Computational Study of Solvent Effects on the Molecular Self-Assembly of Tetrolic Acid in Solution and Implications for the Polymorph Formed from Crystallization

Jie Chen and Bernhardt L. Trout*

Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, E19-502b, Cambridge, Massachusetts 02139

Received: November 6, 2007; Revised Manuscript Received: March 9, 2008

The solution behavior of a model compound, tetrolic acid (TTA), is studied via molecular dynamics simulations in four organic solvents. The results suggest that strong interactions between TTA and solvent molecules (ethanol or dioxane) prevent the formation of carboxylic acid dimers in solution and thus promote the crystallization of TTA in a catemer-based form or a solvate form. Weak interactions, however, between TTA and solvent molecules (carbon tetrachloride or chloroform) facilitate the formation of carboxylic acid dimers in solution and thus promote the crystallization of a dimer-based crystal. Detailed solvent structure plays an important role in determining the relative stability of various growth synthons in solution and also the barriers along the pathway connecting them. This work illustrates the potential of molecular simulations to aid in the rational selection of solvents to obtain desired polymorphs during crystallization.

1. Introduction

The solvent plays an important role in crystallization, a commonly used separation technique in the pharmaceutical, chemical, and food industries.[1–4] It affects the size, morphology, and polymorph, and therefore has consequences for the downstream processing of the final product.[5] Current solvent selection techniques for solution crystallization remain ad hoc, and typically do not have a theoretical underpinning. Elucidation of the interactions between solvent and solute molecules and the mechanism underlying the solvent effects on each aspect of the crystallization process as well as the properties of the final crystal product would be a major aid for rational solvent selection. In this work, we focus on the role that solvent plays in determining the polymorph of the final crystal product.

Polymorphism is the ability of a solid material to exist in more than one form or crystal structure, while retaining the same chemical composition. Polymorphism exists as a result of the difference in crystal packing or the existence of different conformers of the same molecule, which are called packing polymorphism and conformational polymorphism, respectively. The existence of other crystal types as the result of hydration or solvation is called pseudopolymorphism. This phenomenon leads to significant variability in the properties of products in the pharmaceutical, chemical, and food industries and continues to pose a challenge to scientists and engineers in producing crystal products of consistent quality.[1,6] The burgeoning field of crystal engineering now widely views a crystal as a supramolecular entity created from structural synthons, or molecular building blocks, which may be single or intermolecularly bonded groups of molecules.[7] With the mediation of solvent, solute molecules tend to self-assemble into different structural motifs in solution, referred to as growth synthons.[2,8] In certain systems that exhibit polymorphism, such as tetrolic acid, there exists a clear link between the growth synthon formed in solution and the structural synthon packed in the crystal. This was, for example, experimentally verified by Davey et al. through FTIR spectroscopy.[9] Carboxylic acid dimers (growth synthon) formed by a pair of tetrolic acid molecules were detected in a tetrolic acid/chloroform solution, from which a dimer-based (structural synthon) crystal was also obtained. In addition, the dimer motifs were absent in a tetrolic acid/ethanol solution, from which a catemer-based (structural synthon) crystal was obtained. The "link" hypothesis[9] is schematically shown in Figure 1. This hypothesis is successful in explaining the crystal output of various systems like glycine,[10] 2,6-dihydroxylbenzoic acid,[11] as well as tetrolic acid, although exceptions were reported as well, like mandelic acid[12] and anthranilic acid,[13] where the link between solution chemistry and crystal output is absent.

Currently, effort has been dedicated to the study of the relationship between solvent and polymorph of the crystal product. However, understanding is still limited. In practice, the selection of a solvent to obtain a crystal product with the desired polymorph is time-consuming and expensive due to the lack of rational guidance. In this work, we illustrate the potential of using molecular simulation techniques to facilitate solvent selection in solution crystallization. Molecular dynamics (MD) is a powerful tool that can track the motion of each atom in the simulation. It has been used previously to study molecular self-assembly and precursor formation in solution.[14–16] This approach can provide understanding of the solute–solvent interactions at the atomic level and also aid in the rational selection of solvents for solution crystallization. Moreover, MD can be implemented to the study of those processes which are still very difficult to access with use of experimental approaches.[17–19] Hence, it is a good complementary tool for experimental techniques.

In this work, we study the self-assembly of a model compound in four organic solvents using MD. Tetrolic acid (TTA) is selected as the model compound because it is a polymorphic system that exhibits solvent effects and is also a good example for the "link" hypothesis.[9] TTA crystallizes in both a dimer-based, centric P_1 structure (form I) and a catemeric $P2_1$ chain structure (form II).[20] The form II crystal is the stable form at room temperature, while the form I crystal is more stable at higher temperatures, above 354 K.[21] TTA is also reported to

* Author to whom correspondence should be addressed. E-mail: trout@mit.edu. Phone: (617)-258-5021. Fax: (617)-253-2272.

10.1021/jp7106582 CCC: $40.75
Published on Web 06/05/2008

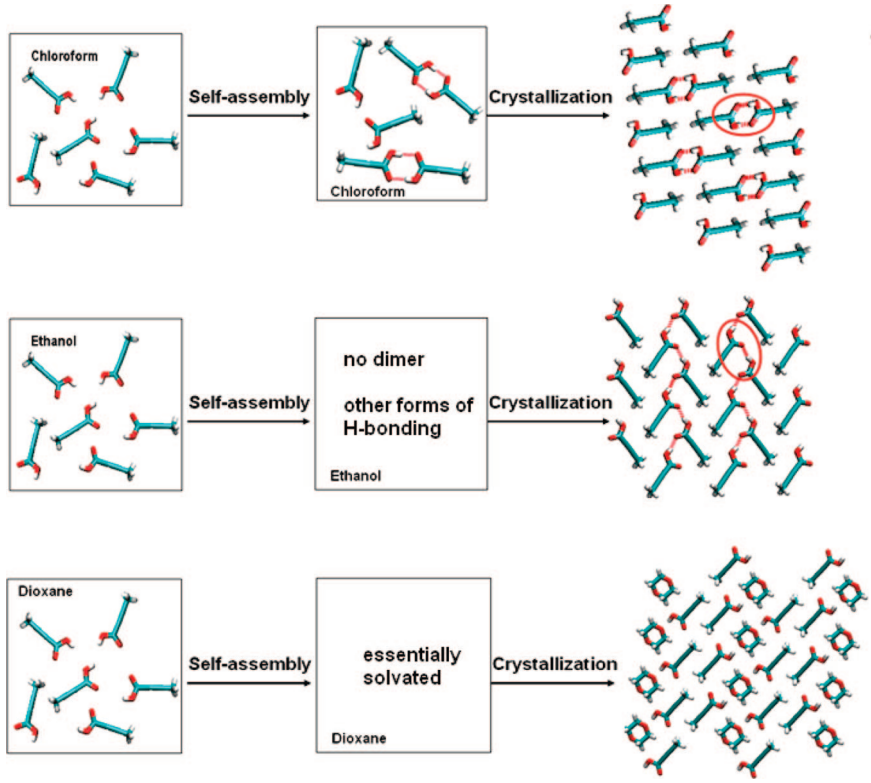


Figure 1. "Link" hypothesis. Red circles highlight the structural synthons in the crystal phase. The TTA molecule is in Licorice representation: O in red, H in white, C in cyan, and H bonds as red dashed lines.

