Molecular Aggregation of Acetic Acid in a Carbon Tetrachloride Solution: A Molecular Dynamics Study with a View to Crystal Nucleation

A. Gavezzotti*[a]

Abstract: Molecular dynamics simulations in the nanosecond time range were conducted on a system initially consisting of 50 widely separated acetic acid solute molecules within a box of 1659 solvent units. Solute aggregation was promoted by neglecting the usual periodic boundary conditions, and by forced withdrawal of solvent molecules from the resulting droplet. Organization of solutes into cyclic oligomer structures is readily observed by the use of different force fields and different computational setups, and their partly fluxional behaviour was investigated in some detail. Molecular diffusion bringing prospective interaction partners into short-range distance is seen to be sufficient for aggregation, while the effect of long-range forces is suggested to be minor. As the droplet becomes smaller and the solute mole fraction increases, the formation of liquidlike solute micelles is observed. A crystalline cluster of a size comparable with that of these micelles almost instantaneously converts into a liquid structure in the dynamic simulation, both in vacuo and solvated. Within the qualitative limitations of the model, the emerging picture suggests the formation of a microemulsion of liquidlike particles as the first step of crystal nucleation. The diffusive (pico- to nanosecond) regime, comfortably accessible to present-day computing, can and should be more and more expoloited in the preliminary study of crystal nucleation.

Keywords: computer chemistry · crystal nucleation · molecular dynamics · molecular recognition

Introduction

The crystal chemistry of organic compounds is a well-developed and quantitative branch of science as far as structure is concerned, interatomic distances within crystals being known with high accuracy thanks to X-ray diffraction. The energetics of most types of intermolecular interactions commonly occurring in organic crystals is also well-described by several kinds of intermolecular empirical potentials. However, very little is known at a molecular level about the mechanisms of crystal formation; nucleation theory has been called one of the few areas of science in which agreement of predicted and measured rates to within several orders of magnitude is considered a major success.[1] In solution, the process must involve elementary molecular recognition steps, followed by nucleation and growth; none of these stages is easily amenable to experimental investigation (although succesful experiments have been reported for crystalline films[2]), so that the field is open for theoretical simulation methods. The molecular dynamics (MD) simulation of pure liquids has attained a high level of sophistication for organic molecules[3] and for water,[4] a particularly difficult case owing to polarization effects; enthalpies of solvation have been estimated by the use of a single solute molecule in water.[5] With regard to the first molecular recognition steps, several MD studies have been conducted on biological or supramolecular dimers,[6] and, from free energy estimation methods, on benzene dimerisation[7, 8] and on hydrophobic effects in water,[9, 10] but very little has been done with an explicit focus on crystal precursor states.[11] A number of elegant MD or MC studies have been carried out in conjuction with sophisticated experiments, on the melting, recrystallisation and phase transitions of molecular clusters,[12] and crystallisation from solvent has been tackled computationally with monoatomic sample systems.[13] No attempts have so far appeared to monitor directly the behaviour of numerous organic solute molecules in a solvent as a function of concentration, with the aim to simulate, if not to describe properly, the coagulation process that may lead to solute clustering into oligomers and micelles, to partial demixing, and eventually to crystal development in solution.

This paper is an attempt to tackle the above problem by standard classical molecular dynamics simulations. The GROMOS96[14] package was employed. Acetic acid as a solute, and carbon tetrachloride as a solvent, provide computationally convenient and reasonably well parameterized

[a] Prof. A. Gavezzotti
Dipartimento di Chimica Strutturale e Stereochimica Inorganica
University of Milano, via Venezian 21, Milano (Italy)
Fax: (+39) 02-70635288
E-mail: gave@stinch12.csmtbo.mi.cnr.it

  0947-6539/99/0502-0567 $ 17.50+.50/0

systems—the solvent, in particular, is nonpolar and can be reliably modelled in the single-atom approximation. Their high bulk miscibility (the mole fraction of acetic acid at the solid–solution equilibrium is around 0.5 at 273 K[15]) is an obvious obstacle that has to be circumvented by computational expedients. In molecular dynamics simulations, bulk effects are usually modelled by applying periodic boundary conditions, which understandably are detrimental when attempting to simulate phase transitions or heterogenous processes.[12] Therefore, our main modelling assumption was the use of a finite-size box, actually a droplet. The first part of the paper deals with the preliminary recognition process that leads to hydrogen bonding, mainly into dimers, starting from an artificial, dilute solution of isolated acetic acid molecules. The influence of temperature and of coulombic interactions is investigated, and the importance of short- and long-range attraction is probed by manipulating the interaction cutoff distance. Two force fields, one developed for crystals (the UNI force field[16, 17]), and one for liquids (the OPLS-united atom force field[18]) are also compared. In the second part, solute concentration by evaporation is hinted at by artificially depleting the solution of solvent molecules, thus computationally increasing the solute mole fraction, and, what is even more important, reducing the droplet size, thus increasing the surface and interface tension. These artifices, it appears, force solute molecules to condense into multimolecular aggregates and allow a preliminary study of the phenomena of interest here. Finally, looking from the opposite end, the evolution of a small crystalline nucleus in solution was simulated. The results are discussed with an effort to provide some insight into the otherwise elusive phenomenon of molecular recognition and crystal nucleation in solution.

Abstract in Italian: *La dinamica molecolare classica nella scala dei tempi del nanosecondo è stata usata per studiare i primi stadi del riconoscimento molecolare durante l'incipiente nucleazione di acido acetico da un solvente apolare. Diversi campi di forza portano a conclusioni simili: l'aggregazione appare promossa in prima istanza da processi diffusivi, fino alla formazione di micro-micelle a struttura liquida. La microemulsione risultante è probabilmente lo stadio precursore su scala molecolare della nucleazione cristallina.*

Computational Methods

Force fields

The UNI model: Molecular dimensions for acetic acid were taken from its crystal structure determination,[19] with C–H bond lengths renormalised to 1.08 Å, HCH angles at the tetrahedral value, and the O–H bond length renormalised to 1.0 Å. Internal degrees of freedom in solute molecules were computationally frozen by applying the SHAKE constraining procedure[14] on bond lengths, and by use of a very high force constant (600 kJ mol^{-1}) on all bond angles over the bending potential $E = 1/2\,k_B(\cos\theta - \cos\theta^0)^2$. An improper dihedral restraint was imposed to enforce planarity at the trigonal carbon with the GROMOS force constant of 0.051 kJ mol^{-1} deg^{-2}. Torsional libration around the single C–O bond was restrained by the GROMOS potential $E = 16.7(1 - \cos 2\phi)$ kJ mol^{-1}, corresponding to a barrier of 33.4 kJ mol^{-1}. All calculations were started with $\phi = 180^\circ$ (*cis* configuration), and practically no jumps over the barrier were observed. It may be worth mentioning that test calculations performed without this restraint gave much worse results in all respects. CCl_4 molecules were modelled in the united atom approximation. In the UNI force-field calculations, nonbonded intermolecular interaction parameters (Table 1) for the 12-6 potentials were obtained from an adaptation[20] of crystalline potentials. The 12-6 recasting of the original 6-exp potential form results[20] in excess repulsion at short interatomic distance especially for H···H and O···O potentials. These were therefore corrected by multiplying the well depths by 1.08 or 1.06, and the equilibrium distances by 0.97 or 0.98, respectively. Nonbonded intramolecular contributions were always excluded. Literature solvent–solvent parameters were used,[21] and the usual square-root combination rule was applied to obtain all solute–solvent interaction parameters.

