Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemical Physics Letters 389 (2004) 373–378
CHEMICAL PHYSICS LETTERS
www.elsevier.com/locate/cplett

Mimicking electrostatic interactions with a set of effective charges: a genetic algorithm

Roberto Berardi, Luca Muccioli, Silvia Orlandi, Matteo Ricci, Claudio Zannoni *

Dipartimento di Chimica Fisica ed Inorganica and INSTM, Università di Bologna, Viale Risorgimento 4, 40136 Bologna, Italy

Received 2 January 2004; in final form 22 March 2004
Available online 20 April 2004

Abstract

We present a genetic algorithm apt to determine a set of effective charges that approximate the electrostatic field around a molecule. We show that these charges provide a reasonably good approximation to the pair electrostatic interaction and argue that the method should provide a valuable tool in computer simulations of condensed phases, particularly liquid crystals.

1. Introduction

The electrostatic interactions between two molecules represent a significant contribution to the total pair energy but, more importantly, their anisotropy plays a significant role in molecular recognition [1,2]. An appropriate modelling of this fundamental molecular feature relies on the possibility of determining correct electronic densities through sufficiently accurate quantum mechanics (QM) calculations, which should allow assigning a fractional charge [3] or, as further improvement, a set of distributed multipoles [4,2] to all atomic centres. While this is a non trivial task in itself, if the model potential is to be used in computer simulations of the condensed phases arising from a large set of molecules at given thermodynamic conditions, there are considerable additional problems related to the sheer number of atomic charges, even without considering, as it is normally the case, higher local multipoles or direct integration methods involving thousand of sample points per molecule [5]. Indeed simulation methods are based on repeated energy calculations and the large number of Coulomb pair contributions coupled to their long range character causes a major consumption of resources, even if computationally efficient methods like Particle Mesh Ewald [6] are employed. In some condensed phases and, particularly, for liquid crystal studies, fairly large samples have to be simulated at a number of state points in order to investigate the structure of the phases formed and their transitions [7]. For these systems a fully atomistic study is particularly demanding [8] and can rarely be performed. A classical modelling approach employed for these systems consists in translating the exceedingly detailed atomistic description into a molecular level one, where particles are replaced with simple objects, e.g., hard convex particles plus multipoles [9,10], or attractive–repulsive Gay–Berne ellipsoids [11,12] with an embedded point dipole [13,14] or quadrupole [15] (see also [16,17] for recent reviews). While this molecular resolution approach has proved to be very useful to investigate trends, e.g., changes in liquid crystals phase behaviour, as the molecular dipole position [13] or strength [18] is changed, setting up a molecular level model corresponding to a specific complex molecule with its charge distribution is far from obvious. It is also worth stressing that from a more formal point of view the single centre multipolar expansion is not justified for closely placed anisometric molecules, where the distance between, say, the tips of two adjacent molecules can be even smaller than the tip–centre distance on the molecule itself. This problem is particularly severe in the simulation of the liquid crystalline phases we are interested in, where the

* Corresponding author. Fax: +39-051-644-7012/+390512093690.
E-mail address: claudio.zannoni@cineca.it (C. Zannoni).

0009-2614/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.cplett.2004.03.119

competition between different contributions to the energy can generate a variety of molecular organisations, e.g., different smectic phases with a bilayer, interdigitated or striped molecular arrangements [13]. Some of these structures and of the even more complex phases based on polyphilic mesogens recently discovered [19] would probably be impossible to obtain without considering a realistic distribution of the charges.

Here, we wish to propose a simple alternative between the fully atomistic and the embedded multipole approaches, suitably placing a reduced set of effective charges that produce a field equivalent to the one coming from the full charge distribution. We shall treat as an example two strongly anisotropic mesogenic molecules: a rodlike one, *trans*-4-(*trans*-4-*n*-pentylcyclohexyl)-cyclohexyl carbonitrile (5CCH) [20], and a discotic one, hexa-thio-triphenylene (HTT) [21,22] and we shall assume, for the present purpose, that they can be approximated with rigid objects.