form a solvate with dioxane.[9] Previous studies about TTA precursors in solution by Gavezzotti et al. using molecular simulation tools[15] showed that carboxylic acid dimers were the most persistent configuration in carbon tetrachloride, while the dimers broke immediately when the solvent was switched from carbon tetrachloride to water. This preliminary study was carried out in a very dilute system without accounting for the multibody interactions between solute molecules. Moreover, the relative stability of various synthons in solution and the mechanism through which solvent molecules mediate the self-assembly process are still not clear. Hence, the objective of this study is to gain fundamental understanding of (1) the interplay between TTA and solvent molecules, (2) the effect of solvent on the formation of various growth synthons and their relative stabilities, and (3) the relationship between solution chemistry and the polymorph formed during crystallization. Additionally, we provide some guidance for rational solvent selection in solution crystallization to obtain the desired polymorph.

2. Computational Details

2.1. System Description. Tetrolic acid (TTA) is chosen as a model solute. It has two polymorphic forms[21] and a solvate form.[9] Detailed crystallographic information is listed in Table 1. The conformational difference between TTA molecules in form I and form II is very small, thus the polymorphism studied in this work is packing polymorphism. TTA is modeled by using the CHARMM[22] force field, supplemented by partial atomic charges for the carboxylic acid group obtained from Gavezzotti's previous work.[15] This modification gives better replication of both the form I and form II X-ray crystal structures (smaller rmsd values) in potential energy optimizations. Detailed results are not shown here. Bond and angle equilibrium parameters are taken from averaging the experimental X-ray values of the form I and form II crystals. The four solvents studied are carbon tetrachloride, chloroform, ethanol, and dioxane. TTA is known to crystallize into form I, form II, and a solvate form from chloroform, ethanol, and dioxane, respectively. Carbon tetrachloride is chosen due to its strong apolarity, although there is currently no TTA crystal experimentally obtained from this solvent. The dielectric constants of these solvents are listed in Table 2 as an indictor of their polarity.[23] Hydrogen-bond parameters[2,24] for these four solvents are listed in Table 2 as well. The α parameter, as defined by Abraham,[24] is a measure of the molecule's hydrogen bond donor ability, which is related to the free energy change of the hydrogen bond formed between the molecule of interest and some reference base; the β parameter is a measure of the hydrogen accepter ability and is related to the free energy change of the hydrogen bond formed between the molecule of interest and some reference acid. For example, ethanol ($\alpha = 2.7$, $\beta = 5.8$) is a strong H-bond donor as well as a strong H-bond acceptor, while dioxane ($\alpha = 0.9$, $\beta = 5.3$) is only a strong H-bond acceptor. Since chloroform

TABLE 1: Polymorph Information of Tetrolic Acid

	crystal type	lattice parameters	reported solvents
form I[a] (dimer form)	triclinic	$a = 7.320(1)$, $\alpha = 83.91$ $b = 5.099(1)$, $\beta = 117.46$ $c = 7.226(1)$, $\gamma = 112$[c]	chloroform
form II[a] (catmer form)	monoclinic	$a = 7.887(1)$, $\alpha = 90$ $b = 7.121(1)$, $\beta = 100.18(1)$ $c = 3.937(1)$, $\gamma = 90$	ethanol
solvate [b]	triclinic	$a = 4.1673(2)$, $\alpha = 97.557(3)$ $b = 6.5063(4)$, $\beta = 92.948(3)$ $c = 12.4424(6)$, $\gamma = 91.670(2)$	dioxane

[a] Crystal structures of form I and form II are from ref 21. [b] Crystal structure of the solvate form is from ref 9. [c] Crystal data are directly obtained from the Cambridge Crystallographic Data Center (CCDC). Thus, the number of significant digits is not consistent.

TABLE 2: Physical and Chemical Properties of Carbon Tetrachloride, Chloroform, Ethanol, and Dioxane

		H-bond parameter		density (g/mL)		ΔH_{vap} (kJ/mol)	
	dielectric constant[a]	α[b]	β[b]	calcd	exptl[c]	calcd	exptl[d]
ethanol	24.3 (298K)	2.7	5.8	0.767 ± 0.001	0.789	42.01 ± 0.17	42.30
chloroform	5.0 (273K)	2.2	0.8	1.462 ± 0.001	1.498	31.38 ± 0.08	31.25
carbon tetrachloride	2.24 (293K)	1.4	0.6	1.618 ± 0.002	1.594	32.17 ± 0.13	32.43
dioxane	2.2 (298K)	0.9	5.3	1.002 ± 0.001	1.033	42.09 ± 0.08	40.42

[a] Experimental values of dielectric constant from ref 23. [b] Obtained from refs 2 and 24. The α parameter is a measure of the H-bond donor ability, the β parameter is a measure of the H-bond acceptor ability. [c] Experimental values of density at 298 K and 1 atm from ref 28. [d] Experimental values of heat of vaporization at 298 K and 1 atm from ref 29.

and carbon tetrachloride are not well-parametrized in the CHARMM[22] force field, all solvents are modeled by using OPLS[25] force field with one modification: arithmetic averages instead of geometric averages are used for combining the Lennard-Jones (LJ) radii, thus the same potential energy functional forms can be applied for both TTA and solvents. All molecules are flexible in the simulations.

This modified CHARMM-OPLS forced field is carefully tested. The density and the heat of vaporization of these four solvents at 298 K and 1 atm are calculated by using MD and compared to experimental values to validate the modification of the combination rule for the LJ radii. A 2 ns MD trajectory of a cubic solvent box is obtained with a time step of 1 fs in the NPT ensemble at 298 K and 1 atm. Periodic boundary conditions are applied. The cutoff for long-range nonbonded interactions is 1.4 nm and accounted for using the smooth particle mesh Ewald method.[26] This setting is used in all MD simulations in this study if periodic boundary conditions are present. All atoms are fully flexible during the simulations, and all calculations are performed with the CHARMM[22] program package. The results of the last 1.5 ns are used to calculate the density and error is estimated by using the method described by Allen.[26] The heat of vaporization is calculated by using the formula:

$$\Delta E_{vap} = E(g) - E(l) \quad (1)$$

$$\Delta H_{vap} = \Delta E_{vap} + RT \quad (2)$$

where E(g) and E(l) are the potential energies for the gas and liquid. The gas phase is treated as an ideal gas.[27] E(g) represents the average potential energy of a single molecule in vacuum (essentially the intramolecular energy for the model compound), computed from a 1 ns vacuum MD simulation. Langevin dynamics is applied, with a friction coefficient of 500.0. The calculated and experimental values[28,29] of the density and heat of vaporization of these solvents are listed in Table 2. There are only minor differences between the calculated and experimental values, which confirms that the modified OPLS force field is adequate to model these four organic solvents.

