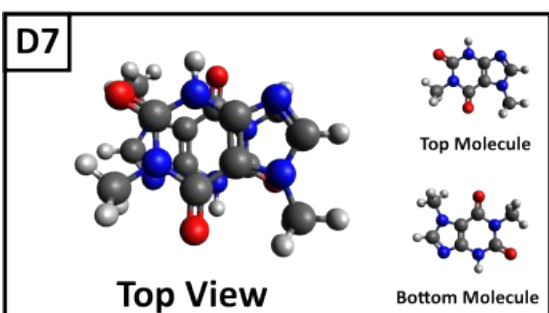
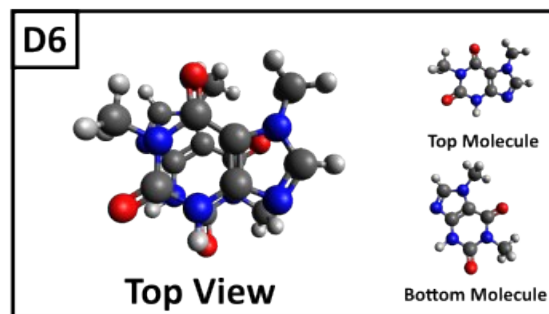
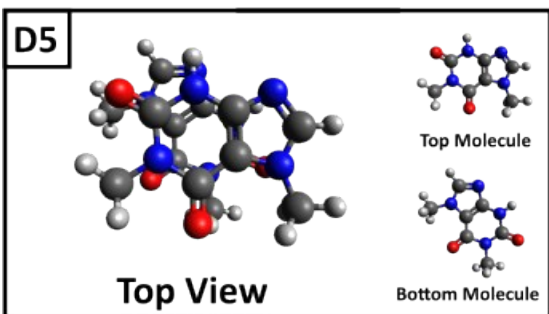
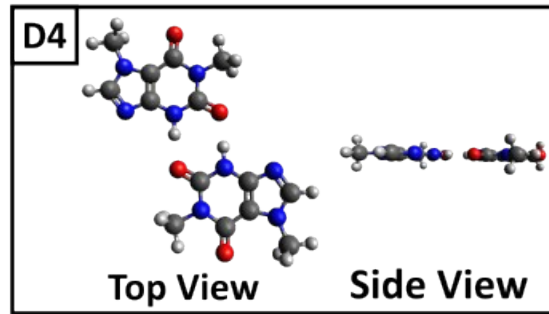
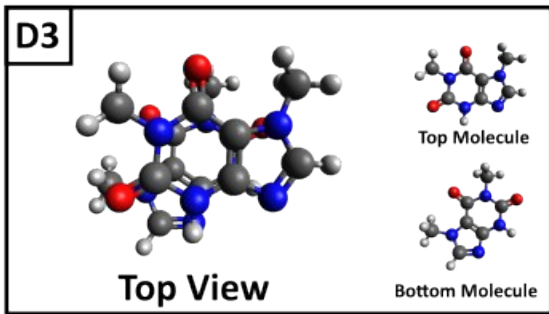
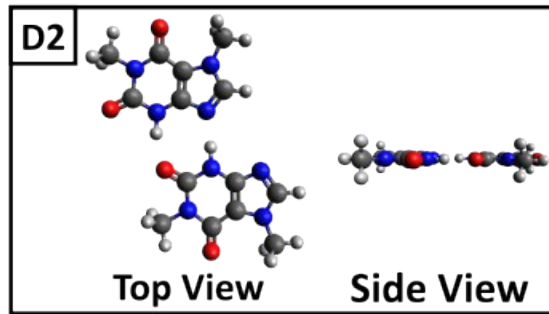
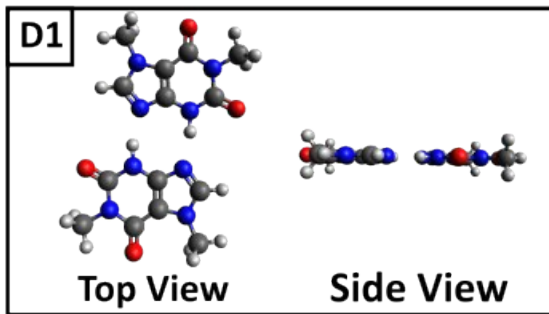
Table S.5 Gibbs free energies of caffeine dimerization (ΔG_D) in different solvents with static dielectric constant ϵ . Values obtained from the Boltzmann average of the free energies of the low-lying (CAF)₂ isomers. Values in kcal·mol⁻¹.