2. Fitting methodology

Our approach consists in obtaining first a QM level description of the charge distribution and of the corresponding reference electrostatic potential around the molecule [23]. Having assumed a certain number N_Q of effective charges, we then determine their optimal positions and values by fitting their electrostatic potential surface to the reference one. This problem is characterised and complicated by the presence of a large number of local minima; to converge to the global minimum we have thus adopted as a minimisation tool a genetic algorithm (GA) [24] based on the classes of the GALib C++ library [25]. Other minimisation procedures like simulated annealing might be used in principle, although GA has proved superior in relatively similar molecular problems [26]. In practice, we have used the genetic algorithm only to explore the space of the possible charge positions, while the actual charge values have been determined with a linear least square fit to the reference potential. We define our N trial solutions or genomes g^i as sets of N_Q charge positions, kept ordered according to their x component. If the molecule possesses symmetry elements, each genome contains only N_Q independent charge coordinates, while the complete set of Q positions is generated applying the symmetry operators to the minimal set

$$g^i = (\mathbf{r}^i_1, \ldots, \mathbf{r}^i_{N_Q}), \quad r^i_{1,x} < \cdots < r^i_{N_Q x}, \tag{1}$$

where for each individual g^i, the corresponding charge vector $\mathbf{q}^i$ is computed equating the least square solution to the overdetermined linear equation system $\mathbf{D}(g^i)\mathbf{q}^i = \mathbf{V}$, resulting from the electrostatic potential generated by the trial charges and the QM potential calculated in M predefined points $\mathbf{p}_j$ outside the molecular surface. A constraint linear equation with statistical weight w (here we have used $w = M/50$) was added to keep the total charge close to the molecular charge C

$$\mathbf{D}(g^i)\mathbf{q}^i = \begin{bmatrix} \frac{1}{|\mathbf{r}^i_1 - \mathbf{p}_1|} & \cdots & \frac{1}{|\mathbf{r}^i_Q - \mathbf{p}_1|} \\ \cdots & \cdots & \cdots \\ \frac{1}{|\mathbf{r}^i_1 - \mathbf{p}_M|} & \cdots & \frac{1}{|\mathbf{r}^i_Q - \mathbf{p}_M|} \\ w & \cdots & w \end{bmatrix} \begin{pmatrix} q^i_1 \\ \cdots \\ q^i_Q \end{pmatrix} = \begin{pmatrix} V_1 \\ \cdots \\ V_M \\ wC \end{pmatrix} = \mathbf{V}. \tag{2}$$

The reference points $\mathbf{p}_j$ have been chosen to lay on an envelope of surfaces generated from the van der Waals (vdW) spheres positioned in the atomic centres of the equilibrium molecular geometry. For each atom we have used seven spheres with radius equal to the vdW radius augmented of $\Delta r = 0.0$, 0.5, 1.0, 1.5, 2.0, 2.5, 5.0 Å. On each spherical surface we have generated a number of points $N = 220/(1 + \Delta r)^2$ and retained only those not belonging to other spheres (in our test cases we have typically considered a total number of points $M \approx 4000$). The fitness function $\mathscr{F}$, whose value determines the survival or extinction of a certain individual, is given by the squared difference between the effective potential and the ab initio one, times a constraint function that forces the charge–charge distances r^i_{jk} to have values greater than $r_{\min} = 0.5$ Å

$$\mathscr{F}^i = \frac{||\mathbf{D}(g^i)\mathbf{q}^i - \mathbf{V}||^2}{M - 1} \prod_{\substack{j<k \\ r^i_{jk} < r_m}}^{Q} \exp\left(\frac{r_m - r^i_{jk}}{r_m}\right). \tag{3}$$

