ELSEVIER

Biophysical Chemistry 51 (1994) 235–241

Biophysical Chemistry

# Molecular mechanics and electrostatic effects

Francisco Figueirido, Gabriela S. Del Buono, Ronald M. Levy *

*Department of Chemistry, Wright-Rieman Laboratories, Rutgers, The State University of New Jersey, Piscataway, NJ 08855-0939, USA*

Received 3 February 1994; accepted 20 February 1994

**Abstract**

Continuum solvent models predict a quadratic charge dependence (linear response) of the free energy of a system of charged solutes. The relation between this prediction and the structure of the solvation shell around the solutes is discussed. Studies of the derivative of the free energy with respect to the charges for different reference states are shown to be a convenient way of testing the linear response assumption without resorting to the standard free energy perturbation method. We illustrate this with a system of two oppositely charged ions in aqueous solution, where nonlinearities are observed before the full charging process is completed. Since molecular mechanics (MM) simulations preserve the full nonlinearity of the problem, they are well suited to the investigation of the conditions under which linear response accurately reflects the behavior of the system. The error when using linear response theory to calculate the free energies of charging is estimated to be as large as 10–20%.

*Key words:* Molecular mechanics; Electrostatics; Solvation; Linear response theory

In simulations of biomolecular systems the long-range electrostatic forces play a predominant role. Not only do proteins and nucleic acids have nonzero partial charges on their surfaces, which sometimes dictates the introduction of a few counterions to achieve net zero charge, but the solvent medium *par excellence*, water, is highly polar.

There are essentially two complementary approaches to biomolecular simulations. On the one hand, the so-called *continuum model* (CM) [1,2] replaces the solvent, which accounts for the bulk of the simulation system, by a dielectric continuum. On the other hand, *molecular mechanics* (MM) simulations [3,4] consider the biomolecule surrounded by a large number of solvent molecules. There are substantial differences between the two approaches which make each better suited to the calculation of different properties. There exists also some hybrid models which treat part of the solvent explicitly while, at the same time, attempting to deal with the effect of the bulk solvent in some sort of mean field approximation [5–9]. Depending on the properties under study these models might be considered as CM or MM; we will not discuss them further except to note that they share with CM the use of a sharp boundary to enclose a region with explicit atoms.

An advantage, in principle, of MM over CM is that it relies less on empirical parameters; in particular the dielectric response of the solvent is obtained from the simulation itself, and is not

* Corresponding author.

0301-4622/94/$07.00 © 1994 Elsevier Science B.V. All rights reserved
*SSDI* 0301-4622(94)00044-K

assumed to be known. It has been proved for some systems that MM simulations with Ewald summation produce the correct dielectric response [10–12]; there are also indications that it might work for biomolecules in solution, provided a large enough number of water molecules is included [13,14].

Since CM assumes that the response of the dielectric medium is known and local, it is computationally faster and the bulk properties of the solvent are better taken into account. However, the water molecules close to the solute are replaced by an unphysical sharp dielectric boundary, and it is not clear how much error is introduced by this approximation. There is evidence that the dielectric response function is not charge independent close to the solute [15,16]. Jayaram et al. [15] have shown by explicit MM calculations that the solvent around a spherical ion responds in a peculiar nonlinear way, becoming highly structured when the ionic charge is moderately high.

Both MM and CM treat solutes as a collection of point charges, which are renormalized to take into account, at least in a mean field way, the effect of the solvent on the electronic wavefunctions. This approach, however, neglects electronic polarizability, which is a truly many-body effect and has been lately recognized as a potentially significant source of errors [17–19]. Recent work has been directed, with some success, to the inclusion of electronic polarizability in MM simulations [18,20–29], as well as in CM calculations [30,31].