The suitability of using the modified CHARMM force field to model TTA is verified by the percentage changes of lattice parameters (PCLPs) and the root mean squared difference (rmsds) of a crystal supercell in a 2 ns MD simulation at 20 K and 1 atm. Low-temperature MD simulation in the NPT ensemble is similar to a potential energy optimization in allowing the system to relax to a local minimum. The initial X-ray crystal structure is used as the reference state. This simulation is performed for both TTA polymorphic form I and form II, and the results are plotted in panels a and c of Figure 2. The validation of mixing CHARMM and OPLS force field is tested by performing the same MD simulation for the solvate form of TTA (shown in Figure 2e). The interactions between TTA and the other three solvents are not tested here due to the lack of comparable experimental data. The PCLPs for all three forms of TTA crystal are within 6.5%, and the rmsd values over the whole MD simulations are less than 1.4 Å, which suggests that the modified CHARMM-OPLS force field is adequate to model TTA crystals. We also perform the same simulations using purely the OPLS force field (geometric average combination rule used for LJ radii) for TTA and dioxane, and the results are plotted in panels b, d, and f of Figure 2. Compared to the modified CHARMM-OPLS force field, the pure OPLS force field gives similar rmsd values, with decreases in some PCLP values (like form II lattice parameter a and solvate form lattice parameter b), but increases in others. Thus, we conclude that it is better to use the modified CHARMM-OPLS force field to model TTA and solvents in the following studies.

2.2. Solvation Free Energy Calculation. The solvation free energy of TTA in different solvents from its gas phase at 298 K and 1 atm is used to characterize how strongly the solute molecules interact with solvent molecules. Thermodynamic integration (TI) is applied to calculate the solvation free energy. The final state (with $\lambda = 1$) is defined as one TTA molecule fully solvated in a $30 \times 30 \times 30$ Å^3 solvent box. The system is equilibrated at the target temperature and pressure. The initial state ($\lambda = 0$) is defined as turning off the electrostatic and VDW interactions between solute and solvent molecules. Ten windows are used from the initial state to the final state with $\Delta\lambda = 0.1$. For each window, the system is equilibrated for 100 ps and sampled for 500 ps.

2.3. Free Energy Change of the Dimerization Reaction. The formation of a carboxylic acid dimer (product) from two fully solvated TTA molecules (reactant) is studied in these four different solvents. The free energy change of the dimerization reaction is used to characterize the stability of the carboxylic acid dimer, which is calculated by using the MD umbrella sampling[30–32] method. The order parameter (OP) used to study this reaction is defined by averaging two distances between the hydroxyl hydrogen and carboxyl oxygen of a pair of TTA molecules, which is schematically shown in Figure 3. Two TTA molecules are inserted into a $30 \times 30 \times 30$ Å^3 pre-equilibrated solvent box. To overcome diffusion, one TTA molecule is confined in the center of the simulation box, using a harmonic restraint with a force constant of 836.8 kJ/(mol·Å^2). A harmonic functional form of an umbrella potential is used, as shown below

$$U = k_u(\delta - \delta_0)^2 \quad (3)$$

where k_u is the harmonic force constant with units kJ/(mol·Å^2), and δ_0 is the equilibrium point of the sampling window.

2.4. Radial Distribution Function Calculations. Radial distribution functions (RDFs)[16] are used to provide more detailed structural information about the interactions between TTA and solvent molecules. To calculate the RDFs, one tetrolic acid molecule is inserted into a pre-equilibrated solvent box. (More

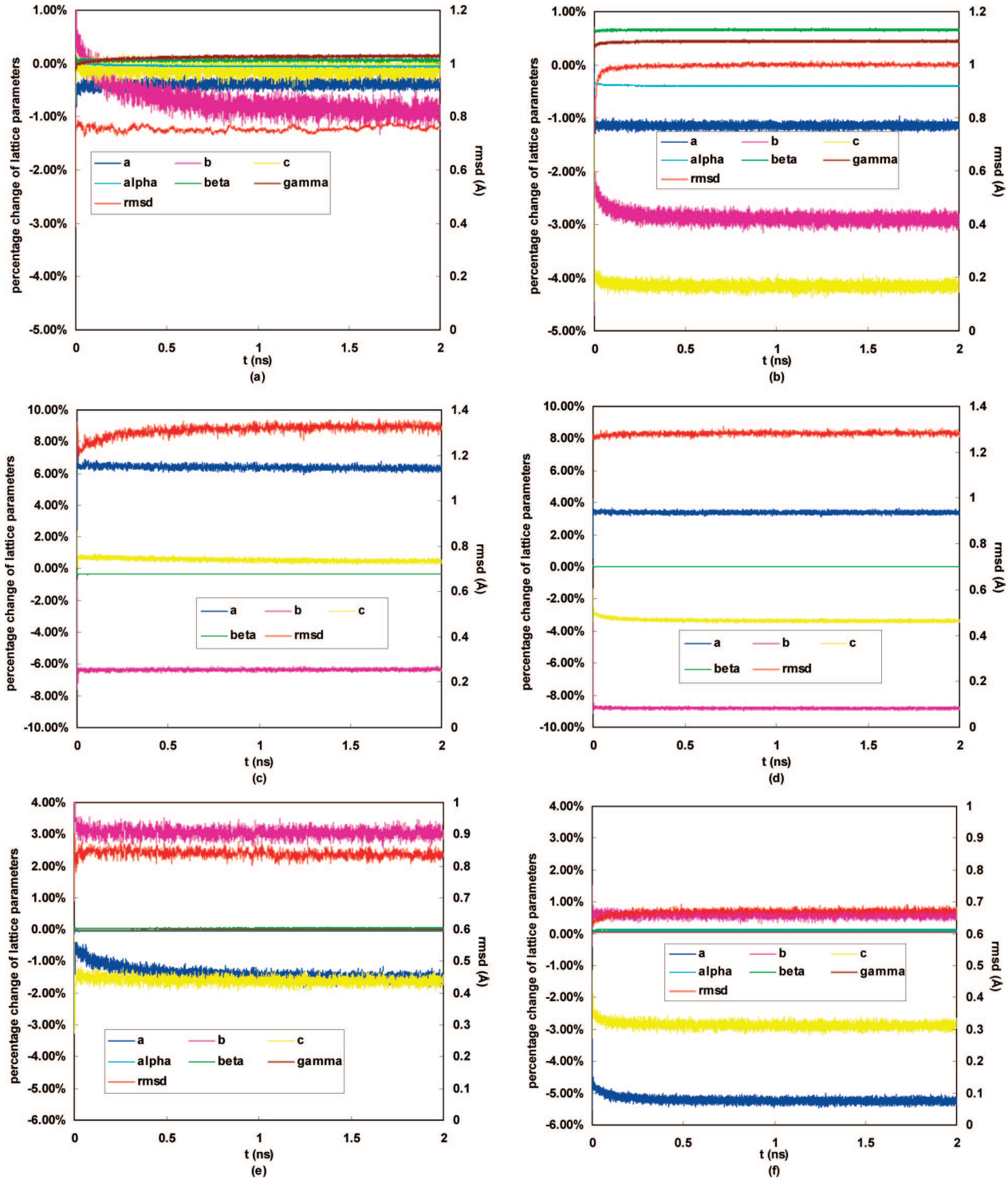
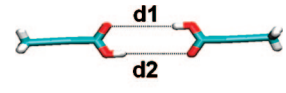


