ChemComm
FEATURE ARTICLE

Self-organization of small organic molecules in liquids, solutions and crystals: static and dynamic calculations

A. Gavezzotti[a]† and G. Filippini[b]

[a] *Dipartimento di Chimica Strutturale e Stereochimica Inorganica, University of Milano, via Venezian 21, 20133 Milano, Italy*
[b] *Centro CNR per lo Studio delle Relazioni tra Struttura e Reattività Chimica, University of Milano, via Golgi 19, 20133 Milano, Italy*

The phase behaviour of organic compounds is at the same time rich, appealing and complicated. From the properties of the pure liquid, to molecular recognition, aggregation and nucleation, both as a liquid and in solution, a number of different paths can be followed whose thermodynamics and kinetics are still to a large extent a mystery. Very little is known at a molecular level about the selective process that causes some nuclei to grow into crystals, while other aggregates are unproductive; similarly, the molecular details of the melting and dissolution processes are missing. Since the dimensions of objects and the timescale of events are scarcely or not at all accessible to experiment, computational simulation seems a viable alternative for the investigation of these phenomena.

Introduction

Modern chemistry has developed an extraordinary capability in the manipulation of molecular structures, and the synthesis of compounds of unusual or extreme chemical bonding (*e.g.* cubane), or stereochemical fine tuning (*e.g.* the total synthesis of taxol) are now within the range of possible, if not common, achievement. However, much less progress has been made in the control of molecular aggregation, and very little is known about the mechanisms of such commonplace events as melting, nucleation, growth and crystallization from the pure liquid or from solution. The main reason for this uneven development is that ordinary chemical bonding occurs in the range 10^2–10^3 kJ mol^{-1}, while molecular recognition and condensation lie in the range 10^0–10^1 kJ mol^{-1}; transformations among bonding patterns have therefore a much more predictable and reproducible course than intermolecular rearrangements, whose energetic landscape is but slightly undulated.

Of course, X-ray diffraction on single crystals has provided a tremendous amount of direct structural information on solids, and has fostered great advances also in our understanding of intermolecular interactions. Liquids and solutions are much less amenable to such detailed analysis, and, correspondingly, our knowledge of their intimate structure and properties is much less developed. It is perhaps obvious that studying intermolecular phenomena requires a knowledge of intermolecular potential energies and forces, out of which the fabric of macroscopic objects is woven. It is perhaps less frequently appreciated that thermal (*i.e.* kinetic) energy and entropy are also involved in determining the behaviour of matter. A 1985 textbook[1] appropriately stated: 'there is a big gap between knowing what the forces between two isolated molecules are and understanding how an assembly of such molecules will behave . . . even today there is no simple formula for deriving the properties of condensed phases from their intermolecular potentials . . .'. Much of this review will be devoted to the use of computer resources to bridge that gap and understand that derivation.

A first conceptual divide is the distinction between methods which explicitly include kinetic energy and account for thermal motion, and those which do not. For example, quantum chemical (QC) calculations can reproduce or predict structures, energies and activation barriers, but not the effects of thermal libration. The same applies to empirical calculations dealing, for example, with lattice energies of crystals and energy differences between polymorphs, using crystal structures frozen in the configuration derived from X-ray diffraction analyses or from some geometrical structure-guessing procedure. The effects of molecular librations may be somehow incorporated in the force field parameters, by calibrating them to reproduce specific volumes, but the description of libration itself is missing in the entirely static setup of both parametrization and modeling: all such calculations formally refer to a temperature of 0 K. Monte Carlo (MC)-type calculations include to some extent thermal energies, in that the Metropolis algorithm introduces a temperature-dependent Boltzmann factor for acceptance or rejection of phase space sampling steps. A full account of the dynamical evolution of a system under the action of potentials and including kinetic energies is given only by molecular dynamics (MD) calculations, which, at least in principle, allow a complete sampling of phase space and a *de novo* derivation of structural, thermodynamic and kinetic parameters.

The availability of accurate and easily applicable potential formulae and parameters is one of the key points in any theoretical treatment of condensed phases. In fact, one could contend that even in quantum chemical calculations the choice of the basis set and of the method for treating electron correlation are the equivalent of empirical parametrizations in classical force fields. However, the subject of potential formulation and optimization will not be considered here, since such a topic, even if schematically treated, would take up all the journal space allotted to this review. The problem of potentials will be addressed, in a cursory fashion, at some places, but the main emphasis will be on a perspective of what can or could be done to simulate, rationalize, predict or control molecular self-organization under the assumption that a suitable potential has indeed been made available. The reader should be conscious that this is not equivalent to saying that suitable potentials can always be derived, or, worse, that the power of MC, MD or quantum chemical algorithms is such that any potential formulation will do.

As is usual in these days of massive publication policies, the literature survey will be representative, rather than exhaustive.

Molecular recognition

Consider two molecules of different species, A and B, in the gas phase or in any condensed phase which allows substantial diffusional freedom. The most elementary definition of molecular recognition is the following aggregation step (1), neglecting the effects of the surrounding medium.

$$A + B \rightarrow AB \quad (1)$$

If this reaction describes hetero-recognition, self-recognition, relevant to condensation, nucleation and crystal growth in pure

substances, is given instead by step (2) where the expression [*n*A] denotes an aggregate of *n* identical molecules of species A.

$$A + A \rightarrow [2A] \qquad (2a)$$

$$[nA] + A \rightarrow [(n+1)A] \qquad (2b)$$

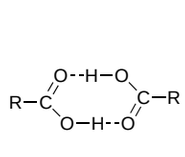
Processes (1) and (2*a*) can be studied quantum chemically, if the sizes of A and B are affordable (a base pair, but not a DNA strand and a ligand) given the required accuracy in the treatment of electron correlation. Process (2*b*) becomes quickly too expensive to be tackled by QC methods for molecular sizes (10–50 atoms) and *n*-values (10–100) of chemical significance.