Despite these shortcomings, studies of biomolecules in solution using CM have yielded very reasonable results [2,31–33]. Jayaram et al. [15] studied the charging free energy of one ion using MM to test the validity of the continuum (Born) treatment of ion hydration. They found that, for their particular system, the continuum prediction of a quadratic charge dependence of the free energy is reproduced for values of the ionic charge up to some limiting value of approximately +1.1 au. Other studies have shown that this limiting value is system dependent, and can be less than 1 au [16,34].

As discussed by Warshel and co-workers [5,35] the calculation of $pK_a$ shifts in proteins is a very stringent test of the model for the electrostatic interactions, besides its being of great interest in the theoretical studies of protein stability [36]. Since proteins have nonzero partial charges on the surface, it is expected that the first solvation shell should be highly structured and thus a CM approach does not in principle seem adequate. Karplus and Bashford [32] have used the finite difference Poisson–Boltzmann [1] method to estimate the $pK_a$s of lysozyme and found relatively good agreement with experiment. Honig and coworkers, however, using the same approach [33], found poor agreement for some of the ionizable groups in T4 lysozyme and traced it to the existence of a salt bridge on the surface, which serves as a seed for solvent structure. To improve their results they added a few explicit waters in this low dielectric region but without much success. Their interpretation that water molecules close to the salt bridge take part in hydrogen bonding suggests once again that continuum predictions can lead to substantial errors when highly structured solvation shells exist around certain charged groups. Other instances when this may be a problem are near active sites of proteins, where water molecules are known to form part of the aqueous hydrogen-bond network [37].

A simple electrostatic argument can also be used to analyze the effect of the artificial boundary on the reaction field felt by charged groups on the surface of a protein. Consider just two charges, say both of charge +1 au, situated at a distance $a$ from the center of a dielectric sphere of dielectric constant $\epsilon_1$ embedded into a continuum of dielectric constant $\epsilon_2 > \epsilon_1$. As the charges approach the dielectric boundary, of radius $b > a$, the effect of the boundary on the electrostatic energy becomes more and more pronounced. Solving the electrostatic problem one finds for the interaction energy of two +1 au charges the expression

$$V = \frac{332}{\epsilon_1 a} \sum_l P_l(\cos\theta) \times \left[1 - \left(\frac{a}{b}\right)^{2l+1} \frac{(l+1)(\epsilon_2 - \epsilon_1)}{(l+1)\epsilon_2 + l\epsilon_1}\right], \quad (1)$$

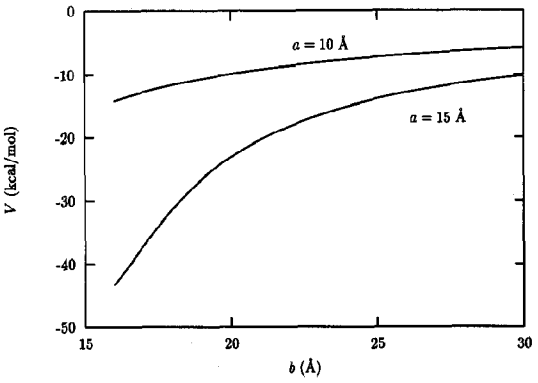


Fig. 1. Reaction-field contribution to the electrostatic energy of two +1 au charges located 5 Å apart at a distance $a$ from the origin as a function of the radius $b$ of the dielectric boundary (assumed spherical). The upper curve corresponds to $a = 10$ Å while the lower corresponds to $a = 15$ Å. The interior dielectric constant was taken to be equal to 4 and the exterior to 78. At a distance of 5 Å with $\epsilon = 4$ the direct Coulomb interaction energy is equal to 16.6 kcal/mol.