Figure 2. Percentage changes of lattice parameters and rmsds of a TTA crystal supercell in a 2 ns MD simulation at 20 K and 1 atm: (a) a box of form I crystal with 6 × 6 × 6 cells, modified CHARMM-OPLS force field applied; (b) a box of form I crystal with 6 × 6 × 6 cells, OPLS force field applied; (c) a box of form II crystal with 6 × 6 × 12 cells, modified CHARMM-OPLS force field applied, α and γ parameters do not change over the whole simulation run, thus, are not shown here; (d) a box of form II crystal with 6 × 6 × 12 cells, OPLS force field applied; (e) a box of solvate form crystal with 8 × 6 × 3 cells, modified CHARMM-OPLS force field applied; and (f) a box of solvate form crystal with 8 × 6 × 3 cells, OPLS force field applied.

system information is shown in Table 3.) Again to overcome diffusion, the TTA molecule is confined in the center of the simulation box by using a harmonic restraint with a force constant of 836.8 kJ/(mol·Å^2). A 2 ns MD trajectory is harvested. Various RDFs are calculated throughout the duration of the production run (lasting 1.5 ns) by evaluating a histogram of the distances between the specific atom in the solvent molecule and the interested atom in TTA every 100 time steps.



$$OP = \frac{d1 + d2}{2}$$

Figure 3. Definition of the order parameter (OP) used in umbrella sampling of the dimerization reaction.

TABLE 3: System Details of the MD Simulations for the RDF Calculations

	carbon tetrachloride soln	chloroform soln	ethanol soln	dioxane soln
no. of solvent molecules	416	390	464	247
no. of solute molecule	1	1	1	1
equilibrated cell size (Å)	40.8	37.4	35.9	32.4

All atoms used in RDF calculations are labeled in Figure 4. RDFs reported are those obtained by averaging all of the histograms and normalizing the RDF value in the bulk to 1.

2.5. Dimer Fraction of the Hydrogen Bond Network. MD simulations for tetrolic acid solutions at various concentrations (Table 4) are conducted to study the structure of the clusters formed in the solution. Previous work from Davey et al.[12] showed that TTA is highly soluble in chloroform, ethanol, and dioxane. The highest TTA concentrations they used in their experiments are 3.0 M in chloroform, 7.9 M in ethanol, and 4.8 M in dioxane, which are all reported to be under saturation. Those TTA concentrations simulated in this work are in line with the concentrations used in Davey's experiments and we still call these TTA−solvent mixtures solutions as Davey does, although some highly concentrated solutions also might be called binary mixtures. Note that no accurate solubility data for TTA have been reported so far. Thus, it is unknown whether the highest concentrations in the MD simulations are above or below the solubility line. The distance cutoff for the hydrogen bond formed between the hydroxyl hydrogen of one TTA molecule and the oxygen (either carboxyl oxygen or hydroxyl oxygen)

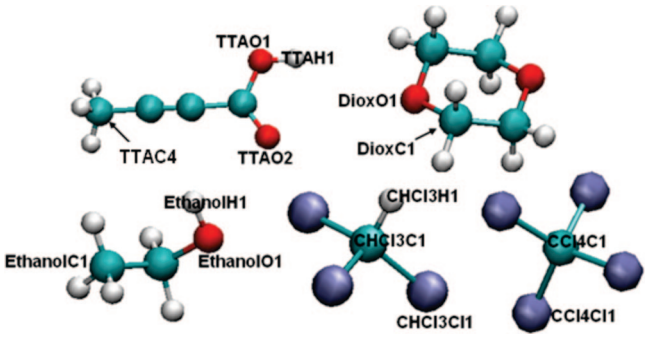


Figure 4. Atom labels in the RDF calculations.

TABLE 4: System Details of the MD Simulations in the Study of Dimer Composition of the H-Bond Network

	solvent	no. of TTA molecules	no. of solvent molecules	concn (mol/L)
1	carbon tetrachloride	339	257	6.17
2	carbon tetrachloride	225	436	3.58
3	carbon tetrachloride	113	632	1.56
4	carbon tetrachloride	75	708	0.99
5	chloroform	339	339	5.87
6	chloroform	225	473	3.84
7	chloroform	113	612	1.88
8	chloroform	75	677	1.21
9	ethanol	339	369	6.47
10	ethanol	225	508	4.44
11	ethanol	113	718	2.12
12	ethanol	75	810	1.35
13	dioxane	339	203	7.07
14	dioxane	225	290	4.89
15	dioxane	113	386	2.51

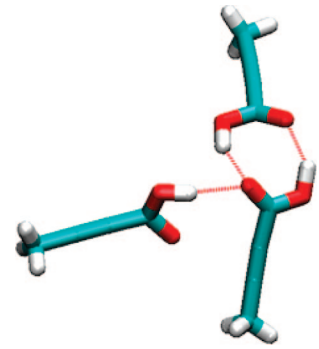


Figure 5. Illustration for dimer fraction calculation.

TABLE 5: Calculated Solvation Free Energies of TTA in Ethanol, Chloroform, Carbon Tetrachloride, and Dioxane at 298 K and 1 atm

	ΔG_{solv} (kJ/mol)
ethanol	−45.95 ± 1.26
chloroform	−41.42 ± 0.92
carbon tetrachloride	−37.78 ± 0.96
dioxane	−47.15 ± 1.13

of a neighboring TTA molecule is 2.35 Å, independent of orientation, which is obtained by adding the average length of the H bond formed in carbon tetrachloride (1.94 Å) and three times its standard deviations. The dimer fraction of the overall hydrogen bonds is used to characterize the H-bond network in various solutions. It is calculated by using the formula:

$$\text{dimer fraction} = \frac{\text{total no. of H bonds in carboxylic dimers}}{\text{total no. of H bonds in tetrolic acid molecules}} \quad (4)$$

For example, if there are only three TTA molecules in the system, and they exist in a conformation shown in Figure 5, there are two H bonds in carboxylic dimer and three H bonds in tetrolic acid molecules. Thus, the dimer fraction is 0.667. The dimer fractions of the form I crystal and the form II crystal are calculated as well, using both the X-ray structure directly obtained from literature and the optimized structures obtained from the low-temperature MD simulation. In both cases, the dimer fraction is 1 for form I crystal (the dimer form) and 0 for form II crystal (the catemer form).