High quality QC calculations on systems like (1) and (2*a*) can be used to derive cohesive energies and equilibrium distances which can then be exploited in the parametrization of empirical potentials. Typical results[2] indicate that, for example, the energies of C–H···C, C–H···S and S···S interactions are −2.9, −1.9 and −1.5 kJ mol^{-1}, respectively, an energy range which is consistent with that indicated in the opening statement of this review.

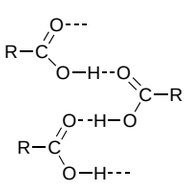
If a more realistic picture of the molecular recognition proceedings is desired, including solvation effects, one can use MC or MD calculations. Potentials here must be empirical, including usual intramolecular force field contributions plus intermolecular contributions, the latter being generally treated as a series of inverse powers of atom–atom distances, *i.e.* terms of the type CR^{-n} where *C*s are calibrating constants and the exponents are integers in the 1–12 range. Such events as solvation, dimerization and further aggregation can be simulated comfortably over systems consisting of a few thousand atoms. Using such techniques, for example, the optimized interaction energies of the benzene dimer in solution have been shown[3] to be within −7.1 and −9.6 kJ mol^{-1}, for widely different dimer structures; differences in stability between these aggregates are, again, in the range of a few to a few tenths of a kJ mol^{-1}.

A substantial drawback in this kind of calculation is that empirical potentials cannot account for molecular polarizability, at least for larger molecules with complex chemical features, since most force fields—at least the ones of more widespread applicability—are invariant with time and chemical environment. Indeed, describing steps (2*a*) and (2*b*) with the same force field for molecule A is unrealistic, and even less realistic is the use of an invariant field for molecule A when it is exposed to a strongly polarizing solvent, water to mention an obvious case.

A bonus of MD is that, besides thermodynamics, a picture of the interaction kinetics is obtained: this is vital if one considers molecular recognition as the preliminary step of condensation, ideally looking for liquid phase or solution precursors to crystal nucleation. Simulations of molecular aggregation by hydrogen-bonding in carboxylic acids and amides prove that in solution an equilibrium exists between cyclic and periodic chain arrangements.



Cyclic dimer



Chain ('catemer') motif

Thus, the relatively high abundance of the chain dimer for tetrolic acid in CCl_4 solution, predicted by MD,[4] is in agreement with the existence of two crystal polymorphs, one with the cyclic and one with the catemer motif. Understandably, such hydrogen-bonded aggregates survive in the apolar solvent, but break apart almost instantly in water. For 2-pyridone in CCl_4, the frequency of events in which one of the two hydrogen bonds of the cyclic dimer is cleaved,[5] most likely due to transfer of kinetic energy from solvent to solvate molecules, agrees with the significant dipole moment observed for the dimer in solution;[6] however, the high stability of the dimer is in contrast to the observed crystal structure, which exhibits the catemer motif. Fig. 1 shows typical time evolution graphs for the relevant hydrogen-bonding distances; such information is in principle very valuable for structural predictions, but clearly yields only a preliminary picture of the subsequent nucleation and growth processes.

Condensation; liquids

The notion that with an appropriate reduction in temperature and/or increase in overall pressure any vapor will condense into a liquid is trivial even in everyday words. Less widespread is the notion that even this familiar process requires a preliminary recognition leading to molecular clustering, a not so intuitive process for matter in the gas phase. Simulation studies of clustering from vapor have been presented,[7] but this topic will not be further pursued in this paper, which is mostly oriented towards molecular aggregation equilibria in condensed phases.

Organic compounds which are in the liquid state at around room temperature are made of small or apolar molecules. For these, the bulk liquid can be studied rather comfortably by MC or MD (*e.g.* ref. 3) with excellent results on the thermodynamics (heats of vaporization and C_p values) and on the structure (radial distribution curves). A recent MD study of liquid ethanol[8] afforded accurate estimates of thermodynamic and transport properties, of molecular conformations and orientational order,

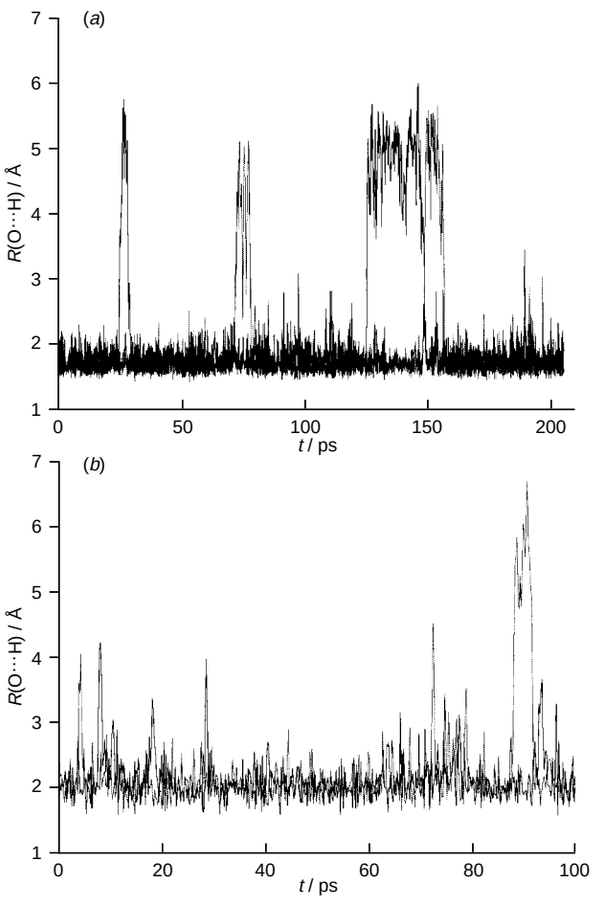


Fig. 1 Time evolution of the hydrogen-bonding O···H distances during MD simulations for cyclic dimers of (*a*) tetrolic acid (ref. 4) and (*b*) 2-pyridone (ref. 5)

as well as a description of the characteristics of hydrogen bonding states. Liquid phenol was studied[9] using an intermolecular potential derived from accurate QC calculations on the isolated molecule or dimers and trimers (remarkably, four dimers of widely different geometries span a cohesive energy range of just 4 kJ mol^{-1}, and the three most stable trimers of just 10 kJ mol^{-1}). MC simulations of the hydrogen bonding patterns in liquid formic acid[10] reveal that cyclic dimers occupy only 7% of the phase space, while in methanol clusters of 5 to 256 molecules,[11] small monocyclic aggregates were found to exhibit considerable persistency, presumably correlated with the high percentage of alcohol crystals with multimolecular asymmetric units composed of such cycles.[12]

