THE JOURNAL OF
PHYSICAL CHEMISTRY A

Subscriber access provided by Queen Mary, University of London

Article

Density Functional Theory Study of the Oligomerization of Carboxylic Acids

Devis Di Tommaso, and Ken Leo Watson

J. Phys. Chem. A, **Just Accepted Manuscript** • DOI: 10.1021/jp509100u • Publication Date (Web): 30 Oct 2014

Downloaded from http://pubs.acs.org on November 10, 2014

Just Accepted

"Just Accepted" manuscripts have been peer-reviewed and accepted for publication. They are posted online prior to technical editing, formatting for publication and author proofing. The American Chemical Society provides "Just Accepted" as a free service to the research community to expedite the dissemination of scientific material as soon as possible after acceptance. "Just Accepted" manuscripts appear in full in PDF format accompanied by an HTML abstract. "Just Accepted" manuscripts have been fully peer reviewed, but should not be considered the official version of record. They are accessible to all readers and citable by the Digital Object Identifier (DOI®). "Just Accepted" is an optional service offered to authors. Therefore, the "Just Accepted" Web site may not include all articles that will be published in the journal. After a manuscript is technically edited and formatted, it will be removed from the "Just Accepted" Web site and published as an ASAP article. Note that technical editing may introduce minor changes to the manuscript text and/or graphics which could affect content, and all legal disclaimers and ethical guidelines that apply to the journal pertain. ACS cannot be held responsible for errors or consequences arising from the use of information contained in these "Just Accepted" manuscripts.

ACS Publications
High quality. High impact.

The Journal of Physical Chemistry A is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Published by American Chemical Society. Copyright © American Chemical Society. However, no copyright claim is made to original U.S. Government works, or works produced by employees of any Commonwealth realm Crown government in the course of their duties.

Density Functional Theory Study of the Oligomerization of Carboxylic Acids

Devis Di Tommaso* and Ken L. Watson

School of Biological and Chemical Sciences, Queen Mary University of London, 20 Gordon Street, London WC1 H0AJ, United Kingdom

30 October 2014

Corresponding author: Dr. Devis DI TOMMASO, School of Biological and Chemical Sciences, Queen Mary University of London, 20 Gordon Street, London WC1 H0AJ, United Kingdom.

Email: d.ditommaso@qmul.ac.uk. Phone: +44 (0)20 7882 6226

Abstract

We present a density functional theory [M06-2X/6-31+G(d,p)] study of the structures and free energies of formation of oligomers of four carboxylic acids (formic acid, acetic acid, tetrolic acid and benzoic acid) in water, chloroform and carbon tetrachloride. Solvation effects were treated using the SMD continuum solvation model. The low-lying energy structures of molecular complexes were located by adopting an efficient search procedure to probe the potential energy surfaces of the oligomers of carboxylic acids $(CA)_n$ (n = 2-6) The free energies of the isomers of $(CA)_n$ in solution were determined as the sum of the electronic energy, vibrational-rotational-translational gas-phase contribution and solvation free energy. The assessment of the computational protocol adopted in this study with respect to the dimerization of acetic acid, $(AA)_2$, and formic acid, $(FA)_2$, locates new isomers of $(AA)_2$ and $(FA)_2$ and gave dimerization constants in good agreement with the experimental values. The calculation of the self-association of acetic acid, tetrolic acid and benzoic acid shows the following: (i) Classic carboxylic dimers are the most stable isomer of $(CA)_2$ in both the gas phase and solution; (ii) Trimers of carboxylic acid are stable in apolar aprotic solvents; (iii) Molecular clusters consisting of two interacting classic carboxylic dimers $(D+D)_T$ are the most stable type of tetramers but their formation from the self-association of classic carboxylic dimers is highly unfavourable; (iv) For acetic acid and tetrolic acid the reactions $(CA)_2 + 2CA \rightarrow (D+D)_T$ and $(CA)_3 + CA \rightarrow (D+D)_T$ are exoergonic but these aggregation pathways go through unstable clusters that could hinder the formation of tetrameric species; (v) For tetrolic acid the pre-nucleation species that are more likely to form in solution are dimeric and trimeric structures that have encoded structural motifs resembling the α- and β-solid forms of tetrolic acid; (vi) Stable tetramers of benzoic acid could form in carbon tetrachloride from the aggregation of trimers and monomers; (vii) Higher order clusters such as acetic acid pentamers and tetrolic acid hexamers are highly unstable in all solvation environments.

Keywords: Molecular self-association; Free-energy computations; Quantum mechanical continuum solvation models;; Carboxylic oligomers.

1. Introduction

Thermodynamically stable pre-nucleation solvated clusters can have encoded structural motifs resembling the solid state form crystallizing from solution.[1-3] Consequently, the determination of the structure and stability of molecular complexes is of importance to understand and predict the process of polymorph selection during crystallization from solution, that is, the ability of a molecule to crystallize in one specific solid state form depending on the solution composition (nature of the solvent and/or presence of additives in the solution).[4-5] However, in order to determine which molecular clusters are thermodynamically stable in solution, a method for the quantification of their free energy of formation is required.

Computational studies of the aggregation of organic molecules have largely relied on *unbiased* molecular dynamics, by means of which the structural information of molecular aggregates containing between two (dimer) and five (pentamer) units could be attained.[5-8] However, it is not possible using this technique to determine the free energy of formation for molecular clusters in solution. Free energy profiles or free energy surfaces associated with the process of molecular aggregation could be obtained by means of sampling techniques such as umbrella sampling[9,10] and metadynamics,[11,12] but the application of these methods require the identification of a suitable set of collective variables capable of describing the mechanism of self-association of the solute molecules.[9,13,14] However, even for simple aromatic molecules such as benzoic acid, which could interact during the process of self-association *via* OH/O and OH/π hydrogen (H) bonding as well as π-π interactions, the choice of collective variables is not obvious.

An alternative computational approach that can be adopted for the calculation of free energies of molecules and clusters in solution is based on quantum mechanical continuum solvation models,[15-17] in which the solute is described at a quantum mechanical level (*ab initio* or DFT) and the solvent is treated as a polarizable continuum. In particular, the quantum mechanical treatment of molecular associates using modern density functionals specifically developed for non-bonding interactions[18,19] can give an accurate description of H-bonding and van der Waals forces that are important in weakly interacting systems at an accessible computational cost. Moreover, polarizable continuum models represent a practical approach to simulate the solvation environment and determine the effect of the medium to the structure and stability of solutes in solution. This approach has been extensively applied to model the dimerization of organic molecules,[15,20-24] but it has been hardly ever used to quantify the thermodynamic stability in solution of higher order molecular clusters such as trimers, tetramers, pentamers or hexamers. In fact, one of the main challenges in modelling the formation of weakly interacting molecular aggregates is the complexity of their potential energy surfaces (PES), which can be characterized by a large number of local minima separated by low-energy saddle points. Consequently, to obtain an accurate evaluation of the thermodynamic stability of molecular oligomers in solution, the most stable isomers on the PES of these species should be located.

Here we report an approach to calculate the low-lying energy structures and free energies in solution of molecular clusters, where the solute is described at the DFT [M06-2X/6-31+G(d,p)] level of theory,[18,25] the solvent is treated using the SMD continuum solvation model,[26] and an efficient approach is employed to locate the low-lying energy structures on the PES of molecular complexes. We have applied this method to a selected set of molecular systems characterized by a systematic increase in the complexity of the molecular structure and inter-molecular interactions: formic acid, acetic acid, tetrolic acid and benzoic acid (see Table 1). We have chosen formic acid and acetic acid because previous theoretical[13,23,24,27] and experimental[28] investigations of the dimerization of these two molecules can be used as a benchmark to verify the accuracy of the methodology employed in the present work. Tetrolic acid is an example of a molecule that crystallizes in different polymorphic forms, and that exhibits selective crystallization depending on the solvent: the α-form, based on a classic carboxylic motif, crystallizes from apolar aprotic solvents (CCl_4 and $CHCl_3$) and the β-form, based on a catameric chain structure, crystallizes from aqueous and alcoholic solutions.[4] Benzoic acid is a model system that can form complex intermolecular interactions, i.e. OH/O and OH/π hydrogen bonds, all of which are important generally in polymorph formation.

2. Methods

2.1 Electronic structure calculations

All calculations were carried out with the NWChem (version 6.3)[29]and Gaussian09[30] codes using the Minnesota 06 global hybrid density functional with 54% HF exchange (M06-2X) along with the 6-31+G(d,p) basis set. We have chosen M06-2X because its assessment against representative databases showed that this method is one of the most accurate density functionals for a combination of main-group thermochemistry, kinetics and noncovalent interaction.[18,25,31] We have used the 6-31+G(d,p) basis set because when evaluating gas-phase binding energies the mere addition of diffuse functions on nonhydrogenionic atoms to a doubly polarized valence double zeta basis set provides a good compromise between accuracy and computational cost.[18,32] Thermal contributions were calculated using the gas-phase harmonic frequencies at the optimised geometries with the temperature and pressure fixed at the values of 298 K and 1 atm, respectively. The frequencies were scaled by a factor of 0.97 in order to account for systematic errors in the density functional and for anharmonicity.[15,33] Free energies of solvation were calculated using the SMD solvation model[26] and the optimized gas-phase geometries. The SMD model was originally tested over a set of 2892 solvation free energies and transfer free energies for neutral solutes and ions in water and non-aqueous solutions, and the mean unsigned error (MUE) over 26 combinations of various basis sets and density functionals was 0.8 kcal mol^{-1} for neutral solutes, and 4.3 kcal mol^{-1} for ions. In particular, the MUE in calculated solvation free energies of carboxylic acids was only 0.25 kcal mol^{-1} in water and 0.55 kcal mol^{-1} in organic solvents using the M06-2X/6-31+G(d,p) level of theory.[26] The SMD model together with the M06-2X density functional was also applied to predict the free energies of aqueous solvation for 61 drug-like molecules in the SAMPL1 test set and the authors reported a MUE of 2.0 kcal mol^{-1}.[34] The SMD/M06-2X level of theory was also used by Ribeiro and co-workers to compute the partition coefficients of nucleobases between chloroform and water with a MUE of 0.8 kcal mol^{-1}.[15] We have also verified the re-optimization of the carboxylic acid molecules and clusters in the SMD solvation model but the effect on the free energies of association was minor. For example, the Boltzmann average values of the trimerization constants of benzoic acid in water, chloroform and carbon tetrachloride computed according to the aggregation reactions 3BA → $(BA)_3$ and $(BA)_2$ + BA → $(BA)_3$ changed by only 0.1 pK units when the structures of the monomer, dimers and trimers were optimized in the solution.

2.2 Free energies of association in solution

The free energy of formation of carboxylic acid clusters were computed according to the following equation:

$$\Delta G^{*}_{ass} = G^{*}_{AB} - G^{*}_{A} - G^{*}_{B} \quad (1)$$

where G^{*}_{X} is the total Gibbs free energy of the species X ($X = AB$, A or B) in the liquid at 298 K, which was determined (see equation 2) by the addition of

$$G^{*}_{X} = E_{e,gas} + \delta G^{\circ}_{VRT,gas} + \Delta G^{*}_{solv} + RTln[\tilde{R}T] \quad (2)$$

the gas-phase total electronic energy of the solute ($E_{e,gas}$), the vibrational-rotational-translational contribution to the to the gas-phase Gibbs free energy ($\delta G^{\circ}_{VRT,gas}$) at T = 298 K under a standard-state partial pressure of 1 atm, the solvation free energy of the solute corresponding to transfer from an ideal gas at a concentration of 1 M to an ideal solution at a liquid-phase concentration of 1 M (ΔG^{*}_{solv}), and the free energy change of 1 mol of an ideal gas from 1 atm to 1 mol L^{-1} ($RTln[\tilde{R}T]$ = 1.89 kcal mol^{-1} at 298.15 K where $\tilde{R}$ = 0.082052 K^{-1}).[15,35,36] For multiple stationary points of a molecular cluster the free energy was determined as the Boltzmann ensemble average

$$\langle G(X) \rangle = \sum_{i=1}^{N} f_i G(X_i) \quad (3)$$

where f_i is the Boltzmann factor corresponding to the i^{th} isomer, $G(X_i)$ is the corresponding free energy and N is the number of low-lying energy structures located on the PES of X. The Boltzmann factor was determined according to

$$f_i = \frac{e^{-G(X_i)/RT}}{\sum_j e^{-G(X_j)/RT}} \quad (4)$$

where R is the universal gas constant, T is the absolute temperature (T = 298 K) and the index j runs over all isomers.

2.3 Localisation of stationary points on the PES of carboxylic acid clusters

The potential energy surfaces of weakly interacting carboxylic acid complexes, $(CA)_n$, can be characterized by a very large number of minima. However, according to eqns. 3 and 4 only the low-lying energy structures of $(CA)_n$ will have a significant contribution to the Boltzmann average value of the Gibbs free energy. To locate the most stable isomers of a molecular cluster we have adopted the following computational strategy:

i. For each carboxylic acid cluster $(CA)_n$ we generated several thousands of candidate structures using the Granada software,[37] which has been designed to randomly distribute one or more molecules around a central unit (a monomer, dimer, trimer etc.) placed at the centre of a cube of defined side length, and extensively applied to study the microhydration and noncovalent interaction of molecular complexes.[38-43] For example, to generate candidate structures for the trimer $(CA)_3$ we considered the random distribution of one CA molecule around the most stable carboxylic acid dimer (the cyclic dimeric structure), and from the random distribution of two CA molecules around a central carboxylic acid molecule. To sample the potential energy surface of the homologous clusters of the smaller carboxylic acids, formic acid and acetic acid, we simply generated a very large number of configurations (between 7500 and 10000). For tetrolic acid and benzoic acid instead, we initially generated, using the Granada code, 80'000 configurations in a cube of approximately 10 Å, and then selected only those configurations such that the minimum distance between the central molecular unit (a monomer, dimer, trimer, etc.) and the randomly distributed molecule(s) was less than 4.0 Å. To increase the sampling we rounded in excess the number of configurations by gradually increasing the cutoff distance used. The molecular units used to generate the candidate structures were fully optimized at the M06-2X/6-31+G(d,p) level of theory.

ii. The energies of the candidate structures were evaluated at the M06-2X/6-31+G(d,p) level of theory using the electronic structure code NWChem and the Boltzmann factor f_i corresponding to the i^{th} configuration was determined as

$$f_i = \frac{e^{-(E_i - E_0)/RT}}{\sum_j e^{-(E_j - E_0)/RT}} \quad (5)$$

where E_i was the energy of the i^{th} candidate structure and E_0 was the energy of the most stable candidate structure.

iii. We considered the candidate structures with a Boltzmann factor $f_i \geq 0.01$ and, to increase our sampling, ten to fifteen randomly selected structure such that $3 \leq E_i - E_0 \leq 15$ kcal mol^{-1}. These systems were subject to full geometry optimization using the Gaussian 09 code. The thermochemical properties and solvation energies of the optimized gas-phase geometries were also evaluated using Gaussian 09.

The gas phase energies, thermal corrections to the Gibbs free energy, energies in the solvation model (including the SMD non-electrostatic terms), and the geometries of the 772 low-lying energy structures of the carboxylic acids clusters located in this study are listed in Supporting Information.

3 Results and Discussion

3.1 Assessment of the method

In this section we report the structures and energetics of formation of the dimers of acetic acid, $(AA)_2$, and formic acid, $(FA)_2$, in both polar (water) and apolar (chloroform and carbon tetrachloride) solvents. The values of absolute dimerization constants of $(AA)_2$ and $(FA)_2$ have been obtained experimentally,[28] and the potential energy surfaces of these two dimers have been subject to extensive theoretical investigation.[13,23,24,27] The aim of this section is therefore to assess the searching procedure described in section 2.3 to locate the stationary points on the PES of $(CA)_n$ complexes, and the SMD/M06-2X/6-31+G(d,p) level of theory to evaluate the thermodynamic properties in solution of carboxylic acid clusters.