Solvent	Classification	ϵ	ΔG_D^*
Water	Polar protic	78.36	0.98
Methanol	Polar protic	32.61	2.48
Acetone	Polar aprotic	20.49	6.06
Chloroform	Non polar	4.71	5.66

S.4 Dimers of paraxanthine

The B97-D2 functional together with the 6-31+G(d,p) was employed to study the stability of paraxanthine (1,7-dimethylxanthine) dimers.

The same procedure applied in the case of caffeine was employed. In the following picture the twenty stable structures of paraxanthine dimer found are reported.



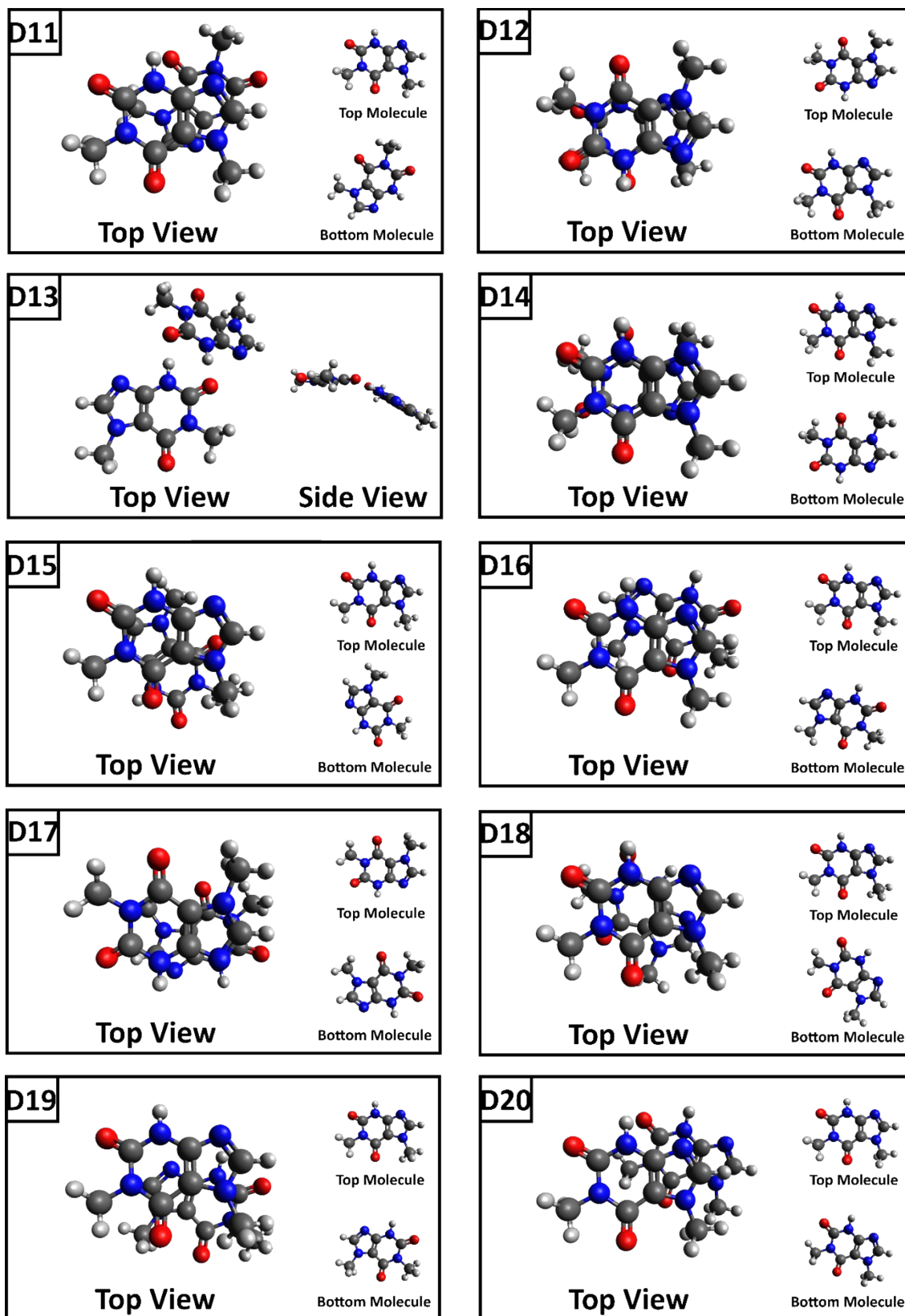


Figure S.6 Top view of the twenty low-lying energy isomers of paraxanthine dimer. To clarify the reciprocal orientation of the two molecules in the dimers, the structure of the isolated top and bottom molecules are also reported.