where $P_l(x)$ is the Legendre polynomial of order $l$ and $\theta$ is related to the distance $d$ between the charges by the equation

$$2a(1 - \cos\theta) = d. \tag{2}$$

The first term inside the brackets gives the interaction energy in the absence of the dielectric boundary, and the second gives the contribution from the reaction field. Fig. 1 shows this contribution for two different choices of $a$ as a function of the dielectric boundary radius $b$ (the distance between the charges was fixed at 5 Å; the inner dielectric constant was set to 4 and the outer to 78, as is customary in the CM approach). Since the dielectric boundary is not physical but rather an artifact of the CM approach the results should not depend strongly on the position of this boundary [38]. For groups that are buried inside a protein the typical effect of the boundary would be as in the upper curve ($a = 10$ Å or less) and it is clear from the figure that the position of the boundary does not have a great effect in this case. However, for groups on the surface the situation is more like that depicted in the second curve; since the boundary cannot be chosen too far away from the surface of the protein (otherwise the effect of the first water shells would not be taken into account), the choice of its position will have a large effect on the interaction energy. For instance, moving the boundary from 16 to 16.5 Å changes the energy by 3.32 kcal/mol. In comparison, the interaction energy in the absence of the boundary is about 16.6 kcal/mol.

CM, as a particular example of a linear response model, predicts a quadratic dependence on the ionic charges of the free energy, $F$. To test the range of validity of this assumption it is helpful to use a relation, derived from microscopic considerations, between the derivative of $F$ with respect to the charges and the mean value and fluctuations of the electrostatic potentials at the positions of the ions. Looking at the derivative instead of $F$ is convenient because it amplifies the deviations; however, it should be borne in mind that large errors in $\partial F/\partial q_i$ do not necessarily imply a large error in $F$. Let $Z(\{q_i\})$ be the partition function of the collection of ions of charges $\{q_i\}$ immersed in a solvent. We assume that the ions are fixed in space and denote the coordinates of the solvent molecules by $\{X\}$. Then the partition function can be written as

$$Z(\{q_i\}) = \int \mathrm{d}\{X\}\, \exp\left[-\beta \sum_i q_i V_i(X)\right] \mathrm{e}^{-\beta H(X)}, \tag{3}$$

where $V_i(X)$ is the electrostatic potential produced by the solvent at the position of ion $i$ and $H(X)$ collects all other interaction terms. Writing

$$\ln Z(\{q_i\}) = -\beta F(\{q_i\}) \tag{4}$$

and taking the logarithmic derivative of $Z$ with respect to $q_i$ we find

$$\frac{\partial F}{\partial q_i} = \langle V_i \rangle. \tag{5}$$

CM gives a mean value $\langle V_i \rangle$ that is proportional to the charges thus giving rise to the relation

$$F = \tfrac{1}{2}E \tag{6}$$

between the free energy of charging the ions and the internal energy, $E$.

It is convenient to rewrite $Z$ in a way that does not explicitly mention the solvent coordinates. This can be done by replacing the integration

over $\{X\}$ by an integration over the values $\{v_i\}$ of the electrostatic potentials at the ions. To this end we introduce the integral

$$d\mu(\{v_i\}) \equiv \int_{V_k(X)=v_k} d\{X\}\ e^{-\beta H(X)}, \tag{7}$$

where the integral is taken over all solvent coordinates that satisfy the constraints $V_k(X) = v_k$ for all $k$. In terms of $d\mu$ we can write

$$\langle f(V_1,\ldots,V_n)\rangle = \frac{\int d\mu(\{v_i\}) f(v_1,\ldots,v_n)\ \exp\left(-\beta \sum_k q_k v_k\right)}{\int d\mu(\{v_i\})\ \exp\left(-\beta \sum_k q_k v_k\right)} \tag{8}$$

for any function $f(\ )$ that depends only on the values of the electrostatic potentials $\{V_i\}$ at the positions of the charges. From this it follows that

$$\frac{d\mu(\{v_i\})\ \exp\left(-\beta \sum_k q_k v_k\right)}{\int d\mu(\{v_i\})\ \exp\left(-\beta \sum_k q_k v_k\right)} \tag{9}$$

is the probability distribution of the values of the electrostatic potentials at these positions. Note that the effect of the solvent is completely embodied in $d\mu$ but the probability distribution depends on the values of the charges. In particular, using Eq. (5), we obtain the expression

$$\frac{\partial F}{\partial q_j}(\{q\}) = \frac{\int d\mu(\{v_i\}) v_j\ \exp\left(-\beta \sum_k q_k v_k\right)}{\int d\mu(\{v_i\})\ \exp\left(-\beta \sum_k q_k v_k\right)} = \langle v_j \rangle. \tag{10}$$