Figure 1 displays the optimized structures and relative energies in the gas-phase of acetic acid dimers. The analysis of the vibrational spectrum of these structures confirmed that they are minima on the PES of $(AA)_2$. Chochouloušová and co-workers previously applied a molecular dynamics / quenching procedure to investigate the PES of the acetic acid dimer and located six low-lying energy isomers of $(AA)_2$ (**I** to **V** and **IX** in Figure 1).[23] Using our computational protocol we could locate isomers **I-V** but not **IX**, which structure was taken from reference 23. On the other hand, we have found three new isomers of $(AA)_2$ (**VI**, **VII** and **VIII** in Figure 1) that are lower in energy than **IX**. The relative energy of **IX** with respect to the global minima **I** is 14 kcal mol^{-1} and we have not locate this isomer probably because only some of the candidate structures with a relative energy lower than 15 kcal mol^{-1} were subject to geometry optimization.

The structures and gas-phase energies of the low-lying structures of the formic acid dimer are shown in Figure 2. We located five of the six isomers of formic acid dimers previously reported by Chochouloušová et al.[27,44] (**I-V**) but we could not determine isomer **VII** ($\Delta E = 13$ kcal mol^{-1}), which structure was taken from reference 27. However, the contribution of **VII** to the Boltzmann average value $\langle G(FA)_2 \rangle$ is negligible. Figure 2 also reports a new isomer of $(FA)_2$ (structure **VI**) that is lower in energy than **VII**. We can therefore conclude that the strategy adopted in this study to probe the potential energy surfaces of molecular clusters is capable of locating the most stable isomers of the oligomers of carboxylic acid.

In Table 2 are compared the values of dimerization constant (pK_D) of acetic acid and formic acid computed at the SMD/M06-2X/6-31+G(d,p) level of theory with those obtained experimentally[45-47,] and by means of free energy (umbrella sampling) molecular dynamics simulations of the formation of $(AA)_2$ and $(FA)_2$ in explicit water.[13] In this work the dimerization constants in solution were computed *via* $K_D = -\log[\exp(\Delta G^*_{ass}/RT)]$, where ΔG^*_{ass} is the free energy of dimerization computed according to eqn. 1 using the Boltzmann averaged free energies of $(AA)_2$ and $(FA)_2$. In water both SMD/M06-2X/6-31+G(d,p) and potential mean force (PMF) methods give values of pK_D that are 0.2 – 1.2 units lower than the experiment, and in chloroform the dimerization constant of acetic acid computed at the M06-2X/6-31+G(d,p) level is 0.5 units lower than the experimental value. Using the methodology employed in this work, the uncertainty in the evaluation of the free energy of association (ΔG^*_{ass}) can be written as the sum of

$$\varepsilon\,(\Delta G^*_{ass}) = \varepsilon\big(\Delta E_{e,gas}\big) + \varepsilon\big(\Delta\,\delta\,G^{\circ}_{VRT,gas}\big) + \varepsilon(\Delta\,\Delta\,G^*_{solv}) \qquad (6)$$

the error in the evaluation of the gas phase interaction energy in the gas phase ($\Delta E_{e,gas}$), the error in the vibrational-rotational-translational contribution to the gas-phase reaction free energy ($\Delta\,\delta\,\mathrm{G}^{\circ}_{\mathrm{VRT,gas}}$), and the difference between the error in the evaluation of the solvation free energies of the product and reactants ($\Delta\triangle\mathrm{G}^{*}_{\mathrm{solv}}$). In particular, for the dimerization process the error in the solvation free energy contribution to the reaction free energy is

$$\varepsilon(\Delta\triangle\mathrm{G}^{*}_{\mathrm{solv}}) = \varepsilon[\triangle\mathrm{G}^{*}_{\mathrm{solv}}(\mathrm{D})] - 2\varepsilon[\triangle\mathrm{G}^{*}_{\mathrm{solv}}(\mathrm{M})] \qquad (7)$$

where $\varepsilon\left[\triangle\mathrm{G}^{*}_{\mathrm{solv}}(\mathrm{D})\right]$ and $\varepsilon\left[\triangle\mathrm{G}^{*}_{\mathrm{solv}}(\mathrm{M})\right]$ are the mean unsigned errors in the calculated solvation free energies of carboxylic acids, which, as discussed in Section 2.2, have been found to be 0.25 kcal mol^{-1} in water and 0.55 kcal mol^{-1} in organic solvents. Therefore, we expect the uncertainty associated with the SMD/M06-2X/6-31+G(d,p) solvation model to be in the range of 0 – 0.75 kcal mol^{-1}. The differences between the calculated and experimental values of the free energy of dimerization of acetic acid $(\mathrm{AA})_2$ and formic acid $(\mathrm{FA})_2$ (0.3 – 1.9 kcal mol^{-1} in Table 2) are therefore consistent with the errors in the evaluation of the free energies of solvation of the monomers and dimers of formic acid and acetic acid.Table 2 also reports the values of pK_D obtained with the counterpoise correction (CPC) of Boys and Bernardi to account for the basis set superposition error (BSSE)[48] during geometry optimization. The constants obtained with (“cpc”) and without (“no-cpc”) the counterpoise method differ by less than 0.2 units. As the cost of calculating the CPC is typically greater than that of calculating the original gas-phase interaction energy,[49] we have not included this term in the evaluation of the gas-phase energy of carboxylic acid clusters. These results indicate that the SMD/M06-2X/6-31+G(d,p) method slightly over-stabilizes the formation of $(\mathrm{AA})_2$ and $(\mathrm{FA})_2$ in solution but is still sufficiently accurate to provide a quite accurate prediction of the thermodynamic stability in solution of carboxylic acid clusters.

3.2 Acetic acid

The free energies of formation in the gas-phase (ΔG°_{ass}) and in polar (water) and non-polar (chloroform and carbon tetrachloride) solvents (ΔG^{*}_{ass}) of acetic acid (AA) clusters consisting of two (dimer), three (trimer), four (tetramer) and five (pentamer) molecular units are listed in Table 3.

The cyclic structure **I**, which is reported in Figure 1, is the most stable dimeric species in both the gas phase and solution. When chloroform and carbon tetrachloride are used the formation of the cyclic dimer is highly exoergonic (ΔG^{*}_{ass}= -3.20 kcal mol^{-1} in $CHCl_3$ and ΔG^{*}_{ass}= -5.87 kcal mol^{-1} in CCl_4). In water the free energy of reaction is -0.18 kcal mol^{-1} but as discussed in section 3.1 our calculation slightly overestimates the interaction between acetic acid molecules. Isomer **VIII** of $(AA)_2$ could also be stable in carbon tetrachloride (ΔG^{*}_{ass}= -0.33 kcal) but all other types of dimers have positive values of the free energy of dimerization (see Table 3). Therefore, according to our simulations, and in agreement with previous studies,[23,24,35] in apolar solvants acetic acid molecules tend to form stable cyclic carboxylic dimers **I**.

To locate the low-lying energy structures of the acetic acid trimer we have generated 7500 candidate $(AA)_3$ structures by considering the aggregation of three monomers [trimers of type $(AA)_{3,(3M)}$] and of the cyclic dimer **I** with one acetic acid molecule [trimers of type $(AA)_{3,(D+M)}$]. The gas-phase energies of these two types of structures, which are reported in Figure 3, are scattered in two different regions and the trimers of type $(AA)_{3,(D+M)}$ are approximately 20 kcal mol^{-1} lower than the trimers of type $(AA)_{3,(3M)}$. From the trimeric structures of type $(AA)_{3,(D+M)}$ we have distinguished eight possible low-lying isomers of $(AA)_3$, which all consist of an acetic acid molecule interacting *via* H-bonding with the carboxylic or methyl groups of the cyclic dimeric unit (see Figure 4). In Table 3 we have listed the energetics of formation of the acetic acid trimers **I-VIII** computed according to the trimerization reactions (i) $(AA)_2$ + AA → $(AA)_{3,(D+M)}$ and (ii) 3AA → $(AA)_{3,(D+M)}$. In chloroform and carbon tetrachloride, acetic acid trimers are stable with respect to non-associated monomers [reaction (ii)] and the lowest energy isomers (structures **I** and **V** in Figure 4) correspond to the favourite aggregation modes for $(AA)_3$ found during molecular dynamics simulations of acetic acid in explicit carbon tetrachloride solvent molecules.[6] Moreover, the thermodynamic stability of acetic acid trimers in CCl_4 supports the findings of Bulmer and Shurvell,[50] who proposed that species larger than dimers should be present in appreciable concentration in solution in order to account for infrared data on the acetic acid – carbon tetrachloride system.

For tetramers we located the isomers of $(AA)_4$ by evaluating the energies of 10000 candidate structures of type $(AA)_{4,(D+D)}$, where D is the cyclic dimer **I** in Figure 1 [structures of type $(AA)_{4,(D+D)}$were found to be considerably lower in energy than those of type $(AA)_{4,(Tr+M)}$ and $(AA)_{4,(D+2M)}$]. As the number of units in a molecular cluster increases, it becomes more difficult to classify an optimized structure to a specific type of isomer, as previously done for the dimeric and trimeric structures of acetic acid. Consequently, the energetics of formation of $(AA)_4$ in Table 3 are the Boltzmann average values of the energies and free energies of 47 optimized tetramers, which have been computed according to the following reactions: (i) $(AA)_2$ + $(AA)_2$ → $(AA)_{4,(D+D)}$; (ii) $(AA)_3$ + AA → $(AA)_{4,(D+D)}$; $(AA)_2$ + 2AA → $(AA)_{4,(D+D)}$. Representative low-lying structures of type $(AA)_{4,(D+D)}$ of acetic acid tetramers are reported in Figure 5. The free energy of self-association of two cyclic dimers **I** [reaction (i)] is endoergonic (ΔG > 0) in all media but reactions (ii) (ΔG^*_{ass}= -1.36 kcal mol^{-1}) and (iii) (ΔG^*_{ass}= -0.56 kcal mol^{-1}) are exoergonic (ΔG < 0) in carbon tetrachloride, which would suggest that acetic acid tetramers exist in apolar solution. However, the aggregation pathways $(AA)_3$ + AA → $(AA)_{4,(D+D)}$ and $(AA)_2$ + 2 AA → $(AA)_{4,(D+D)}$ do not take into account that in solutions containing monomers, dimers and trimers we could have first the formation of intermediate tetrameric structures of type $(AA)_{4,(D+2M)}$ or $(AA)_{4,(Tr+M)}$, followed by the process of reorganization of the molecules in the clusters to give the lower free energy species $(AA)_{4,(D+D)}$. The free energy differences between the tetramers of type $(AA)_{4,(Tr+M)}$, $(AA)_{4,(D+2M)}$ and $(AA)_{4,(D+D)}$ are reported in Figure 6, where the values were obtained from the Boltzmann average free energies of the isomers of $(AA)_{4,(Tr+M)}$, $(AA)_{4,(D+2M)}$ and $(AA)_{4,(D+D)}$, respectively. Figure 6 shows that the formation of tetramers of type $(AA)_{4,(D+2M)}$ (ΔG^*_{ass}= 2.65 kcal mol^{-1} in CCl_4) and $(AA)_{4,(Tr+M)}$ (ΔG^*_{ass}= 1.17 kcal mol^{-1} in CCl_4) is not favourable, and, consequently, the formation of the tetrameric structures $(AA)_{4,(D+D)}$could be hindered by the thermodynamic instability of the intermediate species $(AA)_{4,(D+2M)}$and $(AA)_{4,(Tr+M)}$. Moreover, there could also be an activation barrier associated with the reorganization of acetic acid molecules from the molecular arrangement in the tetramers $(AA)_{4,(D+2M)}$ or $(AA)_{4,(Tr+M)}$to the arrangement observed in the $(AA)_{4,(D+D)}$species (see Figure 5).

The energetics of formation of pentamers in Table 3, which representative most stable structures are reported in Figure 7, is highly endoergonic in all solvation environments and this higher order species is unlikely to be found in solution. To probe the PES and locate the low-lying isomers of $(AA)_5$ we evaluated the energies of 10000 candidate structures of type $(AA)_{5,(Tr+D)}$, and the energetics in Table 3 are the Boltzmann average values of the energies of 46 optimized pentameric structures.

3.3 Tetrolic acid

The free energies of formation of dimers, trimers, tetramers and hexamers of tetrolic acid (TTA) are listed in Table 4. We located nine isomers on the PES of the tetrolic acid dimer, which structures are reported in Figure 8. The cyclic structure **I** is the only dimeric species that is stable in chloroform (ΔG^{*}_{ass} = -1.80 kcal mol^{-1}) and carbon tetrachloride (ΔG^{*}_{ass} = -4.62 kcal mol^{-1}) but not in water (ΔG^{*}_{ass} = 0.85 kcal mol^{-1}). Isomer **I** corresponds to the structural synthon found in the polymorph of tetrolic acid that crystallizes from chloroform and carbon tetrachloride (α-TTA).[4,51] The formation of all other dimers is endoergonic in both the gas-phase and solution but it is interesting to notice that the second most stable isomer of $(TTA)_2$ (structure **II** in Figure 8) corresponds to the catemeric unit found in the β form of TTA.[51]

The energetics of formation of the trimers in Table 4 represent the Boltzmann average of the energies and free energies of 40 low-lying structures, which were selected from 6000 candidate structures of type $(3M)_{Tr}$ and 3000 candidate structures of type $(D+M)_{Tr}$, where D and M are the optimised geometries of the cyclic dimer **I** and of the tetrolic acid molecule, respectively. The most stable structures of $(TTA)_3$, which are reported in Figure 9, show that the isomers of the tetrolic acid trimer consist of a cyclic dimer **I** interacting with a tetrolic acid molecule and, similarly to what found for acetic acid, in the most stable isomer of $(AA)_3$ (structure **I** in Figure 9) the carboxylic acid molecule forms both an O-H···O and an C-H···O bond to the dimer. Notice also that species **I** corresponds to the structural unit found in the β form of TTA.[51]

Tetramers of tetrolic acid could form from the aggregation of two cyclic dimer **I**, $(TTA)_{4,(D+D)}$, a cyclic dimer **I** with two tetrolic acid molecules $(TTA)_{4,(D+2M)}$, and the trimer **I** with one tetrolic acid molecule $(TTA)_{4,(Tr+M)}$. However, Figure 10 shows that tetramers of type $(TTA)_{4,(D+D)}$are significantly lower in energy than tetramers of type $(TTA)_{4,(D+2M)}$ and $(TTA)_{4,(Tr+M)}$. Some representative $(TTA)_{4,(D+D)}$structures are reported in Figure 11 where we have highlighted some of the H-bonding between the cyclic dimeric units that stabilize the clusters. The energetics of formation of $(TTA)_{4,(D+D)}$in Table 4, which have been obtained from the Boltzmann average of the energies of 70 minima on the PES of $(TTA)_{4,(D+D)}$, indicate that the self-association of tetrolic acid dimers [reaction $(TTA)_2 + (TTA)_2 \rightarrow (TTA)_4$ in Table 4] is endoergonic, but the formation of tetramers could occur in carbon tetrachloride only along the aggregation pathway $(TTA)_3 + (TTA) \rightarrow (TTA)_4$ (ΔG^*_{ass} = -1.35 kcal mol^{-1}). However, this reaction does not take into account the energetics required for the formation of intermediate clusters of type $(TTA)_{4,(Tr+M)}$, which could form when a trimer **I** interacts with a tetrolic acid molecule. In Figure 12 the free energy profiles along the reaction pathways (i) $(TTA)_3 + (TTA) \rightarrow (TTA)_4$ and (ii) $(TTA)_2 + 2\ (TTA) \rightarrow (TTA)_4$ indicate that species of type $(TTA)_{4,(Tr+M)}$and $(TTA)_{4,(D+2M)}$are unstable in solution and they could therefore hinder the formation of the thermodynamically stable $(TTA)_{4,(D+D)}$tetramers.

Table 4 also reports the Boltzmann average values of the energetics of formation of hexamers, which low lying energy structures have been located by considering 8000 structures of type $(TTA)_{6,(3D)}$, where D was the cyclic dimer **I**. The values in Table 4 indicate that the formation of $(TTA)_6$ species is highly endoergonic in all solvation environments.