The vast majority of organic compounds of molecular weight over 100 are solid at room temperature, and experimental measurements of the properties of their liquids can be awkward. Computer simulation can be very helpful here: test MD calculations[13] show that even simple 6–12 or 6–exponential atom–atom potentials, calibrated using static crystal structure data, can reliably estimate the density and heats of vaporization of comparatively large compounds, *e.g.* benzamide or coumarin. The complete calculation of specific volume as a function of temperature for liquid and crystal phases is also feasible (*e.g.* Fig. 2). Note that no timescale problem exists in the MD simulation of equilibrium bulk liquids, since experimental time constants for molecular diffusion span the 90 ps–20 fs range,[14] quite within the possibilities of today's MD. Presumably, the computational estimation of the thermochemical properties of small to medium-size organic molecules successfully competes, on the basis of cost : quality ratio, with experimental determinations.

Nucleation and the solid–liquid interface

A bulk liquid at a temperature much higher than its freezing point should ideally be completely disordered and hence perfectly homogeneous. Quite a different situation must arise when the temperature is lowered to the freezing point, and, even more, below it. An undercooled liquid must experience fluctuations towards the thermodynamically stable crystalline phase, and, if the size of nuclei formed during these fluctuations exceeds a certain critical threshold, evolution to the crystal is observed; otherwise, the subcritical nuclei merge back into the bulk fluid.[15] The theoretical study of these nucleation phenomena can proceed at an almost entirely macroscopic level, using comprehensive descriptors without regard to molecular structure,[15–17] or through several kinds of computational guinea pigs such as the Lennard–Jones system or other, similarly oversimplified (at least, with respect to ordinary organic chemistry) interaction sites.[18–20] Macroscopic theories of crystal growth have been developed mostly using cell models (like the Burton–Cabrera–Frank model), and models intermediate between the macroscopic and the molecular level exploit simple concepts such as the relationships between the growth speed of different crystal faces and the interplanar spacing (the Hartmann–Perdok model) or molecular attachment energies; however, these approaches lack resolution on the detail of molecular recognition, and hence are beyond our present scope. Heterogeneous nucleation, induced by interactions with random impurities, and hence orders of magnitude more complicated than homogeneous nucleation, will not be considered here for obvious reasons.

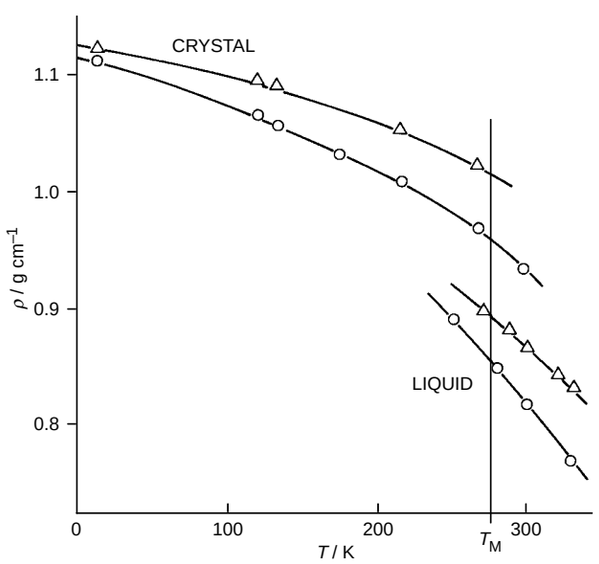


Fig. 2 Specific volume *versus* temperature for benzene from MD simulations of the liquid and crystal phases (from ref. 13). (○) calculated; (△) observed.

Progress on systems composed of real molecules has been made by joint experimental and theoretical studies of molecular clusters.[21] Clusters can be prepared in such conditions as to have nucleation rates as high as 10^{30} m^{-3} s^{-1}, perhaps 15 orders of magnitude larger than those in bulk liquids, and this is what makes them especially attractive targets for nucleation studies. For example, MD simulations on clusters including a few hundred molecules correctly reproduced solid–liquid transitions, and, even more significantly, an MD simulation on a cluster of 188 *tert*-butyl chloride molecules showed stability in the observed plastic crystalline tetragonal phase as well as spontaneous transformation to a low-temperature ordered, monoclinic crystalline phase, in agreement with electron and neutron diffraction data.[21] *tert*-Butyl compounds have a rich phase behaviour with several solid–solid and solid–liquid features, which have been studied by thermal analysis and theoretical dynamical methods (see ref. 22 for results and related literature).

While a complete phase diagram in the P–T plane can be obtained for a system of Lennard–Jones spheres,[23] solid–liquid equilibrium in molecular systems of moderate complexity is also almost within reach of present day computer capabilities. The nucleation and melting of linear C_n alkanes has been studied[24] by MD using an *n*-site Lennard–Jones representation of the molecules, in the NPT ensemble, *i.e.* at constant pressure, in a box with periodic boundary conditions resembling the bulk more than a cluster. The time evolution of thermodynamic and structural parameters could be monitored on line, in what can be considered an ideal computational experiment in nucleation. Similarly, encouraging results have been obtained for the liquid–solid phase transition of cyclohexane, with a six-site molecular model.[25] Using MD, the freezing of supercooled water, induced by an electric field, has been successfully simulated[26] over a period of a few hundred picoseconds, a quite affordable timescale for present-day computers.

Simulations of this kind, when extended to larger molecules, and analyzed for the identification of crystallization precursors, can be of extreme value in the progress of our understanding of nucleation from the melt.

Nucleation and crystallization from solvent

Nucleation of organic molecules within a solution is at present an essentially unaccessible phenomenon. The nuclei are too small to be seen by direct or scattered light, and their size distribution and dynamic growth properties cannot be determined experimentally. Computer experiments are therefore the only means to probe this elusive reality.