The evolution takes place by a steady state GA with overlapping populations between generations, and a genome replacement probability of 80%. The crossover operator $\hat{\mathscr{X}}^k$, i.e., the operation that generates two children from two parent genomes, has been chosen to act on 80% of the population and to be of single point type, i.e., to cut and splice two genomes only in one random gene position k

$$\hat{\mathscr{X}}^k\{g^i, g^j\} = \{\hat{\mathscr{R}}(\mathbf{r}^i_1, \ldots, \mathbf{r}^i_k, \mathbf{r}^j_{k+1}, \ldots, \mathbf{r}^j_{N_Q}), \hat{\mathscr{R}}(\mathbf{r}^j_1, \ldots, \mathbf{r}^j_k, \mathbf{r}^i_{k+1}, \ldots, \mathbf{r}^i_{N_Q})\}, \tag{4}$$

where we have introduced the ordering operation with the symbol $\hat{\mathscr{R}}$. The mutations are attempted with a 25% probability and performed by randomly changing the fitting charge positions, with a maximum displacement of 0.25 Å and constraining the new position to be inside a reduced vdW molecular volume (we have used a scaling factor of 0.8). Using this algorithm, a certain number of solutions (i.e., a population $G^N = \{g^1, \ldots, g^N\}$) evolves through random mutations and exchanges of part of the genomes, until a termination criterion is achieved. The termination may occur either upon lack of diversity, i.e., when the differences between

the genomes fall below a chosen threshold value, or upon convergence, that is when the average fitness function does not vary for a chosen number of steps. The initial genome points have been randomly chosen inside the vdW scaled molecular volume.

3. Results and discussion

As test cases to verify the effectiveness of the technique, we have chosen, as already mentioned, two prototype rodlike and discotic molecules in the field of liquid–crystalline materials: 5CCH and HTT (see Figs. 1 and 2). HTT is of interest here also because it represents a particularly unfavourable case for our approximation, since its D_{3h} symmetry reduces the number of independent charge positions available. It is, however, worth noticing that in real HTT based discotics [21], or in the numerous similar families of discotics [22,27], alkyl or alkoxy chains are normally attached to the core of the mesogen. Starting from ab initio equilibrium molecular geometries we have calculated the atom–centred Electrostatic Potential (ESP) charges [23], following the typical approach used to parametrise atomistic Molecular Dynamics simulations. The potential corresponding to the ESP charges represents our reference system: we have then applied our charge–modelling methodology to approximate it with a lower number of effective charges, ranging from $N_Q = 2$ to $N_Q = 15$ for 5CCH and from $N_Q = 1(3)$ to $N_Q = 4(24)$ for HTT, where N_Q is the number of symmetry independent charges and we give in brackets the total number of charges obtained by applying the molecular symmetry operators to the N_Q charges. In Figs. 1 and 2 we show a rendering of the reference and effective electrostatic potentials for 5CCH and HTT at the molecular surface defined by the vdW radius. We see at once (with the help of a colour coding) that the potential map obtained from the set of effective charges is quite similar to the reference one, indicating that the approximation is providing the correct charge distribution across the molecules. On a more quantitative side, as well as to test the effect of varying the number of charges, we report in Table 1 (third column) the Root Mean Square error $\sqrt{\mathscr{F}}$ for the fit when employing different N_Q. It is interesting to note that the effective independent charge positions required to approximate the reference potential at the level shown in the figures are just 5 for 5CCH and 2 for HTT (which become 12 when symmetry is applied). In the limiting case that the number of fitting charges equals the number of atoms the atomic positions and charges are recovered. Moreover the positions of the charges (Fig. 3) make these systems very different from the corresponding models with a point dipole or quadrupole in the molecular centre. Notice that we do not constrain the charges in any specific way so as to optimize the point multipoles of the molecule. Thus it is not surprising that, although in the case of 5CCH the dipole and the quadrupole are well reproduced for all choices of N_Q, for the more symmetric HTT at least three independent charges are needed to have dipoles and quadrupole components of the correct sign and magnitude (see Table 2).