Based on our simulations we therefore suggest that in homogeneous tetrolic acid aqueous solutions the monomers of TTA dominate, whereas in apolar aprotic solvents, like carbon tetrachloride or chloroform, tetrolic acid molecules form stable dimeric (**I** in Figure 8) and trimeric (**I** in Figure 9) species that have structural features resembling the α and β polymorphic forms of TTA.

3.4 Benzoic acid

The energetics of formation of the dimers, trimers and tetramers of benzoic acid (BA) in the gas-phase and in solution are reported in Table 5. Molecules of benzoic acid have the ability to form H-bonds, π–π and OH/π interactions and therefore it would have been difficult to locate the most stable isomers of $(BA)_n$ clusters only by means of chemical intuition.

Using the computational protocol adopted in this work we located eighteen isomers on the PES of $(BA)_2$ and the structure of the six most stable benzoic acid dimers are displayed in Figure 13. The energetics of the dimerization reaction indicate that in chloroform (ΔG^*_{ass} = -2.97 kcal mol^{-1}) and carbon tetrachloride (ΔG^*_{ass} = -7.77 kcal mol^{-1}) the classic carboxylic structure **I** is thermodynamically very stable. Isomers **III** (ΔG^*_{ass} = -0.75 kcal mol^{-1}) and **IV** (ΔG^*_{ass} = -0.56 kcal mol^{-1}), which consist of two molecules interacting *via* a single H-bonding, could also form in carbon tetrachloride. For all other dimeric structures, where the benzoic acid molecules interact *via* weaker π-π and OH/π interactions (**II**, **V** and **VI** in Figure 13), the energetics of dimerization is positive in both the gas phase and solution (see Table 5).

To locate the low-lying energy structures of benzoic acid trimers we generated and computed the energies of 10000 candidate structures of type $(3M)_{Tr}$ and $(D+M)_{Tr}$. The energetics of trimerization reaction in Table 5 represents the Boltzmann average values of 36 low-lying $(BA)_3$ isomers. The aggregation of the cyclic dimer **I** with a benzoic acid molecule $[(BA)_2 + BA \rightarrow (BA)_3]$ is not thermodynamically favourable, but in carbon tetrachloride the free energy change for the tetramerization reaction $3(BA) \rightarrow (BA)_3$ is highly exoergonic (-5.20 kcal mol^{-1}). This result agrees with experiments of apolar (benzene) solutions of benzoic acids by Goralski and Taniewska, who found that the presence of monomers, dimers as well as trimers were required to account for osmometric data.[52] We have verified the stability of benzoic acid trimers in benzene by computing the energetics of the benzoic acid – benzene system, and the resulting Boltzmann average free energy of trimerization reaction 3(BA) → $(BA)_3$ is negative (-1.94 kcal mol^{-1}). On the other hand, recent density functional theory [PCM/B3P86/6-31G(d,p)] calculations of the trimerization of benzoic acid in benzene did not support the formation of substantial amount of the trimeric phase.[53] However, in this study the authors considered a single $(BA)_3$ structure, which gas-phase stabilization energy ($\Delta E_{e,gas}$ = -22.58 kcal mol^{-1})[53] is lower than the Boltzmann average value reported in Table 5 ($\Delta E_{e,gas}$ = -28.20 kcal mol^{-1}). In fact, several $(BA)_3$ isomers located in this work are significantly more stable than the structure reported by Pham and co-workers,[53] where three benzoic molecules interacting *via* the three carboxylic groups, such as isomer **V** in Figure 14, which consist of a benzoic acid molecule interacting *via* H-bonding as well as π–π interaction with the cyclic dimer **I**. We would like also to mention that using our searching technique we determined several isomers with a gas-phase stabilization energy in the range [-15.6, -23.9] kcal mol^{-1} [the structures and energetics of the trimers of type $(BA)_{3,(D+M)}$ and $(BA)_{3,(3M)}$ reported in Supporting Information]. Moreover, it is also important to stress that when using DFT methods to compute relative energies of weakly bounded complexes such as $(CA)_n$ the inclusion of diffuse functions on non-hydrogenionic atoms is crucial,[32,54] and the 6-31G(d,p) basis set employed by Pham and co-workers to evaluate the gas-phase reaction energies could have not provided a sufficient accuracy for the evaluation of the energetics of dimerization and trimerization reactions of benzoic acid.

For the tetramers we determined the structure and free energies of 22 $(BA)_4$ low-lying energy structures, which were selected from 5000 candidate structures of type $(BA)_{4,(D+D)}$, where D is the benzoic acid cyclic dimer **I** in Figure 13. The isomers of $(BA)_4$ consist of two dimeric units arranged in a stacked configuration, which allows the formation of both H-bonding and/or π-π interactions between the molecules of benzoic acid (see Figure 15). In chloroform and carbon tetrachloride, where benzoic acid molecules form stable classic carboxylic species, the aggregation of two dimers [$(BA)_2 + (BA)_2 \rightarrow (BA)_{4,(D+D)}$] is endoergonic ($\Delta G^{\circ}_{ass,CHCl_3}$ = 1.65 kcal mol^{-1} and $\Delta G^{\circ}_{ass,CCl_4}$ = 3.04 kcal mol^{-1}), but according to the values reported in Table 5, the $(BA)_{4,(D+D)}$tetramers could form along the reaction pathways $(BA)_2 + 2\ BA \rightarrow (BA)_4$ and $(BA)_3 + BA \rightarrow (BA)_4$. Figure 16 reports the relative free energies of formation of the molecular aggregates $(BA)_{4,(D+D)}$, $(BA)_{4,(D+2M)}$ and $(BA)_{4,(Tr+2M)}$, where the values have been obtained from the Boltzmann average of the free energies of the molecular clusters of type $(BA)_{4,(D+D)}$(tetramers consisting of two benzoic acid cyclic dimer **I** units), $(BA)_{4,(Tr+2M)}$ (tetramers containing the most stable trimer in CCl_4, isomer **I** in Figure 14), and $(BA)_{4,(D+2M)}$. Structures of type $(BA)_{4,(D+D)}$ represent the lowest energy molecular aggregates but stable intermediate clusters of type $(BA)_{4,(Tr+2M)}$could also form from the aggregation of trimers and monomers [$(BA)_3 + BA \rightarrow (BA)_4$] in apolar solents. On the other hand, the formation of the tetramers $(BA)_{4,(D+D)}$along the reaction pathway $(BA)_2 + 2\ BA \rightarrow (BA)_4$ is thermodynamically hindered by the instability of species of type $(BA)_{4,(D+2M)}$.

4. Summary and Conclusions

Density functional theory [M06-2X/6-31+G(d,p)] continuum solvation (SMD) calculations were used to determine the structure, thermodynamic stability and free energy pathways for the formation of carboxylic acid clusters: formic acid $(FA)_n$ (n = 2), acetic acid $(AA)_n$ (n = 2-5), tetrolic acid $(TTA)_n$ (n = 2-6) and benzoic acid $(BA)_n$ (n = 2-4). The low-lying energy structures of weakly interacting molecular complexes were located by adopting a search procedure to probe the potential energy surface of $(CA)_n$. The free energies of the low-lying energy isomers of $(CA)_n$ in solution were determined as the sum of the electronic energy, vibrational-rotational-translational gas-phase contribution and solvation free energy. Based on the results we conclude the following:

i. The application of the computational approach adopted in the present study to compute the dimerization of acetic acid and formic acid resulted in the location of three new low lying isomers of $(AA)_2$ and one new isomer of $(FA)_2$. The Boltzmann averaged dimerization constants (pK_D) of $(AA)_2$ and $(FA)_2$ were in good agreement with the values of pK_D obtained experimentally.

ii. The most stable isomer of the carboxylic acid dimer, in both the gas phase and solution, is the classic carboxylic species, which formation is endoergonic in apolar aprotic solvents like chloroform and carbon tetrachloride.

iii. For all carboxylic acids we found that the formation of trimers according to reaction $3CA \rightarrow (CA)_3$ is thermodynamically spontaneous in chloroform and carbon tetrachloride. The stability of $(CA)_3$ species in solution support experimental data of acetic acid – carbon tetrachloride[50] and benzoic acid – benzene benzoic[52] systems, which suggested that trimers are present in appreciable concentrations in apolar aprotic solvents.

iv. The structure of the most stable trimer of tetrolic acid consist of the cyclic dimeric unit, corresponding to the structural synthon found in the α polymorph of tetrolic acid, which forms both an O-H···O and an C-H···O bond to a tetrolic acid molecule and that corresponds to the catameric unit found in the β-TTA.[28]

v. For all carboxylic acid, the most stable tetrameric structures consist of two interacting cyclic dimers, $(CA)_{4,(D+D)}$, but the self-association reaction $(CA)_2 + (CA)_2 \rightarrow (CA)_{4,(D+D)}$is endoergonic in all solvation environments.

vi. Tetramers of type $(CA)_{4,(D+D)}$could form in carbon tetrachloride and chloroform along the reaction pathways $(CA)_2 + 2\ (CA) \rightarrow (CA)_4$ and $(CA)_3 + CA \rightarrow (CA)_4$. For acetic acid and tetrolic acid these aggregation reactions involve the formation of intermediate clusters of type $(CA)_{4,(D+2M)}$and $(CA)_{4,(Tr+M)}$that are unstable in solution. For benzoic acid tetrameric structures could also form from the aggregation of trimers and monomers in carbon tetrachloride and chloroform $[(BA)_3 + BA \rightarrow (BA)_4]$.

The free energy pathways for the formation of the dimers, trimers and tetramers of acetic acid, tetrolic acid and benzoic acid as computed at the SMD/M06-2X/6-31+G(d,p) level of theory are summarized in Figure 17.

vii. Higher order clusters such as acetic acid pentamers and tetrolic acid hexamers are highly unstable in all solvation environments.

In conclusion, this work proposes the application of a computational approach for the determination of the structure and stability in solution of molecular clusters. We conducted all calculations using an accurate, but computationally demanding, hybrid density functional method (M06-2X). As all calculation of the thermodynamic properties of weakly interacting carboxylic acid molecular clusters should consider the location of the low-lying isomers other than the deepest minimum structures, the application of this method to large and/or conformationally flexible molecules would require prohibiting computational resources. However, a hybrid methodology could be adopted where the gas-phase contributions to the free energy ($E_{e,gas} + \delta G^{\circ}_{VRT,gas}$ in eqn. 2) is computed using a forcefield model, which accuracy could be assessed with respect to the structures and energetics reported in this study, and the solvation free energy term is evaluated using the accurate SMD model (ΔG^{*}_{solv} in eqn. 3).

Acknowledgements

D.D.T. thanks the U.K.'s Royal Society for the award of a Royal Society Industry Fellowship. This research utilised Queen Mary's MidPlus computational facilities, supported by QMUL Research-IT and funded by EPSRC grant EP/K000128/1. Via our membership of the UK's HPC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), this work made use of the facilities of HECToR and ARCHER the UK's national high-performance computing service, which is funded by the Office of Science and Technology through EPSRC's High End Computing Programme. The authors would like to thank Prof. Luis A. Montero-Cabrera and Dr. Cercis Morera-Boado (Universidad de La Habana, Havana) for providing us with a copy of the Granada code.

Supporting Information Available

The gas phase energies, thermal corrections to the Gibbs free energy, energies in the solvation model (including the SMD non-electrostatic terms), and the geometries of the 772 low-lying energy structures of the carboxylic acids clusters located in this study are listed in Supporting Information. This information is free of charge via the Internet at http://pubs.acs.org

References

(1) Gebauer, D.; Kellermeier, M.; Gale, J. D.; Bergstrom, L.; Colfen, H. Pre-nucleation Clusters as Solute Precursors in Crystallisation. *Chem. Soc. Rev.* **2014**, *43*, 2348-2371

(2) Davey, R. J.; Blagden, N.; Righini, S.; Alison, H.; Quayle, M. J.; Fuller, S. Crystal Polymorphism as a Probe for Molecular Self-assembly During Nucleation from Solutions: The Case of 2,6-dihydroxybenzoic Acid. *Cryst Growth Des* **2001**, *1*, 59-65.

(3) Davey, R. J.; Schroeder, S. L. M.; ter Horst, J. H. Nucleation of Organic Crystals - A Molecular Perspective. *Angew. Chem. Int. Edit.* **2013**, *52*, 2-16.

(4) Parveen, S.; Davey, R. J.; Dent, G.; Pritchard, R. G. Linking Solution Chemistry to Crystal Nucleation: The Case of Tetrolic acid. *Chem. Commun.* **2005**, 1531-1533.

(5) Hamad, S.; Moon, C.; Catlow, C. R. A.; Hulme, A. T.; Price, S. L. Kinetic Insights into the Role of the Solvent in the Polymorphism of 5-fluorouracil from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2006**, *110*, 3323-3329.

(6) Gavezzotti, A. Molecular Aggregation of Acetic Acid in a Carbon Tetrachloride Solution: A Molecular Dynamics Study with a View to Crystal Nucleation. *Chem. Eur. J.* **1999**, *5*, 567-576.

(7) Hamad, S.; Hughes, C. E.; Catlow, C. R. A.; Harris, K. D. M. Clustering of Glycine Molecules in Aqueous Solution Studied by Molecular Dynamics Simulation. *J. Phys. Chem. B* **2008**, *112*, 7280-7288.

(8) Gavezzotti, A.; Filippini, G.; Kroon, J.; vanEijck, B. P.; Klewinghaus, P., The Crystal Polymorphism of Tetrolic Acid: A Molecular Dynamics Study of Precursors in Solution, and a Crystal Structure Generation. *Chem. Eur. J.* **1997**, *3*, 893-899.

(9) Chen, J.; Trout, B. L., Computational Study of Solvent Effects on the Molecular Self-assembly of Tetrolic Acid in Solution and Implications for the Polymorph Formed from Crystallization. *J. Phys. Chem. B* **2008**, *112*, 7794-7802.

(10) Kästner, J., Umbrella Sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011**, *1*, 932-942.

(11) Laio, A.; Parrinello M. Escaping Free-energy Minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562-12566.

(12) Laio, A.; Gervasio, F. L. Metadynamics: a Method to Simulate Rare Events and Reconstruct the Free Energy in Biophysics, Chemistry and Material Science. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(13) Chen, J. H.; Brooks, C. L.; Scheraga, H. A. Revisiting the Carboxylic Acid Dimers in Aqueous Solution: Interplay of Hydrogen Bonding, Hydrophobic Interactions, and Entropy. *J. Phys. Chem. B* **2008**, *112*, 242-249.

(14) Salvalaglio, M.; Vetter, T.; Giberti, F.; Mazzotti, M.; Parrinello, M. Uncovering Molecular Details of Urea Crystal Growth in the Presence of Additives. *J. Am. Chem. Soc.* **2012**, *134*, 17221-17233.

(15) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G., The Solvation, Partitioning, Hydrogen Bonding, and Dimerization of Nucleotide Bases: a Multifaceted Challenge for Quantum Chemistry. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10908-10922.

(16) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999-3093.

(17) Cramer, C. J.; Truhlar, D. G., Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chem. Rev.* **1999**, *99*, 2161-2200.

(18) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-class Functionals and 12 Other Functionals. *Theoretical Chemistry Accounts* **2008**, *120*, 215-241.

(19) Klimeš, J.; Michaelides, A., Perspective: Advances and Challenges in Treating van der Waals Dispersion Forces in Density Functional Theory. *J Chem Phys* **2012**, *137*, 12.

(20) Friant-Michel, P.; Ruiz-Lopez, M. F., Glycine Dimers: Structure, Stability, and Medium Effects. *ChemPhysChem* **2010**, *11*, 3499-3504.

(21) Ramondo, F.; Bencivenni, L.; Caminiti, R.; Pieretti, A.; Gontrani, L. Dimerisation of Urea in Water Solution: a Quantum Mechanical Investigation. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2206-2215.

(22) Di Tommaso, D. The Molecular Self-association of Carboxylic Acids in Solution: Testing the Validity of the Link Hypothesis Using a Quantum Mechanical Continuum Solvation Approach. *CrystEngComm* **2013**, *15*, 6564-6577.

