NJC

Cite this: *New J. Chem.*, 2011, **35**, 1360–1368

www.rsc.org/njc

PAPER

Efficient computer modeling of organic materials. The atom–atom, Coulomb–London–Pauli (AA-CLP) model for intermolecular electrostatic-polarization, dispersion and repulsion energies†

Angelo Gavezzotti*

Received (in Montpellier, France) 14th December 2010, Accepted 21st February 2011
DOI: 10.1039/c0nj00982b

An atom–atom intermolecular force field with subdivision of interaction energies into Coulombic-polarization, dispersion (London) and repulsion (Pauli) terms is presented. Instead of using fixed interaction functions for atomic species, atom–atom potential functions are calculated for each different molecule on the basis of a few standard atomic parameters like atomic numbers, atomic polarizability and ionization potentials, and of local atomic point charges from Mulliken population analysis. The energy partitioning is conducted under guidance from the more accurate evaluation of the same terms by the PIXEL method, also highlighting some intrinsic deficiencies of all atom–atom schemes due to the neglect of penetration energies in Coulombic terms on localized charges. The potential energy scheme is optimized for H, C, N, O, Cl atoms in all chemical connectivities and can be extended to F, S, P, Br, I atoms with minor modifications. The scheme is shown to reproduce the sublimation heats of 154 organic crystal structures, to reproduce about 400 observed crystal structures without distortion, and to reproduce heats of evaporation and specific gravities of 12 common organic liquids. It is therefore suitable for both static and evolutionary (Monte Carlo) molecular simulation. Fine tuning of the four terms for specific systems can be easily performed on the basis of chemical intuition, by the introduction of one overall damping factor for each of them. The scheme is embedded in a suite of Fortran computer programs portable on any platform. For reproducibility and general use, source codes are available for distribution.

Introduction

Intermolecular forces are the determinants of structure and properties of materials. The availability of a reliable and affordable method for the calculation of intermolecular potentials is the prerequisite for a successful molecular simulation of liquids, mesophases, crystals and composites. Computational chemistry and computing resources have developed to the point where fully *ab initio*, basis-set-limit calculations of recognition energies including electron correlation are feasible for dimers of medium-sized molecules;[1–5] the interaction energies thus obtained are more reliable than the very scarce experimental results. Hybrid *ab initio*-semiempirical methods are being used for massive crystal structure generation and possible prediction.[6] Quantum chemical methods are however computer-intensive and do not (yet) open a path to large scale simulation, as is necessary for the dynamic study of relative phase stabilities and of phase transitions:[7,8] for dynamic simulation, the cohesive energy of a computational multimolecular system must be obtainable in times of the order of fractions of a second so that only empirical atom–atom schemes qualify.[9]

Intermolecular attractive-stabilizing or repulsive-destabilizing interaction stems from diffuse interactions between nuclei and electron clouds, susceptible of partly classical and partly quantistic interpretation. A useful approximation takes the following conceptual steps: (a) electron densities of interacting molecules are calculated separately by *ab initio* methods and are considered as undeformed on intermolecular approach; (b) Coulombic and linear-polarization terms are evaluated by classical formulae[10] using charge densities and their electric fields; (c) dispersion, or quantum hyperpolarization terms due to electron correlation are evaluated in the London approach;[11] and (d) repulsion is evaluated in terms of Pauli spin avoidance,[12] depending on the amount of overlap between proximal charge densities. This Coulomb–London–Pauli (CLP) assumption is often criticized, not without reason, but it has served the chemical community for a long time and

Department of Structural Chemistry, University of Milano, via Venezian 21, 20133 Milano, Italy.
E-mail: angelo.gavezzotti@unimi.it
† Electronic supplementary information (ESI) available: Details of Monte Carlo methods, force fields, geometries. Data points for Fig. 1–4; supplementary Fig. S1–S9. See DOI: 10. 1039/c0nj00982b

promises to do so for a long time yet.[13] The partitioning corresponds to elementary chemical ideas about electron distribution and polarization, and gives a more immediate access to structure-based chemical interpretation than is afforded by just one number for total intermolecular energy.

The CLP partitioning can be cast in first-principles form in Stone's IMPT (InterMolecular Perturbation Theory) approach,[14] with subsequent elaborations.[15] A semiempirical scheme called PIXEL[16] for the same purpose is based on numerical integrations over electron densities, and gives reliable results using a few parameters and quite affordable computing times.[17] The purpose of this paper is to present a form of the CLP manifold in the atom–atom approach, in which the total interaction energy is a sum over pairwise terms centered at atomic nuclear positions. In the traditional approach,[9] empirical atom–atom parameters are calibrated for each pair of atomic species. In the present approach the interactive nature of each atomic center depends not only on intrinsic atomic properties (ionization potential, polarizability) but also on an estimate of the relative electron population in its actual atomic basin in a given molecule, represented by a Mulliken population analysis on an Extended Hückel wavefunction.[18] The parameterization procedure goes through the following stages: (a) preliminary adjustment of protocols and parameters to reproduce the experimental sublimation enthalpies of 154 organic crystals, comparing separate energy terms with those from the well established PIXEL energy partitioning; (b) lattice energy minimization of about 500 organic crystal structures of general chemical composition, ensuring that the relaxation does not produce major structure distortion; (c) Monte Carlo simulation of liquids, to reproduce experimental vaporization enthalpies and specific gravities. Simultaneous calibration over static and dynamic simulation should overcome the well known problems encountered when statically optimized potentials are transferred to dynamic simulation.[19]

The atom–atom Coulomb–London–Pauli (AA–CLP) force-field model

Analytical form of the force field

A molecular structure is described by coordinates for the location of the nuclei. Hydrogen-atom positions from X-ray diffraction determinations must be renormalized (this is a vital step).[20] Table 1 gives a summary of the atom types considered. For a given i–j atom pair on different molecules the Coulomb–London–Pauli (CLP) interaction function has four terms, for Coulombic, polarization, dispersion and repulsion:

$$E_{i,j} = 1/(4\pi\varepsilon^{\circ})\,(q_i q_j)R_{i,j}^{-1} - F_{\mathrm{P}}P_{i,j}R_{i,j}^{-4} - F_{\mathrm{D}}D_{i,j}R_{i,j}^{-6} + F_R T_{i,j}R_{i,j}^{-12} \quad (1)$$

In this equation $R_{i,j}$ is an internuclear distance, $q_i = F_Q q_i^{\circ}$ is the rescaled net charge population on atom i, and F_Q, F_P, F_D, F_R are empirical, disposable scaling parameters. The polarization term depends on the inverse fourth power of distance as does the square of the electric field in the classical linear polarization formula; the inverse sixth-power and twelfth-power dependence of dispersion and repulsion terms are the most widely accepted. P, D and T are coefficients calculated on the basis of the local environment of an atom in a given molecule, as follows:

Table 1 Atomic species considered in the CLP intermolecular energy scheme. See the text for the explanation of atomic parameters, to be used as are in eqn (1). B and $H_{Bd,a}$ are dimensionless numbers

	α/Å^3	IP/eV	B	$H_{Bd,a}$
Hydrogen	0.39	0.500		
Acetylene CH			0.60	0.20
=CH_2, arom CH			0.62	0.10
Aliphatic (C)–H			0.64	0.05
R–OH, R–SH alcohol, thiol			0.75	0.99
COO–H acid			0.80	0.99
CON–H amide			0.80	0.90
R_2NH, RNH_2, $(R_3N^+)H$			0.80	0.99
H_2O (water)			0.80	0.99
Carbon		0.414	1.00	0.00
Carbonyl C=(O)	1.05			
≡C–	1.35			
sp^2 or allene C	1.35			
sp^3 C	1.05			
Aromatic core C	1.90			
Nitrogen	1.00	0.534		
$(R_nH_{4-n})N^+$			0.63	0.00
$(R_nH_{3-n})N$			0.63	−0.97
Aromatic N, R=N(H)	1.05		0.58	−0.99
–C≡N, –N=N			0.70	−0.70
Nitro N			0.63	0.00
Amide N			0.63	−0.85
Oxygen	0.75	0.500		
–O–			0.45	−0.90
H_2O (water)			0.70	−0.99
C=O, COO^-			0.50	−0.99
(C=O)–OH			0.50	−0.90
R–OH			0.45	−0.99
N=O			0.75	−0.95
S=O			0.75	−0.90
P=O			0.75	−0.90
Chlorine	2.50	0.477	2.40	−0.20

$$\alpha_{\mathrm{eff},i} = \alpha_i\,(Z_{\mathrm{V},i} - q_i)/Z_{\mathrm{V},i}) \quad (2a)$$

$$\alpha_{i,j} = [(\alpha_{\mathrm{eff},i})(\alpha_{\mathrm{eff},j})]^{1/2} \quad (2b)$$

$$P_{i,j} = \alpha_{i,j}\mathrm{abs}(q_i\,q_j) \quad (3)$$

$$D_{i,j} = \alpha_{i,j} n_i n_j (I_i\,I_j)^{1/2} \quad (4)$$

The reference polarizability of atomic species, α_i, is estimated with few adaptations from standard surveys,[21] and is corrected to an effective value by a local charge factor, eqn (2a), where $Z_{V,i}$ is the number of valence electrons, considering that a negatively charged atom with a richer electron cloud should be more polarizable, and *vice versa*. The polarization energy coefficient, eqn (3), depends on polarizabilities and on the absolute magnitude of the involved atomic charges, as in the expression for the electric field. The dispersion energy coefficient, eqn (4), depends on polarizability and also on the delocalization of the electron cloud, as represented by the quantum number n_i of the valence electrons, and on ionization potentials I_i as in the London formula.

The repulsion energy coefficient is assumed to depend on the space extension of the involved electron clouds, as represented by the number of valence electrons corrected by the atomic charge (a negatively charged atom is supposedly more diffuse) corrected by an empirical "diffusion factor" B_i, which for fine tuning may be different for same atomic species in different

molecular environments. The overlap/repulsion relationship may not hold in the same form for a marginal overlap (the approach of two benzene molecules) and for a large overlap (hydrogen bonding). Therefore additional parameters H_{Bd} (hydrogen-bonding donor ability) or H_{Ba} (acceptor ability) represent the hydrogen-bonding propensity of each atomic species. These parameters are given a value between 0 and 1, positive for donors and negative for acceptors, and the correction is applied when the two indicators have opposite sign. The final form is:

$$T_{i,j} = (1 + H_{Bd}H_{Ba})\,(Z_{V,i} - q_i)\,(Z_{V,j} - q_j)\,(B_iB_j)^{1/2} \quad (H_{Bd}H_{Ba} = 0 \text{ if } > 0) \qquad (5)$$

Table 1 collects optimized values of all the above described empirical parameters. As appears, and as typical of atom–atom formulations, the whole scheme has the character of an expert system of general efficiency rather than of a theoretical approach to the real physics of the interaction.

Evaluation of atomic point charges

The key parameter in the above formulation is q_i^o, the charge in each atomic basin. These numbers are estimated from a Mulliken population analysis on an Extended Hückel Theory (EHT)[18] molecular orbital wavefunction. The EHT Hamiltonian includes only valence orbital ionization potentials (VOIP) for diagonal terms, and an overlap-weighted average VOIP for off-diagonal terms; therefore, the EHT charge distribution essentially reflects electronegativity differences, and is well suited for the present purpose. The hydrogen-atom VOIP was reset to −10 eV (instead of the standard −13.6); EHT point charges must then be rescaled by an appropriate damping factor, and F_Q in eqn (1) is actually one of the adjustable parameters of the AA–CLP formulation. Such rescaled EHT point charges are remarkably similar to Mulliken point charges from an MP2/6-31G** calculation, obviously for an infinitesimal fraction of the computing time. MP2 and rescaled EHT charges are nearly identical over C–H groups, but show some fluctuations over hydrogen bonding X–H groups. Table S1 (ESI†) shows some typical examples and Fig. S1 (ESI†) shows the details of the comparison.

Calibration of parameters using crystal data

The essential innovation of the present calibration procedure is that the orders of magnitude of separate Coulombic, polarization, dispersion and repulsion contributions were primarily established under guidance of the corresponding PIXEL partitioned energies. The PIXEL energy partitioning is physically realistic[17] and compares favorably with dispersion-corrected Density Functional Theory results.[22]

The main database for parameter optimization consists of 154 crystals for which both the crystal structure and sublimation enthalpy are known (all structural data were retrieved from the Cambridge Database,[23] thermochemical data from critical compilations).[24] As appears from long experience, the accuracy of experimental ΔH(s)'s is generally low due to the intrinsic difficulty of the experiment: purity of the material, calibration of the instrumentation, control of the composition of the gas phase, control of partial decomposition are some of the factors that adversely influence the results even to the most careful experimenter. In addition, our calculated lattice energies refer to the hypothetical process of extraction of a rigid molecule from the crystal structure neglecting all intra- or intermolecular processes that may intervene in the phase transition from crystal to gas. A coarse estimate suggests that the expected uncertainty of experimental values is around 5 kJ mol^{-1} or 10%, whichever is larger, so pushing the calibration of parameters beyond this limit is probably pointless. Intramolecular energies are not considered, as is common practice in such static calculations.