A first attempt to study the behaviour of elementary nuclei in solution was made[4] on dimers of tetrolic acid, as discussed previously in this paper. For a more comprehensive and significant test, an MD simulation was run on a computational box initially consisting of $9 \times 9 \times 9$ cells about 7 Å wide, 20 of which, picked at random, were occupied by (solute) acetic

acid molecules, the others being occupied by (solvent) CCl_4 molecules; no solute–solute distances below 15 Å were present in the starting configuration. Crystal potentials[27] were used for intermolecular interactions of the solute, and standard potentials[28] for the united-atom solvent molecules, applying the usual averaging rules for cross interactions and a 30 Å cutoff in intermolecular summations. The intramolecular geometry of the acetic acid molecules was frozen, except for the wagging motions of the acidic proton. Several runs were conducted at constant temperature between 200 and 300 K; the 200 K results are discussed, but the results did not change significantly with temperature. The GROMOS96 package[29] was employed.

The results are, if not conclusive, at least stimulating. Since no periodic boundaries were imposed, the starting box quickly relaxed into a pseudospherical shape [Fig. 3 and 4(*a*)], and at the same time a very fast condensation of solute molecules into small clusters made of 2–5 molecules was observed [Fig. 4(*b*) and 5], with geometries ranging from head-to-head dimers over the COOH function, to small hydrogen-bonded oligomers bound into cyclic structures. Not unexpectedly, no trace of intermolecular symmetry was found within these prototypical droplets, in which the molecular centers of mass stayed within cohesive distance during the simulation, while the detailed structure exhibited a highly fluxional behaviour. The solute potentials were of the 6–exp type,[27] transformed to 6–12 for compatibility with GROMOS; their well depth corresponds to a hydrogen-bonding energy (about 30 kJ mol^{-1}), but they fall off very quickly with distance. Therefore, to probe the efficiency of molecular attraction at long range, the computer experiment was then repeated supplementing the 6–exp potential with coulombic terms computed with the GROMOS96 charge distribution over the COOH group, thus increasing artificially, and even unrealistically, the intermolecular attractive forces. The stronger potentials bring molecules together more rapidly and hold them more tightly together within the droplets; although energies and trajectories differ somewhat, however, the final results in terms of structure and the total number of hydrogen bonds were essentially the same. One possible interpretation is that solute condensation is helped by solvent reorganization, besides solute–solute attractive interactions; in other words, some of the driving force for the segregation of solute molecules may come from the tendency of the solvent to squeeze out the disturbance to its own structure. If and how these conclusions can stand the trial of changes in the solute potential, in the length of the simulation, or in boundary conditions, remains to be seen; the example proves the feasibility, and hints to a possible usefulness, of this kind of simulation.

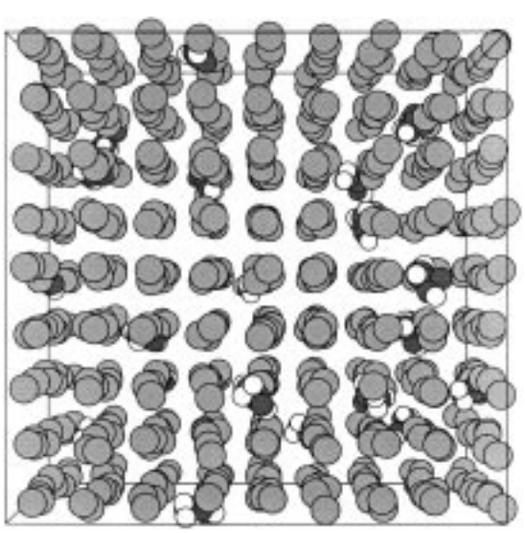


Fig. 3 Starting configuration within the computational box for 20 AcOH and 709 CCl_4 molecules; large circles are united-atom solvent molecules

The other relevant aspect of the solute–solvent equilibrium is the growth of macroscopic crystals once the nucleation stage has been overcome, and template crystal surface(s) are available within the solution. The timescale of events involved in crystal growth goes from (presumably) picoseconds for structural relaxation of molecules adsorbed on growing surfaces, to nanoseconds for molecular diffusion over the surface, to seconds for the growth of several monolayers; the first two steps are easily manageable by MD, but not the third. The so-called kinetic MC method can be employed, in which, roughly speaking, transition probabilities among configurations are weighted by estimated rates of the transition events, rather than by the Metropolis test as is the case in thermodynamic MC. Apparently, KMC simulations span formal times of the order of hundreds of seconds (see ref. 30 for further description and a review).

For chemical purposes, growth from solution is the method of choice. The concept of a relationship between a given crystal structure and the structure of precursor nuclei in solution has