Even if a good reproduction of the ESP around a molecule is encouraging, since we are mainly interested in using our effective charge representation in computer simulations, we have compared, as a more severe test, the intermolecular electrostatic potential obtained from the full and the reduced sets of charges, also to examine how the effect of the approximations propagate into the

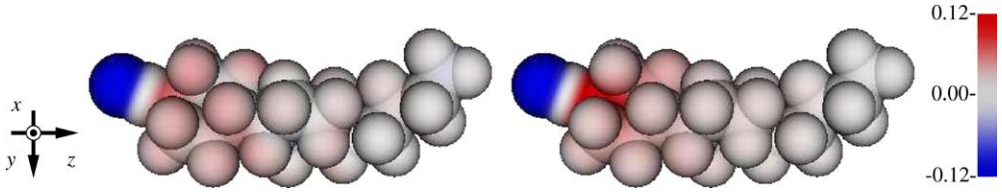

Fig. 1. The ab initio (ESP, left) and effective ($N_Q = 5$, right) electrostatic potential at the molecular surface for the 5CCH molecule, colour coded as from the palette (units $4\pi\epsilon_0 e$ Å). View along the y molecular axis. Geometry and charges obtained at B3LYP/6-31G* level.

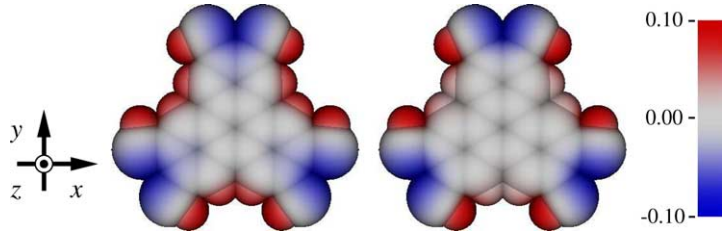

Fig. 2. HTT ab initio (left) and $N_Q = 2(12)$ (right) electrostatic potential at the molecular surface, colour coded as from the palette (units $4\pi\epsilon_0 e$ Å). View along z molecular axis. Geometry and charges obtained at MP2/STO-3G and MP2/6-311G+ level, respectively.

Table 1
Electrostatic potential fit quality indicators as a function of the number of symmetry independent charges N_Q in the reduced sets for 5CCH and HTT

	N_Q	$\sqrt{\mathscr{F}}$	MAE	RMSE	MAPE
5CCH	50*	0.0000	0.0000	0.0000	0.0
	2	0.0406	0.0233	0.0519	26.2
	3	0.0373	0.0112	0.0321	13.6
	4	0.0359	0.0109	0.0297	13.0
	5	0.0334	0.0095	0.0262	13.9
	10	0.0285	0.0089	0.0239	12.8
	15	0.0265	0.0082	0.0233	10.9
HTT	6 (36)*	0.000	0.000	0.000	0.0
	1 (3)	0.330	0.320	0.788	100.0
	1 (6)	0.280	0.295	0.659	97.1
	2 (12)	0.095	0.056	0.149	30.9
	3 (18)	0.032	0.038	0.090	24.5
	4 (24)	0.006	0.006	0.014	4.3

The star (*) marks the reference value for the full set of charges and, where relevant, we give in brackets the total number of symmetry replicated charges. We show: the root mean square error $\sqrt{\mathscr{F}}$ (kcal mol^{-1} C^{-1}) for the ESP around the single molecule and the mean absolute error (MAE) (kcal mol^{-1}), root mean square error (RMSE) (kcal mol^{-1}), mean absolute percentage error (MAPE) (%) relative to the pair interaction energy.