Other databases were prepared for the lattice relaxation checks on crystal structures at low temperature, with $T < 200$ K, one molecule in an asymmetric unit, one chemical residue, crystallographic R-factor <5%, number of atoms <40, screening out by hand highly flexible molecules, for crystals containing C, H (hydrocarbons), C, H, O but no OH or COOH groups (oxohydrocarbons), C, H, Cl (chlorohydrocarbons), and carboxylic acids; a database of 185 crystals of small rigid molecules in a wide range of chemical functionalities, with crystal structure determination at low temperature, was also collected.

In all AA–CLP calculations, crystal structures are represented by clusters of molecules built over crystal symmetry operations to a cutoff distance between centers of mass of 40 Å. Lattice energies are evaluated as the corresponding lattice sums; an approximate correction for kinetic energy differences, eqn (7), may be applied:

$$U(\text{lattice}) = U(\text{Coul}) + U(\text{pol}) + U(\text{disp}) + U(\text{rep}) = 1/2\Sigma_i\Sigma_j E(i,j) \qquad (6)$$

$$\Delta H(\text{subl}) = -U(\text{lattice}) - 2RT \qquad (7)$$

As the clusters contain several thousand molecules, and since the sums are carried out over entire neutral molecules without imposing a distance cutoff, it has been verified that Coulomb sums, as implied by the first term of eqn (1), converge in all cases considered. Truncation effects are nonzero only for a very few crystal structures in polar space groups and with a strong molecular dipole oriented nearly parallel to the polar direction, and even so, corrections are almost negligible. Crystal structure relaxation allows variation of cell parameters and of rigid-body positional parameters, preserving the space-group symmetry. The optimization uses the Symplex method that requires no gradients and is particularly suitable when the target function is computationally inexpensive. For a crystal structure with cell parameters a, b, c, α, β, γ with rigid-body positional displacements x, y, z, θ, ϕ, χ after relaxation, the deformation is quantified by a structure-drift index S:

$$S = 100(\Delta a/a + \Delta b/b + \Delta c/c)/3 + 100(\Delta\alpha/\alpha + \Delta\beta/\beta + \Delta\gamma/\gamma)/N_{var} + 100(x^2 + y^2 + z^2)^{1/2}/N_{var,tras} + 100(\theta^2 + \phi^2 + \chi^2)/^{1/2}\,N_{var,rot} \qquad (8)$$

In this index, a 0.1 Å translational shift weighs as much as a 1% cell edge variation. The N_{var} factors are needed because the number of variable cell angles and translational and rotational degrees of freedom varies with each structure.

S should be zero for a perfect potential scheme and for a crystal at zero degrees: for an order of magnitude, $S < 5$ corresponds to scarcely significant structure changes. For crystal structures determined at room temperature, the temperature-less (formally 0 K) structure relaxation inevitably produces a cell shrinkage.

The form of the atom–atom potential, eqn (2)–(5), was marginally modified during optimization, while the four main rescaling coefficients and the disposable parameters listed in Table 1 were varied by trial and error until a satisfactory agreement with experimental thermodynamic and structural data was obtained at $F_Q = 0.41$, $F_P = 235$, $F_D = 650$, $F_R = 77\,000$. Notably, optimization of parameters for F, Br, I, S was not carried out and is postponed to future work.

Monte Carlo computational procedures

A Monte Carlo computer program was developed *ex novo* for application with the AA–CLP scheme. Simulations can be carried out in the isotropic or anisotropic NPT (constant number of particles, pressure and temperature) or NVT (constant volume) ensembles. Molecules of one or two kinds are enclosed in a computational box with periodic boundary conditions. Each molecule is described by (a) a number of atoms in a fixed "core" rigid configuration, (b) a number of "slave" atoms built from the core atoms by specifying appropriate values for bond distances, bond angles and torsion angles, (c) a chirality index that allows transformation into an enantiomer if necessary, and (d) once a full molecular model has been built, by three positional parameters for the center of mass and three rigid-body rotation angles. This allows all possibilities between a fully flexible, a partly flexible-torsionally free, or a fully rigid molecular representation. Further details can be found in ESI.†

The total configurational energy is a sum of intramolecular stretching–bending, torsional and non-bonded terms, and of intermolecular non-bonded terms. In general:

$$E(\text{tot}) = \Sigma F(\text{stretch, bend}) + \Sigma f(\tau) + \Sigma u(R, \text{intra}) + \Sigma u(R, \text{inter}) \quad (9)$$

where each summation runs over the allowed degrees of freedom. Intramolecular energy terms are indispensable in Monte Carlo, contrary to static crystal structure simulations. Stretching and bending were neglected because high-frequency vibrations are less relevant in the simulation of equilibrium liquids. The torsional part $f(\tau)$ can either be a polynomial in τ or a trigonometric function in τ ($0 < \tau < 180^\circ$):

$$f(\tau) = a + b\tau + c\tau^2 + d\tau^3 + e\tau^4 \quad (10)$$

$$f(\tau) = K\{1 - \cos n[\tau - \tau^\circ]\} \quad n = 1, 2, 3 \quad (11)$$

In both cases the employed functions were fitting an energy profile coming from *ab initio* calculations on the free molecule. Table S2 (ESI†) has details of all molecular models and force fields.

The intermolecular non-bonded potential is the atom–atom Coulomb–London–Pauli form, *i.e.* $u(R,\text{inter}) = E(\text{AA–CLP})$ as in eqn (1)–(5). The cutoff in the summations applies to distances between centers of coordinates rather than to distances between single atom–atom pairs, so that the summations always extend over entire molecules (neutral charge units), thus reducing if not eliminating completely the truncation effects, as previously discussed. The cutoff involves both molecules within the central box and molecules within the surrounding boxes of the periodic boundaries. Intramolecular non-bonded contacts may involve non-neutral units and hence the AA–CLP scheme cannot be applied. UNI atom–atom potentials[25] in 6-exp form are used for the purpose:

$$u(R,\text{intra}) = A\exp(-BR) - CR^{-6} \quad (12)$$

where R is an intramolecular non-bonded distance and the A, B, C coefficients for each pair of atomic species are shown in Table S3 (ESI†).