Table 1. Nonbonded energy parameters for $E = A \times R^{-12} - C \times R^{-6}$ in kJ mol^{-1}, in which R is any intermolecular distance in Å. ε is the well depth, R^0 the distance at the minimum.

	A	C	ε	R^0	force field
C···C	1.3590×10^7	4.8868×10^3	0.439	4.209	OPLS[18][a]
C–solvent	2.9630×10^8	3.9822×10^4	1.338	4.960	
O···O	1.5899×10^6	2.3639×10^3	0.879	3.322	
O···solvent	1.0135×10^8	2.7696×10^4	1.892	4.406	
OH··OH	1.5120×10^6	2.0741×10^3	0.711	3.367	
OH···solvent	9.8833×10^7	2.5943×10^4	1.702	4.436	
Me··Me	3.4189×10^7	9.5682×10^3	0.669	4.389	
Me–solvent	4.6997×10^8	5.5722×10^4	1.652	5.064	
H···H	6.5671×10^4	1.0901×10^2	0.045	3.262	UNI[16, 17]
H···C	3.3632×10^5	5.2566×10^2	0.205	3.295	
H···O	1.1776×10^5	4.8773×10^2	0.505	2.801	
C···C	4.6636×10^6	2.6882×10^3	0.387	3.891	
C···O	3.2987×10^6	2.9812×10^3	0.674	3.610	
O···O	1.3627×10^6	1.3926×10^3	0.356	3.537	
HB···O	8.2694×10^3	9.8208×10^2	29.16	1.601	
H···solvent	2.0597×10^7	5.9476×10^3	0.429	4.366	
C···solvent	1.7357×10^8	2.9535×10^4	1.256	4.786	
O···solvent	9.3825×10^7	2.1258×10^4	1.204	4.546	
solvent	6.4600×10^9	3.2450×10^5	4.075	5.844	Ref. [21]

[a] Cross interactions by the square-root rule.

The UNIq model: The UNI force field in its general formulation needs no charge parameters, but can be supplemented by some, in special cases.[17] For calculations involving coulomb-type contributions (CTC) in acetic acid, the carboxyl group was considered as a single charge group with the GROMOS96 charge distribution,[14] 0.530, –0.380, –0.548 and 0.398 e$^-$ for C, =O, O and H, respectively. The UNI formulation is self-sufficient in describing molecular crystals even without R^{-1} terms, because the hydrogen-bonding 12-6 potential already incorporates a substantial attractive energy (well depth 29 kJ mol^{-1} at R^0 of 1.6 Å). The CTC superimpose a presumably exaggerated attractive effect over the hydrogen-bonding system, but this setup permits a computational experiment in which these long-range interactions are deliberately excluded or introduced. Artifacts in the detail of structures or of interaction energies cannot be ruled out, but the results of UNIq calculations will be used strictly for comparison with chargeless models, and not as carriers of genuine structural or energetic information.

The OPLS model: In this case, solute molecular dimensions and the intermolecular potential were taken from the OPLS force field,[18] in which the methyl group is described in the united-atom approximation. Solute–solvent interaction parameters were obtained by the square-root combination rules. The whole molecule was taken as a single-charge group with OPLS charges of 0.55, –0.50, –0.58, 0.45 and 0.08 electrons for C, =O, O, H and the methyl group, respectively. The intramolecular part of the field was kept the same as in the UNI model. The cutoff distance in intermolecular nonbonded energy calculations was chosen to ensure that long-range effects, if any, would be taken into account. The GROMOS twin-range method was used: a nonbonded distance pairlist was constructed at 25 Å,

0947-6539/99/0502-0568 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 2

interactions were calculated at 20 Å, updating the pairlist and calculating the full interaction at 25 Å every 100 MD steps. To test the consequences of using a shorter cutoff distance, simulations were also conducted with cutoffs of 17 (versus 25) and 15 Å (versus 20). No correction for truncation of the coulombic sums was applied to these systems of relatively few polar entities immersed in a nonpolar medium.

The systems

Pure crystals and liquids: As a check of the UNI and OPLS models, standard MD simulations of pure liquid and crystalline acetic acid were conducted at 250 K and 1 atm, in an N/P/T ensemble, with 108- and 192-molecule boxes, respectively, with periodic boundary conditions.

Solutions: The starting box for the simulation of solutions was constructed as follows. A 2196-atom solvent box was equilibrated at 250 K with periodic boundary conditions, and a smaller (1709 atoms) box was extracted by deleting the outermost solvent atoms. 50 solute molecules were then inserted at random locations and in a random orientation, replacing as many solvent molecules. A few energy minimisation cycles ensured that no strongly repulsive contacts were present, while a 0.2 ps MD runs served the purpose of assigning initial velocities at each temperature. The coordinate and velocity output from this calculation will be referred to in the following as the zero set. No hydrogen bonds were present in this setup, and only a few intermolecular distances were shorter than 10 Å (an intermolecular distance is defined as the distance between the centres of coordinates of two solute molecules). All subsequent MD runs were at constant temperature, ensured by the standard GROMOS T-coupling procedure (internal and external, solvent and solute degrees of freedom separately coupled to a bath, coupling constant 0.1 ps). Since no periodic boundary conditions were applied, surface solvent and/or solute molecules are in principle free to leave the droplet, the computational equivalent of true evaporation. A number of MD runs were performed, starting from the zero set, to check the first stages of the molecular aggregation process against the effects of changes in force field, in cutoff distance and in temperature. Longer, subsequent simulations of the evaporation stages were formally conducted at constant mass, but at chosen times the run was stopped, and outermost solvent and solute molecules were manually removed from the system, while preserving coordinates and velocities of all remaining molecules. When the solute mole fraction had reached 0.1, this procedure was discontinued and the simulation proceeded on a droplet of nearly constant mole fraction. The number of genuinely evaporated solute or solvent molecules, having spontaneously drifted away from the droplet, was always very small.

Crystal nuclei in solution: A crystal nucleus of 37 molecules was extracted from the acetic acid crystal structure by taking the 36 nearest-neighbours to a reference molcule; this crystal nucleus was simulated both in vacuo and surrounded by 310 solvent molecules, in the latter case forming a droplet with the same mole fraction as the end product of the final steps as above. The solvent molecules were added to the crystal nucleus by first forming a regular grid, avoiding highly repulsive contacts, and then running a short energy minimisation cycle. The conditions were the same (UNI model, 280 K, 25 Å cutoff) as in the former simulation; a 2 ps preliminary MD run was used to assign velocities, initially at 150 K and then warming to 280 K.