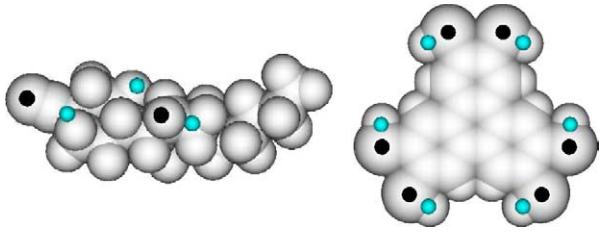

Fig. 3. Optimal charge positions (positive gray, negative black spheres) superimposed on the molecular framework for 5CCH (left view along x axis) and HTT (right view along z axis).

intermolecular potential. We have considered configurations of two identical molecules with the first one fixed in the centre of the coordinate system with the inertia axes aligned with the cartesian axes. The second one has the inertial frame rotated of Euler angles [28] (α, β, 0) using 4 grid steps for α and 5 intervals for β, and has been translated at the points (r, θ, ϕ) obtained using 8 and 5 steps for θ and ϕ, and varying r from 0 to $r_{\max} = 55$ Å with a step of 0.2 Å. At each point laying in the attractive or weakly repulsive region of the Lennard–Jones intermolecular potential (where $E_{\mathrm{LJ}} < 1$ kcal mol^{-1}), the electrostatic intermolecular energy has been computed using the reference and the reduced set of charges. We have computed the mean absolute error MAE, the root mean square error RMSE and the mean absolute percentage error MAPE only for the points having $|E_{\mathrm{ref}}| > 0.01$ kcal mol^{-1}, since these are the regions more meaningful from a chemical point of view.

Table 1 displays the overall quality of the fit obtained by a different number of charges. As expected, increasing the number of charges the fit improves and the deviations MAE, RMSE and MAPE decrease; we can notice that the improvement is not linear and a number of point charges considerably lower than the number of

Table 2
Dipole moment μ_i (Debye, 1 Debye = 3.33564×10^{-30} C m) and quadrupole components Q_{ii}(Debye Å) from various sets of charges for the 5CCH and HTT molecules

	N_Q	μ_x	μ_y	μ_z	Q_{xx}	Q_{yy}	Q_{zz}
5CCH	50*	0.01	−1.45	4.34	53.6	46.2	−99.8
	2	−0.03	−1.35	3.73	49.2	42.2	−91.4
	3	0.00	−1.38	4.44	52.0	44.1	−96.1
	4	−0.03	−1.29	4.40	51.4	44.4	−95.8
	5	0.02	−1.31	4.49	51.9	45.6	−97.5
	10	0.03	−1.42	4.51	53.1	46.3	−99.5
	15	0.03	−1.37	4.40	52.6	45.3	−98.0
HTT	6 (36)*	0.00	0.00	0.00	−1.27	0.47	0.80
	1 (3)	−0.19	0.14	0.00	0.14	−0.14	0.00
	1 (6)	−0.05	−0.04	0.00	0.98	2.04	−3.02
	2 (12)	−0.07	0.03	0.00	−4.01	−4.77	8.78
	3 (18)	0.01	0.04	0.00	−1.06	0.76	0.30
	4 (24)	0.00	−0.02	0.00	−0.96	0.11	0.85

The quadrupole moments have been calculated in the molecular inertial frame; (* indicates the reference ESP charges).

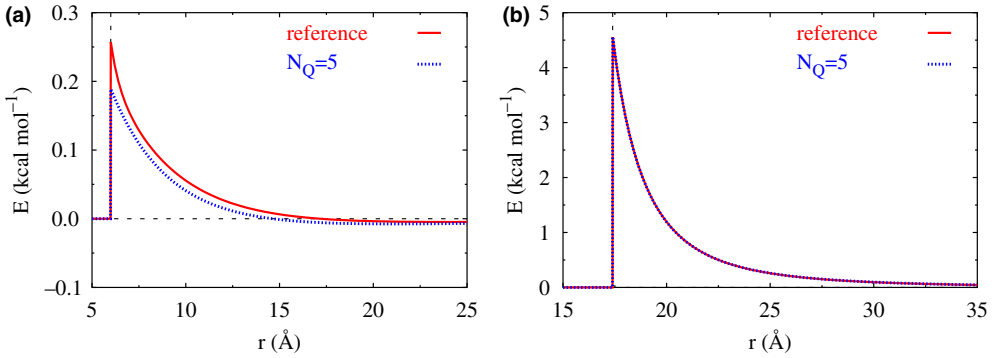

Fig. 4. The pair electrostatic energy calculated using all the atomic charges (red) and the reduced set for two 5CCH molecules approaching along x (a) and z (b) for ($\alpha = 0$, $\beta = \pi$, $\gamma = 0$) orientations. The thin dotted vertical lines represent the repulsive region of the potential, where the electrostatic potential is set to zero.