The degrees of freedom in a MC simulation are in principle the six rigid-body molecular parameters for each molecule in the box, plus slave-atom parameters. As a first approximation, an overlap between rigid-body and torsional degrees of freedom is neglected in the d.o.f. count. Box periodicity can be imposed in one, two or three dimensions; each molecule in the original box has translated counterparts for a ± 1 addition of periodicity vector(s), for a maximum total of 26 repeated boxes in the three-dimensional case. Intermolecular energies are computed over all pairs of molecules within the box or translation-repeated molecules, when the distance between centers of coordinates is below a given threshold. As soon as the center of a molecule moves outside a box boundary, the whole molecule is deleted and re-enters the box at the opposite end. This avoids problems with connectivity breaking that occur when single atoms are moved to comply with box periodicity.

A Monte Carlo move or action can be one of the following types:

(a) Random variation of one molecular parameter P. A random number $0 < r < 1$ is generated and positive or negative steps are taken by using $P' = P + (r - 0.5)$ times a maximum stepwidth, taken of the order of 0.4 Å, 30° and 20°, respectively, for translation, rotation and torsional changes. This provides an acceptance ratio of about 60%.

(b) Box dimensions change. Box dimensions can be changed at selected intervals. The computational box is identified by three variable box edges and three variable box angles. Isotropic change is implemented by varying all box edges by the same percent amount with all box angles 90°. Anisotropic change is implemented by a simultaneous random variation of all six box parameters; this option has been included for extension to simulations of organic crystals and their transitions, the subject of a forthcoming paper. The positions of all molecules then undergo a rigid-body change by the following procedure: (i) calculate fractional coordinates for the centers in the old box metrics; (ii) from these, calculate new orthogonalized coordinates of the centers in the new box metrics; (iii) calculate the components of the displacement of each molecular center, dx, dy, dz, and apply the same displacement vector to all atoms in the computational box.

(c) Pressure control. Whenever box dimensions are changed, pressure control is normally possible by the standard isothermal–isobaric ensemble (IIE) algorithm.[26] The new total

box energy U', the new volume V' and the quantity δH are then evaluated:

$$\delta H = U' - U^\circ + P(V' - V^\circ) - N\,kT\ln(V'/V^\circ) \quad (13)$$

The Metropolis algorithm is implemented on this last quantity. N is the number of "particles" in the system. Pressure checks are carried out every 700–1000 steps, with volume variations of the order of 100–300 Å^3.

The IIE P-control of eqn (13) samples directly from the proper probability density of the NPT ensemble, and is therefore more rigorous from the statistical thermodynamics point of view. Its application is straightforward for isotropic and anisotropic cases, although the number of independent "particles" N is ill-defined in a simulation involving large semi-flexible molecules; it has been taken here in a first approximation as the number of molecules even for non-rigid cases. For isotropic box dimensions change in a rectangular box, a weak-coupling algorithm similar to those used in molecular dynamics simulations[27] can also be used. The algorithm involves the equipartition kinetic energy, the virial, Ξ, calculated from derivatives of the potentials, and the current pressure to estimate a variation factor for box periodicity (see further details in ESI†). The virial P-control is realistic in that it allows pressure fluctuations, as internal pressure can oscillate in a range of several hundred atmospheres even at equilibrium. However, it cannot be applied to anisotropic cases because in Monte Carlo there is no access to molecular velocities or to pressure, kinetic energy and virial components along the three separate directions. Although it has been verified to perform properly in the prediction of densities of many organic liquids, there is no proof that the algorithm samples from the appropriate NPT ensemble. It is considered as a valuable check of the performance of the IIE pressure control algorithm.[26a]

Vaporization (sublimation) enthalpies are estimated as:[28]

$$\Delta H(\text{vap}) = \Delta U(\text{conf}) - U(\text{box})/N + \text{RT} \quad (14)$$

where ΔU(conf) is the intramolecular energy difference between gas and liquid (crystal), U(box) is the total configurational energy of the computational box, and the RT term corrects for the difference between internal energy and enthalpy. It has been verified that in the present cases the assumption $\Delta U(\text{conf}) = 0$ is valid in all cases. Specific gravities are calculated from the average dimensions of the computational box.

Outline of the computer program package

All calculations were carried out using a new program package, CLP, derived from the OPiX program suite.[29] All programs are written in Fortran and include software for retrieval from the CSD *cif*-file format with assignment and/or renormalization of positions of hydrogen atoms, and for calculation of atomic point charges. Further modules are for the calculation of static lattice energies by PIXEL or atom–atom, and for crystal structure relaxation. A separate suite of modules encodes the procedure for Monte Carlo calculations. To ensure reproducibility of the results, all source codes with documentation, user manual and sample inputs and outputs, a list of crystal structures used in the calibration, and details of all the force fields and molecular geometries for the Monte Carlo simulations can be obtained upon request from the author at his e-mail address.

Static simulation: results

Comparison between lattice energies and sublimation enthalpies

Fig. 1 shows the comparison between experimental sublimation enthalpies and calculated lattice energies, while Table S4 (ESI†) has the corresponding numerical values. The PIXEL and atom–atom results are almost undistinguishable, for a computational price of about 100 hours and 5 seconds, respectively. The agreement between experiment and the AA-CLP calculation, see details in Table 2, can be considered excellent: the average percent discrepancy is less than 8%, a

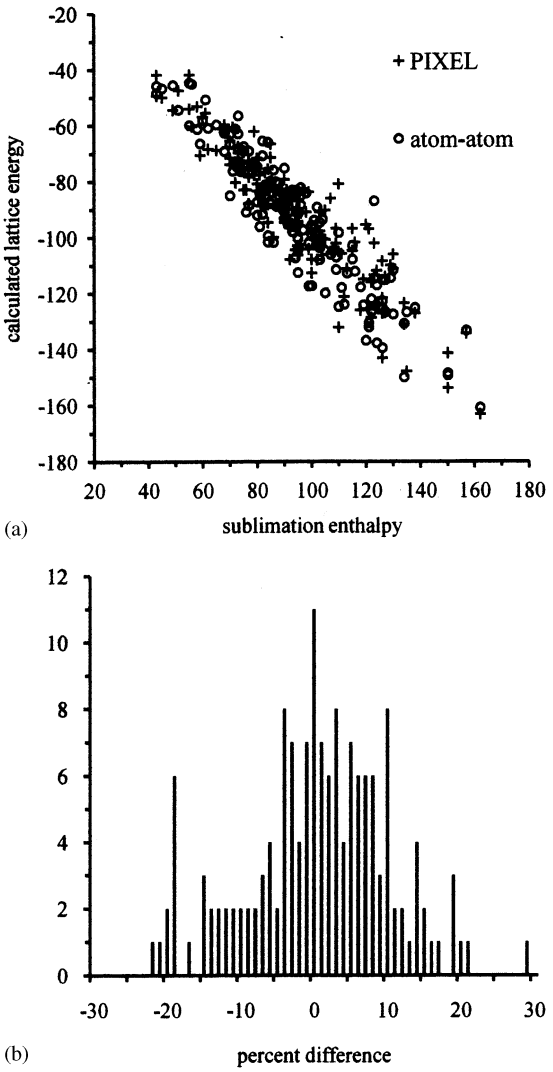

Fig. 1 (a) Comparison between experimental sublimation enthalpies and lattice energies calculated by PIXEL or AA–CLP. The least-squares lines are $y = 0.97x$ ($R^2 = 0.845$) and $y = 1.02x$ ($R^2 = 0.849$), respectively. 154 crystal structures, kJ mol^{-1} units. (b) Histogram of the percent differences between AA–CLP lattice energies and experimental sublimation enthalpies.