3. Results and Discussions

3.1. Interplay between Tetrolic Acid and Solvent Molecules. The solvation free energy (ΔG_{solv}) of TTA in carbon tetrachloride, chloroform, and ethanol is computed to be −37.78, −41.42, and −45.94 kJ/mol, respectively, as listed in Table 5. It is clear that ΔG_{solv} is more negative as the solvent polarity increases, as measured by the increase in the dielectric constant. This is mainly because the carboxylic acid group in tetrolic acid will be better solvated in polar solvents than in nonpolar solvents. However, ΔG_{solv} in dioxane is an exception, being comparable to or even more negative than that in ethanol, although the polarity of dioxane is close to that of carbon tetrachloride. It is clear that solvent polarity cannot fully explain the solvation free energies of TTA. More detailed structural characteristics of the solvent must be considered as well, which is reflected in the RDF plots in Figure 6. The peaks near r = 0.17 nm in Figure 6c TTAH1-EthanolO1 curve and Figure 6d TTAH1-DioxO1 curve indicate that strong interactions are formed between the hydroxyl hydrogen of TTA and the oxygen atom of ethanol and dioxane through hydrogen bonding. Ethanol

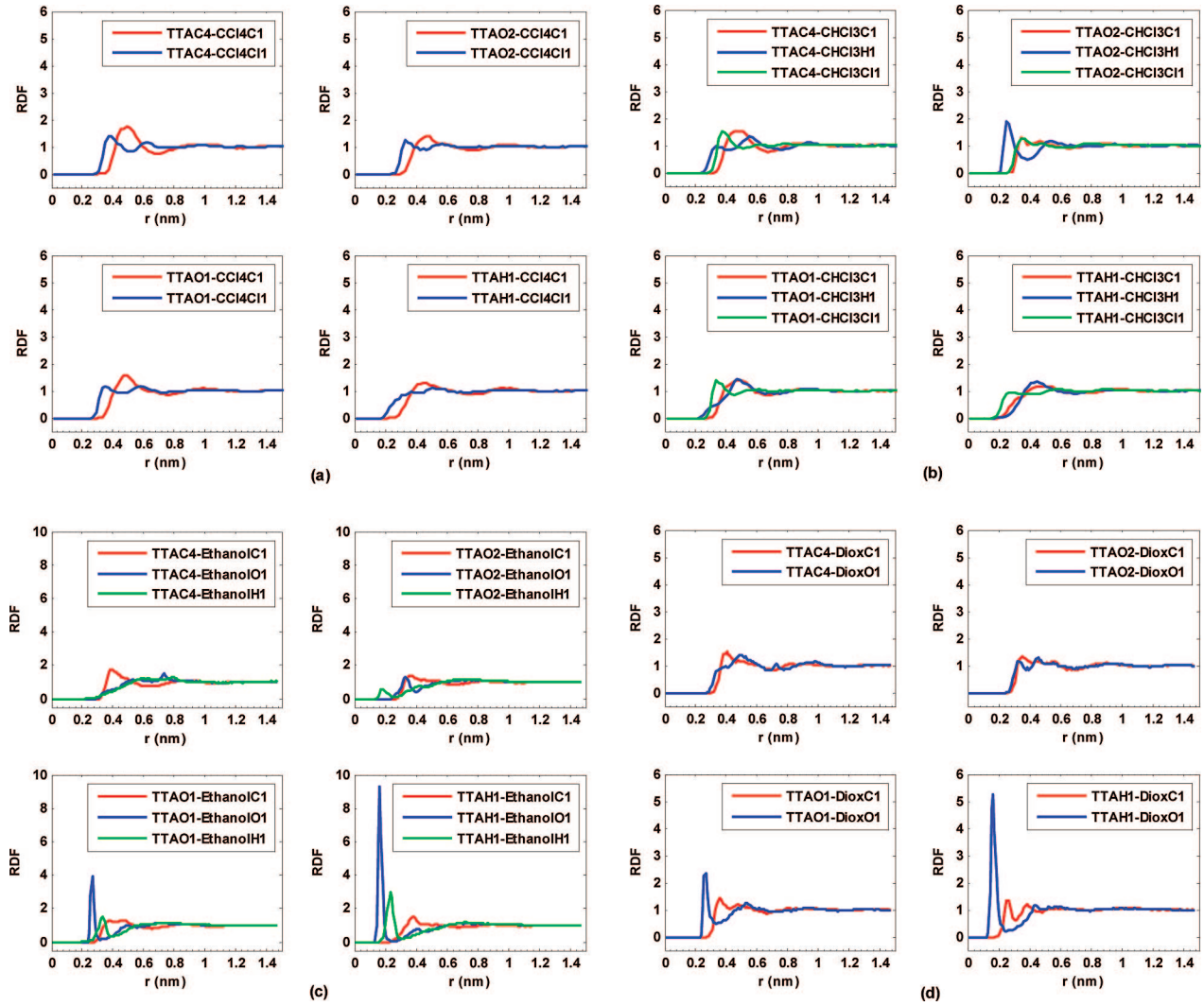


Figure 6. RDFs of the distances between the atoms in TTA and the atoms in solvent molecules: (a) carbon tetrachloride (CCl_4), (b) chloroform ($CHCl_3$), (c) ethanol, and (d) dioxane. Atom labels are the same as those shown in Figure 4. "TTAH1-CCl4C1" denotes the distance between the TTAH1 atom in TTA and the CCl4C1 atom in carbon tetrachloride.

and dioxane are both strong H-bond acceptors as indicated by their β parameters. These interactions contribute greatly to the large negative solvation free energy values of tetrolic acid in ethanol and dioxane. It should also be noted that there is a peak near $r = 0.18$ nm in the TTAO2-EthanolH1 curve, while no peak exists near this position in the RDFs between TTAO2 and dioxane. This peak corresponds to the hydrogen bond formed between the carboxyl oxygen of TTA and the hydroxyl hydrogen of ethanol. This is consistent with the fact that ethanol is a strong H-bond donor ($\alpha = 2.7$), while dioxane is not ($\alpha = 0.9$). The RDF curves obtained from carbon tetrachloride (Figure 6a) are significantly broader than those obtained from dioxane and ethanol. There are no significant preferential interactions between TTA and carbon tetrachloride. The carboxylic acid group is poorly solvated in carbon tetrachloride, which has both a poor H-bond donor ability and H-bond acceptor ability indicated by its small α parameter and β parameter values (Table 2). In chloroform (Figure 6b), there is a clear peak near $r = 0.25$ nm on the TTAO2-CHCl3H1 curve, which is shifted to the right compared to the first peak on the TTAO2-EthanolH1 curve. This peak corresponds to the weak hydrogen-bonding interactions between the carboxyl oxygen of TTA and the hydrogen of chloroform, as a result of the weak H-bond donor ability of chloroform. In addition, there are no strong interactions between TTAO1 and chloroform, as indicated by the absence of clear peaks on all the TTAO1-CHCl3 curves. All these suggest that the solute–solvent interactions in chloroform are stronger than those in carbon tetrachloride, but weaker than those in ethanol and dioxane, which is consistent with their solvation free energy values.