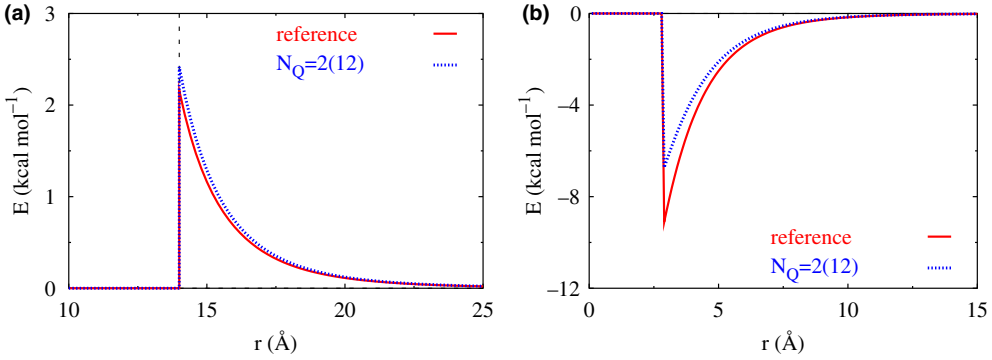

Fig. 5. The pair electrostatic energy calculated using all the atomic charges (red) and the reduced set for two HTT molecules approaching along x (a) and z (b) for ($\alpha = \pi$, $\beta = 0$, $\gamma = 0$) orientations. The thin dotted vertical lines represent the repulsive region of the potential, where the electrostatic potential is set to zero.

atoms is sufficient to describe the intermolecular electrostatic potential with a RMSE of the order of 10% or lower. The quality of the fitted potential depends on the distance and orientation, but it is important to notice that, at least for our purposes, it is sufficiently good even at short distances and for configurations that can be expected to be particularly important, like the side–side one for the elongated molecule 5CCH and the face–face for the discotic HTT. In Figs. 4 and 5 we show as an example the results for two typical approach directions. The good agreement of the pair potential over the whole region of close approach and for a variety of approach directions is important for the simulation of condensed phases, where densities are such that molecules are likely to be in near contact.

As a further test of the effectiveness of the approximation in a realistic situation, we have calculated the electrostatic energy for a system of $N = 1000$ Gay–Berne [11,12] oblate ellipsoids modelled after the HTT mesogen and we have found that, when calculated with the reduced set of 12 charges, it deviates from that obtained from the full set of 36 charges by less than 5% both in the isotropic and in the columnar phases.

4. Conclusions

We have proposed a simple approach to represent the charge distribution of complex molecules with a small set of effective charges (in the cases studied here only about a third of the original atomic charges) and we have described a genetic algorithm to determine these reduced charges and their locations. We have shown that the method can reproduce the electrostatic potential of a molecule and that the pair potential is also well approximated. The method should present the advantage of large savings of computer time when compared with the standard one assigning a charge to each atomic centre. Reduced set of charges can be useful in coarse grained methodologies and to add local interactions to rigid molecular models of purely repulsive or Gay–Berne type. In this case the presence of regions of different charge localised away from the centre should allow the simulation of phases where interdigitation, or in general, local structuring occurs [16]. We thus believe that the method can be a useful addition to the set of tools used in the simulations of complex phases, both low molar mass and polymeric ones, particularly the

many recently discovered ones based on polyphilic molecules [19], where it is essential to allow for the inhomogeneous distribution of attractive and repulsive sites in the constituent molecules.