Results

Pure acetic acid, UNI model: For the liquid at 250 K after 40 ps, the simulation box reached a steady behaviour of energies and density, and the estimated heat of vaporisation is 47 kJ mol^{-1} (expt 54[18]), while the estimated density is 1.12 g cm^{-3} (expt 1.04 at 298 K). The crystal of acetic acid is a particularly difficult test for the potentials, since cooperative effects are present in the infinite chain arrangement of hydrogen bonds, which are beyond the capabilities of simple atom–atom two-body potentials. Nevertheless, equilibration at 250 K after 40 ps yielded an estimated enthalpy of sublimation of 54 kJ mol^{-1} (expt 66[17]), and a density of 1.23 g cm^{-3} (expt 1.27 at 278 K[19]). The chargeless UNI potentials describe the molecular interactions in the condensed phases of acetic acid in a way that was considered satisfactory for the present purposes.

Acetic acid in CCl_4: first stages of aggregation. From the zero set, five simulations, each 200 ps long, were carried out: i) UNI model, 250 K, cutoff 25/20 Å; ii), UNI model, 250 K, cutoff 17/15 Å; iii), UNI model, 280 K, cutoff 25/20 Å; iv), UNIq model, 250 K, cutoff 25/20 Å, and v), OPLS model, 280 K, cutoff 25/20 Å.

The choice of temperatures resulted from several considerations. The main aim being a study of the self-aggregation of initially faraway molecular entities, thermal motion had to be kept within a suitable window, neither too high (too much contrast to intermolecular attraction) nor too low (too little diffusive freedom for molecules to reach their recognition partners). In the absence of better information, 250 K was chosen as close to the freezing temperature of the solvent, 280 K as slightly below that of the solute.

In all simulations, the nearly cubic starting box, whose shape results from equilibration of the solvent with periodic boundary conditions (Figure 1a), quickly collapsed to a

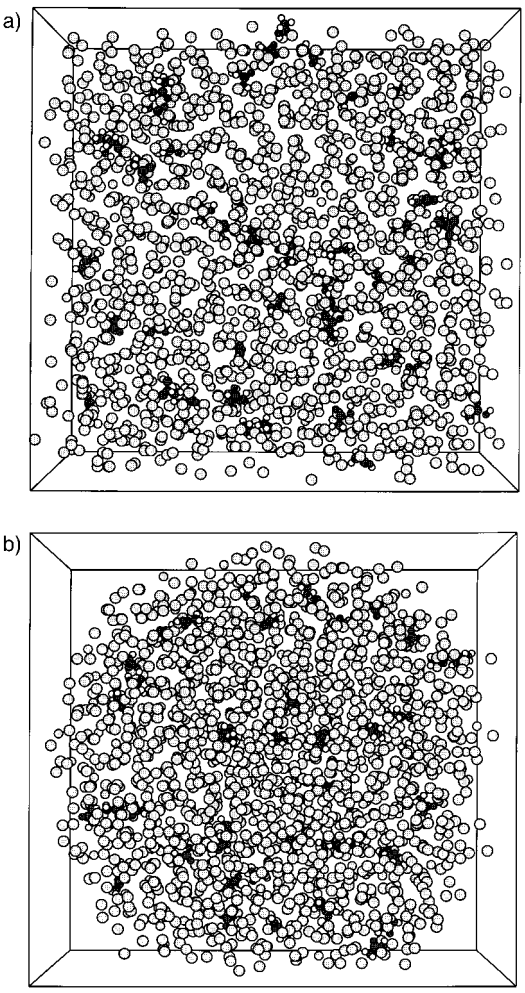

Figure 1. a) The starting system for the simulation of solutions, containing 50 acetic acid and 1659 carbon tetrachloride units; b) the system after 50 ps (UNI force field, 250 K, but this result is representative of all other simulations). Black dots are oxygen atoms.

droplike arrangement (Figure 1b). Total energies underwent only minor evolution to increased stability, suggesting that the starting system was not too unrealistic. Intramolecular potential energies were small, indeed negligible with respect to intermolecular ones, confirming that intramolecular motion had been indeed quenched by the applied restraints.

To monitor the aggregation process, three geometric entities were defined: i) the hydrogen bond, an O···H contact shorter than 2.6 Å, a generous threshold in view of thermal motion; ii) the cluster, any straight, branched or cyclic sequence of molecules connected by intermolecular distances less than 5.4 Å; and iii) the histogram of solute intermolecular distances, an indicator that becomes significant only at higher aggregation stages. The relevant results for the five preliminary simulations are collected in Table 2. Figure 2 shows the comparative evolution of the intermolecular interaction energies.

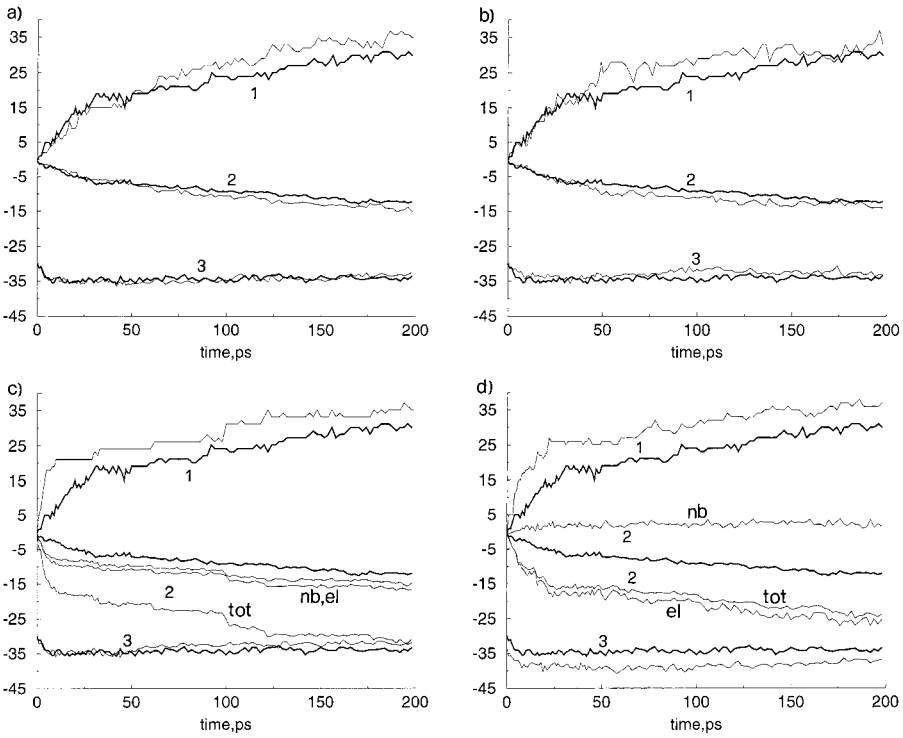

Figure 2. Preliminary simulations: first 50 ps, curves sampled at 1 ps intervals, then at 2 ps intervals. Heavy lines: UNI force field, 250 K, 25 Å cutoff; curve 1, number of hydrogen bonds; curve 2 solute – solute interaction energy per solute molecule (kJ mol^{-1}); curve 3 solute – solvent interaction energy per solute molecule (kJ mol^{-1}). a) Comparison with 1.7 Å cutoff; b) comparison with $T = 280$ K; c) comparison with UNIq force field: solute – solute energy curves split into nonbonded (nb), electrostatic (el), and total (tot) contributions; d) comparison with OPLS force field, solute – solute curves split as in c).