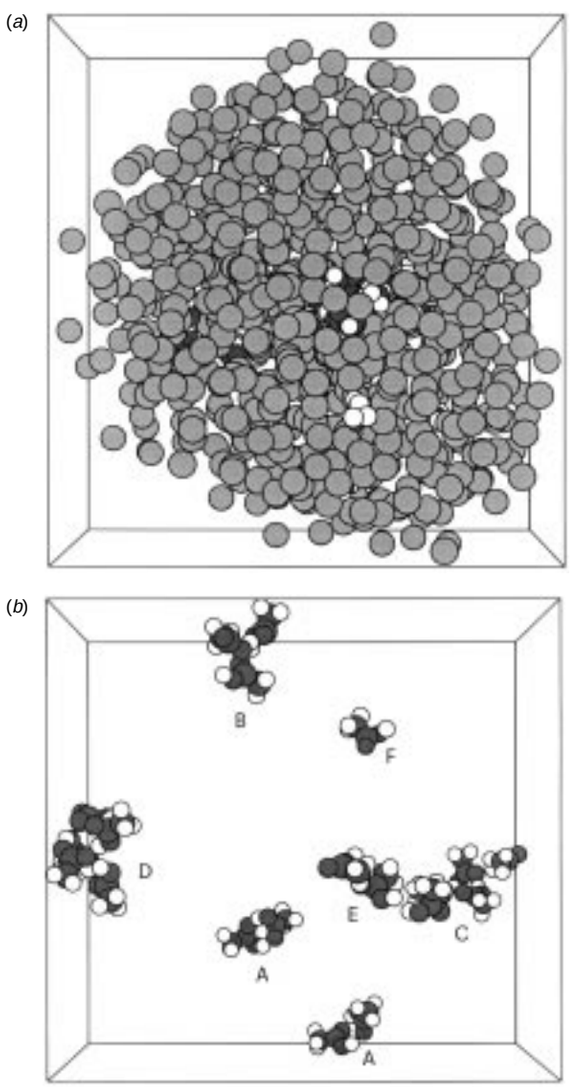


Fig. 4 (*a*) Overall view of the AcOH–CCl_4 system after 400 ps; 10 solvent molecules (not appearing) have departed from the drop, simulating evaporation. (*b*) Detail of the arrangement of AcOH molecules. A denotes cyclic dimers, B a chain trimer, C a four-molecule structure with bifurcated H-bonding and D a cyclic pentamer. In E, two molecules are close together but not H-bonded. F is an isolated molecule.

been introduced and exploited by the Weizmann school in a series of beautiful experiments, in which nucleation inhibitors were designed, on the basis of molecular structure, to bind stereospecifically at the surface of the stable polymorph, thus preventing its growth and enhancing the growth of metastable polymorphs (see ref. 31 for a recent account on the control of glycine polymorphism). At the other extreme, holistic analyses of the nucleation and growth kinetics produce phenomenological equations which lack molecular detail, but can be used[32] to study the rate-controlling processes in solvent-mediated phase transformations (Ostwald cycles).

Several experimental techniques are nowadays available for the *in situ* monitoring of crystal growth at atomic level; among others, atomic force microscopy[33] and laser interferometry.[34] The ideal computational experiment in crystal growth and dissolution involves the preparation of a computational box in which one or more crystal faces are exposed to either the pure solvent or a solution of the crystallizing substance, and the monitoring of the evolution in time of the system by MD. Problems of potentials (polarizability) and of timescale concur in making this experiment a very awkward one, *e.g.* given an elongation rate of 0.3 μm min^{-1} on a needle of the diameter of 10 μm,[35] a simple calculation shows that on a computational surface 100 × 100 Å wide, one should wait 3×10^8 ps to observe the attachment of a single molecule. This experimental result obviously incorporates an unknown time lag due to diffusional barriers within the solution; attachment and detachment events within the microscopic layer in intimate contact with the crystal surface may proceed at a much higher rate. In fact, MD simulations of the interface between a saturated urea solution and the urea crystal[36] revealed, at least, interesting preorganization details within the adsorbed layers. Well within the range of MD is instead the equilibration of water over crystal surfaces, or an essentially surface phenomenon, revealing details of the interaction with relevance to morphology[37] and wettability.[38]

Crystals

The final product of molecular recognition can be comfortably examined in the crystal structure, thanks to X-ray diffraction—at least in the majority of occurrences, but not always, since for yet unknown reasons a small but significant percentage of organic molecules refuse to organize into suitable single crystals. Present-day diffraction facilities, using stronger sources and two-dimensional detectors, have considerably reduced the number of inaccessible crystalline materials.

The dynamics of molecules in crystals far from transition temperatures is essentially harmonic, and hence rather uninspiring. Intermolecular librations can be successfully modeled by harmonic lattice dynamics, while extensive MD simulations are rather a tool for a more accurate calibration of potentials than for the discovery of new facts.[39]

Long before MD methods were devised, extensive computational work had been conducted on molecular crystals, dealing successfully with crystal packing analysis and crystal thermodynamics (see ref. 40 for a historical perspective). The background was thus laid for tackling more ambitious goals, like the enforcement of close-packing in translationally symmetrical molecular assemblies, and eventually, the computational prediction of crystal structures.

In the last five years or so, computational methods have been developed for guessing the crystal structure a compound will adopt, starting from the bare molecular constitution, and not without some success.[41–48] Such methods completely overlook all preliminary, dynamic molecular recognition stages, and rely on astute algorithms and shortcuts to assemble the crystal structure like a sort of molecular LEGO puzzle, the guiding concept being that the predicted crystal structure must be the one with the most stabilizing lattice energy. These procedures rely sometimes on random, Monte Carlo-type, or even brute force searches through the potential energy hypersurface, or on energy minimizing algorithms, and sometimes on symmetry considerations with an exploitation of close-packing principles; they are therefore essentially static in nature, although at some stages dynamic reshuffling in the form of simulated annealing may be applied, to facilitate the crossing of barriers and the unification of apparently different valleys. Typically, hundreds or thousands of plausibly close-packed structures are generated, and clustering of equivalent ones is problematic—consider, for example, just the problem of cell reduction in triclinic space groups. The basic result, common to all of these procedures, is a portrait of the potential energy landscape in the proximity of its minima, a picture that invariably reveals shallow regions among which the recognition of absolute stability is impossible. Fig. 6 and Tables 1–4 demonstrate this assertion.

Things could not be different, given the unavoidable physical nature of weak intermolecular interactions, and their energy range (see the introduction to this paper). The energetic resolution of empirical intermolecular potentials is presumably of the same order of magnitude as the energy differences they try to gauge; entropy differences are neglected. Besides, the starting molecular geometry is assumed *in vacuo*, and a reliable account of the interplay between the intra- and inter-molecular force field is awkward. Nevertheless, these structure prediction

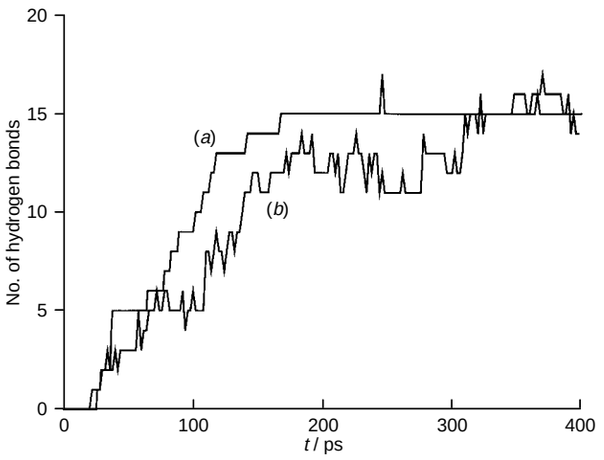


Fig. 5 AcOH in CCl_4: number of hydrogen bonds formed during the MD simulation, starting from the configuration in Fig. 3. Curve (*a*), with charges, curve (*b*), without charges (see text). T = 200 K.