3.2. Dimerization Reaction. As discussed in the Computational Details section, MD umbrella sampling is used to obtain the free energy profile of the dimerization reaction (shown in Figure 7). Figure 8 shows the free energy profiles of the dimerization reaction obtained from four different solvents. In the free energy curve obtained from the carbon tetrachloride solution, there is a dip near OP = 1.94 Å, which corresponds to the carboxylic dimer basin. For convenience, it is labeled state 1. When the OP is greater than 5.0 Å, the free energy curve is almost flat. This corresponds to two completely separated (fully solvated) TTA molecules, labeled state 3. The dip near OP = 3.6 Å corresponds to the intermediate catemer motif, where only one hydrogen bond is formed between two tetrolic acid molecules, labeled state 2. The basin positions do not shift significantly when switching solvents. In carbon tetrachloride, state 1 is the most thermodynamically stable state among the three states. Noting that the OP's are not necessarily the true reaction coordinates, the barriers reported for the

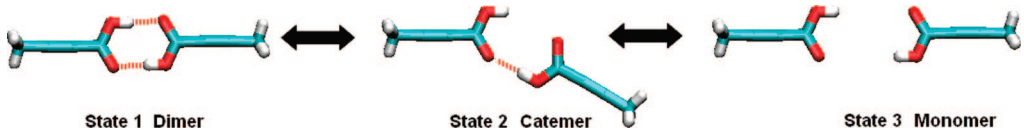


Figure 7. Dimerization reaction.

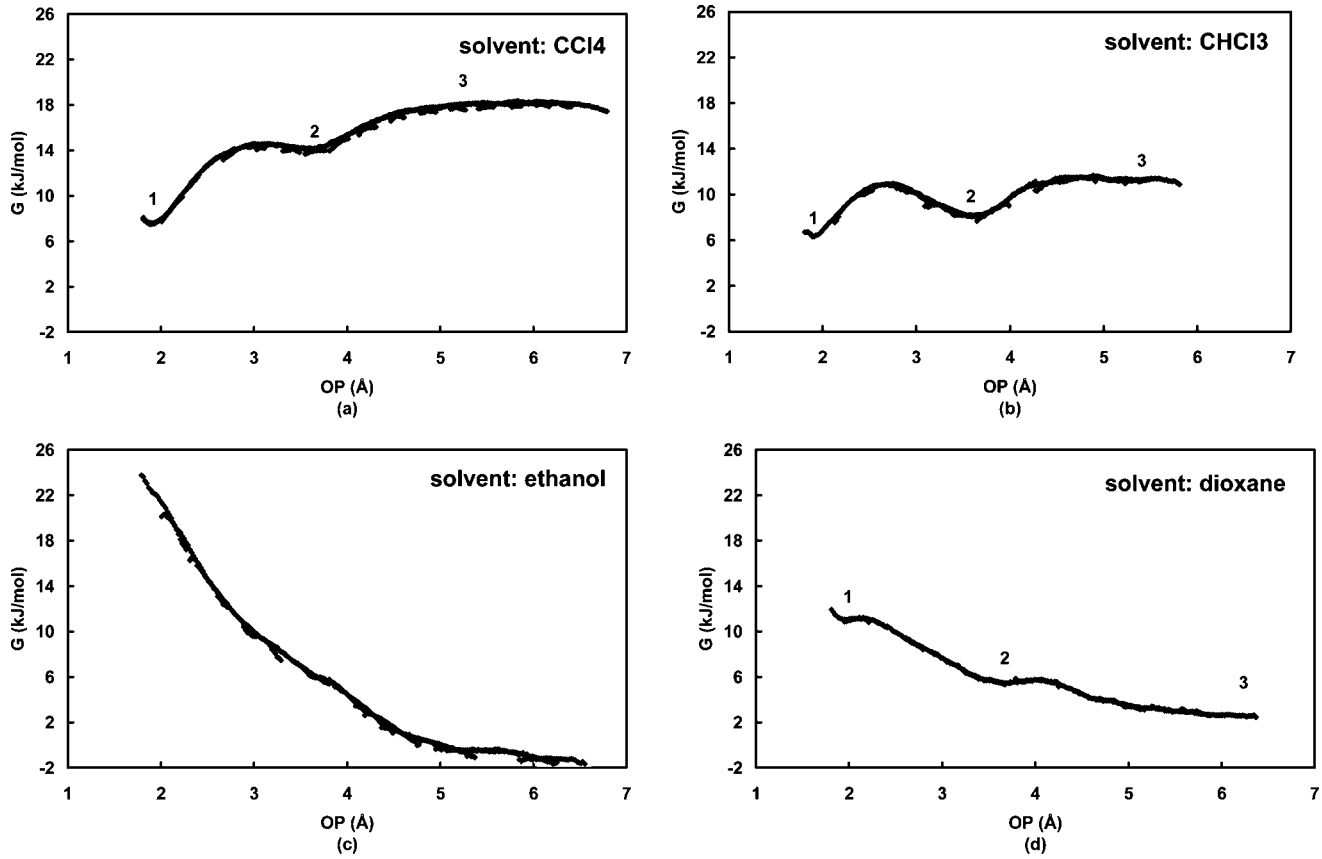


Figure 8. Free energy profiles for the TTA dimerization reation in various solvents at 298 K and 1 atm obtained by using MD umbrella sampling: (a) carbon tetrachloride, (b) chloroform, (c) ethanol, and (d) dioxane.

dimerization reactions may be underestimated. The true reaction coordinate could be obtained by using likelihood maximization,[33,34] an approach that uses information theory to determine the best reaction coordinate from a list of trial coordinates. However, this would be another study in itself, and we think that important qualitative insight can be gleaned from the OP results. There is almost no free energy barrier from state 3 to state 1.

In chloroform (Figure 8b), state 1 and state 2 are nearly thermodynamically equally stable with a free energy difference of only 1.3 kJ/mol. There is no energy barrier for the formation of catemers from fully solvated molecules, and there is an energy barrier of around 2.9 kJ/mol for the formation of carboxylic dimers from catemers, which suggests that the dimerization reaction for TTA in chloroform is less favorable than that in carbon tetrachloride. This may be due to the weak hydrogen bond between the hydrogen atom in chloroform and the oxygen atom in tetrolic acid. However, the 2.9 kJ/mol energy barrier is not very high and leaves state 1 and state 2 both kinetically accessible. The polymorph resulting from crystallization of TTA in chloroform might be a kinetically controlled process. At 298 K, the dimer-based crystal (form I) is the metastable form, nucleates first, and does not transform into the more stable catemer-based crystal (form II) in the time scale of the crystallization experiment.

In ethanol (Figure 8c), there are no clear dips to define the dimer and catemer basins. If we use the same order parameter value to define the carboxylic acid dimer, we can see that the free energy barrier to go from state 3 to state 1 is approximately 23.4 kJ/mol. State 1 becomes thermodynamically unfavorable, as does state 2. As discussed in the Solvation Free Energy Calculation section, the tetrolic acid molecule is well-solvated in ethanol. The strong interactions between ethanol and tetrolic acid make the solute–solute interactions unfavorable, and thus forming the catemer motif is unfavorable and forming a dimer even more so.