The first 10 ps or so in each simulation involve the rearrangement of the droplet from the approximately cubic into a spherical shape, with an overall stabilisation. Lowering the cutoff distance (Figure 2a) reduces the importance of solvent – solvent interactions relative to solute – solute interactions, since solvent particles are a single, large entity, and this is presumably the reason for the increased aggregation rate after 80 ps or so. Raising the temperature (Figure 2b) significantly enhances the aggregation rate in the medium and in the long run. The introduction of CTCs (Figure 2c) substantially increases the initial aggregation rate and somewhat reduces the fluctuations in the number of hydrogen bonds, as expected from the much stiffer hydrogen-bonding potential, but in the long run results do not differ substantially from those of other calculations. The OPLS force field results (Figure 2d) show a largely increased rate of aggregation in the very first stages, but in the long run the structural results do not differ significantly from those from the UNIq model or from the chargeless UNI model.

The evolution of solute – solute interaction energies parallels that of the number of hydrogen bonds; these cohesive energies increase steeply in the very first part of the recognition process, as isolated molecules agglomerate into oligomers. They are much larger in calculations including coulombic terms, and, as expected, are larger in the UNIq than in the OPLS model. Solute – solvent and solvent – solvent interaction energies, always stabilising, level off at -35 and -22 kJ mol^{-1}, respectively, after a short induction period and are independent of the degree of solute aggregation and insensitive to changes in the force field. Indeed, major structural rearrangements after the first 100 ps occur at constant energy. Comparisons with experimental heats of mixing were not attempted, in view of the difficulty of their definitions, mostly related with the formation of solute oligomers.

Table 2. Acetic acid in solution: final frames of 200 ps simulations. Distances [Å], energies [kJ mol^{-1}].

Force field\T\cutoff	H-bonds			Clusters[a]	Energies[b]		
	number	distance range	average		Ac–Ac	Ac–S	S–S
UNI\250\25	30	1.63 – 1.91	1.77	9(2), 4(3)	– 12	– 22	– 34
UNI\250\17	36	1.60 – 1.94	1.75	12(2), 4(3)	– 15	– 22	– 33
UNI\280\25	34	1.59 – 2.40	1.81	12(2), 2(3), 1(5)	– 14	– 22	– 34
UNIq\250\25	35	1.57 – 1.83	1.67	12(2), 3(3)	– 34	– 23	– 31
OPLS\280\25	36	1.46 – 2.10	1.71	11(2), 3(3), 1(5)	– 25	– 21	– 36

[a] H-bonded clusters only. $n(m)$ means n clusters each with m molecules. [b] Ac: acetic acid, S: solvent.

0947-6539/99/0502-0570 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 2

All simulations, in all conditions, show the aggregation of solute molecules in some patterns, among which are recognisable cyclic dimers (two molecules joined by two hydrogen bonds), cyclic dimers plus one appended hydrogen-bonded molecule, and cyclic trimers (see Table 2). Larger aggregates are very sporadic within the 200 ps time range. Cyclic dimers occasionally undergo rupture of one of the two double bonds, in what have been called catemer jumps,[22] and this is one source of fluctuations in the total number of hydrogen bonds; hydrogen bond exchanges are, however, very frequent, the lifetime of an oligomer being typically of the order of a few tens of picoseconds. Molecules and/or dimers occasionally come within the dispersive interaction range without condensing by hydrogen bonding, but these structures are fleeting at the concentration used, with picosecond lifetimes. There seem to be no really significant differences in the number and type of hydrogen-bonding structures or clusters formed in each of the calculations with different options, at least within the 200 ps timescale (see Table 2). The hydrogen bonding distances, as expected, are marginally shorter in calculations including CTCs.

The main conclusions that can be drawn from these preliminary calculations are: i) although with largely different cohesive energies, all force fields lead to essentially the same structural results; ii) the effect of cutoff is marginal in the first stages of clustering, dominated by hydrogen-bond formation; iii) the inclusion of CTCs enhances the aggregation rate of single solute molecules, but is less effective in enhancing further aggregation, when a significant number of dimers has been formed; this is presumably the result of the neutralisation of the molecular dipole that results from nearly centrosymmetric pairing; iv) the increase in temperature from 250 to 280 K speeds up the solute aggregation process; v) the inclusion of CTCs is not indispensable for molecular clustering as long as it mainly involves hydrogen bonding, and the UNI model is as efficient as the UNIq or OPLS models in this respect. Significantly, a simulation at 250 K with the UNI model on the 50 solute molecules in the starting box, without solvent, resulted in complete dispersion in space of monomers after a few picoseconds; use of the OPLS model led to formation of a few dimers, followed by dispersion. Thus, no solute force field is in itself sufficient to bring about extensive clustering from the starting configuration. The solvent is indispensable, and acts as a cohesive medium that provides a chance for solute molecules to come into short-range contact (intermolecular distance < 10 Å) and consequently to aggregate.

Taken together, these results suggest that aggregation of acetic acid molecules towards hydrogen bonding in an nonpolar solution is more controlled by diffusion than driven by long-range forces. The very fast formation and the persistence of cyclic dimer structures, recorded in calculations with or without CTCs, is not in contrast with standard pictures of acetic acid solubilisation.

Acetic acid in CCl_4—long-run simulations: On the basis of the above results, longer simulations were planned using the pure UNI model without CTCs. The temperature was kept at the more favourable value of 280 K, and the cutoff distance was kept at 25 Å, since a longer-range interaction might be effective in further recognition, based on dispersive forces.

Starting from the output of the first 200 ps, several successive 200 ps simulations were run. At the end of each of these time periods, coordinates and velocities of peripheral solute and solvent molecules having less than four nearest neighbours (defined by an intermolecular distance of less than 6 Å) were removed from the trajectory.

The removal of mass from the system, both by spontaneous evaporation and by forced depletion, produced rather large effects on the distribution of kinetic energies; in particular, the analysis of starting temperatures of each new system revealed that solutes migrating into vacuum had drawn internal, and to a lesser extent translational, kinetic energy from the corresponding baths, lowering the internal and translational temperature of solutes left behind. These disturbances were quickly quenched by the temperature-coupling mechanism, but they are presumably the cause of the discontinuities observed in the energy curves (Figure 3a). Care was taken to remove any motions of the overall centre of mass of the system.