Table S.6 Energetics [kcal/mol] for the formation of the 20 low-energy minima dimer configurations of paraxanthine in the gas-phase and aqueous solution as computed at the B97-D2/6-31+G(d,p) theory level .

Dimer	ΔE	ΔG_{ass}°	ΔG_{ass}^*
D1	-16.35	-4.72	2.25
D2	-14.89	-3.76	2.11
D3	-17.51	-3.30	0.45
D4	-14.76	-3.23	1.52
D5	-17.49	-3.10	0.67
D6	-17.03	-2.11	1.87
D7	-16.83	-1.88	1.88
D8	-16.31	-1.80	1.62
D9	-17.00	-1.56	2.34
D10	-16.40	-1.49	1.77
D11	-14.89	-1.39	1.05
D12	-15.34	-1.27	2.30
D13	-14.12	-0.79	3.36
D14	-15.16	-0.74	1.81
D15	-13.9	-0.57	0.14
D16	-14.89	-0.91	-0.92
D17	-14.89	-0.89	-0.60
D18	-10.29	3.17	-0.44
D19	-14.97	-0.34	-0.24
D20	-6.99	5.22	-0.12

S.5 Experimental methods

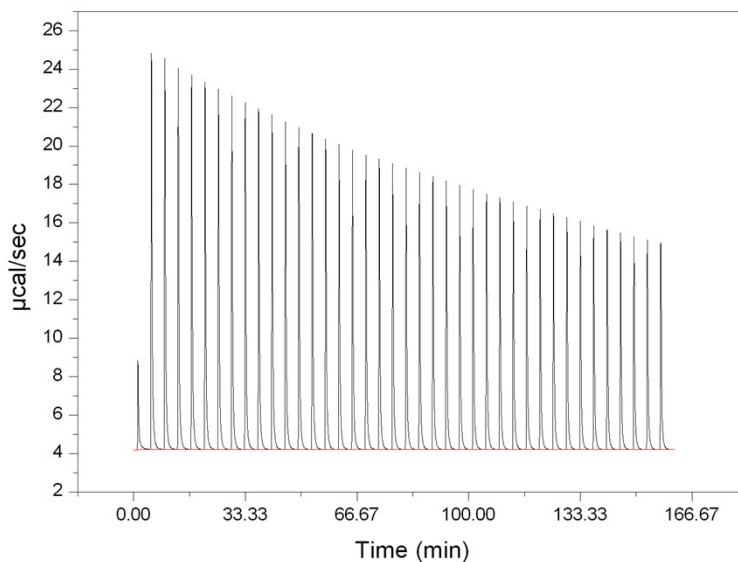


Figure S.7 Raw calorimetric data for the dissociation experiment of caffeine. An aqueous solution of caffeine (82.4mM) in the syringe was stepwise injected in water, at 25°C.