(23) Chocholousova, J.; Vacek, J.; Hobza, P. Acetic Acid Dimer in the Gas Phase, Nonpolar Solvent, Microhydrated Environment, and Dilute and Concentrated Acetic Acid: Ab Initio Quantum Chemical and Molecular Dynamics Simulations. *J. Phys. Chem. A* **2003**, *107*, 3086-3092.

(24) Pasalic, H.; Aquino, A. J. A.; Tunega, D.; Haberhauer, G.; Gerzabek, M. H.; Georg, H. C.; Moraes, T. F.; Coutinho, K.; Canuto, S.; Lischka, H. Thermodynamic Stabillity of Hydrogen-Bonded Systems in Polar and Nonpolar Environments. *J. Comput. Chem.* **2010**, *31*, 2046-2055.

(25) Zhao, Y.; Truhlar, D. G., Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J. Chem. Theory Comput.* **2008**, *4*, 1849-1868.

(26) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Model Based on Solute Electron Density and on a Continuum Model of the Solvent Defined by the Bulk Dielectric Constant and Atomic Surface Tensions. *J. Phys. Chem. B* **2009**, *113*, 6378-6396.

(27) Chocholousova, J.; Vacek, J.; Hobza, P. Potential Energy and Free Energy Surfaces of the Formic Acid Dimer: Correlated Ab Initio Calculations and Molecular Dynamics Simulations. *Phys. Chem. Chem. Phys.* **2002**, *4*, 2119-2122.

(28) Schrier, E. E.; Robinson, R. A. Study of Free Energy Relationships in Some Amino Acid-Sodium Chloride-Water Systems. *J. Biol. Chem.* **1971**, *246*, 2870-2874.

(29) Valiev, M.; Bylaska, E. J.; Govind, N.; Kowalski, K.; Straatsma, T. P.; Van Dam, H. J. J.; Wang, D.; Nieplocha, J.; Apra, E.; Windus, T. L. et al. NWChem: A Comprehensive and Scalable Open-source Solution for Large Scale Molecular Simulations. *Comput. Phys. Commun.* **2010**, *181*, 1477-1489.

(30) Gaussian 09, Revision C.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A. et al. Gaussian, Inc., Wallingford CT, **2009**.

(31) Riley, K. E.; Pitonak, M.; Jurecka, P.; Hobza, P. Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories. *Chem. Rev.* **2010**, *110*, 5023-5063.

(32) Lynch, B. J.; Zhao, Y.; Truhlar, D. G. Effectiveness of Diffuse Basis Functions for Calculating Relative Energies by Density Functional Theory. *J. Phys. Chem. A* **2003**, *107*, 1384-1388.

(33) Alecu, I. M.; Zheng, J. J.; Zhao, Y.; Truhlar, D. G. Computational Thermochemistry: Scale Factor Databases and Scale Factors for Vibrational Frequencies Obtained from Electronic Model Chemistries. *J. Chem. Theory Comput.* **2010**, *6*, 2872-2887.

(34) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Performance of SM6, SM8, and SMD on the SAMPL1 Test Set for the Prediction of Small-Molecule Solvation Free Energies. *J. Phys. Chem. B* **2009**, *113*, 4538-4543.

(35) Colominas, C.; Teixido, J.; Cemeli, J.; Luque, F. J.; Orozco, M. Dimerization of Carboxylic Acids: Reliability of Theoretical Calculations and the Effect of Solvent. *J. Phys. Chem. B* **1998**, *102*, 2269-2276.

(36) Tang, E.; Di Tommaso, D.; de Leeuw, N. H. Accuracy of the Microsolvation-Continuum Approach in Computing the pK(a) and the Free Energies of Formation of Phosphate Species in Aqueous Solution. *Phys Chem Chem Phys* **2010**, *12*, 13804-13815.

(37) Montero, L. A. GRANADA and Q programs for PC computers. 1996.

(38) Codorniu-Hernandez, E.; Mesa-Ibirico, A.; Montero-Cabrera, L. A.; Martinez-Luzardo, F.; Borrmann, T.; Stohrer, W. D. Theoretical Study of Flavonoids and Proline Interactions. Aqueous and Gas Phases. *J. Mol. Struc. (Theochem)* **2003**, *623*, 63-73.

(39) Montero, L. A.; Molina, J.; Fabian, J. Multiple Minima Hypersurfaces of Water Clusters for Calculations of Association Energy. *Int. J. Quantum. Chem.* **2000**, *79*, 8-16.

(40) Sanchez-Garcia, E.; Mardyukov, A.; Studentkowski, M.; Montero, L. A.; Sander, W. Furan-Formic Acid Dimers: An Ab Initio and Matrix Isolation Study. *J. Phys. Chem. A* **2006**, *110*, 13775-13785.

(41) Sanchez-Garcia, E.; Mardyukov, A.; Tekin, A.; Crespo-Otero, R.; Montero, L. A.; Sander, W.; Jansen, G. Ab Initio and Matrix Isolation Study of the Acetylene-Furan Dimer. *Chem Phys* **2008**, *343*, 168-185.

(42) Sanchez-Garcia, E.; Montero, L. A.; Sander, W. Computational Study of Noncovalent Complexes Between Formamide and Formic Acid. *J. Phys. Chem. A* **2006**, *110*, 12613-12622.

(43) Sanchez-Garcia, E.; Studentkowski, M.; Montero, L. A.; Sander, W. Noncovalent complexes between dimethyl ether and formic acid - An ab initio and matrix isolation study. *ChemPhysChem* **2005**, *6*, 618-624.

(44) Madeja, F.; Havenith, M.; Nauta, K.; Miller, R. E.; Chocholousova, J.; Hobza, P., Polar Isomer of Formic Acid Dimers Formed in Helium Nanodroplets. *J. Chem. Phys.* **2004**, *120*, 10554-10560.

(45) Pistolis, G.; Paleos, C. M.; Malliaris, A. Molecular Recognition in Organic-Solvents - the Importance of Excimer Fluorescence Spectroscopy. *J. Phys. Chem.* **1995**, *99*, 8896-8902.

(46) Sellini, H.; Maurizot, J. C.; Dimicoli, J. L.; Helene, C. Hydrogen-Bonding of Amino-Acid Side Chains to Nucleic-Acid Bases. *Febs. Lett.* **1973**, *30*, 219-224.

(47) Schrier, E. E.; Pottle, M.; Scheraga, H. A. The Influence of Hydrogen and Hydrophobic Bonds on the Stability of the Carboxylic Acid Dimers in Aqueous Solution. *J. Am. Chem. Soc.* **1964**, *86*, 3444-3449.

(48) Boys, S. F.; Bernardi, F. Calculation of Small Molecular Interactions by Differences of Separate Total Energies - Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553-566.

(49) Deng, J.; Gilbert, A. T. B.; Gill, P. M. W. Communication: Efficient Counterpoise Corrections by a Perturbative Approach. *J. Chem. Phys.* **2011**, *135*, 081105.

(50) Bulmer, J. T.; Shurvell, H. F. Factor-Analysis as a Complement to Band Resolution Techniques .1. Method and Its Application to Self-Association of Acetic-Acid. *J. Phys. Chem.* **1973**, *77*, 256-262.

(51) Benghiat, V.; Leiserow. L Molecular Packing Modes. 6. Crystal and Molecular Structures of 2 Modifications of Tetrolic Acid. *J. Chem. Soc. Perk. Trans 2* **1972**, 1763-1768.

(52) Goralski, P.; Taniewskaosinska, S. On the Analysis of Calorimetric Data in the Study of Dimerization of Carboxylic-Acids in Non-Polar Solvents. *Can. J. Chem.* **1983**, *61*, 2209-2213.

(53) Pham, H. H.; Taylor, C. D.; Henson, N. J. First-Principles Prediction of the Effects of Temperature and Solvent Selection on the Dimerization of Benzoic Acid. *J. Phys. Chem. B* **2013**, *117*, 868-876.

(54) Di Tommaso, D.; de Leeuw, N. H. The Onset of Calcium Carbonate Nucleation: A Density Functional Theory Molecular Dynamics and Hybrid Microsolvation/Continuum Study. *J. Phys. Chem. B* **2008**, *112*, 6965-6975.

Tables

Table 1. Simulation systems investigated in the present work.

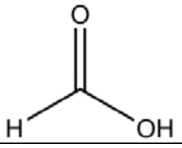
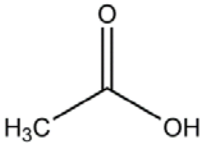
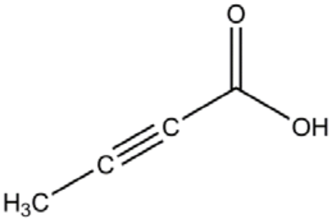
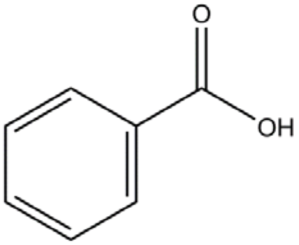
Formic acid (FA)	
Acetic acid (AA)	
Tetrolic acid (TTA)	
Benzoic acid (BA)	

Table 2. Calculated and experimental values of the dimerization constants (pK_D) of acetic acid $(AA)_2$ and formic acid $(FA)_2$.

	solvent	M06-2X (no-cpc)[a]	M06-2X (cpc)[b]	PMF[c]	exptl
$(AA)_2$	H_2O	-0.25	-0.32	0.038 ± 0.01	0.0-1.2[d]
	$CHCl_3$	-2.33	-2.52		-1.8[e]
	CCl_4	-4.27	-4.46		
$(FA)_2$	H_2O	0.79	0.76	0.25 ± 0.02	1.2-2.1[f]
	$CHCl_3$	-1.13	-1.14		
	CCl_4	-3.18	-3.19		

[a] We use “no-cpc” to denote calculations without counterpoise correction for the basis set superposition error (BSSE). [b] We use “cpc” to denote geometry optimization with counterpoise correction for BSSE. [c] Taken from Ref [13]. [d] Taken from Ref [45]. [e] Taken from Ref [46]. [f] Taken from Ref [47].

Table 3. Energetics [kcal mol^{-1}] of formation of acetic acid (AA) clusters as computed at the M06-2X/6-31+G(d,p) level of theory: $\Delta E_{e,gas}$ is the gas phase interaction energy; ΔG°_{ass} is the standard state (1 atm) gas-phase free energy of association at 298 K; ΔG^{*}_{ass} is the standard state (1 mol L^{-1}) free energies of reaction in the liquid-phase.

species	**reaction**	**type**	$\Delta \boldsymbol{E}_{e,gas}$	$\Delta \boldsymbol{G}^{\circ}_{ass}$	$\Delta \boldsymbol{G}^{*}_{ass}$		
					H_2O	**$CHCl_3$**	**CCl_4**
dimer	AA + AA → $(AA)_2$	I	-17.70	-6.56	-0.18	-3.20	-5.87
		II	-9.19	1.06	4.69	3.00	1.13
		III	-9.96	0.43	3.81	2.30	0.71
		IV	-7.16	3.03	5.09	4.14	3.00
		V	-7.16	1.97	2.93	2.14	1.26
		VI	-10.05	0.63	3.79	2.30	0.91
		VII	-6.36	4.76	5.37	5.20	4.77
		VIII	-10.07	-0.10	2.61	1.22	-0.33
		IX	-4.66	5.08	5.97	5.66	5.20
trimer	$(AA)_2$ + AA + → $(AA)_3$ [a]	I	-9.36	0.59	3.02	2.02	0.75
		II	-9.06	2.63	4.71	2.88	2.72
		III	-5.73	4.21	4.52	2.69	3.52
		IV	-7.78	3.14	4.66	2.83	2.87
		V	-7.99	2.60	5.17	3.34	2.55
		VI	-8.81	4.56	5.03	3.21	5.04
		VII	-7.99	3.11	5.24	3.41	3.21
	3 AA → $(AA)_3$ [a]	I	-27.07	-5.91	2.90	-1.13	-5.07
		II	-26.77	-3.88	4.58	-0.27	-3.10
		III	-23.44	-2.29	4.40	-0.45	-2.29
		IV	-25.49	-3.37	4.53	-0.32	-2.95
		V	-25.69	-3.91	5.05	0.20	-3.27
		VI	-25.51	-1.95	4.91	0.06	-0.78
		VII	-25.69	-3.40	5.12	0.27	-3.16
tetramer	$(AA)_2$ + $(AA)_2$ → $(AA)_4$ [b]		-10.40	4.70	3.69	4.78	5.26
	$(AA)_3$ + AA → $(AA)_4$ [b]		-18.78	-2.12	0.46	-0.46	-1.36
	$(AA)_2$ + 2 AA → $(AA)_4$ [b]		-28.10	0.00	3.57	1.63	-0.56
pentamer	$(AA)_3$ + $(AA)_2$ + → $(AA)_5$ [c]		-11.82	5.50	4.32	5.46	6.12
	2 $(AA)_2$ + AA → $(AA)_4$ [c]		-21.14	6.13	7.43	7.55	6.93

[a]Trimers generated using as building units the gas phase geometries of the acetic acid molecule and the cyclic dimer **I**. [b]Tetramers generated using as building units the cyclic dimer **I**. [c]Pentamers generated using as building units the cyclic dimer **I** and the trimeric structure **I**.

Table 4. Energetics [kcal mol^{-1}] of formation of tetrolic acid (TTA) clusters as computed at the M06-2X/6-31+G(d,p) level of theory: $\Delta E_{e,gas}$ is the gas phase interaction energy; ΔG°_{ass} is the standard state (1 atm) gas-phase free energy of association at 298 K; ΔG^{*}_{ass} is the standard state (1 mol L^{-1}) free energies of reaction in the liquid-phase.

species	**reaction**	**type**	$\Delta \boldsymbol{E}_{e,gas}$	$\Delta \boldsymbol{G}^{\circ}_{ass}$	$\Delta \boldsymbol{G}^{*}_{ass}$		
					H_2O	**$CHCl_3$**	**CCl_4**
dimer	TTA + TTA → $(TTA)_2$	I	-17.56	-5.38	0.85	-1.80	-4.62
		II	-9.61	2.03	5.13	4.19	2.79
		III	-9.33	2.21	5.34	4.45	2.99
		IV	-7.06	5.25	7.23	7.27	6.47
		V	-6.79	4.00	6.41	6.03	4.52
		VI	-6.17	5.22	5.98	4.42	4.60
		VII	-5.04	5.96	6.30	6.78	6.37
		VIII	-5.42	9.30	8.73	9.28	9.11
		IX	-1.02	11.03	11.14	10.65	10.26
trimer	$(TTA)_2$ + TTA → $(TTA)_3$ [a]		-8.90	1.48	3.50	3.97	2.37
	3 TTA → $(TTA)_3$ [a]		-26.47	-3.90	4.30	2.19	-2.23
tetramer	$(TTA)_2$ + $(TTA)_2$ → $(TTA)_4$ [b]		-8.11	3.72	3.30	5.76	5.61
	$(TTA)_3$ + TTA → $(TTA)_4$ [b]		-16.53	-1.54	-0.30	-0.05	-1.35
	$(TTA)_2$ + 2 TTA → $(TTA)_4$ [b]		-25.68	0.22	4.09	3.83	0.84
hexamer	3 $(TTA)_2$ → $(TTA)_6$ [c]		-20.23	10.28	3.77	10.61	11.81