Numerical results are collected in Table 3, while Figures 4–7 show some structural detail. The solute cohesive energy decreases steadily (Figure 3a), although at 1350 ps it is still far from that of the pure liquid, or 54 kJ mol^{-1}. The total number

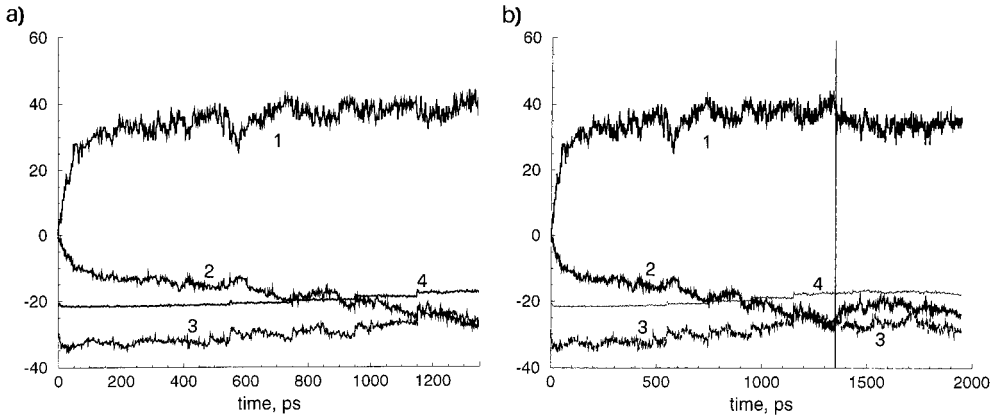

Figure 3. Long-run simulations (UNI force field, 280 K, 25 Å cutoff). a) First 1350 ps, with forced evaporation-concentration procedure; b) the complete run, with constant concentration after 1350 ps. Curve 1, number of hydrogen bonds; curve 2, solute–solute interaction energies per solute molecule (kJ mol^{-1}); curve 3, solute–solvent interaction energies per solute molecule (kJ mol^{-1}); curve 4, solvent interaction energy (kJ mol^{-1}).

Table 3. Acetic acid in solution: composition and structures in the final frames of each section of the simulation. UNI force field, $T = 280$ K, 25 Å cutoff.

Time [ps]	Solute molecules	Solvents molecules	Number of H-bonds	Clusters[a]	Total clusters	Single molecules
0	50	1659	0	none	0	50
50	50	1659	25	9(2), 2(3)	11	26
200	49	1644	34	10(2), 2(3), 1(4), 1(6)	14	14
350	46	1402	32	9(2), 1(3), 3(4)	13	16
550	44	1032	33	9(2), 5(3), 1(4)	15	9
750	41	735	42	7(2), 1(3), 1(4), 2(5),1(7)	12	6
950	40	514	40	5(2), 2(3), 1(5), 1(7), 1(11)	10	2
1150	38	336	39	3(2),1(4),1(5),2(11)	7	3
1350	37	328	40	1(2), 2(8), 1(9), 1(10)	5	1
1550	37	315	32	5(2), 1(3), 1(6), 1(7), 1(8)	9	3
1750	36	303	34	1(2), 1(3), 2(4), 1(7), 1(13)	6	3
1950	35	294	36	2(2), 2(3), 1(4), 1(19)	6	2

[a] Clusters held together by hydrogen bonds or by dispersive-electrostatic interactions; $n(m)$ means n clusters each with m molecules.

of hydrogen bonds formed rises asymptotically until, at about 1200 ps, it becomes greater than the number of solute molecules; visual inspection of selected frames confirmed that at that stage very nearly all of the remaining solute was engaged in hydrogen bonding. Solute–solvent interaction energies decrease sluggishly, in parallel with solute segregation; the solvent cohesive energy stays constant, except for a spurious and quantitatively insignificant long-run decrease due to the reduction of the system size.

The data in Table 3 show the evolution of the clustering process, with many small clusters at the beginning, replaced by a few larger clusters towards the end, while the number of isolated molecules becomes negligible. Figure 4 shows the

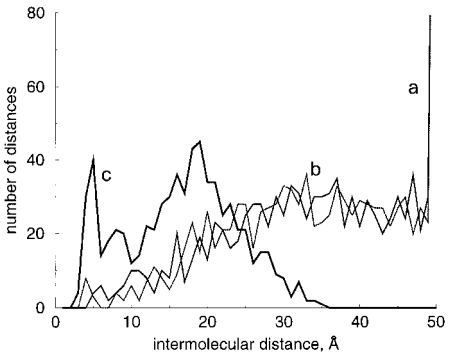

Figure 4. Distribution of solute intermolecular distances in a) the starting system of Figure 1a; b) the system in Figure 1b after 50 ps; c) the droplet (Figure 5) after 1350 ps.

distribution of solute intermolecular distances, with a prominent peak at the hydrogen-bonding distance around 5 Å at the end of the simulation. Solute aggregation modes pass through several stages, with cyclic dimers growing into larger and larger clusters, until 10–15-molecule micelles are formed within the solution. Figure 5 shows the last frame of this part of the simulation after 1350 ps, with one large multimolecular aggregate, two medium-size ones, and a few cyclic dimers still floating within the solution. Thus, the agglomeration process has not led to a single micelle comprising all solute molecules.

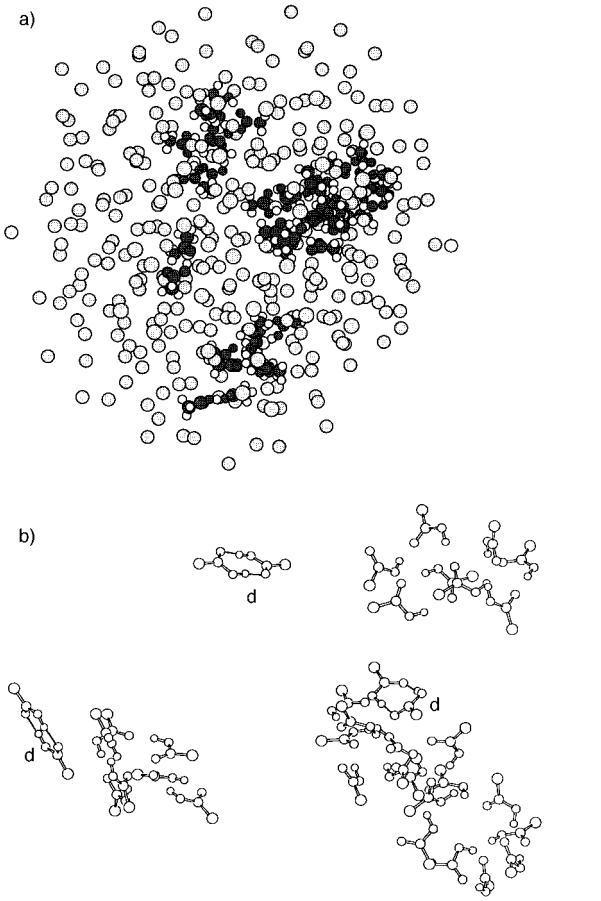

Figure 5. Structural detail in the 1350 ps droplet: a) whole system; b) solutes only, methyl hydrogens omitted for clarity. Cyclic dimers are labelled **d**.

At least one cyclic dimer is clearly embedded in a larger cluster, bound to it by dispersive interactions only.