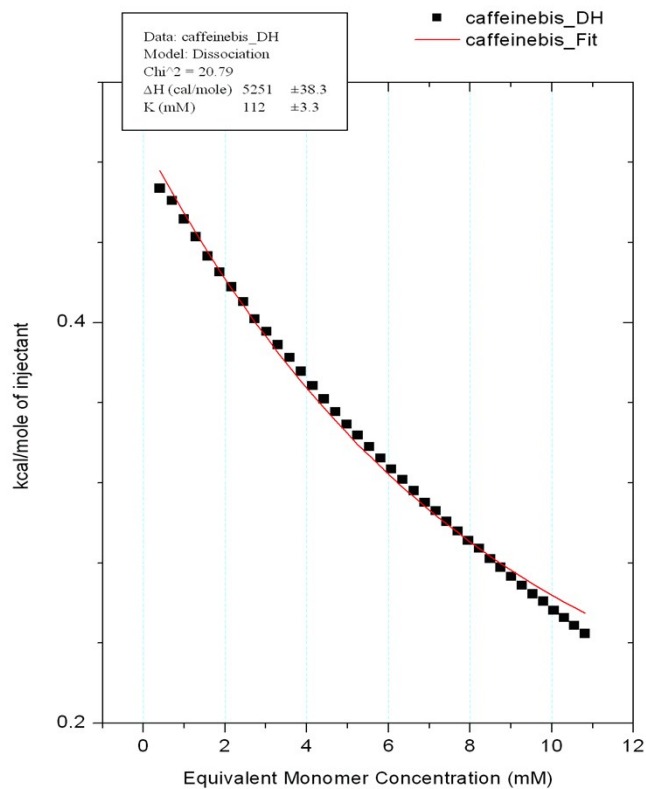


Figure S.8 Integrated heat plots (change in enthalpy against concentration) for the dissociation of caffeine in water. The solid line corresponds to the theoretical curves with $K_{\text{diss}} = 112 \pm 3.3$ mM and $\Delta H_{\text{diss}} = 5.251 \pm 0.038$ kcal/mol.

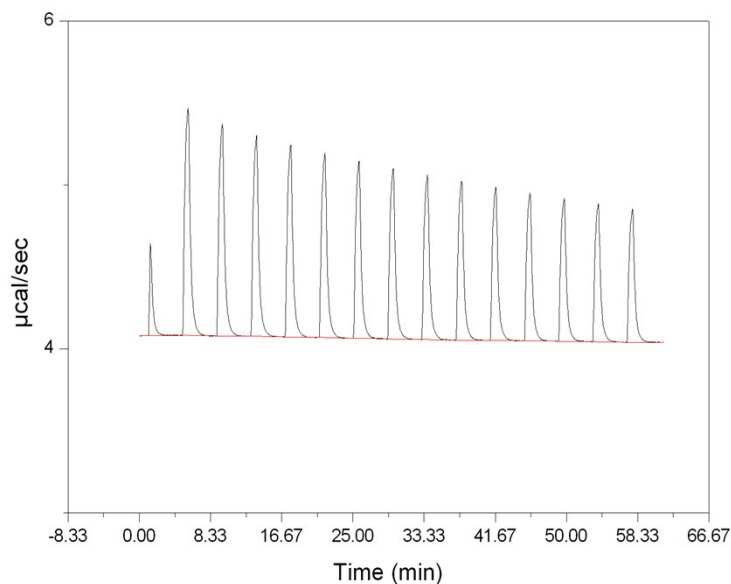


Figure S.9 Raw calorimetric data for the dissociation of paraxanthine. An aqueous solution of paraxanthine (5.55mM) in the syringe was stepwise injected in water, at 25°C.