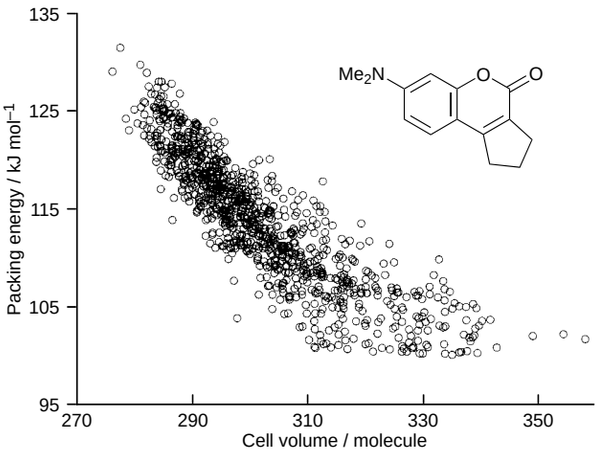
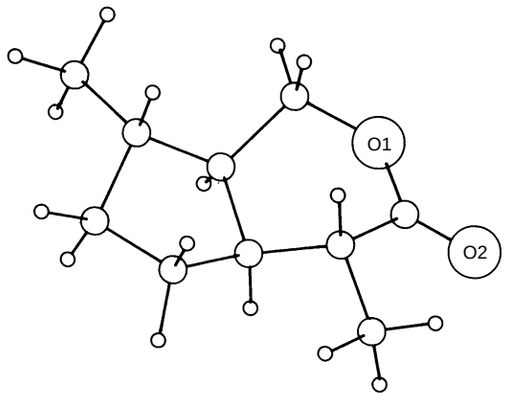


Fig. 6 Scatterplot of the cell volumes and packing energies of crystal structures generated in six different space-groups for the molecule shown in the inset (after ref. 48).

Table 1 Results of crystal structure prediction for isoiridomyrmecin (CSD refcode ISIRIN) (ref. 42)



Space group	Z	E/kJ mol^{-1}	V_{cell}/Å^3	b axis
$P2_1$	2	103.0	473.8	6.42
		115.3	474.5	12.4
		111.3	471.9	9.02
		112.2	497.0	7.35
$P1$	1	103.5	241.1	—

Table 2 Predicted crystal structures for 1,8-dinitronaphthalene

Space group		(V/Z)/Å^3	E/kJ mol^{-1}	Ref.
$P2_12_12_1$	Exp	233.31	105.1	43
$P\bar{1}$		233.25	102.5	43
$P2_1$		237.04	103.4	43
$P2_1$/c		231.29	108.5	43
$P2_12_12_1$	Exp opt	219.5	123.8	49
$I2/a$		227.9	119.7	49
$P2_1/c$		221.4	123.4	49
$Pbca$		227.4	120.5	49

Table 3 First 10 structures predicted for durene (1,2,4,5-tetramethylbenzene, ref. 45)

Space group	E/kJ mol^{-1}	Z
$P\bar{1}$	89.64	1
$P\bar{1}$	83.74	1
$Pca2_1$	83.31	4
$Pbca$	82.81	4
$P2_1/a$	82.23	4
$P2_1/c$	81.68	4
$Pna2_1$	80.93	4
$P2_1$	80.64	2
$P2_1/a$	80.47	2
$P2_1/c$	79.76	2

Table 4 Crystal structure prediction for AcOH (ref. 47). For each space group N^o is the total number of structures generated, N' the number after clustering

	N^o	N'	Best E/kJ mol^{-1}
$P\bar{1}$	37 705	7	69.2
$P2_1$	2294	10	67.3
$P2_1/c$	60 683	113	69.7
$C2/c$	69 908	139	69.9
$P2_12_12_1$	7035	19	69.8
$Pna2_1$	3874	47	69.5
$Pbca$	9410	49	69.1
$Pnma$	15 100	40	63.5

algorithms are extremely valuable in that they usually can narrow down the choice to a few, typically 5–10 possible crystal structures, and this is an important result which should not be overlooked. Besides, when coupled with partial X-ray diffraction data (*e.g.* from faulty crystals or from powders) they lead to consistent structure prediction, the discrimination between computationally generated structures being taken care of by the reliable (often indisputable) experimental crystallographic information.[49,50] These facts are more than enough to foresee a wide impact of computational techniques, in the near future, on the design of novel molecular materials with desirable structure–property relationships. And, after all, with some luck and a modicum of chemical or crystallographic intuition, some of these methods can actually lead to a real *de novo* crystal structure prediction for some molecules; the trouble is, success or failure depends on still uncontrolled factors, and the success rate cannot be assessed in a systematic way, so that weak spots cannot be clearly identified and ameliorations cannot be rationally planned, at least thus far.

Phase behaviour of organic compounds

The study of organic molecules in the 100–1000 Da molecular weight range[51] is full of richness and fascination, although more so as regards their aggregation properties than the bare molecular structure, *i.e.* more in the inter- than in the intramolecular arena. The definition and modeling of weak recognition forces pose a great theoretical challenge, but also the applicative side is full of promise.

The *in vacuo* theoretical chemistry of small molecules includes a design step, done by sketching a molecular composition and connectivity on paper, and a structural exploration stage, in which quantum chemical or force field methods are used to define the conformation(s) and overall shape of the molecule in the absence of surrounding fields. If a high-quality wavefunction is available, the main features and even some detail of the molecular electrostatic field can be obtained for each conformation.