The analysis of the detailed structure of moderately persistent (not completely fleeting) solute aggregates is feasible up to clusters containing about five molecules, after which the aggregation mode becomes too complicated for single pictures to be sorted out. From extensive spot checks during the simulations, the favourite aggregation mode in the short to medium run is clearly the cyclic dimer with or without appended molecules (Figure 6a), which, when the structure is observed over a few to a few tens of picoseconds, sometimes are inserted into the cycle (Figure 6b). Tetramers (Figure 7) sometimes arise from dimer pairs followed by an insertion mechanism quite similar to the evolution from Figure 6a to Figure 6b. Pentamers already show an increasing tendency of

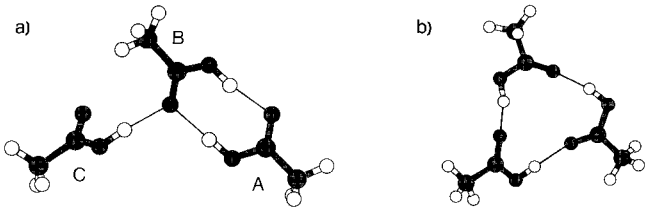

Figure 6. Elementary aggregation modes: a) cyclic dimer (A, B) plus a catemer precursor molecule (C), cohesive energy 22 kJ mol^{-1} per molecule; b) cyclic trimer, cohesive energy 23 kJ mol^{-1} per molecule.

0947-6539/99/0502-0572 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 2

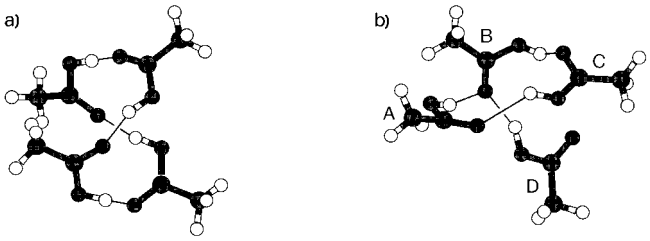

Figure 7. a) Twisted chain tetramer, cohesive energy 24 kJ mol^{-1} per molecule; b) cyclic trimer (A–B–C) plus one appended molecule (D), cohesive energy 22 kJ mol^{-1} per molecule. Note the hydrogen-bond bifurcation at the carbonyl oxygen of molecule B.

the solutes to segregate polar groups out of the solvent, compatibly with the requirement of steric avoidance of the methyl groups. Figure 8 shows some of the expedients,

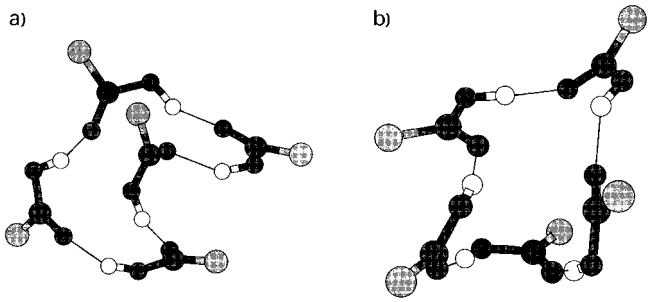

Figure 8. Cyclic pentamers: a) UNI force field (methyl hydrogens omitted for clarity), coehsive energy 28 kJ mol^{-1} per molecule; b) OPLS force field, cohesive energy 41 kJ mol^{-1} per molecule.

actuated, it appears, irrespective of the force-field employed; 20–30 ps is the upper threshold for the lifetime of such cyclic structures. One minor difference between OPLS and UNI is that in the former force field, the O(H) oxygen can act as a hydrogen bond acceptor, and indeed does so in a minor fraction of occurrences here, as was the case in OPLS liquid acetic acid.[18] No experiment is available to resolve this ambiguity. Figure 9 shows this phenomenon, along with another example of the cyclisation–insertion mechanism that brings about the formation of larger cycles.

Larger aggregates (7–20 molecules) show complicated and largely fluxional structures, of which it can only be said that they presumably approach those to be found in the pure liquid. Acetic and formic acid are known[18, 23] to form fluxional chain structures rather than dimers in their liquids, and visual analysis of our clusters conforms to this view, although a quantitative statistical analysis is difficult owing to their small size. Some snapshots may illustrate a few of the instantaneous modes of these aggregation units: Figure 10a shows a cyclic dimer stitching together two short chains; Figure 10b

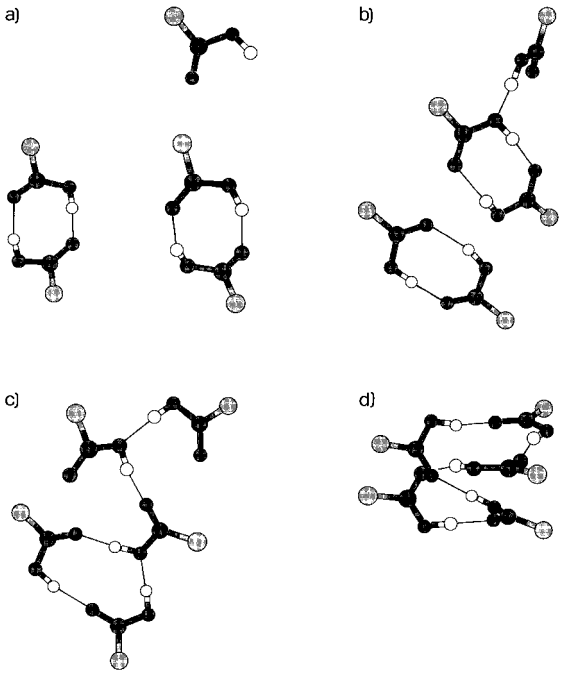

Figure 9. The formation of a cyclic pentamer within the OPLS model. a) Initial structure; b) 48 ps, lone molecule captured by O(H) acceptor; c) 58 ps, insertion into a dimer by O(H) acceptor, d) 130 ps, cyclised pentamer with only carboxyl oxygens as acceptors (more stable structure). The cohesive energy per molecule increases from 29 to 31 to 35 and to 40 kJ mol^{-1}.

shows a parallel arrangement of dimers in dispersive contact; Figure 10c shows two hydrogen-bonded units spliced together by dispersion forces into a micelle. If only by somewhat subjective inspection of the trajectories, it appears that linear chain oligomers, mimicking the structural motif present in the acetic acid crystal, are very seldom observed in our simulations. Presumably, the enthalpic gain due to segregation of polar groups within a cycle exceeds the entropic loss due to cyclisation; Figure 11 indeed

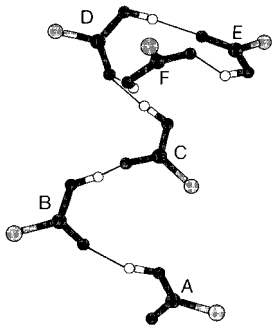

Figure 11. A chain (ABCD) curling at the end into a cyclic trimer (DEF). Note the hydrogen bond bifurcation at the carbonyl oxygen of molecule D (joined to C and F).