In dioxane, there are dips near 1.94 and 3.60 Å, which correspond to the dimer motif and catemer motif, respectively. The dimer motif is less favorable than the catemer motif, and the catemer motif is less favorable than two fully solvated molecules. Unlike those for the ethanol solution, the free energy differences between the three labeled states are much smaller, although ΔG_{solv} of TTA in dioxane is comparable to that in ethanol. The reason is that both oxygen atoms of TTA form strong interactions with solvents in ethanol due to ethanol's strong H-bond donor and acceptor abilities, while only the hydroxyl oxygen of TTA forms strong interactions with the solvent in dioxane, due to dioxane's strong H-bond acceptor ability but poor donor ability.

In conclusion, the interactions between solute and solvent molecules determine the relative thermodynamic stability of different growth synthons as well as the barriers on the pathways connecting these synthons. This analysis uses certain OP's in

TABLE 6: Free Energy Differences between the Three States on the Dimerization Pathway and Their Relative Population [a]

	ΔG_{32}	K_{32} ([D]/[C])	$\Delta G_{32}^{\ddagger}$	ΔG_{21}	K_{21} ([C]/[M]^2)	$\Delta G_{21}^{\ddagger}$	ΔG_{31}	K_{31} ([D]/[M]^2)
carbon tetrachloride	−3.8	4.6	0	−6.7	14.9	~0	−10.5	68.0
chloroform	−3.3	3.9	0	−1.3	1.7	2.9	−4.6	6.4
ethanol	7.1	6×10^{-2}	7.1	16.3	1×10^{-3}	16.3	23.4	8×10^{-5}
dioxane	2.9	0.3	2.9	5.4	0.1	~5.4	8.4	0.03

[a] State 1 carboxylic acid dimer (2 H bonds formed), state 2 catemer (1 H bond formed), state 3 two fully solvated TTA monomers (0 H bond formed). [D], [C], and [M] represent the concentrations of dimer, catemer, and monomer respectively. ΔG_{ij} is the free energy difference between state i and state j, and $\Delta Gij^{\ddagger}$ is the free energy barrier from state i to state j. They all have units of kJ/mol. K_{32}, K_{21}, and K_{31} have units of 1, L/mol, and L/mol, respectively.

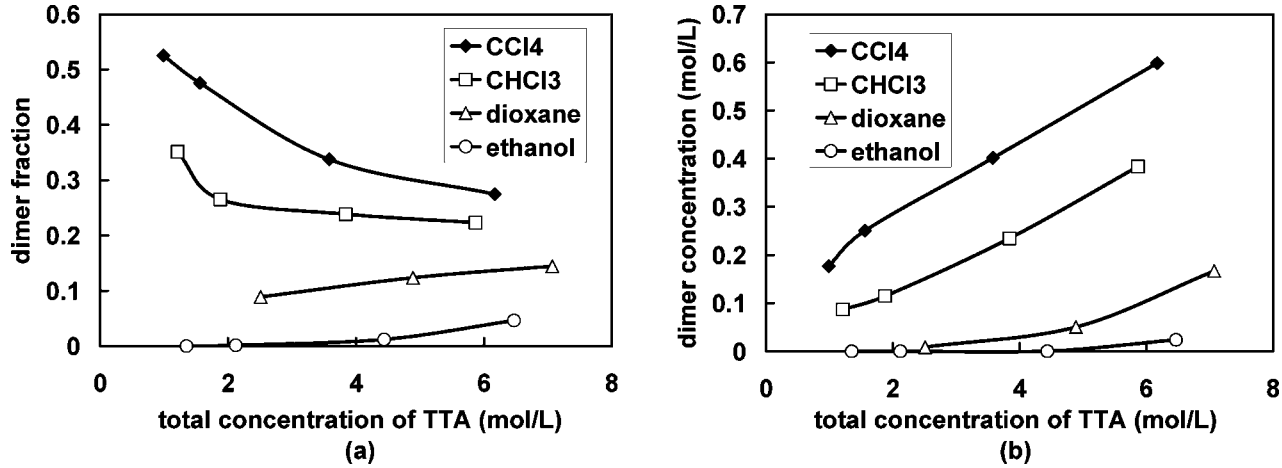


Figure 9. (a) Dimer fraction of the H-bond network in various tetrolic acid solutions. (b) Dimer concentration in various tetrolic acid solutions.

the dimerization reaction, but does not assume that the OP's are the true reaction coordinates. The free energy differences between those three states on the dimerization pathway estimated by using the OP approach and their relative populations are listed in Table 6.

3.3. Dimer Fraction of Hydrogen Bond Network. As shown in Figure 9a, dimer growth synthons exist in all of the tetrolic acid solutions, no matter what is the polymorph resulting from crystallization. However, the choice of solvent clearly affects the dimer composition of the hydrogen bond network of tetrolic acid. The dimer fraction is the highest in carbon tetrachloride solution and the lowest in ethanol solution, with the values in chloroform and dioxane in between. This is consistent with the order of the ease of carboxylic acid dimer formation in these four solvents (as shown in Figure 8). Moreover, the dimer fraction curves decrease to the right in chloroform and carbon tetrachloride, where a dimer-based crystal is obtained, but increase to the right in ethanol and dioxane, where catemer-based crystal or solvate is obtained. This fact can be used to help select the correct solvent to get the desired polymorph. One possible reason for this phenomenon is that stronger interactions between TTA molecules in chloroform and carbon tetrachloride cause the formation of larger clusters (trimer, tetramer, etc.), which reduces the opportunity to form a carboxylic acid dimer and thus decreases its composition in the entire hydrogen bond network. In other words, the H-bond network expands faster than the dimer motifs. Some snapshots of those conformations that form instead of the dimers are shown in Figure 10. Similar conformations were also reported previously by Gavezzotti in his work on acetic acid in carbon tetrachloride.[14] In ethanol and dioxane, the interactions between TTA molecules and the solvent molecules are much stronger. The chance for two TTA molecules to meet and form a dimer is very low when the solution concentration is low, but it increases when the total concentration is higher. In conclusion,

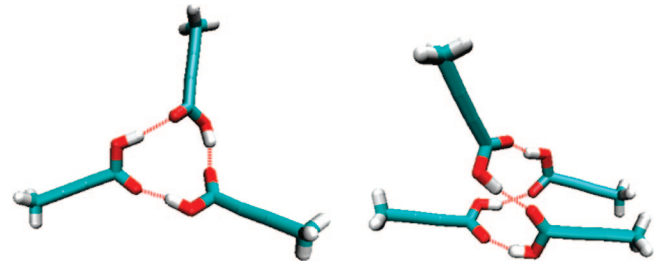


Figure 10. Clusters of TTA molecules which reduce the potential to form carboxylic acid dimer (obtained from MD simulations of TTA in carbon tetrachloride at 6.17 mol/L). H bonds are shown as red dashed lines.

the downward dimer fraction curves relate to the dimer-based polymorph. This is due to solute–solute interactions overcoming solute–solvent interactions. The opposite is the case for the upward dimer fraction curves, which relate to the catemer-based polymorph or solvate. This is the result of solute–solvent interactions overcoming solute–solute interactions. While the dimer fraction curve of TTA from chloroform and carbon tetrachloride shows a downward trend, the absolute concentration of dimer motifs still increases with the total TTA concentration in all four solvents, as shown in Figure 9b. Moreover, the dimer motifs are more abundant in carbon tetrachloride and chloroform than in ethanol and dioxane at any TTA concentration level, an observation that supports the link hypothesis.