Table 2 Comparison of experimental enthalpies of sublimation with AA–CLP lattice energies, and of experimental crystal structures with energy-optimized structures

	Room-T crystal structures					Low-T crystal structures		
			Structure relaxation					
Class of compounds	n	Average ΔE_{latt}[a]	n	S[b]	Δcell[c]	n	S[b]	Δcell[c]
Hydrocarbons	29	6.2	26	8	−10	53	6	−7
Oxohydroc.	21	6.0	19	7	−7	71	6	−2
Acids	24	8.5	21	7	−5	26	5	−1
Amides	25	10	21	9	−7	—	—	—
Chlorohydroc.	11	6.3	8	5	−9	31	6	−8
Nitro compds.	12	7.3	10	6	+1	—	—	—
Azahydroc.	15	7.6	11	10	−10	—	—	—
Varia	11	7.1	11	10	−7	—	—	—
Alcohols	6	11	—	—	—	—	—	—
All	154	7.7	—	—	—	—	—	—
Low-T structures	—	—	—	—	—	185	6	−2

[a] Average percent difference between calculated lattice energy and experimental sublimation enthalpy, eqn (7) with T = 293 K. [b] Structure-drift index upon energy minimization, eqn (8). [c] Percent cell volume variation. Each entry is preceded by the number of crystal structures considered.

deviation within the error bars is due to the previously discussed uncertainties. Fig. 1b shows that in only 19 cases the deviation is larger than 15% (for some of these the experimental value is suspicious). The AA–CLP atom–atom result is less satisfactory for alcohols and amides, classes of substances for which the position of H-bonding hydrogen atoms may be questionable. Note that the same overall agreement between experiment and calculation can be obtained with exclusion of the $-2RT$ correction in eqn (7), and a minor decrease of the dispersion coefficient F_D to 600 instead of 650; a further indication that the $-2RT$ correction is just an overall rescaling that may well merge into numerical noise. Future work will assess whether the above conditions are sufficient for accurate polymorph screening among typical energy differences of 5 kJ mol^{-1} or less.

PIXEL Coulombic energies are not parametric and have been shown to be almost error-free.[30] They take due care of "penetration" effects, that is effects arising from interaction between diffuse electron clouds and nuclear charges, which are intrinsically absent in localized models operating with net point-charges at nuclear locations. Due to the neglect of penetration energies, AA–CLP Coulombic energies are usually about one half of the PIXEL ones (Fig. 2). Use of a larger value of F_Q in eqn (1) does not solve the problem: aside from the unrealistic atomic point charges, when close contact between atoms with charge of same sign is inevitable, as is often the case in crystals, the localized point-charge representation may result in repulsive terms where attractive ones should be (see the upper right corner of Fig. 2). This is the case in crystals of saturated hydrocarbons, and, for an extreme example, in the crystals of trichlorobenzene, pentachlorobenzene, hexachlorobenzene and perchlorobiphenyl, where atom–atom Coulombic energies are all destabilizing by up to 6 kJ mol^{-1}, whereas PIXEL Coulombic energies are stabilizing by −20 to −25 kJ mol^{-1}. Using a larger F_Q would only emphasize this wrong result. In the present AA–CLP model, however, this shortcoming is partly mitigated by the polarization term, always stabilizing even over charges of the same sign. The AA–CLP estimation of polarization terms is however a weak point of the scheme, because it neglects many-body field

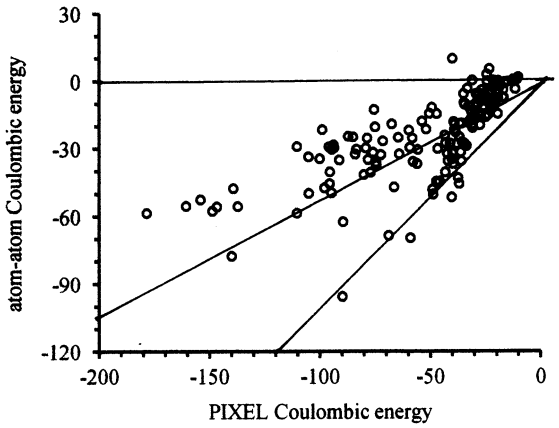

Fig. 2 Comparison of PIXEL (delocalized) and AA–CLP (point-charge) Coulombic energies. The two lines are for E(AA–CLP) = E(PIXEL) and E(AA–CLP) = 0.5 E(PIXEL). Same database as in Fig. 1 (kJ mol^{-1} units).

effects. In fact the accurate PIXEL polarization energy and the AA polarization term show little agreement (Fig. S2, ESI†). PIXEL and atom–atom dispersion energies are on the other hand remarkably similar (Fig. 3) as expected since the two models use a quite similar approach.

As a consequence of the underestimation of Coulombic energies, atom–atom repulsion energies are largely reduced with respect to PIXEL ones (Fig. 4). The excellent agreement between AA calculations and experiment results from a cancellation of two large underestimation errors.

Crystal structure relaxation

Table 2 shows the average structural shift factors (eqn (8)) for relaxation over separate groups of crystal structures, with the corresponding average percent cell variation. Relaxation of room-temperature structures produces the expected minor cell shrinkages, in the range up to 10%. For low-T structures, relaxation produces much smaller, mostly almost insignificant structure drifts. The structure drift and cell shrinkage tests confirm a lack of adverse effects rather than being carriers of

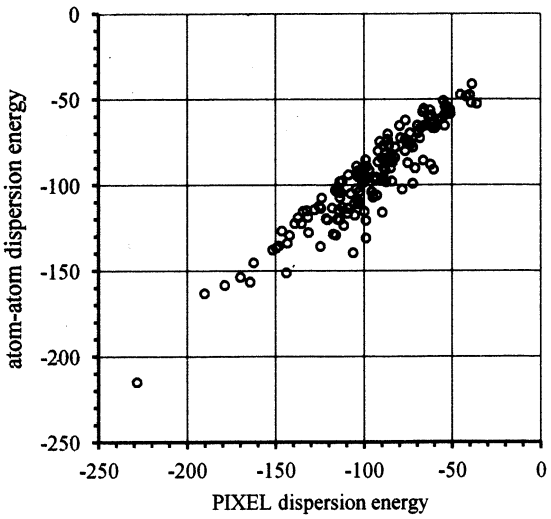

Fig. 3 Comparison of PIXEL and AA–CLP dispersion energies. The least-squares line is E(AA–CLP) = 0.962 E(PIXEL), R^2 = 0.824. Same database as in Fig. 1 (kJ mol^{-1} units).