Acknowledgements

We thank University of Bologna, EU (FULCE TMR) and MIUR for financial support through PRIN 'Cristalli Liquidi' and FIRB projects, as well as V. Barone (Napoli) and M. Cecchini (Zürich) for useful discussions.

References

[1] A. Buckingham, Adv. Chem. Phys. 12 (1967) 107.
[2] S. Price, Intermolecular forces – from the molecular charge distribution to the molecular packing, in: A. Gavezzotti (Ed.), Theoretical Aspects and Computer modeling of the Molecular Solid State, Wiley, New York, 1997, pp. 31–60 (Chapter 2).
[3] C. Cramer, Essentials of Computational Chemistry, Wiley, New York, 2002.
[4] A. Stone, The Theory of Intermolecular Forces, Oxford University Press, Oxford, 1996.
[5] A. Gavezzotti, J. Phys. Chem. B 106 (2002) 4145.
[6] T.A. Darden, D.M. York, L.G. Pedersen, J. Chem. Phys. 98 (1993) 10089.
[7] P. Pasini, C. Zannoni (Eds.), Advances in the Computer Simulations of Liquid Crystals, Kluwer, Dordrecht, 2000.
[8] R. Berardi, L. Muccioli, C. Zannoni, ChemPhysChem 5 (2004) 104.
[9] T. Kihara, Adv. Chem. Phys. 5 (1963) 147.
[10] A. Gil-Vilegas, S. McGrother, G. Jackson, Mol. Phys. 92 (1997) 723.
[11] J.G. Gay, B.J. Berne, J. Chem. Phys. 74 (1981) 3316.
[12] R. Berardi, C. Fava, C. Zannoni, Chem. Phys. Lett. 297 (1998) 8.
[13] R. Berardi, S. Orlandi, C. Zannoni, Chem. Phys. Lett. 261 (1996) 357.
[14] M. Houssa, L. Rull, S. McGrother, J. Chem. Phys. 109 (1998) 9529.
[15] I. Withers, C. Care, M. Neal, D. Cleaver, Mol. Phys. 100 (2002) 1911.
[16] C. Zannoni, J. Mater. Chem. 11 (2001) 2637.
[17] L. Longa, H.-R. Trebin, G. Cholewiack, in: W. Haase, S. Wrobel (Eds.), Relaxation Phenomena, Springer, Berlin, 2003, pp. 204–236.
[18] R. Berardi, S. Orlandi, D. Photinos, A. Vanakaras, C. Zannoni, Phys. Chem. Chem. Phys. 4 (2002) 770.
[19] C. Tschierske, Ann. Rep. Prog. Chem., C 97 (2001) 191.
[20] D. Dunmur, A. Fukuda, G. Luckhurst, Physical Properties of Liquid Crystals: Nematics, vol. 25 of EMIS datareview series, IEE, 2000.
[21] E. Gramsbergen, H. Hoving, W. De Jeu, K. Praefcke, B. Kohne, Liq. Cryst. 1 (1986) 397.
[22] S. Chandrasekhar, S. Prasad, Contemp. Phys. 40 (1999) 237.
[23] B.H. Besler, K.M. Merz, P.A. Kollman, J. Comput. Chem. 11 (1990) 431.
[24] D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley, Reading, MA, 1989.
[25] M. Wall, Galib 2.4.5. Available from: <http://lancet.mit.edu/ga/>, a C++ Library of Genetic Algorithm Components; Copyright 1995–1996 MIT; 1996–1999 Matthew Wall.
[26] Y. Zeiri, Phys. Rev. E 51 (1995) R2769.
[27] O. Roussel, G. Kestemont, J. Tant, V. de Halleux, R. Aspe, J. Levin, A. Remacle, I. Gearba, D. Ivanov, M. Lehmann, Y. Geerts, Mol. Cryst. Liquid Cryst. 396 PN 1 (2003) 35.
[28] M.E. Rose, Elementary Theory of Angular Momentum, Wiley, New York, 1957.