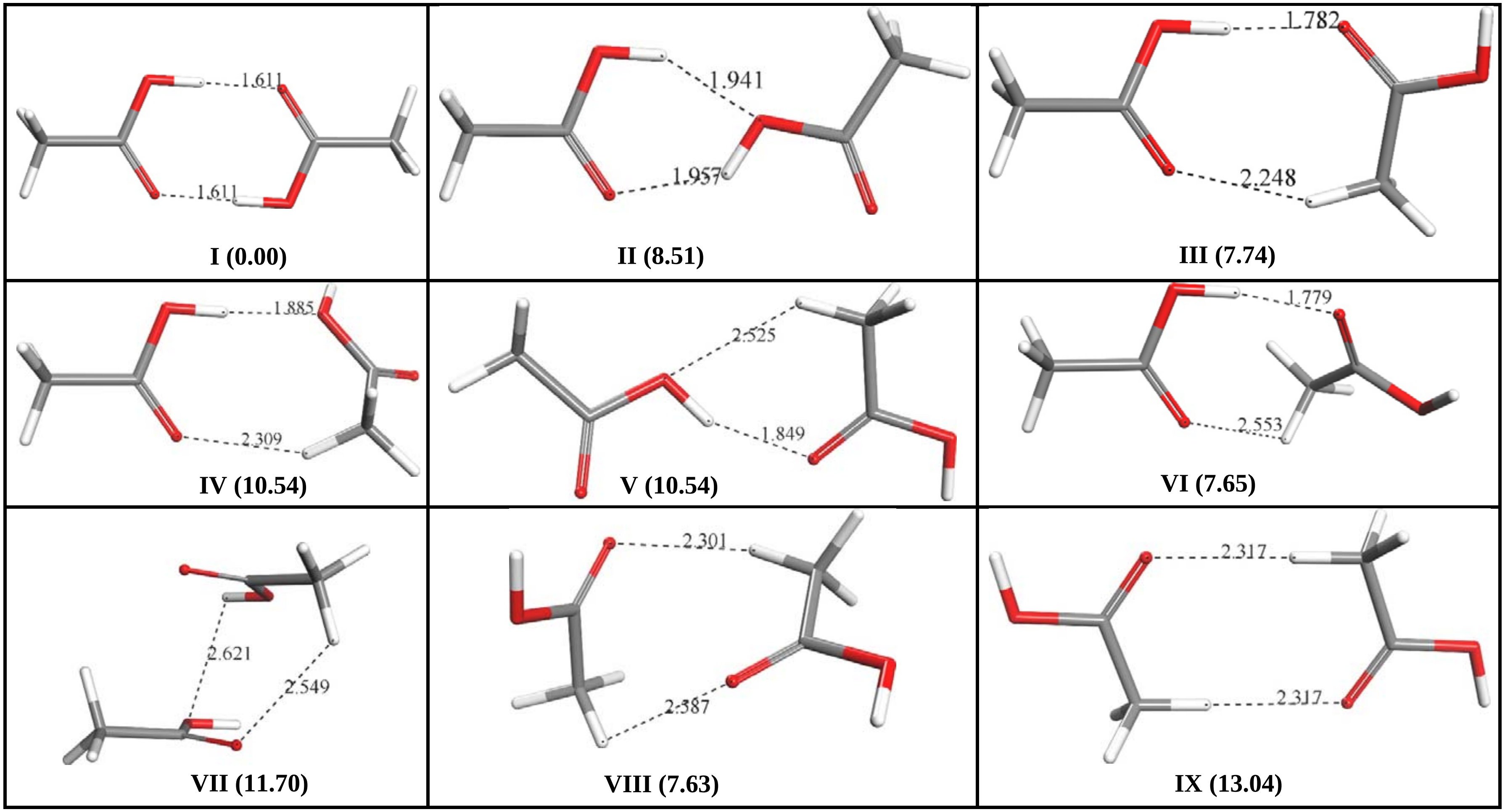
[a] Trimers generated using the gas phase geometries of the tetrolic acid molecule and the cyclic dimer **I**. [b] Tetramers generated using two acetic acids cyclic dimers **I**. [c] Hexamers generated using three acetic acid cyclic dimers.

Table 5. Energetics [kcal mol^{-1}] of formation of benzoic acid (BA) clusters as computed at the M06-2X/6-31+G(d,p) level of theory: $\Delta E_{e,gas}$ is the gas phase interaction energy; ΔG°_{ass} is the standard state (1 atm) gas-phase free energy of association at 298 K; ΔG^{*}_{ass} is the standard state (1 mol L^{-1}) free energies of reaction in the liquid-phase.

species	**reaction**	**structure**	$\Delta \boldsymbol{E}_{e,gas}$	$\Delta \boldsymbol{G}^{\circ}_{ass}$	$\Delta \boldsymbol{G}^{*}_{ass}$		
					H_2O	**$CHCl_3$**	**CCl_4**
dimer	BA + BA → $(BA)_2$	I	-18.51	-6.48	0.21	-2.97	-7.77
		II	-9.46	2.44	2.40	3.63	1.15
		III	-9.23	1.33	2.64	2.25	-0.75
		IV	-9.20	1.34	4.50	3.30	-0.56
		V	-7.98	3.68	2.33	4.26	2.14
		VI	-7.041	4.24	1.26	4.09	2.43
		VII	-7.36	4.96	2.51	4.80	3.12
		VIII	-6.90	3.98	1.06	3.90	2.21
		IX	-6.83	4.53	1.54	4.44	2.80
		X	-6.55	4.72	1.77	4.56	2.93
		XI	-6.54	5.89	2.63	5.56	3.96
		XII	-6.47	5.70	3.40	5.56	3.90
		XIII	-5.73	6.10	2.33	5.58	3.09
		XIV	-5.56	5.34	2.42	4.66	4.03
		XV	-5.29	6.06	3.03	5.54	3.88
		XVI	-4.48	5.97	4.96	5.89	3.45
		XVII	-4.21	5.10	4.71	5.15	2.29
		XVIII	-3.48	7.33	5.41	6.93	4.93
trimer	$(BA)_2$ + BA → $(BA)_3$ [a]		-9.73	3.44	0.31	3.61	2.67
	3 (BA) → $(BA)_3$ [a]		-28.20	-3.17	0.43	0.55	-5.20
tetramer	$(BA)_2$ + $(BA)_2$ → $(BA)_4$ [b]		-14.14	1.19	-4.07	1.65	3.04
	$(BA)_2$ + 2 (BA) → $(BA)_4$ [b]		-32.65	-5.38	-3.55	-1.41	-4.83
	$(BA)_3$ + (BA) → $(BA)_4$ [b]		-22.96	-8.78	-4.07	-5.03	-7.50

[a] Trimers generated using the gas phase geometries of the benzoic acid molecule and the cyclic dimer **I**. [b] Tetramer generated using two benzoic acids cyclic dimers **I**.

Figures

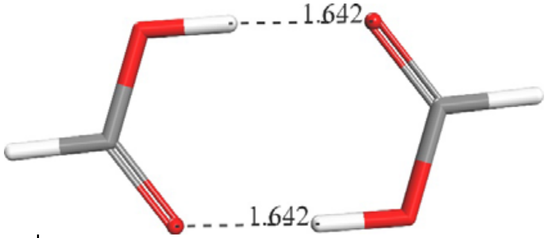


I (0.00) II (8.51) III (7.74)

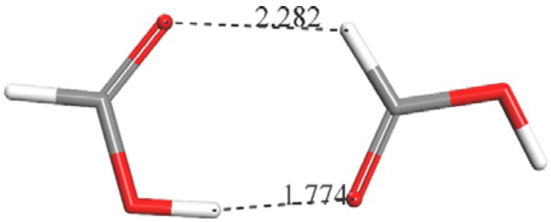
IV (10.54) V (10.54) VI (7.65)

VII (11.70) VIII (7.63) IX (13.04)

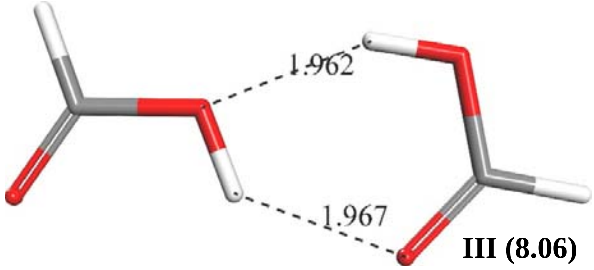
Figure 1. Structures of acetic acid dimers. In bold are the structures that have been located in this study. In parenthesis are given the energy differences with respect to structure **I**. Distances in Å and energies in kcal mol^{-1}.



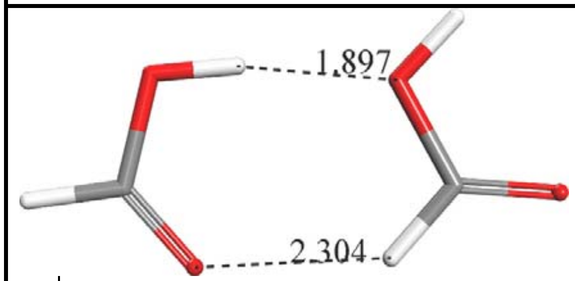
I (0)



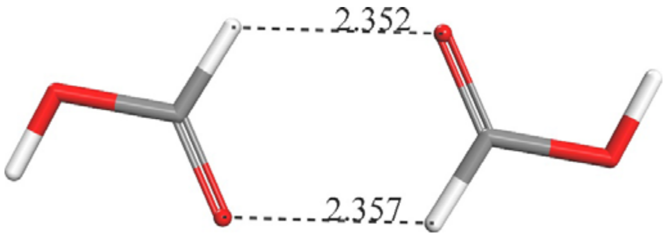
II (6.13)



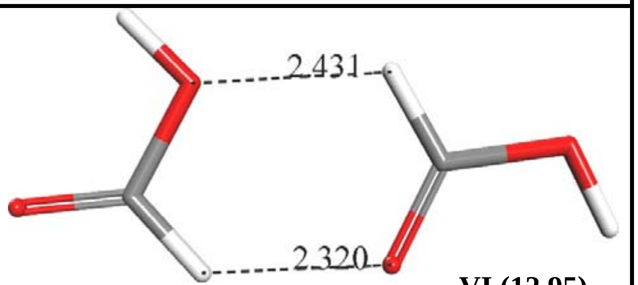
III (8.06)



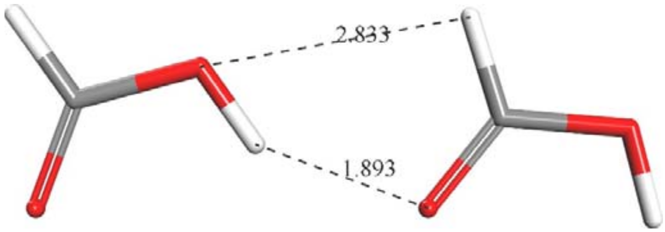
IV (9.60)



V (11.86)



VI (12.95)



VII (10.56)

Figure 2. Structures of formic acid dimers. In bold are the structures that have been located in this study. In parenthesis are reported the energy differences with respect to structure **I**. Distances in Å and energies in kcal mol^{-1}.

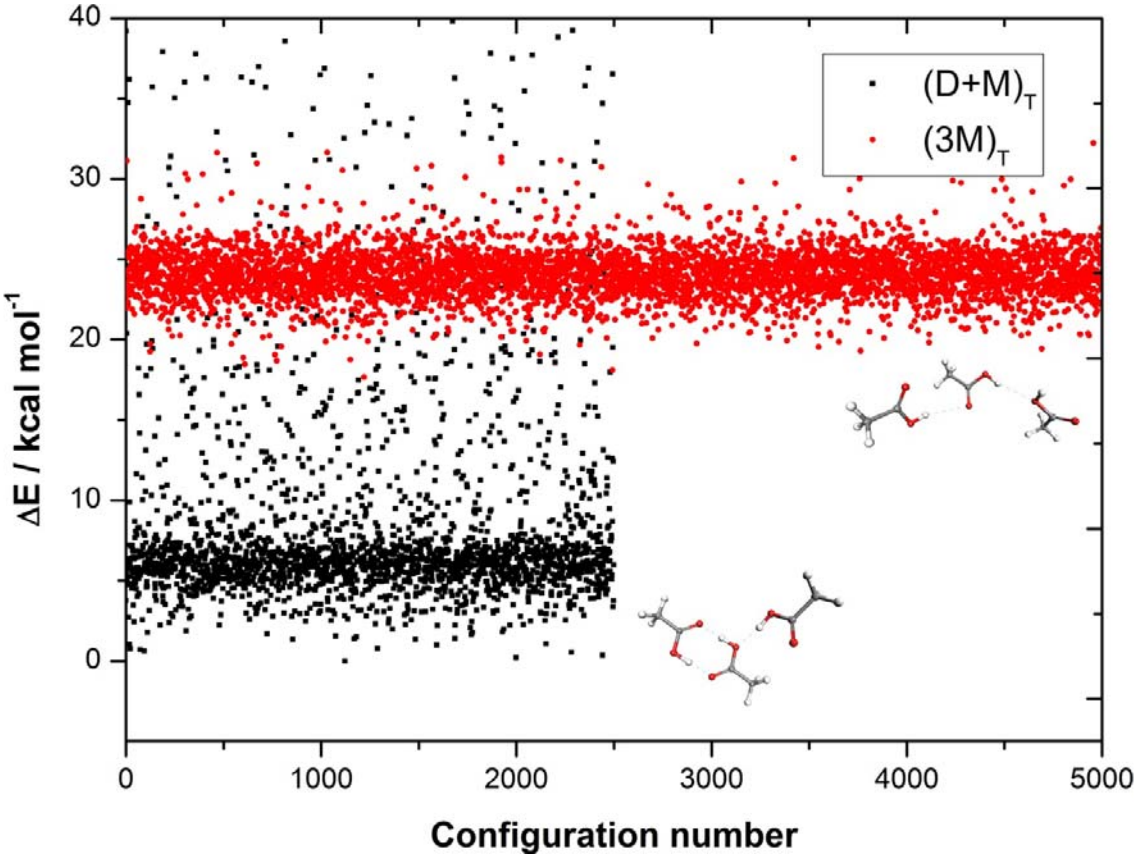


Figure 3. Relative energies of candidate structures of acetic acid trimers generated from the aggregation of three acetic acid molecules, $(3M)_{Tr}$, and of the cyclic dimer **I** with on acetic acid molecule, $(D+M)_{Tr}$. Values computed at the M06-2X/6-31+G(d,p) level of theory.

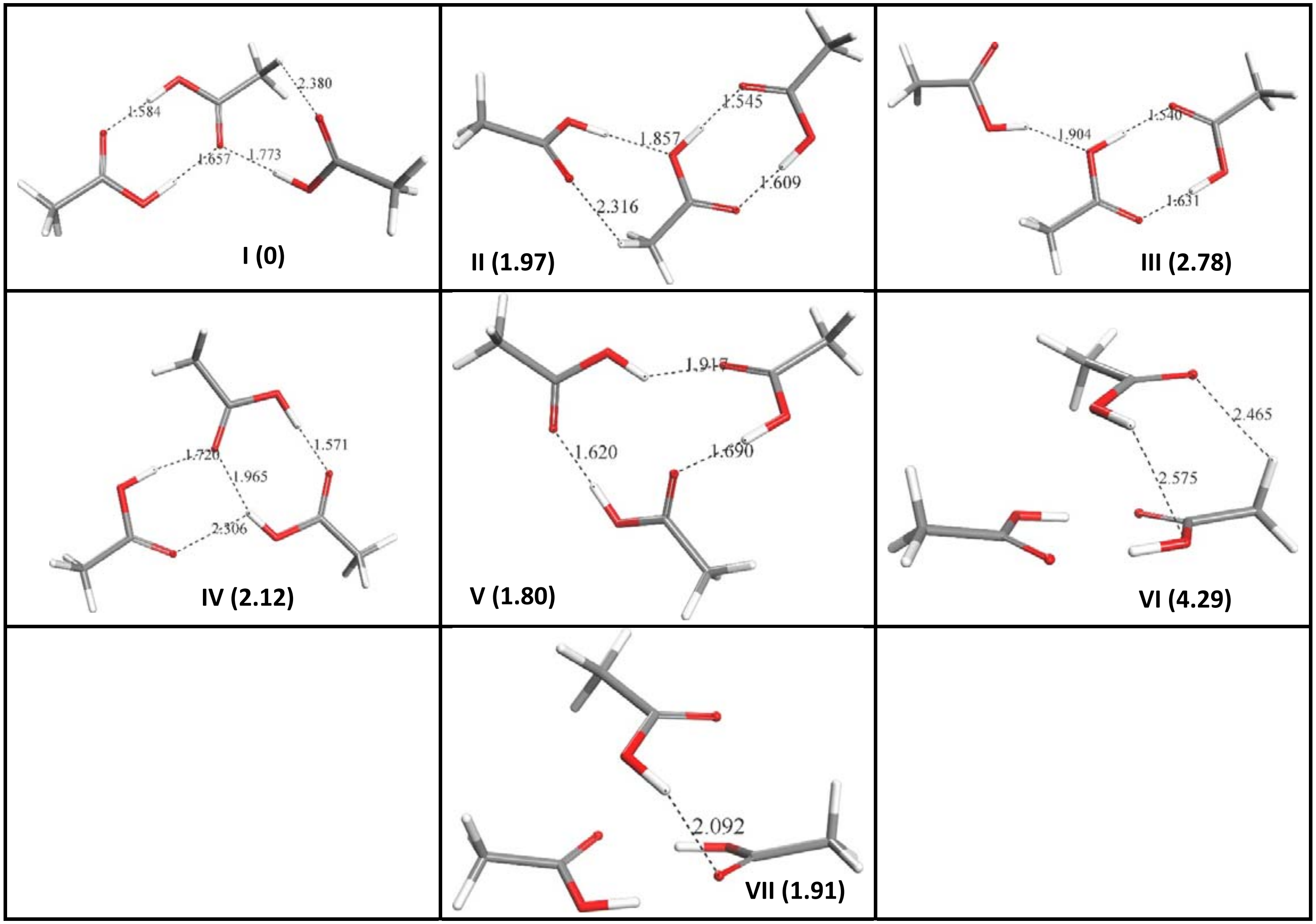

2.380
1.584
1.657
1.773
I (0)
1.545
1.857
1.609
2.316
II (1.97)
1.540
1.904
1.631
III (2.78)
1.720
1.571
1.965
2.306
IV (2.12)
1.917
1.620
1.690
V (1.80)
2.465
2.575
VI (4.29)
2.092
VII (1.91)

Figure 4. Structures of acetic acid trimers. In parenthesis are given the energy differences with respect to structure **I**. Distances in Å and energies in kcal mol^{-1}.