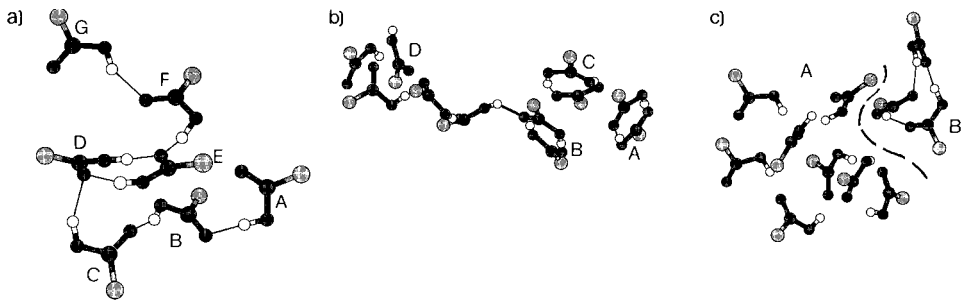

Figure 10. Higher complexes in the long-run UNI force field simulation: a) linear chain (ABC), joined to dimer (DE), joined to a chain (FG); b) parallel cyclic dimers (A, B, C) linked to a 5-membered structure (D); c) hydrogen-bonded heptamer (A) conglomerated with a cyclic trimer (B; no hydrogen bonds across the dotted line). Methyl hydrogens omitted.

shows a linear chain curling up into a cycle at one end. In any case, no trace of molecular ordering into crystal precursors was observed, nor could have been over such short timescales. In particular, no persistent traces of the experimental catemer hydrogen-bonding scheme (Figure 12) were observed, and this is not unexpected, in such small aggregates where bulk crystal properties are still a faraway goal.

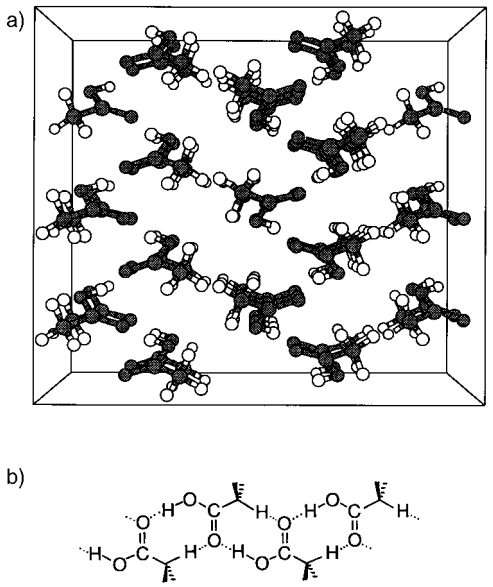

Figure 12. a) The 37-molecule cluster extracted from the crystal structure of acetic acid; b) the catemer structure, comprising a O–H···O hydrogen bond and a weak C–H···O interaction.

Acetic acid in CCl_4—crystal clusters versus micelles: When the solute mole fraction had reached 0.1, in the above described system, the forced evaporation–concentration procedure was discontinued and the simulation was continued on a droplet (37 solute molecules, 328 solvent molecules) from which, at 200 ps intervals, only the few solute and solvent molecules that had spontaneously evaporated were removed.

Results are summarised in Table 3 and in Figure 3b. The droplet, upon continuing the simulation at nearly constant mole fraction, showed a sudden increase in solute dispersion, which confirms that the forced decrease of the system size was indeed promoting solute aggregation. No hope was seen to have this system reach a steady state, due to its small size, although aggregation resumed until a 19-molecule micelle was observed, held together by hydrogen-bonding and by dispersive and/or electrostatic forces. The formation and persistence of these aggregations confirms that the implied forces are effectively incorporated, at least at short range, into the UNI potentials; this is consistent with their good performance in crystal studies.[24]

The problem was then tackled from the other end, that is, observing the evolution of a compact crystal nucleus of about the same size as the micelles, both in vacuo and wetted by the solvent. The 37-molecule crystal nucleus of Figure 12a did not survive and was almost instantaneously disrupted, owing to its exceedingly large surface tensions. After the instantaneous breaking up of the crystal hydrogen-bonding system, Figure 13 shows that further disordering, within the first 20 ps,

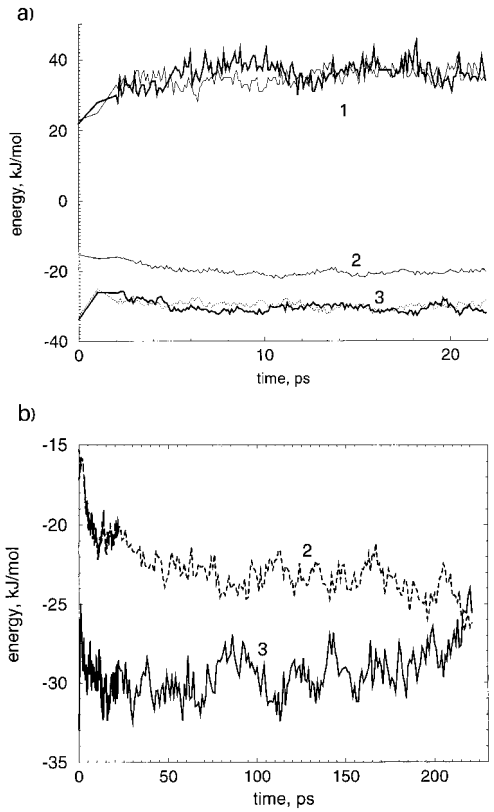

Figure 13. a) Short-run simulations of a crystal cluster of 37 molecules in vacuo (heavy lines) and in a droplet with 310 solvent molecules (see Figure 12); b) longer-run simulation of the solvated crystal cluster. Curve 1, number of hydrogen bonds. Curve 2, acetic acid–solvent interaction energies per molecule. Curve 3, acetic acid–acetic acid interaction energies per molecule.

occurs at constant number of hydrogen bonds and also with a slight increase in solute–solute interaction energy, due to an increase of solute–solute contacts consequent to reshaping of the cluster with a reduction of the number of surface molecules. The end result after 40 ps is not dissimilar to that of the aggregation process after 1350 ps (Figure 5b).

Conclusion

The aggregation of molecules with hydrogen-bonding capability within a nonpolar medium has been investigated by means of classical molecular dynamics, on a system with computationally induced high surface and interface tension, with different potential formulations that allow a schematic appreciation of long- and short-range forces. The intrinsic limitations of the method and of the computational setup are such that the emphasis was, by necessity, on time-dependent structural effects, so that thermodynamic aspects could not be accessed, and qualitative kinetic factors only can be discussed. In doing so, it should be kept in mind that molecular dynamics is a technique for the simulation of chemical events, and that simulation need not coincide with description. In the limit of short timescales and small system size, phase space is swept in a partial and random way, and results provide hints at what could happen, rather than defining what should happen.

0947-6539/99/0502-0574 $ 17.50+.50/0 *Chem. Eur. J.* **1999**, *5*, No. 2

Solute aggregation proceeds by i) formation of hydrogen-bonded cyclic dimers, ii) addition of molecules to the dimers by hydrogen-bond bifurcation, and iii) insertion of added molecules into the cycle to yield cyclic trimers and tetramers. In the longer run, entities formed in step iii) may evolve back into an equilibrium with steps ii) and i), or add further molecules to form entities of higher molecularity like iv) hydrogen-bonded acyclic or cyclic oligomers, the latter here observed up to pentamers, and eventually v) micelles held together by both hydrogen bonding and dispersive or polar forces. Remarkably, major structural rearrangements occur at nearly constant total energy, so that stages iv) and v) show a high degree of fluxionality; however, open-chain structures are less common. The essence of these structural results is independent of the force field used, a fact that adds to their credibility.