The study of the intermolecular theoretical chemistry of organic small molecules hits a big stumbling block in its very first stage, which is the polarization of the *in vacuo* molecular field by the environment. The study of condensed phases should start from the simplest one, a pure liquid; from there on, a wide range of possible paths can be envisaged in the modeling of the phase domain, using temperature/kinetic energy as a guideline. Several degrees and types of structuring and ordering set in as the temperature of the liquid decreases, from molecular clustering, to one- or two-dimensional translational ordering in liquid crystals (LC), to three-dimensional translational ordering with rotational disorder in plastic crystals; further on, or alternatively, a glassy state could be reached. All of these mesophases could be either thermodynamically stable or metastable with respect to the crystalline solid. At this stage the system assumes the status of a material with a macroscopic molecular assembly which can be used for specific purposes, since its texture interferes and specifically interacts with electromagnetic waves (LC display devices, non-linear optics), or with electric and magnetic fields (electrets and organic magnets).

In a conceptually final condensation stage, with further reduction of translational and rotational kinetic energies, the crystalline state is accessed, whose complete anisotropy of course enhances all possibilities for practical uses of the material. Even here, some flexibility is left in the consideration of possible polymorphism.

Binary solutions are just the two-component equivalent of the already immense one-component problem outlined above. Although neither pure liquids nor solutions can compete with solids for applicative purposes, the liquid state holds the premises for our comprehension of solidification. The basic dissolution–segregation process is in fact the key to an

understanding of the mechanisms of formation of the vast majority of organic solids.

Summary

Thus, the phase–mesophase behaviour of organic compounds is an inextricable tangle of kinetics and thermodynamics. We would like to conclude this article with some assessment of present or near-future computational approaches to its prediction and control. We assume that molecular structure and intermolecular potentials have been somehow established using the procedures and within the limitations outlined in the preceding sections.

Fairly accessible would be the prediction of the densities and cohesive energies of the liquid, liquid crystal, glassy and crystalline states of the substance. For the first three, dynamic calculations are necessary, but the corresponding computational boxes could be rather easily set up and equilibrated at any desired temperature. At least for the crystal, prediction could also go through static calculations, using one of the several algorithms described in a former section of this paper. Although the detailed geometrical structure of the crystal may not be accurately and reproducibly predicted, the available methods can usually produce 5–10 structures among which the possible polymorphs would almost certainly be included.

Quite a different problem would be the prediction of relative thermodynamic stabilities and of transition temperatures, for which an accurate evaluation of the enthalpies and entropies of all phases as a function of temperature would be needed. This is, so far, a prohibitive task. The dynamic simulation of the detailed transformation paths and of their kinetics is also so far extremely demanding, but steps are already being taken in this direction, and the promise for very quick development in the near future is high.

Control is a step beyond prediction. It would certainly be desirable to learn how to design the molecular architecture so as to engineer a certain property within the material, *e.g.* its propensity to form liquid crystal phases, or the presence or absence of a center of symmetry in the crystal, up to a fine tuning of molecular orientation in the solid to produce a certain electrooptical effect. A structural approach to crystal structure prediction and control makes use of the concept of crystal synthons, or basic recognition blocks which, when implanted in molecular objects, drive their spatial recognition to pre-established goals.[52] Its success depends on some systematization and much chemical intuition.

There is little that can be done directly, in terms of control of the properties of a material, by pure calculations; a computer computes properties, but cannot be confidently taught to appreciate the influence of molecular chemistry on them, a task which is more appropriate to the human than to the electronic mind (the crystal packing modes of primary amides had been codified and to some extent predicted[53] in times when computers were in their infancy). It is still for the human to line up all the computational information and to organize it towards comprehension. In this respect, what computers can easily do, besides brute force exploration of the phase space, is to gather at a fantastic speed, for the use of a human operator, information that would be too tedious or quite impossible to obtain by hand. Using computer-accessible collections of crystal information, like the Cambridge Structural Database,[54] one learns for instance[55] (after having programmed computers for decades to explore all the cell space) that molecular centers of mass must lie in special positions within the crystal cell, *i.e.* roughly halfway between inversion centers or between screw axes. Also, crystal structure prediction algorithms can be designed to learn from the structural or energetic properties of existing structures.[56] Together with the exponential increase in mere computing power, the near future should see a more and more widespread cooperation of human and electronic mindpower towards a more comprehensive understanding of the phase behaviour of organic compounds.

Angelo Gavezzotti graduated in Chemistry in 1968 at the University of Milano. He started his scientific career in X-ray crystallography and theoretical chemistry, under the supervision of Massimo Simonetta. His interests moved later to the simulation and prediction of organic crystal structures and properties, and, more recently, to computer techniques in the study of the physical chemistry of condensed phases of organic compounds.

Giuseppe Filippini received his degree in Chemistry from the University of Milano with a thesis in Crystallography (1968), under the supervision of Massimo Simonetta. He spent one postdoctoral year in Manchester with Professor D. W. J. Cruickshank. Currently, he is senior research scientist at the CNR Center for the Study of the Relationships between Structure and Chemical Reactivity. His active interests cover correlation and systematics of intermolecular interactions in organic crystals trough lattice dynamics and atom–atom potential calculations.