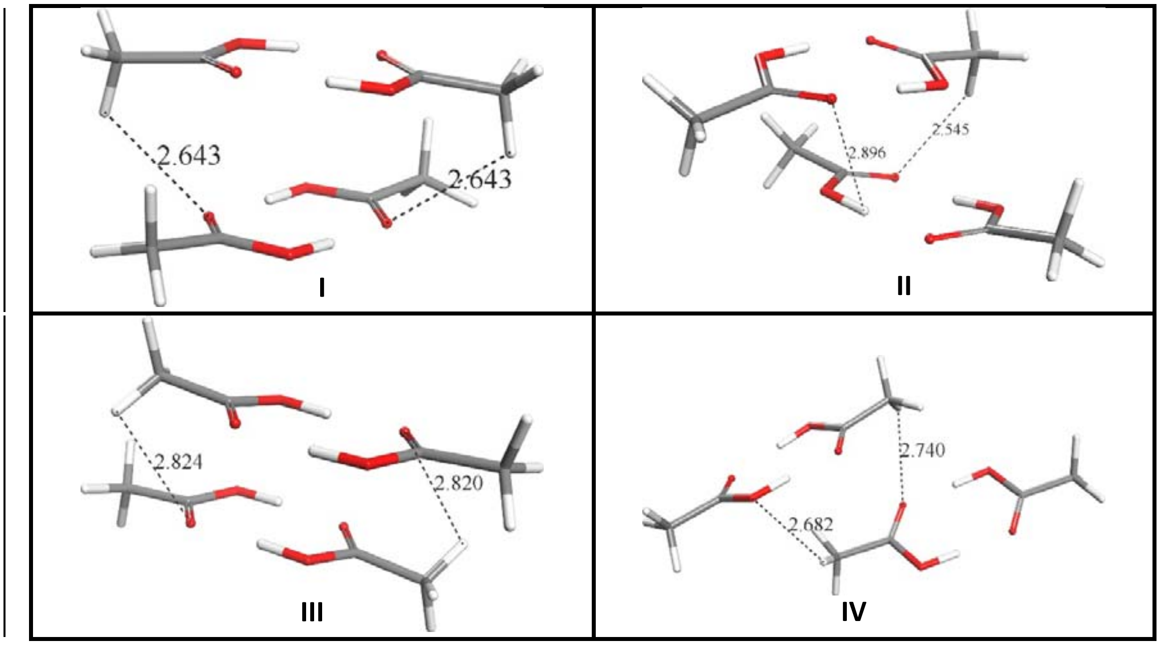


Figure 5. Structures of acetic acid tetramers. Distances in Å.

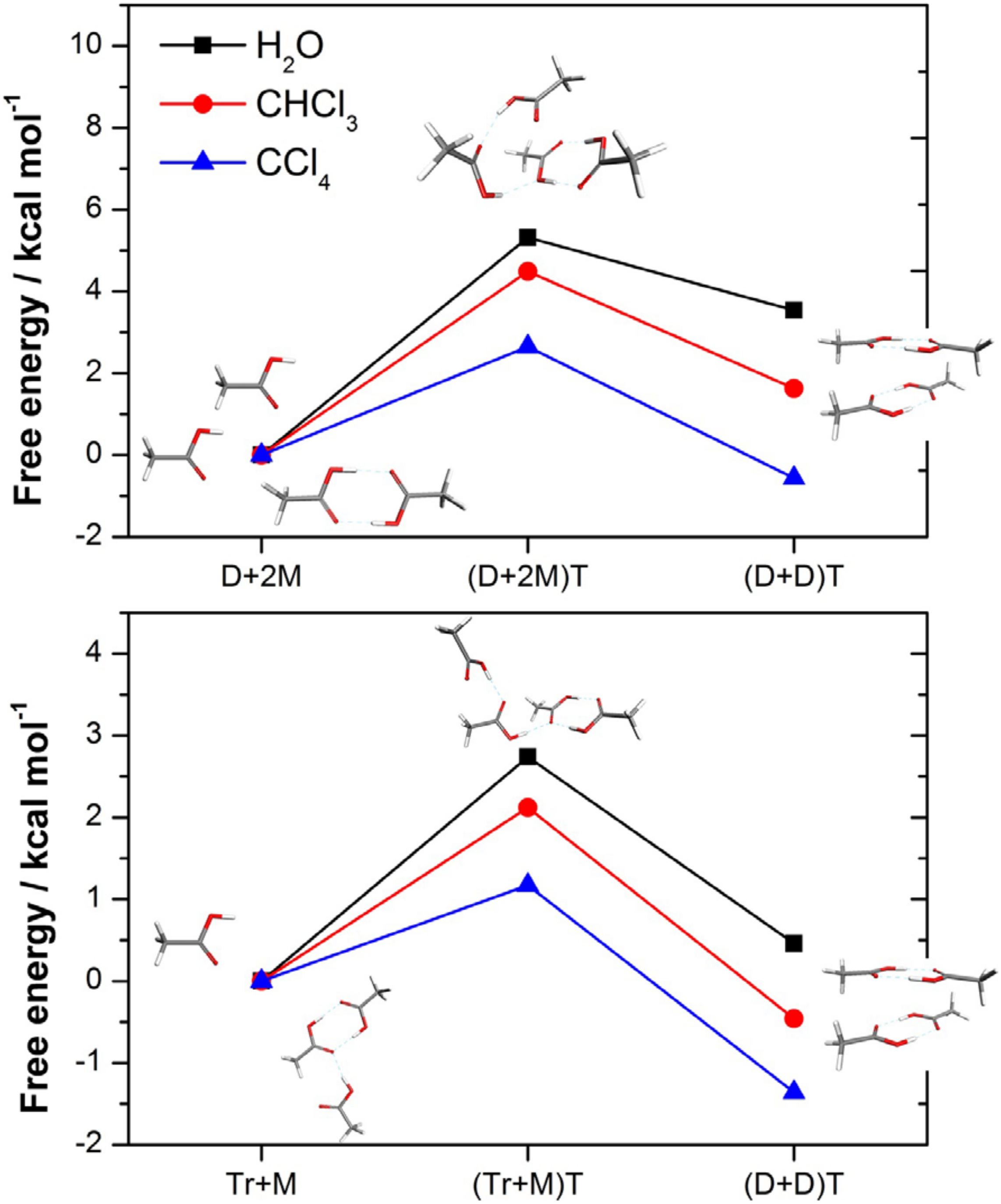


Figure 6. Calculated reaction free energy of acetic acid tetramers.

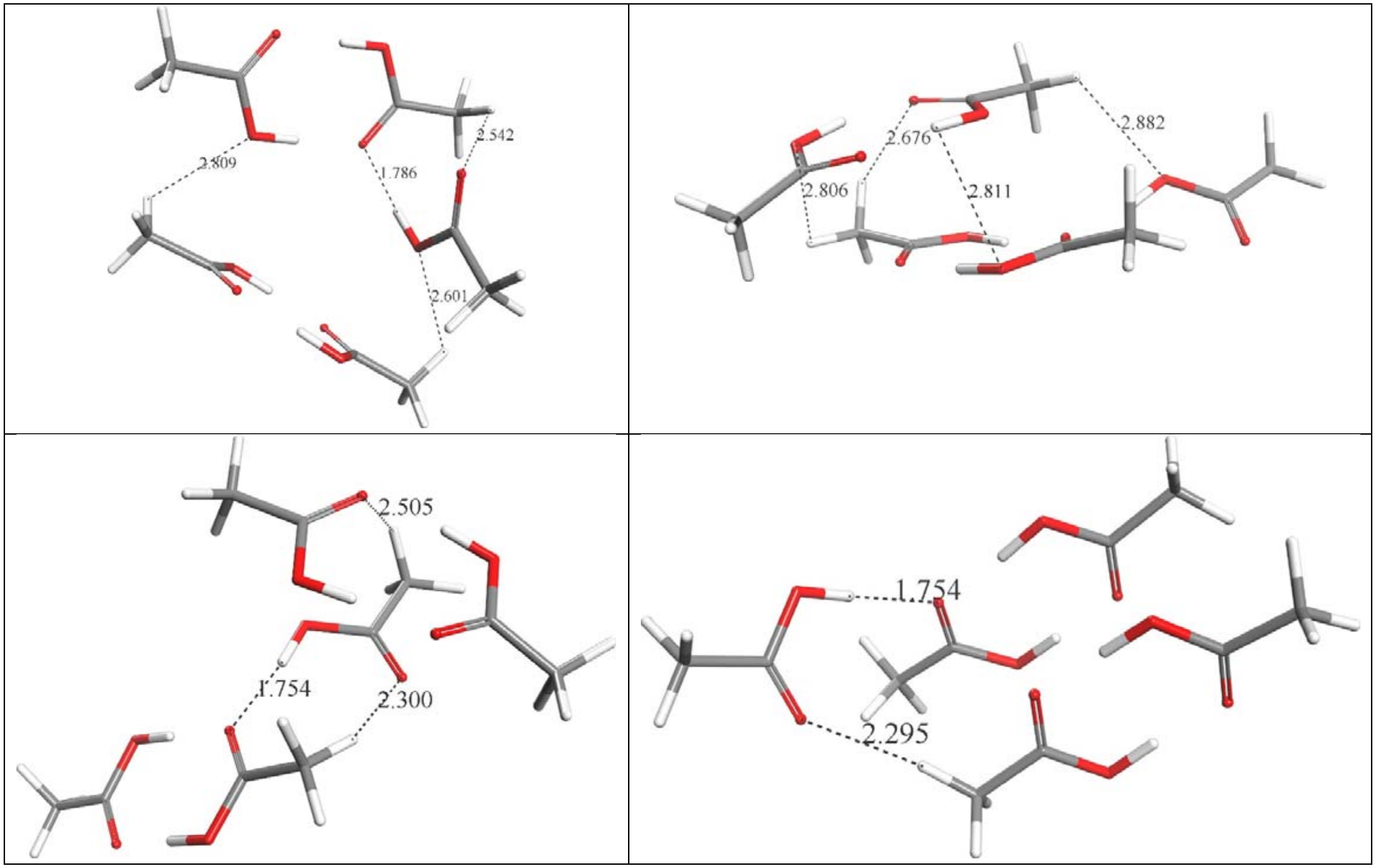

2.809
1.786
2.542
2.601
2.676
2.882
2.806
2.811
2.505
1.754
2.300
1.754
2.295

Figure 7. Structures of the acetic acid pentamers. Distances in Å.

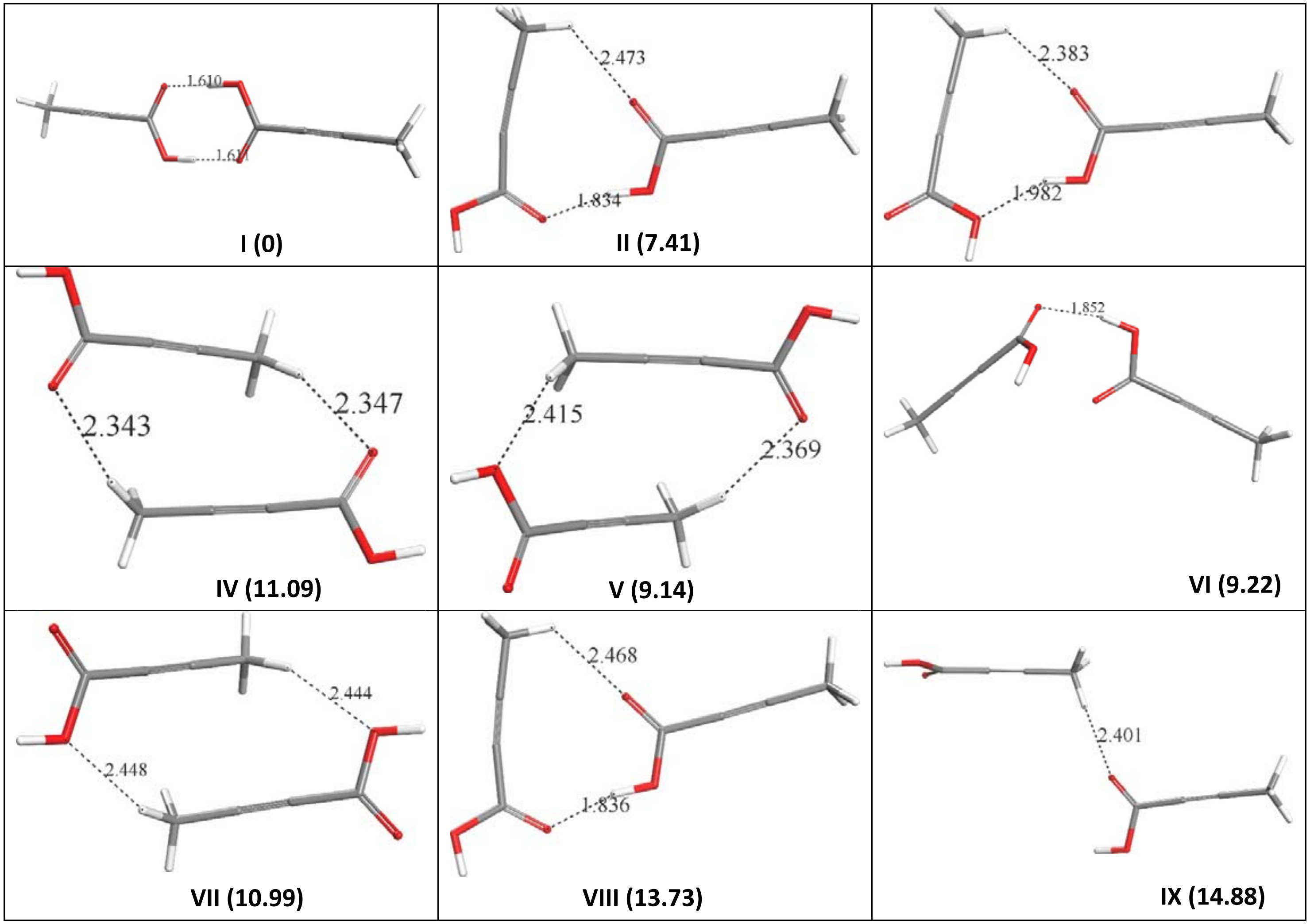
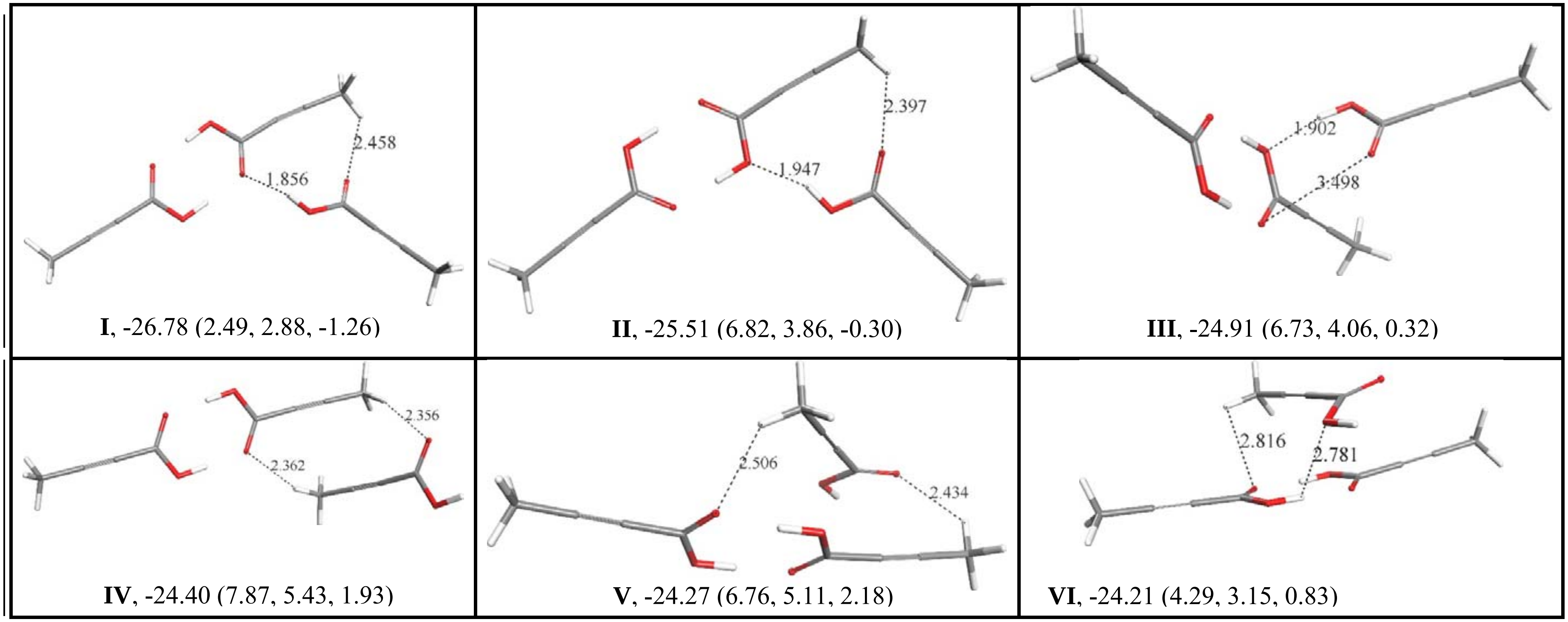