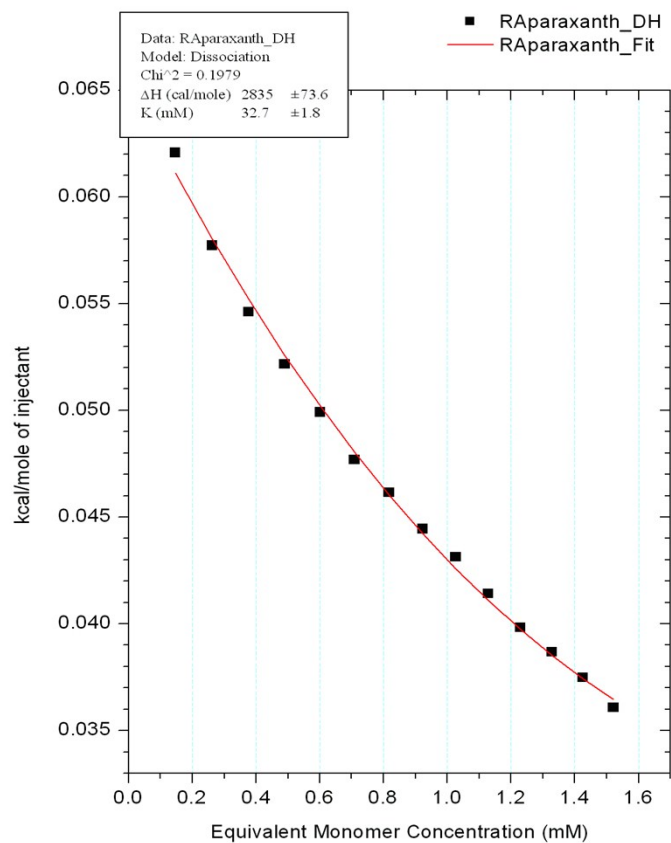


Figure S.10 Integrated heat plots (change in enthalpy against concentration) for the dissociation of paraxanthine in water. The solid line corresponds to the theoretical curve with $K_{\text{diss}} = 32.7 \pm 1.8$ mM and $\Delta H_{\text{diss}} = 2.835 \pm 0.074$ kcal/mol.

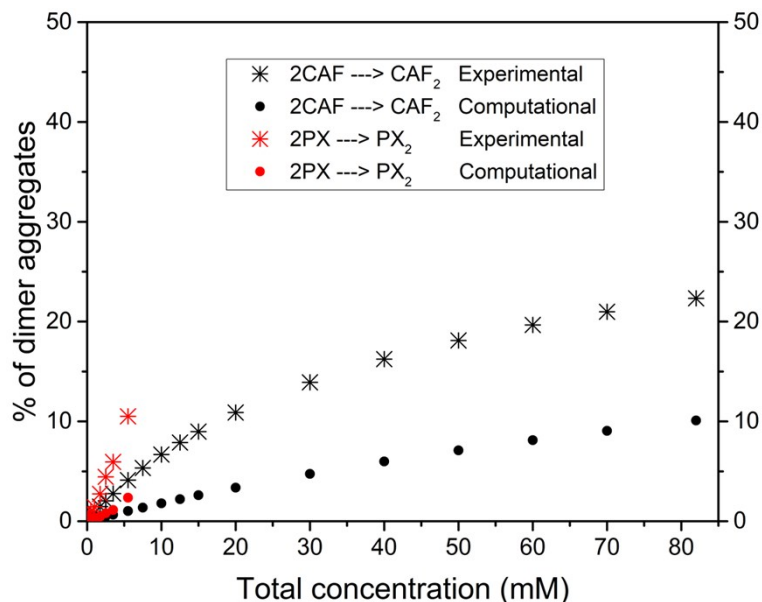


Figure S.11 Comparison between the computational and experimental profiles of the population of dimers in solution at concentrations up to the limit of solubility of paraxanthine (5 mM) and caffeine (82 mM). The computational points were determined using the Gibbs free energy of dimerization in water of the most stable dimers of caffeine (black circles) and paraxanthine (red circles). The experimental points were calculated using the Gibbs free energy of association determined using ITC experiments: caffeine: (black star) and paraxanthine (red star).

S.6 Supplementary references

- 1 J. D. Sutor, *Acta Crystallogr.*, 1965, **19**, 453–458.
- 2 C. W. Lehmann and F. Stowasser, *Chem. A Eur. J.*, 2007, **13**, 2908–2911.
- 3 J. D. Sutor, *Acta Cryst.*, 1958, **11**, 453–458.
- 4 L. A. Burns, Á. Vázquez-Mayagoitia, B. G. Sumpter and C. D. Sherrill, *J. Chem. Phys.*, 2011, **134**, 084107.
- 5 Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2008, **4**, 1849–1868.
- 6 B. J. Lynch, Y. Zhao and D. G. Truhlar, *J. Phys. Chem. A*, 2003, **107**, 1384–1388.
- 7 J. W. Ponder, C. Wu, P. Ren, V. S. Pande, J. D. Chodera, M. J. Schnieders, I. Haque, D. L. Mobley, D. S. Lambrecht, R. A. DiStasio, M. Head-Gordon, G. N. I. Clark, M. E. Johnson and T. Head-Gordon, *J. Phys. Chem. B*, 2010, **114**, 2549–2564.