The simulation suggests that inter-solute forces, which bring about the first stages of recognition, are mainly of short-range type, since solvent-mediated migration of prospective solute interaction partners is instrumental, while electrostatic contributions just marginally increase the rate of aggregation. No solute–solute preferential orientation appears, and no ordering effects are seen within the micelles, apart from the need to fulfil the distance range for hydrogen bonding. Therefore, even a rather strong inter-solute interaction like full hydrogen bonding does not bring about instant order, and it is unlikely that crystal nucleation may proceed directly from isolated or moderately aggregated solutes to crystal nuclei. Rather, liquidlike entities are readily formed within the solution, and, most likely, order arises within them only at later stages, in much larger aggregates, when size and temperature conditions are proper, and on a timescale that would highly strain present-day computing capabilities in MD simulations.[12] There is indeed experimental mass spectrometric evidence for molecular clustering of carboxylic acids in solution.[25]

Alternatively, but concurringly, the simulation showed a 37-molecule crystalline cluster to be unstable with respect to melting even at very low temperatures. Although one could argue that this concerns a cluster of arbitrary shape and composition, and that other structures could be stable and ordered even with that small size, the disruption was so complete, instantaneous and isotropic that such a contention seems improbable. This strongly suggests that molecular aggregates of such small size cannot survive in crystalline form.

The picture that emerges from our results is one in which the first stage of crystal nucleation must go through formation of microdroplets comprising a few to a few tens of molecules, in a definitely liquidlike state. This result, in keeping with the character of the molecular dynamics simulations involved, is to be considered as a working hypothesis rather than a definite and quantitative fact of either thermodynamic or kinetic nature. Further stages of organisation, and in particular the evolution toward the hydrogen-bonded chain structure found in the crystal (Figure 12b), remain obscure: Figure 6a is an illustration of the aggregation dilemma for small carboxylic acids, including at the same time the dimer and catemer options. An explanation or even a simulation of condensation of hydrogen-bonded chains into an ordered three-dimensional structure, involving forces weaker than the hydrogen bond, is even more problematic. The force-field problem, which appears not to be crucial when simulating the first stages of condensation, could indeed become a primary issue when investigating fine detail of structural and energetic differences between crystal nuclei. For example, a catemer structure, quite similar to the one pictured in Figure 11, has been calculated, in accurate modelling based on atomic dipoles obtained from X-ray electron densities, to be less stable than the observed one (Figure 12b).[26] Indeed, a better description of the electrostatics around the carboxyl group might have brought forth aggregations more similar to those observed in the crystal, but definite and consistent ideas on modelling electrostatic forces in empirical force fields have not yet been forthcoming (distributed dipoles[26, 27] are one option). Also, an elusive entity like the C–H $\cdots$ O interaction may not come to the fore in room-temperature dynamic calculations where its energy, about 4 kJ mol^{-1},[26] compares with RT. The representation of subtler structural effects has to wait for much larger systems and much longer timescales, or much later stages of freezing and ordering. The structural results in our pictures (it should be once again reminded) are snapshots and not stabilised averages.

The results presented here were obtained by means of standard, actually almost obsolete, computer resources (an INDY Silicon Graphics workstation), confirming that molecular events in solution within the diffusive regime can comfortably be simulated with present-day computational tools. A path is then suggested for wide development of such studies with an explicit aim at discussions of and comparisons with the crystalline state, and at the transformation of the working hypothesis here put forward into an established chemical model or theory.

Acknowledgments

Thanks are due to W. F. van Gunsteren for releasing the GROMOS package at symbolic cost. Help in the implementation at Milano, by P. Klewinghaus and B. P. van Eijck, with financial support from a European Community Human Capital and Mobility Grant, is gratefully acknowledged.

[1] D. W. Oxtoby, *Acc. Chem. Res.* **1998**, *31*, 92.
[2] S. Lafont, H. Rapaport, G. J. Soemjen, A. Renault, P. B. Howes, K. Kjaer, A. Als-Nielsen, L. Leiserowitz, M. Lahav, *J. Phys. Chem. B* **1998**, *102*, 761.
[3] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225.
[4] L. X. Dang, *J. Phys. Chem. B* **1998**, *102*, 620.
[5] W. L. Jorgensen, C. J. Swenson, *J. Am. Chem. Soc.* **1985**, *107*, 1489.
[6] W. L. Jorgensen, *Chemtracts: Org. Chem.* **1991**, *4*, 91.
[7] C. Chipot, R. Jaffe, B. Maigret, D. A. Pearlman, P. A. Kollman, *J. Am. Chem. Soc.* **1996**, *118*, 11217.
[8] P. Linse, *J. Am. Chem. Soc.* **1992**, *114*, 4366.
[9] R. L. Mancera, A. D. Buckingham, N. T. Skipper, *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 2263.
[10] S. Luedemann, R. Abseher, H. Schreiber, O. Steinhauser, *J. Am. Chem. Soc.* **1997**, *199*, 4206.
[11] For the sulfuric acid-water system, see: I. Kusaka, Z.-G. Wang, J. H. Seinfeld, *J. Chem. Phys.* **1998,** *108*, 6829; for a review, and a

preliminary report on the present work, see: A. Gavezzotti, G. Filippini, *Chem. Commun*. **1998**, 287.
[12] L. S. Bartell, *J. Phys. Chem*. **1995**, *99*, 1080.
[13] J. Anwar, P. K. Boateng, personal communication.
[14] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide,* BIOMOS, Zürich-Groningen, **1996**.
[15] R. Carta, S. Dernini, *J. Chem. Eng. Data* **1983**, *28*, 328.
[16] A. Gavezzotti, G. Filippini, *Acta Crystallogr. B* **1992**, *48*, 537.
[17] A. Gavezzotti, G. Filippini, *J. Phys. Chem*. **1994**, *98*, 4831.
[18] J. M. Briggs, T. B. Nguyen, W. L. Jorgensen, *J. Phys. Chem*. **1991**, *95*, 3315.
[19] I. Nahringbauer, *Acta Chem. Scand*. **1970**, *24*, 453.
[20] A. Gavezzotti, G. Filippini, B. P. van Eijck, unpublished results.
[21] D. W. Rebertus, B. J. Berne, D. Chandler, *J. Chem. Phys.* **1979**, *70*, 3395.
[22] A. Gavezzotti, G. Filippini, J. Kroon, B. P. van Eijck, P. Klewinghaus, *Chem. Eur. J.* **1997**, *3*, 893.
[23] P. Jedlovszky, L. Turi, *J. Phys. Chem. B* **1997**, *101*, 5429.
[24] A. Gavezzotti, *Crystallogr. Rev.* **1998**, *7*, 5.
[25] S. Mochizuki, Y. Usui, A. Wakisaka, *J. Chem. Soc. Faraday Trans.* **1998**, *94*, 547.
[26] Z. Berkovitch-Yellin, L.Leiserowitz, *J. Am. Chem. Soc*. **1980**, *102*, 7677.
[27] S. L. Price, *J. Chem. Soc. Faraday Trans.* **1996**, *92*, 2997.

Received: June 12, 1998 [F1201]

0947-6539/99/0502-0576 $ 17.50+.50/0