Figure 8. Structures of tetrolic acid dimers. In parenthesis are given the energy differences with respect to structure **I**. Distances in Å.



I, -26.78 (2.49, 2.88, -1.26)

II, -25.51 (6.82, 3.86, -0.30)

III, -24.91 (6.73, 4.06, 0.32)

IV, -24.40 (7.87, 5.43, 1.93)

V, -24.27 (6.76, 5.11, 2.18)

VI, -24.21 (4.29, 3.15, 0.83)

Figure 9. Structures of tetrolic acid trimers. The gas phase interaction energies and the standard state (1 mol L^{-1}) free energies of formation in water, chloroform and carbon tetrachloride (values in parenthesis) computed according to the association reaction 3 BA → $(BA)_3$ are also reported. Distances in Angstrom and energies in kcal mol^{-1}.

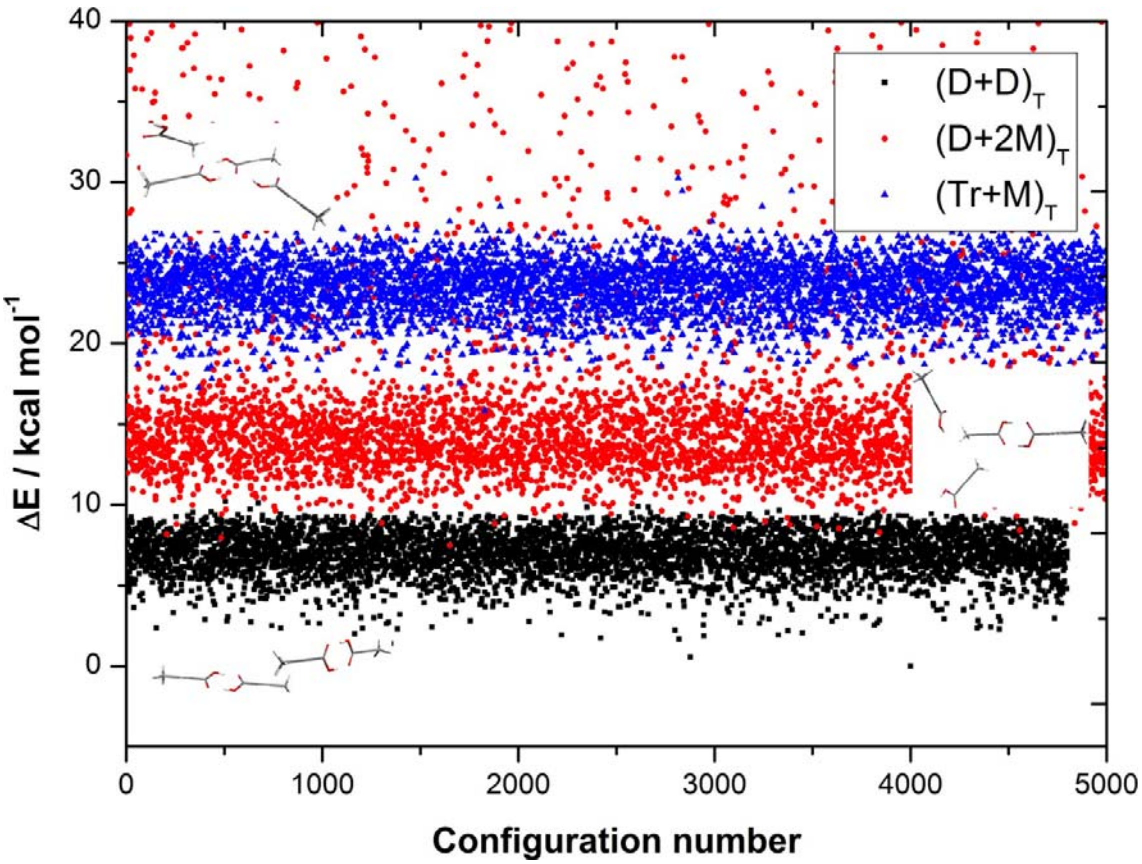


Figure 10. Relative energies of the candidate structures of the tetrolic acid acid tetramer generated using as building units two cyclic dimers **I**, (D+D)$_T$, the cyclic dimer **I** with two tetrolic acid molecules, (D+2M)$_T$, and the most stable trimer **I** with one tetrolic acid molecule, (Tr+M)$_T$. Values computed at the M06-2X/6-31+G(d,p) level of theory.

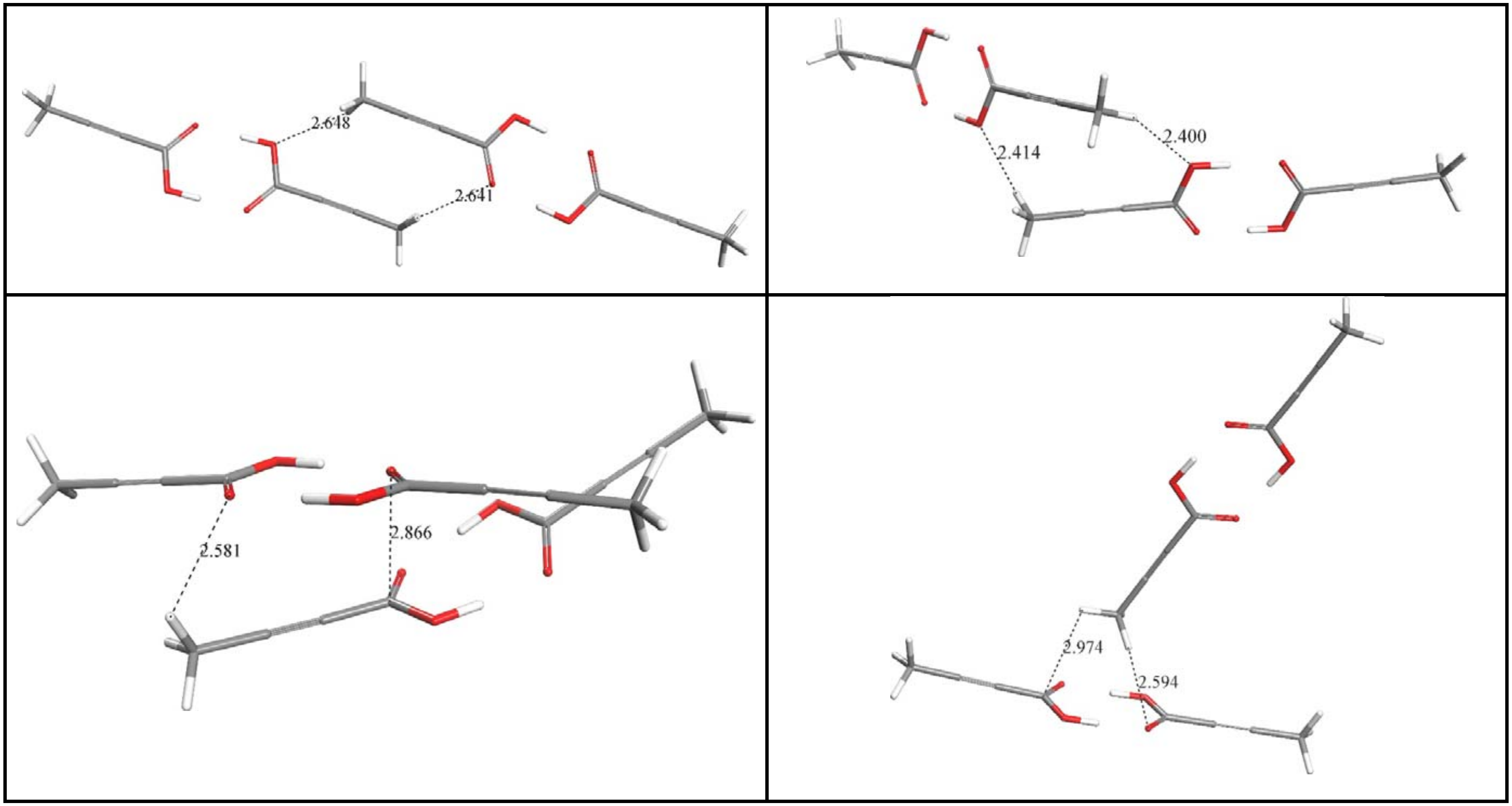


Figure 11. Structures of the tetrolic acid tetramers. Distances in Å.

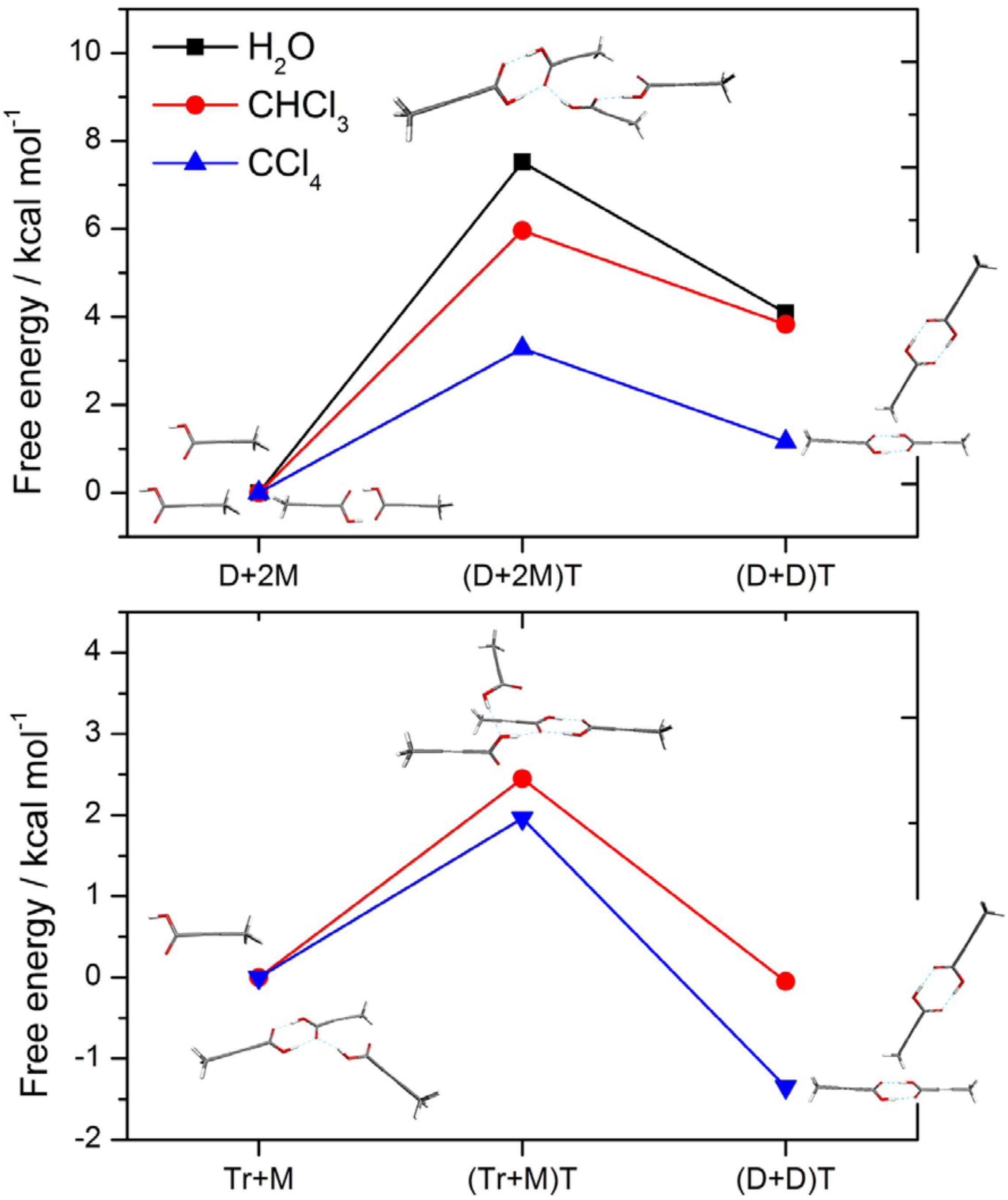


Figure 12. Calculated reaction free energy of tetrolic acid tetramers.