Notes and References

† E-mail: gave@stinch12.csmtbo.mi.cnr.it

1 J. N. Israelachvili, *Intermolecular and Surface Forces*, Academic Press, London, 1985, p. 9.
2 J. J. Novoa, M. Carme Rovira, C. Rovira, J. Veciana and J. Tarres, *Adv. Mater.*, 1995, **7**, 233.
3 W. L. Jorgensen, *Chemtracts: Org. Chem.*, 1991, **4**, 91.
4 A. Gavezzotti, G. Filippini, J. Kroon, B. P. van Eijck and P. Klewinghaus, *Chem. Eur. J.*, 1997, **3**, 893.
5 A. Gavezzotti, *Faraday Disc.*, 1997, **106**, 63.
6 K. DeSmet, P. Kedziora, J. Jadzyn and L. Hellemans, *J. Phys. Chem.*, 1996, **100**, 7662.
7 Z. Li and H. A. Scheraga, *J. Chem. Phys.*, 1990, **92**, 5499.
8 L. Saiz, J. A. Padrò and E. Guardia, *J. Phys. Chem.*, 1997, **101**, 78.
9 K. Sagarik and P. Asawakun, *Chem. Phys.*, 1997, **219**, 173.
10 P. Jedlovszky and L. Turi, *J. Phys. Chem.*, 1997, **101**, 5429.
11 D. Wright and M. S. El-Shall, *J. Chem. Phys.*, 1996, **105**, 11 199.
12 C. P. Brock and L. L. Duncan, *Chem. Mater.*, 1994, **6**, 1307.
13 A. Gavezzotti, G. Filippini and B. P. van Eijck, work in progress.
14 Y. J. Chang and E. W. Castner, *J. Phys. Chem.*, 1996, **100**, 3330.
15 C. K. Bagdassarian and D. W. Oxtoby, *J. Chem. Phys.*, 1994, **100**, 2139.
16 P. Harrowell and D. W. Oxtoby, *J. Chem. Phys.*, 1984, **80**, 1639.
17 H. Hettema and J. S. McFeaters, *J. Chem. Phys.*, 1996, **105**, 2816.
18 J. S. van Duijneveldt and D. Frenkel, *J. Chem. Phys.*, 1992, **96**, 4655.
19 V. Talanquer and D. W. Oxtoby, *J. Chem. Phys.*, 1995, **103**, 3686.
20 P. R. ten Wolde, M. J. Ruiz-Montero and D. Frenkel, *J. Chem. Phys.*, 1996, **104**, 9932.
21 L. S. Bartell, *J. Phys. Chem.*, 1995, **99**, 1080.
22 J. Reuter, D. Buesing, J. L. Tamarit and A. Wuerflinger, *J. Mater. Chem.*, 1997, **7**, 41.
23 R. Agrawal and D. A. Kofke, *Mol. Phys.*, 1995, **85**, 43.
24 K. Esselink, P. A. J. Hilbers and B. W. H. van Beest, *J. Chem. Phys.*, 1994, **101**, 9033.
25 A. Brodka and T. W. Zerda, *J. Chem. Phys.*, 1992, **97**, 5669.
26 I. N. Svihschev and P. G. Kusalik, *J. Am. Chem. Soc.*, 1996, **118**, 649.
27 A. Gavezzotti and G. Filippini, *J. Phys. Chem.*, 1994, **98**, 4831.
28 D. W. Rebertus and B. J. H. Berne, *J. Chem. Phys.*, 1979, **70**, 3395.
29 W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hunenberger, P. Kruger, A. E. Mark, W. R. P. Scott and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, BIOMOS b.v., Zürich-Groningen, 1996.
30 M. Kotrla, *Comput. Phys. Commun.*, 1996, **97**, 82.
31 I. Weissbuch, L. Leiserowitz and M. Lahav, *Adv. Mater.*, 1994, **6**, 952.
32 R. J. Davey, P. T. Cardew, D. McEwan and D. E. Sadler, *J. Cryst. Growth*, 1986, **79**, 648.
33 A. J. Malkin, Y. G. Kutznetsov, W. Glantz and A. McPherson, *J. Phys. Chem.*, 1996, **100**, 11 736.
34 B. Y. Shekunov and R. J. Latham, *J. Phys. Chem.*, 1996, **100**, 5464.

35 R. J. Davey and J. Richards, *J. Cryst. Growth*, 1985, **71**, 597.
36 E. S. Boek, W. J. Briels and D. Feil, *J. Phys. Chem.*, 1994, **98**, 1674.
37 S. Khoshkoo and J. Anwar, *J. Chem. Soc., Faraday Trans.*, 1996, **92**, 1023.
38 J. Anwar and S. Khoshkoo, *Pharm. Res.*, 1996, **13**, 1006.
39 D. C. Sorescu, B. M. Rice and D. L. Thompson, *J. Phys. Chem.*, 1997, **101**, 798.
40 A. Gavezzotti and G. Filippini, in *Theoretical Aspects and Computer Modeling of the Molecular Solid State*, ed. A. Gavezzotti, Wiley, Chichester, 1997, ch. 3.
41 A. Gavezzotti, *J. Am. Chem. Soc.*, 1991, **113**, 4622.
42 H. R. Karfunkel and R. J. Gdanitz, *J. Comput. Chem.*, 1992, **13**, 1171.
43 J. R. Holden, Z. Du and H. L. Ammon, *J. Comput. Chem.*, 1993, **14**, 422.
44 B. P. van Eijck, W. T. M. Mooij and J.Kroon, *Acta Crystallogr., Sect. B*, 1995, **51**, 99.
45 A. M. Chaka, R. Zaniewski, W. Youngs, C. Tessier and G. Klopman, *Acta Crystallogr., Sect. B*, 1996, **52**, 165.
46 M. U. Schmidt and U. Englert, *J. Chem. Soc., Dalton Trans.*, 1996, 2077.
47 W. T. M. Mooij, B. P. van Eijck, S. Price, P. Verwer and J. Kroon, *J. Comput. Chem.*, 1997, **18**, in the press.
48 A. Gavezzotti, *Acta Crystallogr., Sect. B*, 1996, **52**, 201.
49 A. Gavezzotti and G. Filippini, *J. Am. Chem. Soc.*, 1996, **118**, 7153.
50 R.B. Hammond, K. J. Roberts, R. Docherty and M. Edmonson, *J. Phys. Chem.*, 1997, **101**, 6532.
51 A. Gavezzotti, submitted for publication in *Crystallogr. Rev.*
52 G. R. Desiraju, *Angew. Chem., Int. Ed. Engl.*, 1995, **34**, 2311.
53 L. Leiserowitz and A. T. Hagler, *Proc. R. Soc. Lond., Ser. A*, 1983, **388**, 133.
54 F. H. Allen and O. Kennard, *Chem. Des. Automat. News*, 1993, **8**, 31.
55 W. D. S. Motherwell, *Acta Crystallogr., Sect. B*, 1997, **53**, 726.
56 D. W. M. Hofmann and T. Lengauer, *Acta Crystallogr., Sect. B*, 1997, **53**, 225.

7/07818H