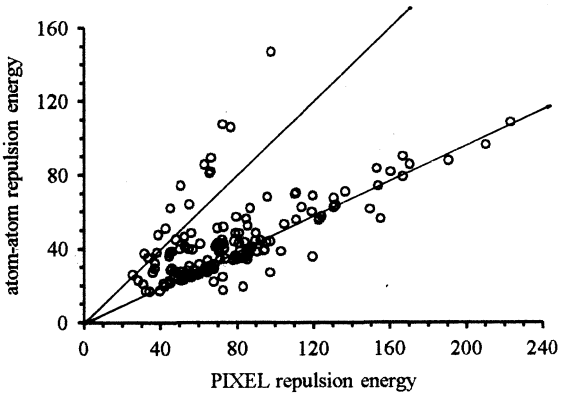

Fig. 4 Comparison of PIXEL and AA–CLP Pauli repulsion energies. The two lines are for E(AA–CLP) = E(PIXEL) and E(AA–CLP) = 0.5 E(PIXEL). The small population for which the two energies are almost equal is for crystals of nitro compounds where Extended Hückel charges, and hence Coulombic terms, are very large. Same database as in Fig. 1 (kJ mol^{-1} units).

positive agreement, but the results are promising also in view of further applications to dynamic simulations of liquids.

The benzene stack dimer and the acetic acid H-bonded dimer

A comparison between the PIXEL and atom–atom description of these two typical dimers helps in the interpretation of the physical basis for the assumptions in the AA model.

Fig. 5 shows the comparison between PIXEL and AA–CLP results for the benzene stacked dimer. Dispersion energies are nearly identical. The AA Coulombic term has the wrong sign at distances <4 Å, and the correct binding energy[31] results from a correspondingly large underestimation of the repulsion term. For the acetic acid cyclic dimer (Fig. 6) both methods give an essentially correct answer for the equilibrium distance, but the AA–CLP binding energy is far too small, as a result of the atom–atom underestimation of Coulombic energy. This underestimation of binding energies of hydrogen-bonded dimers is common to all empirical atom–atom force fields.[32]

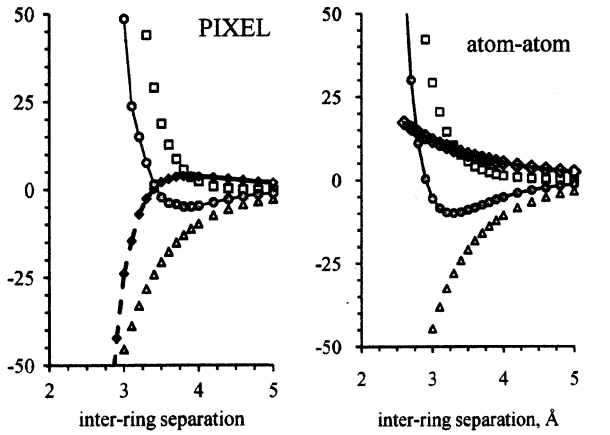

Fig. 5 PIXEL and atom–atom interaction energies for the benzene stacked dimer. The dashed line is Coulombic, the full line total. Triangles, dispersion; squares, repulsion (kJ mol^{-1} units).

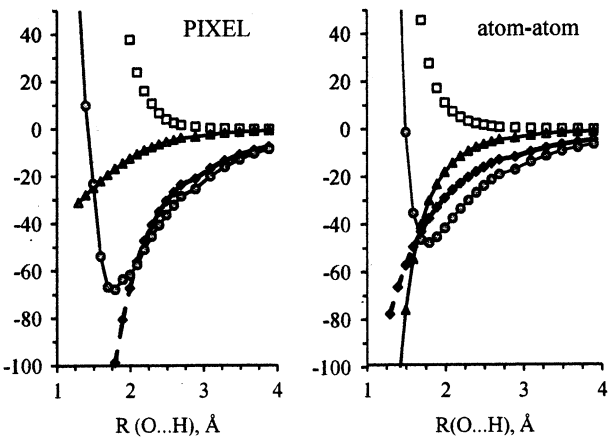

Fig. 6 PIXEL and atom–atom interaction energies for the acetic acid double H-bonded dimer. The dashed line is Coulombic, the full line total. Triangles, dispersion; squares, repulsion (kJ mol^{-1} units).

Remarkably, total lattice energies in crystals are correct as a result of compensations. These results confirm that the atom–atom method is an extremely useful and a versatile expert system, but may well be locally wrong, especially at the short ranges implied by close intermolecular contact between atoms in organic crystals, since it does not rest on sound physics.

Monte Carlo simulation of liquids

A number of common organic solvents were considered. The protocol for the simulations was as follows: (1) generate a computational box with molecules at the nodes of a body-centered cubic lattice whose periodicity is close to the average molecular diameter; boxes of a few hundred molecules were seen to be more than sufficient; (2) perform pre-stabilization runs of typically 1–5 Mmoves (Mmove = 1 000 000 moves) first with T = 0 and no box control, and then at room temperature and with a box periodicity d starting at $\sim$1.1 d°, where d° is the box length that reproduces the observed density of the liquid; no pressure control is yet applied; (3) when the system reaches a steady state with $d \approx d^\circ$, apply the virial pressure control with P = 101 300 Pa in stabilization runs of

Table 3 Results for the simulation of the liquid state of organic compounds[a]