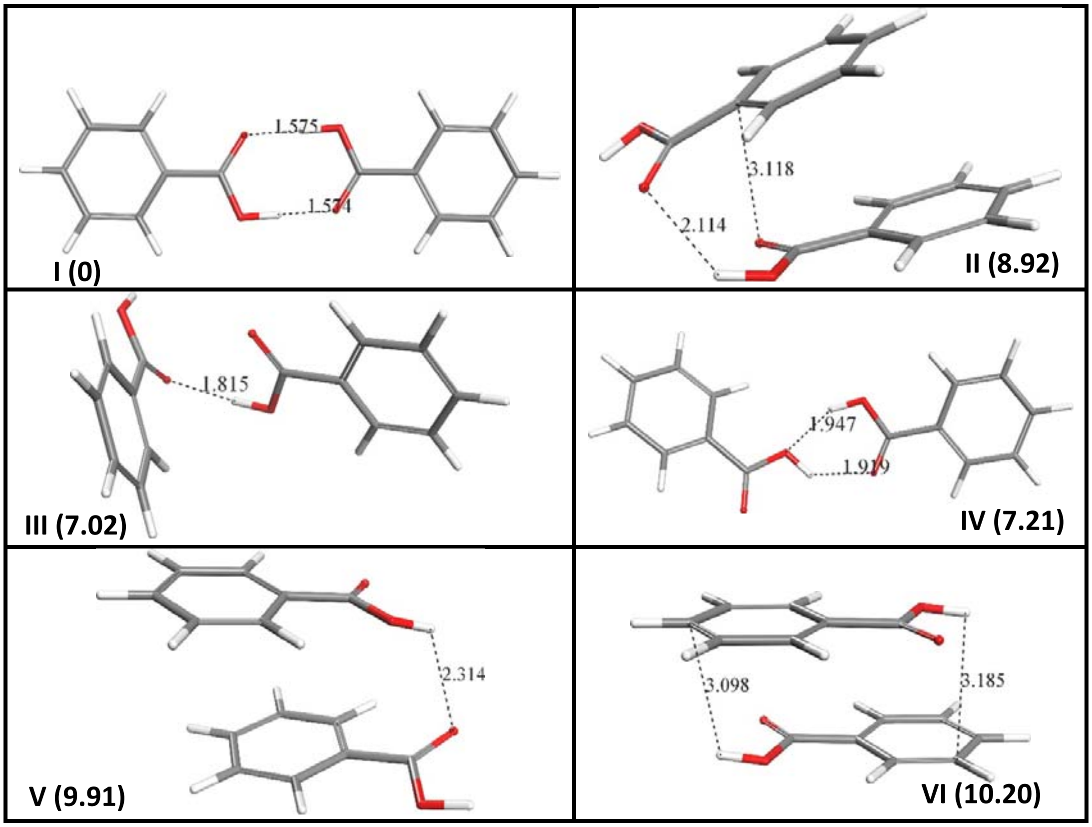


Figure 13. Structures of the six most stable benzoic acid dimers. In parenthesis are given the energy differences with respect to structure **I**. Distances in Å and energies in kcal mol^{-1}.

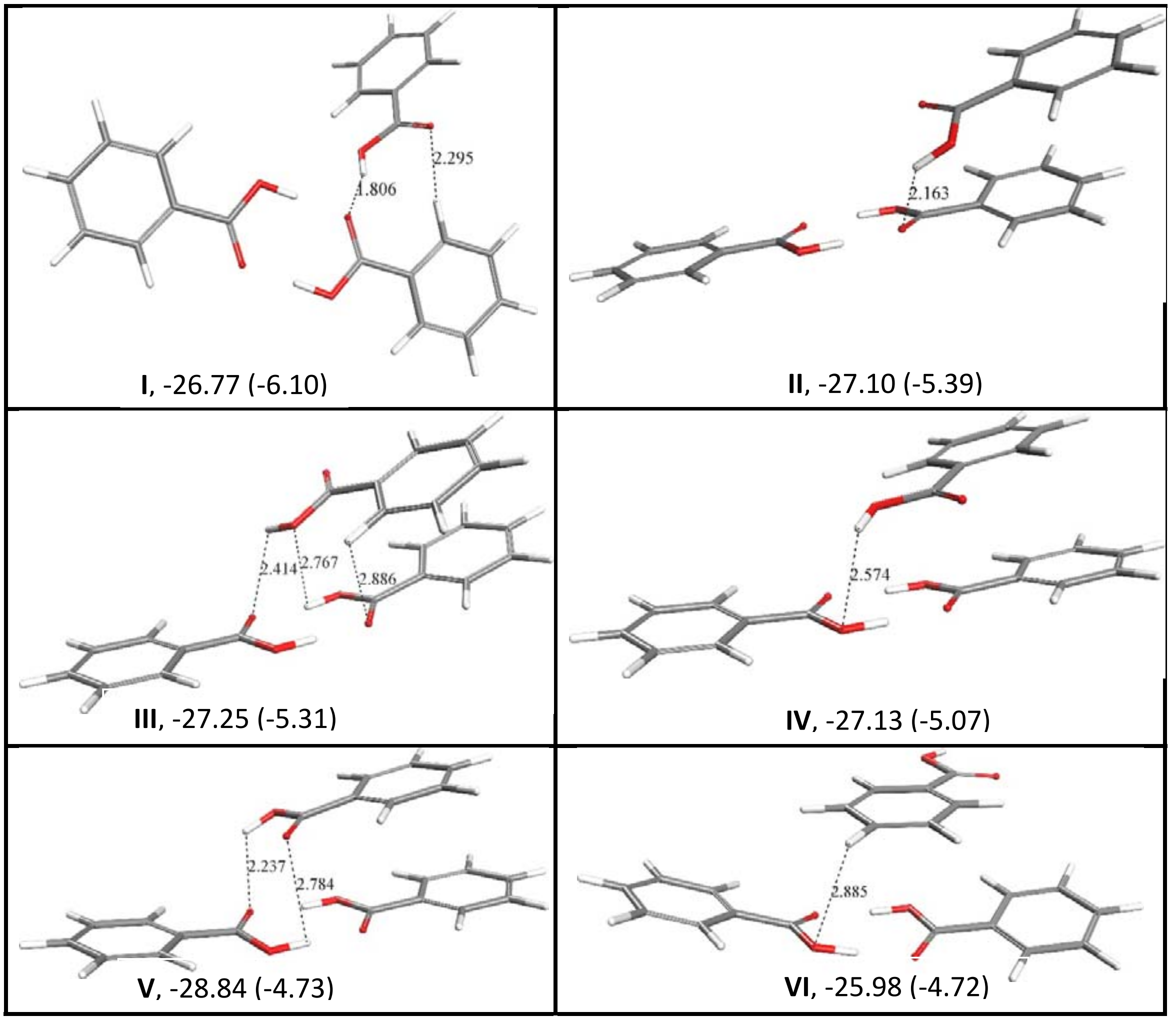


Figure 14. Structures and gas phase interaction energies ($\Delta E_{e,gas}$) of the most stable benzoic acid trimers. In parenthesis are reported the standard state (1 mol L^{-1}) free energies of reaction in carbon tetrachloride (ΔG^{*}_{ass}) computed according to the association reaction 3 BA → (BA)$_3$. Distances in Angstrom and energies in kcal mol^{-1}.

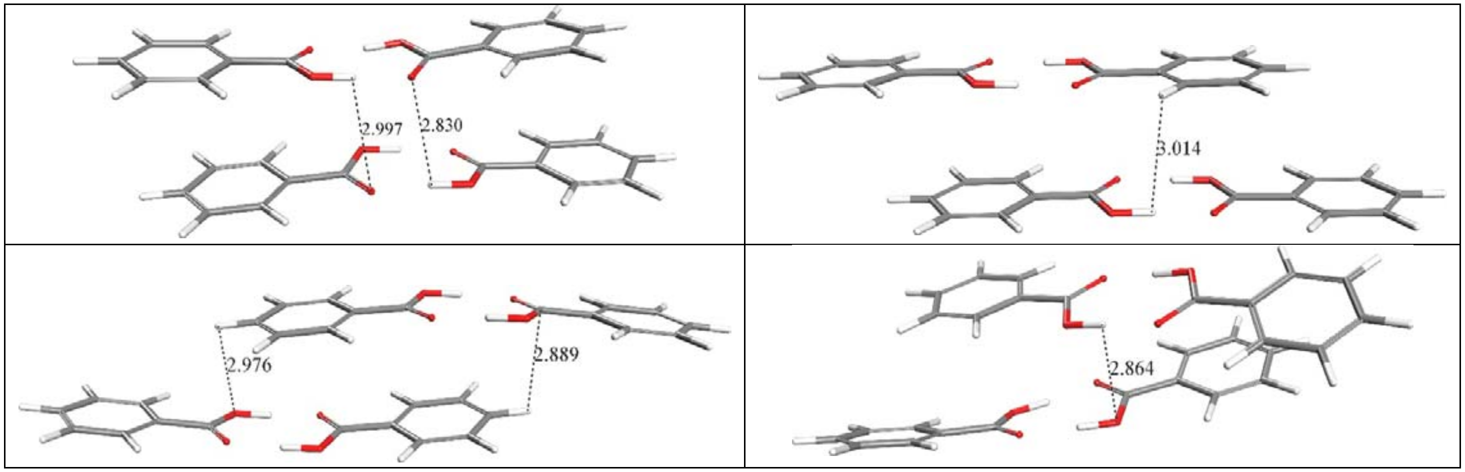


Figure 15. Structures of representative benzoic acid tetramers. Distances in Angstrom.

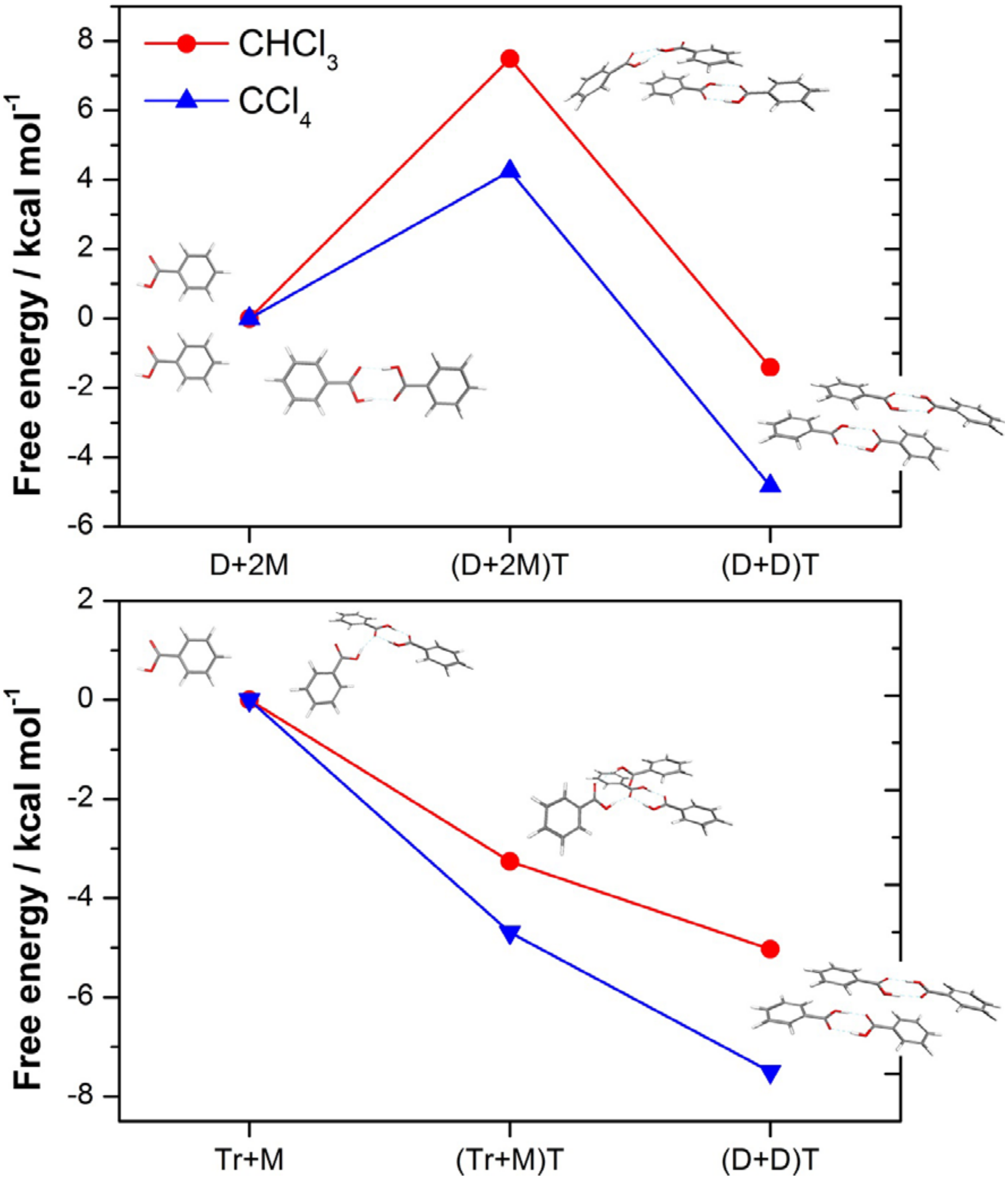


Figure 16. Calculated reaction free energy of benzoic acid tetramers.

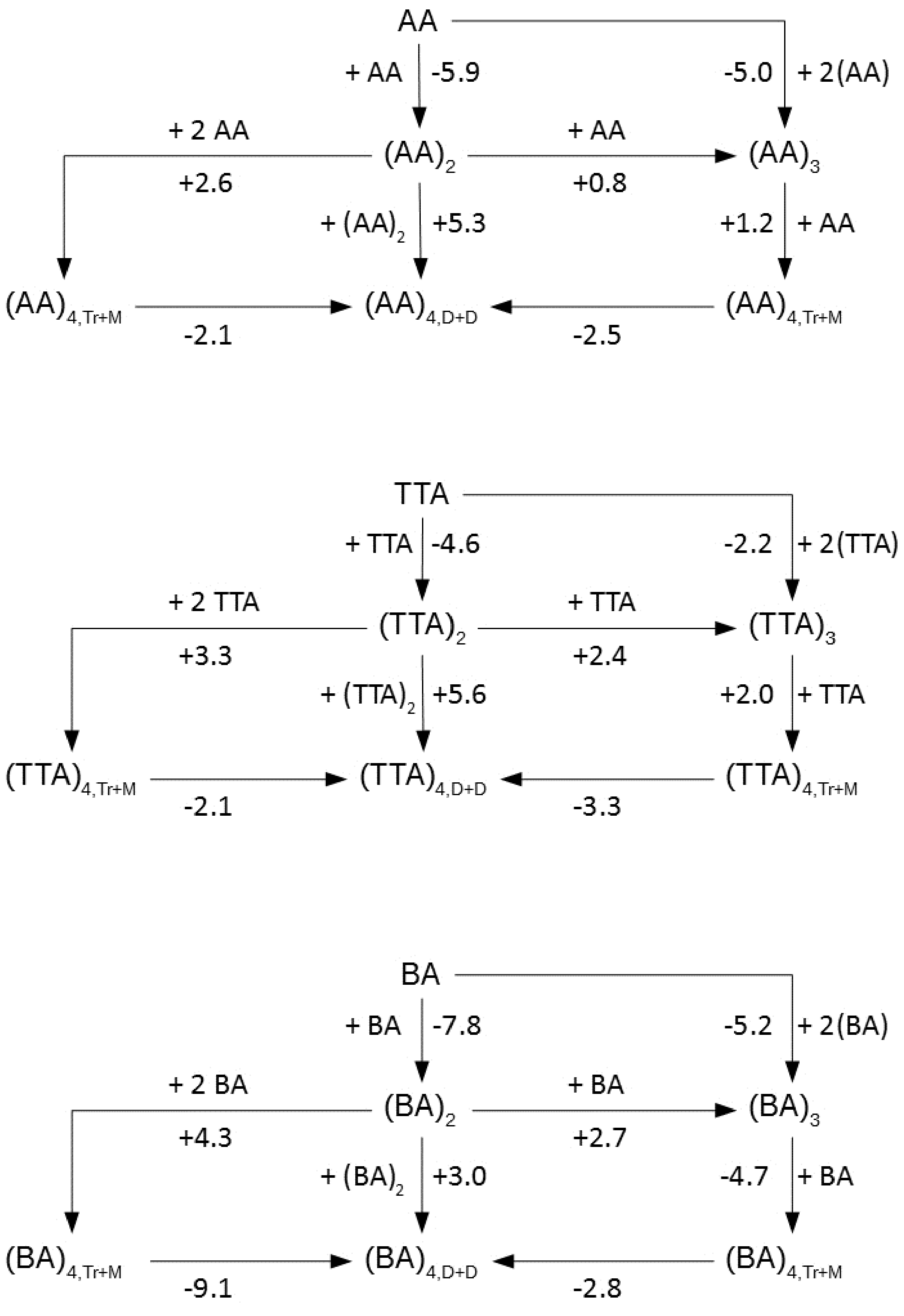


Figure 17. Free energy pathways of the formation in carbon tetrachloride of dimers, trimers and tetramers of acetic acid (AA), tetrolic acid (TTA) and benzoic acid (BA). Values in kcal mol^{-1}.

Table of Contents - Graphic