If we choose now a reference set of charges, $\{q_i^{\mathrm{ref}}\}$, and write $q_k = q_k^{\mathrm{ref}} + \delta q_k$ we can expand Eq. (10) to first order in $\{\delta q_i\}$ to obtain the relation, valid in a neighborhood of the reference charges,

$$\frac{\partial F}{\partial q_i}(\{q\}) = \langle v_i \rangle^{\mathrm{ref}} - \beta \sum_k \langle \delta v_i \delta v_k \rangle^{\mathrm{ref}} \left(q_k - q_k^{\mathrm{ref}}\right), \tag{11}$$

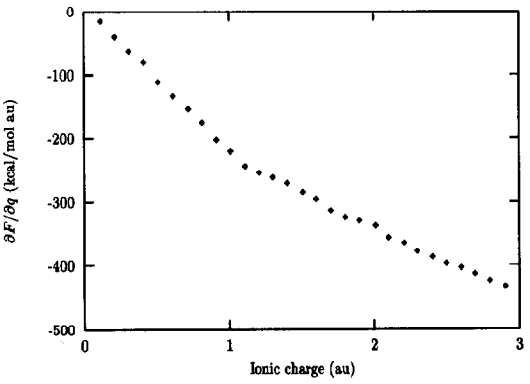


Fig. 2. Plot of the derivative of the free energy of charging one ion with respect to the ion's charge as a function of the charge. The data has been taken from [15].

where $\delta v_s = v_s - \langle v_s \rangle^{\mathrm{ref}}$ and the averages are taken in the reference state ($\delta q_k$ is expressed as $q_k - q_k^{\mathrm{ref}}$). Integrating this equation yields

$$F(\{q\}) = \text{const.} + \sum_i \langle v_i \rangle^{\mathrm{ref}} \left(q_i - q_i^{\mathrm{ref}}\right) - \frac{\beta}{2} \sum_{i,k} \langle \delta v_i \delta v_k \rangle^{\mathrm{ref}} \left(q_i - q_i^{\mathrm{ref}}\right) \times \left(q_k - q_k^{\mathrm{ref}}\right), \tag{12}$$

valid in a neighborhood of the reference state [16,39].

As Jayaram et al. [15] have shown, for an ion in aqueous solution $\langle V_i \rangle$ is well approximated by a piecewise linear function over a broad range of charges. Their Fig. 3, reproduced in part here as Fig. 2, shows that Eq. (11) is a good approximation but with different parameters for $q < 1.1$ au and for $q > 1.1$ au. They argued that therefore the simple CM approach is valid for the calculation of $pK_a$s since the first range ($q < 1.1$ au) follows closely the CM prediction and the change of regime does not occur for interesting charges. The breakdown at 1.1 au reflects an essential nonlinearity in the response function which, interestingly, gives rise to (at least) two regions of linearity. In their article they traced this breakdown to the sudden appearance of structure in the first solvation shell of the ion. Maroncelli and Fleming [34] found that the structure of the water around the solute is very dependent on solute

size. Significant structure is observed for their 'small' solute when the charge is just less than $\frac{1}{2}$ au.