	No. of mols	ΔH_v[b]	ΔH_v calc.	d, exp.	d, calc.	%C[c]
Acetic acid	686	51.6	52.5	1.049	1.125	34
Acetone	686	32	34.9	0.784	0.817	24
			31.9[d]		0.790[d]	
Aceto nitrile	1024	33.2	37.0	0.782	0.850	44
			31.9[e]		0.841[e]	
Benzene	686	34	35.5	0.874	0.926	8
			32.4[f]		0.915[f]	
$CHCl_3$	1024	31	31.9	1.465	1.576	5
			27.8[g]		1.523[g]	6
CCl_4	1024	32.5	39.3	1.584	1.800	0.2
			33.4[g]		1.724[g]	0.3
DMF	686	47.6	44.4	0.95	0.894	39
			45.6[h]		0.898[h]	
Formamide	686	60–65	70.3	1.129	1.261	
			58.7[e]		1.213[e]	
Ethanol	686	42.3	43.0	0.785	0.765	47
Methanol	686	37.4	38.7	0.787	0.788	59
n-Hexane	432	31.3	34.2	0.655	0.641	0.5
Methyl acetate	686	33–36	37.5	0.94	0.919	21
Methyl ether (T = 220 K)	686	19–21	23.8	0.668	0.724	29
Pyridine	686	40.3	41.8	0.979	1.055	18

[a] Energies in kJ mol^{-1}, densities in g cm^{-3}. T = 293 K unless otherwise stated. Averages over the last half of the final IIE P-control run; root-mean-square deviations from the averages for energies 0.1 to 0.3 kJ mol^{-1}, for box edges 0.05–0.10 Å. [b] From ref. 24. [c] Percent of Coulombic energy over the sum of Coulomb, polarization and dispersion. [d] F_Q = 0.39. [e] F_Q = 0.37. [f] F_D = 610. [g] With α(Cl) = 2.3 instead of 2.5 Å^3. [h] F_R = 73 000.

5–10 Mmoves. Final production run using the IIE pressure control gives the final results shown in Table 3, but with minor changes from the virial pressure control calculation. The radial distribution functions calculated on the final frames of simulations for acetic acid, methyl alcohol, chloroform and formamide (ESI†, Fig. S3–S9; results are identical to those from averages over the last equilibrated frames) are almost identical to those obtained from standard top-quality simulations by Monte Carlo or Molecular Dynamics.[9b,33–35]

Discussion and conclusions

A new atom–atom potential energy scheme is presented for the calculation of intermolecular recognition energies in organic condensed phases. Its main features are: (a) the adaptation of potential energy profiles for each atom pair in each molecule, based on an estimation of local charges in each atomic basin, lifting the restriction of fixed parameters for each atomic species; (b) the systematic and inexpensive estimation of local point-charge parameters; (c) the subdivision of total energies into Coulombic-polarization, dispersion and repulsion parts, adding information on the nature of the interaction, albeit within the limits of the so far discussed adherence to physics; in any case, the subdivision is more informative than just one total energy number; (d) the proportional calibration of these terms against the more reliable ones generated by PIXEL calculations; (e) the simultaneous calibration over static simulations of crystal structures and over evolutionary (Monte Carlo) liquid-state simulations; (f) the flexibility allowed by separate fine-tuning of coefficients, starting from universal ones, for each of the four components of interaction energies; (g) extensibility in principle to a good part of the periodic table by extrapolation from the parameters of the core atoms of organic chemistry.

The agreement between calculated lattice energies and experimental sublimation enthalpies is very good, considering that the fitting is carried out on an experimental sample of unprecedented amplitude (154 structures). Computing times for the calculation of lattice energies are negligible; it is expected that a calculated sublimation enthalpy is as reliable as an experimental determination, for the price of a few seconds of computing time against days or weeks of a cumbersome experiment. The optimization of lattice energies does not produce structural distortions in excess of those expected from the static nature of the calculation, *i.e.*, a contraction of cell parameters for room-T structures, very much reduced on low-temperature structures.

As an order of magnitude, atom–atom Coulombic and repulsion energies are about one half of the accurate PIXEL values, while dispersion contributions are identical in the two schemes. The information on the relative weight of these contributions, both in the overall lattice energy and in interaction energies between separate pairs in a crystal, can help in assessing the type of forces at work in a condensed system, and in guiding the substitution patterns to modify each type of interaction (Coulombic or dispersive) on the basis of the polarity of the substituent. However, the description of local interactions can be sometimes inaccurate in the atom–atom scheme due to the intrinsic deficiency of atom–atom schemes, related to point-charge localization and neglect of penetration energies (energies that come from a correct diffuse representation of electron densities). This effect is more relevant in higher halogens, or when an intermolecular overlap is very strong, as in hydrogen bonding.

The flexibility of the scheme, as discussed at point (f) above, is illustrated by some results of Monte Carlo simulations of liquids. Results obtained with the universal set of the four parameters of the theory (F_Q = 0.41, F_P = 325, F_D = 650,

F_R = 77 000) are already moderately satisfactory, with maximum obs–calc differences of 10% (see Table 3), although perhaps not as satisfactory as those of top-level, carefully parameterized simulations.[33] Minor adjustments of the scaling parameters, of the order of less than 10%, can however improve largely the agreement between simulation and experiment, recalling that an increase of F_Q and F_D, or a decrease of F_R, causes larger cohesion energy and density, and *vice versa*. Thus a change of F_Q from 0.41 to 0.39 or 0.37 is effective for acetone, acetonitrile and formamide, while a small reduction of F_D from 650 to 610 improves the agreement for benzene, and a reduction of F_R to 73 000 suits well the case of dimethylformamide. In perchlorocompounds, where the treatment of chlorine atoms is crucial, the wrong estimation of destabilizing Coulombic energies in the crystal can be partly compensated by enhancing the dispersion contribution using an atomic polarizability for chlorine of 2.5 Å^3. In the dynamic calculation however the correct value of 2.3 Å^3 performs better.

These manipulations of parameters raise a methodological problem. On the one hand, an ideal parameterization should be unique and invariable. On the other hand, the attempt to simulate all the chemistry of a dozen or so different atomic species in variable environments by just one set of parameters is close to impossibility. It is probably better to produce an informed view on how to adjust the parameters, with some judicious guess related to the chemical identity of the involved species, rather than promising presumably less reliable, all-comprehensive black-box parameterization. The present approach is a sensible compromise, the adjustment being conveniently performed by easily disposable damping factors.

Acknowledgements

Thanks are due to Jamshed Anwar (Bradford) for useful discussions and advice on some Monte Carlo procedures.