4. Conclusions

The interactions between solvent and solute molecules play an important role in the self-assembly process of solute molecules in solution. The solvent cannot only change the relative thermodynamic stability of different growth synthons

of tetrolic acid, but also modify the energy barriers along the pathways to create these growth synthons. Weak interactions (less negative solvation free energy value) between the solvent and TTA molecules prompt two solute molecules to assemble into a carboxylic dimer, and also increase the dimer fraction of the overall H-bond network formed among all TTA molecules. This explains why TTA crystals obtained from chloroform pack in a dimer-based form. We also expect that the TTA crystal obtained from carbon tetrachloride solution will be in a dimer-based form. As the solute–solvent interactions become stronger (more negative solvation free energy value), the formation of a carboxylic dimer from two TTA molecules becomes both thermodynamically and kinetically unfavorable. In addition, the propensity of TTA molecules to form an extensive H-bond network is reduced, and the dimer fraction of the overall H-bond network decreases significantly. Strong solute–solvent interactions cause TTA molecules to crystallize in either catemer-based crystals or solvates. However, we cannot determine which one will be formed based solely on interactions in solution. To explain fully why TTA forms a solvate with dioxane, further investigations of the interactions of the TTA and solvent molecules in the solid phase are required, a task beyond the scope of the study. In the screening of the solvent for potential polymorphs in pharmaceutical crystallization, a wide range of solvents with different chemical and physical properties, like polarity, hydrogen bond donor ability, hydrogen bond acceptor ability, etc., should be investigated to find potential polymorphs. Such an investigation could be performed in silico, yielding molecular level insights as a complement to experimental results. We have demonstrated that molecular dynamics is a powerful tool to understand the solution behavior of target molecules in various solvents, particularly for aiding in the selection of the proper solvent for crystallization processes which yield the desired polymorph. First, solvation free energies can be computed to indicate whether or not the solvent molecules form strong interactions with the target solute molecule. Second, radial distribution functions can be applied to understand the detailed interactions between solute and solvent molecules. These can help in the selection of a wide range of solvents for solvent screening. Third, molecular dynamics of a mixture of solvent and solute molecules at various concentrations can be used to identify possible growth synthons. Fourth, umbrella sampling or a related method can be applied to compare the relative stability of various growth synthons in solution and find the most stable one. Moreover, molecular dynamics can be applied to study the cluster formation of the solute molecules in various solvents and at various concentrations, leading to a correlation between solution chemistry and solid state outcome, yielding rational guidance for further experiments.

Acknowledgment. This work was supported by the Singapore-MIT Alliance. We especially thank Dr. Ning Shan for helpful discussion. In addition, we thank Dr. Gregg Beckham for both technical and editorial contributions to this manuscript.

References and Notes

(1) Vippagunta, S. R.; Brittain, H. G.; Grant, D. J. W. *Adv. Drug. Delivery Rev.* **2001**, *48*, 3.
(2) Hunter, C. A. *Angew. Chem., Int. Ed.* **2004**, *43*, 5310.
(3) Afoakwa, E. O.; Paterson, A.; Fowler, M. *Trends Food Sci. Technol.* **2007**, *18*, 290.
(4) Fujiwara, M.; Nagy, Z. K.; Chew, J. W.; Braatz, R. D. *J. Process Control* **2005**, *15*, 493.
(5) Myerson, A. S. *Handbook of industrial crystallization*; Butterworth-Heinemann: Oxford, UK, 2002.
(6) Mullin, J. W. *Crystallization*, 4 ed.; Butterworth-Heinemann: Oxford, UK, 2001.
(7) Desiraju, G. R. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 2311.
(8) Steed, J. W.; Atwood, J. L. *Supramolecular Chemistry*; Wiley: New York, 2000.
(9) Parveen, S.; Davey, R. J.; Dent, G.; Pritchard, R. G. *Chem. Commun.* **2005**, 1531.
(10) Chattopadhyay, S.; Erdemir, D.; Evans, J. M. B.; Ilavsky, J.; Amenitsch, H.; Segre, C. U.; Myerson, A. S. *Cryst. Growth Des.* **2005**, *5*, 523.
(11) Davey, R. J.; Blagden, N.; Righini, S.; Alison, H.; Quayle, M. J.; Fuller, S. *Cryst. Growth Des.* **2001**, *1*, 59.
(12) Davey, R. J.; Dent, G.; Mughal, R. K.; Parveen, S. *Cryst. Growth Des.* **2006**, *6*, 1788.
(13) Towler, C. S.; Taylor, L. S. *Cryst. Growth Des.* **2007**, *7*, 633.
(14) Gavezzotti, A. *Chem. Eur. J.* **1999**, *5*, 567.
(15) Gavezzotti, A.; Filippini, G.; Kroon, J.; vanEijck, B. P.; Klewinghaus, P. *Chem. Eur. J.* **1997**, *3*, 893.
(16) Hamad, S.; Moon, C.; Catlow, C. R. A.; Hulme, A. T.; Price, S. L. *J. Phys. Chem. B* **2006**, *110*, 3323.
(17) Moroni, D.; ten Wolde, P. R.; Bolhuis, P. G. *Phys. Rev. Lett.* **2005**, 94.
(18) Hussain, M.; Anwar, J. *J. Am. Chem. Soc.* **1999**, *121*, 8583.
(19) Beckham, G. T.; Peters, B.; Starbuck, C.; Variankaval, N.; Trout, B. L. *J. Am. Chem. Soc.* **2007**, *129*, 4714.
(20) Leiserowitz, L. *Acta Crystallogr. Sect. B* **1976**, *32*, 775.
(21) Leiserowitz, L.; Benghiat, V. *J. Chem. Soc., Perkin Trans. 2* **1972**, 1763.
(22) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
(23) http://www.clippercontrols.com/info/dielectric_constants.html.
(24) Abraham, M. H.; Platts, J. A. *J. Org. Chem.* **2001**, *66*, 3484.
(25) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.
(26) Allen, M.; Tildesley, D. Computer Simulation of Liquids; Clarendon Press: Oxford, UK, 1987.
(27) Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.
(28) http://chemfinder.cambridgesoft.com/.
(29) Majer, V.; Svoboda, V. *Enthalpies of Vaporization of Organic Compounds: A Critical Review and Data Compilation*; Blackwell Scientific Publications: Oxford, UK, 1985.
(30) Kottalam, J.; Case, D. A. *J. Am. Chem. Soc.* **1988**, *110*, 7690.
(31) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187.
(32) Torrie, G. M.; Valleau, J. P. *J. Chem. Phys.* **1977**, *66*, 1402.
(33) Peters, B.; Trout, B. L. *J. Chem. Phys.* **2006**, 125.
(34) Peters, B.; Beckham, G. T.; Trout, B. L. *J. Chem. Phys.* **2007**, 127.

JP7106582