Levy and co-workers [16] have studied the charging free energy of two ions separated by a distance of 10 Å under the assumption that Eq. (12) is valid over a broad range of charge values (the so-called Gaussian approximation). They find that this assumption works well but not for the complete charging process, from 0 to +1 au, implying that the response function must contain some nonlinearity before full charge is attained. For a system of two ions of charge $+q$ and $-q$ Eq. (11) becomes

$$\frac{\partial F}{\partial q}(\{q\}) = \langle v_+ - v_- \rangle^{\mathrm{ref}} - \beta(q - q^{\mathrm{ref}})\langle(\delta v_+ - \delta v_-)^2\rangle^{\mathrm{ref}}, \tag{13}$$

where $v_\pm$ are the electrostatic potentials at the position of the ions. Since Levy et al. [16] computed the parameters entering Eq. (13) for two different reference states, one with both ions uncharged and the other with charges +1 au and −1 au, we can estimate the crossover point by plotting Eq. (13) and finding the value of $q$ at which the two lines intersect (see Fig. 3). It is clear from this figure that nonlinearities in the response function become relevant well before the full charging process is complete.

These studies show that it is quite possible for a system of charges in aqueous solution to show deviations from linear response even for small charges. An interesting question is whether, or under which conditions, $\partial F/\partial q_i$ is piecewise linear. From Eq. (11) it can be seen that to ascertain the breakdown of linear response we do not need to know the free energies of solvation over a wide range of charges; it suffices to compare the linear dependencies of $\partial F/\partial q_i$ in small neighborhoods of some set of charge states chosen *a priori*. Of course, to identify the points where linear response breaks down the full free energy (or its derivative) must be known. Our group has taken

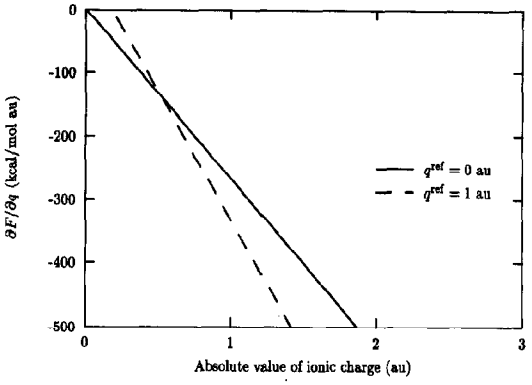


Fig. 3. Plot of the derivative of the free energy of charging two ions with respect to the absolute value of the ions' charge as a function of the charge. The data has been taken from [16]. The solid line corresponds to the reference state with both ions uncharged. The dashed line corresponds to the reference state with charges +1 au and −1 au.

a first step in this direction with a MM study of the $pK_a$s of ionizable groups in lysozyme [39].

Comparing Figs. 2 and 3 one can notice that for one ion the slope decreases in absolute value as the charge increases (a clear signal of dielectric saturation), whereas in the case of two ions it increases with increasing charge, at least in the region of charges that have been studied in [16]. It should be noted that this positive feedback ("anti-saturation") was also observed by Maroncelli and Fleming [34] in their study of a single ion in solution.

From Fig. 3 we can also estimate that the error in the calculation of the charging free energy, going from zero to full charge, when using the linear-response result (the $q^{\mathrm{ref}} = 0$ curve), is about 10–20%. Although this calculation pertains only to the particular system studied in ref. [16], it is an indication that nonlinear effects due to the molecular structure of the solvent are not negligible. Since MM preserves the nonlinear nature of the problem, it is well suited to investigate the magnitude of these effects and their system dependence. This should not be taken, however, to mean that MM and CM differ only in that the latter does not incorporate nonlinearities in the response function, or that the former would be of

value only in nonlinear regimes. In fact, it is not yet clear whether either approach reproduces the correct linear response regime, or whether many-body effects can be taken into account by simply renormalizing two-body interaction parameters.

In our group we are interested in the effects due to the solvent structure around a solute, with a strong emphasis on dielectric properties. In particular, we are focused on understanding the dielectric response of biomolecules in aqueous solutions at a microscopic level. Since electrostatic forces are at the heart of the question, their correct treatment is of the utmost importance.