References

1 P. Jurecka, J. Sponer, J. Cerny and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993; K. E. Riley, M. Pitonak, J. Cerny and P. Hobza, *J. Chem. Theor. Comput.*, 2010, **6**, 66–80.
2 F. O. Kannemann and A. D. Becke, *J. Chem. Theor. Comput.*, 2010, **6**, 1081–1088.
3 Y. Zhao and D. G. Truhlar, *J. Chem. Theor. Comput.*, 2005, **1**, 415–432.
4 R. Podeszwa and K. Szalewicz, *Chem. Phys. Lett.*, 2005, **412**, 488–493.
5 L. Goerigk and S. Grimme, *J. Chem. Theor. Comput.*, 2010, **6**, 107–126.
6 J. van der Streek and M. A. Neumann, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2010, **B66**, 544–558.
7 J. Anwar, S. Tuble and J. Kendrick, *J. Am. Chem. Soc.*, 2007, **129**, 2542–2547.
8 D. Kofke and D. Frenkel, *Perspective Free energies and phase equilibria*, in *Handbook of Materials Modeling*, ed. S. Yip, Springer, Dordrecht, 2005.
9 (*a*) W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197; (*b*) W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236; (*c*) W. R. P. Scott, P. H. Huenenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krueger and W. F. van Gunsteren, *J. Phys. Chem.*, 1999, **A103**, 3596–3607; (*d*) B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187–217.
10 R. P. Feynman, *The Feynman Lectures on Physics*, Addison-Wesley, Reading, Mass., 1989, vol. 2, chap. 11.
11 F. London, *Trans. Faraday Soc.*, 1937, **33**, 8–26.
12 W. Kauzmann, *Quantum Chemistry, An Introduction*, Academic Press, New York, 1957, pp. 305–321.
13 J. D. Dunitz and A. Gavezzotti, *Angew. Chem., Int. Ed.*, 2005, **44**, 1766–1787.
14 A. J. Stone, *The theory of intermolecular forces*, Clarendon Press, Oxford, 1996, (reprinted with corrections, 2000).
15 K. Szalewicz, K. Patkowski and B. Jeziorski, *Intermolecular Interactions via Perturbation theory: from Diatoms to Biomolecules*, in *Structure and Bonding*, Springer-Verlag, Berlin, 2005, vol. 116, pp. 43–117.
16 A. Gavezzotti, *J. Phys. Chem.*, 2002, **B106**, 4145–4154; A. Gavezzotti, *J. Phys. Chem.*, 2003, **B107**, 2344–2353; A. Gavezzotti, *Molecular aggregation*, Oxford University Press, Oxford, 2007, chap. 12. The latest version of the PIXEL method is described in A. Gavezzotti, *Mol. Phys.*, 2008, **106**, 1473–1485.
17 A. Gavezzotti, *CrystEngComm*, 2003, **5**, 429–446; A. Gavezzotti, *J. Chem. Theor. Comput.*, 2005, **1**, 834–840; W. B. Schweizer and J. D. Dunitz, *J. Chem. Theor. Comput.*, 2006, **2**, 288–291; E. K. Gibson, J. M. Winfield, K. W. Muir, R. H. Carr, A. Eaglesham, A. Gavezzotti and D. Lennon, *Phys. Chem. Chem. Phys.*, 2010, **12**, 3824–3833; S. Aldridge, A. J. Downs, C. Y. Tang, S. Parsons, M. C. Clarke, R. D. L. Johnstone, H. E. Robertson, D. W. H. Rankin and D. A. Wann, *J. Am. Chem. Soc.*, 2009, **131**, 2231–2243; J. D. Dunitz and W. B. Schweizer, *Chem.–Eur. J.*, 2006, **12**, 6804–6815; S. Muniappan, S. Lipstman and I. Goldberg, *Chem. Commun.*, 2008, 1777–1779; P. A. Wood, R. S. Forgan, A. R. Lennie, S. Parsons, E. Pidcock, P. A. Tasker and J. E. Warren, *CrystEngComm*, 2008, **10**, 239–251; P. A. Wood, D. A. Haynes, R. Alistair, W. D. S. Motherwell, S. Parsons, E. Pidcock and J. E. Warren, *Cryst. Growth Des.*, 2008, **8**, 549–558; R. S. Forgan, P. A. Wood, J. Campbell, D. K. Henderson, F. E. McAllister, S. Parsons, E. Pidcock, R. M. Swart and P. A. Tasker, *Chem. Commun.*, 2007, 4940–4942; R. M. Ibberson, M. T. F. Telling and S. Parsons, *Cryst. Growth Des.*, 2008, **8**, 512–518.
18 R. Hoffmann, *J. Phys. Chem.*, 1963, **39**, 1397–1412.
19 A. Gavezzotti, *Chem.–Eur. J.*, 2002, **8**, 1710–1718.
20 A. Gavezzotti, *CrystEngComm*, 2008, **10**, 389–398.
21 K. J. Miller, *J. Am. Chem. Soc.*, 1990, **112**, 8533–8542.
22 L. Maschio, B. Civalleri, P. Ugliengo and A. Gavezzotti, *manuscript in preparation.*
23 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **B58**, 380–388.
24 Sublimation enthalpies of crystals and vaporization enthalpies of liquids have been retrieved mainly from the NIST Webbook (http://webbook/nist/gov/chemistry); see also A. Gavezzotti, *Molecular aggregation*, Oxford University Press, Oxford, 2007, p. 190ff, and ref. 25 below. A complete list of these data is in ESI†.
25 G. Filippini and A. Gavezzotti, *Acta Crystallogr., Sect. B: Struct. Sci.*, 1993, **B49**, 868–880; A. Gavezzotti and G. Filippini, *J. Phys. Chem.*, 1994, **98**, 4831–4837.
26 (*a*) M. P. Allen and D. J. Tildesley, *Computer simulation of liquids*, Oxford University Press, Oxford, 1989, pp. 41 and 124; (*b*) W. L. Jorgensen, J. Chandrasekhar and J. D. Madura, *J. Chem. Phys.*, 1983, **79**, 926–935.
27 W. F. van Gunsteren and H. J. C. Berendsen, *Angew. Chem., Int. Ed.*, 1990, **29**, 992–1023.
28 M. M. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, 2000, **112**, 8910–8922.
29 A. Gavezzotti, *Z. Krist.*, 2005, **220**, 499–510.
30 P. Coppens and A. Volkov, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2004, **A60**, 357–364.
31 W. B. Schweizer and J. D. Dunitz, *J. Chem. Theor. Comput.*, 2006, **2**, 288–291; A. L. Ringer and C. D. Sherrill, *Chem.–Eur. J.*, 2008, **14**, 2542–2547.
32 R. S. Paton and J. M. Goodman, *J. Chem. Inf. Model.*, 2009, **49**, 944–955.
33 J. M. Briggs, T. B. Nguyen and W. L. Jorgensen, *J. Phys. Chem.*, 1991, **95**, 3315–3322.
34 I. G. Tironi and W. F. van Gunsteren, *Mol. Phys.*, 1994, **83**, 381–403.
35 E. Tsuchida, *J. Chem. Phys.*, 2004, **121**, 4740–4746.