Until recently MM simulations of biomolecules had to sacrifice the correct treatment of the electrostatic interactions to avoid unacceptably long computer runs. This sacrifice takes several forms: (a) the number of solvent molecules had to be kept small (typically about 3000), forcing the simulation box to be barely larger than the protein itself and producing spurious correlations between water molecules on opposite sides of the protein; and (b) truncating the electrostatic interactions at a finite distance instead of using the commonly accepted Ewald summation method to correctly incorporate the effect of periodic boundary conditions. Schreiber and Steinhauser [14,40] have given evidence that such a truncation can strongly affect the protein dynamics and it is known [41–43] that it affects the orientational correlations of solvent molecules.

Advances in MM algorithms [44–50] and the advent of massively parallel supercomputers are opening the way to large-scale simulations of biomolecules with explicit solvent. Our group is currently testing an implementation of the hierarchical multipole methods that runs on networks of workstations using PVM [51] as well as on a Thinking Machines' CM-5 using their native CMMD message-passing library. As part of our group's ongoing effort to study electrostatic effects on proteins we plan to apply it in the near future to the calculation of the p$K_a$s of lysozyme to investigate the breakdown of linear response as described above. We also plan to apply the program to the calculation of the charging free energy of ions in aqueous solutions, which provides a well-defined framework for the study of the structure of the first solvation shell.

This work was supported by an NIH grant (GM-30580), and by the Columbia University Center for Biomolecular Simulations (NIH P41RR06892).

## References

[1] M.K. Gilson, K.A. Sharp, and B. Honig, J. Comput. Chem. 9 (1987) 327.
[2] A.A. Rashin, Progr. Biophys. Mol. Biol. 60 (1993) 73.
[3] A. Rahman and F.H. Stillinger, J. Chem. Phys. 55 (1971) 3336.
[4] M.P. Allen and D.J. Tildesley, Computer simulation of liquids (Oxford Univ. Press, Oxford, 1991).
[5] S.T. Russell and A. Warshel, J. Mol. Biol. 185 (1985) 389.
[6] G. King, F.S. Lee and A. Warshel, J. Chem. Phys. 95 (1991) 4366.
[7] K. Sharp, J. Comput. Chem. 12 (1991) 454.
[8] A.H. Juffer, Ph.D. Thesis, Rijksuniversiteit Groningen (1993).
[9] H.E. Alper and R.M. Levy, J. Chem. Phys. 99 (1993) 9847.
[10] J.S. Bader and D. Chandler, J. Phys. Chem. 96 (1992) 6423.
[11] M. Neumann and O. Steinhauser, Chem. Phys. Letters 95 (1983) 417.
[12] M. Belhadj, H.E. Alper and R.M. Levy, Chem. Phys. Letters 179 (1991) 13.
[13] P.E. Smith and B.M. Pettitt, J. Chem. Phys. 95 (1991) 8430.
[14] H. Schreiber and O. Steinhauser, Biochemistry 31 (1992) 5856.
[15] B. Jayaram, R. Fine, K.A. Sharp and B. Honig, J. Phys. Chem. 93 (1989) 4320.
[16] R.M. Levy, M. Belhadj and D.B. Kitchen, J. Chem. Phys. 95 (1991) 3627.
[17] J.B. Foresman and C.L.I. Brooks, J. Chem. Phys. 87 (1987) 5892.
[18] D. Van Belle, I. Couplet, M. Prevost and S.J. Wodak, J. Mol. Biol. 198 (1987) 721.
[19] A. Wallqvist, Chem. Phys. Letters 165 (1990) 437.
[20] A. Wallqvist, P. Ahlström and G. Karlström, J. Phys. Chem. 94 (1990) 1649.
[21] B.T. Thole, Chem. Phys. 59 (1981) 341.
[22] J.A.C. Rullman and P.T. van Duijnen, Mol. Phys. 63 (1988) 451.
[23] M. Sprik and M.L. Klein, J. Chem. Phys. 89 (1988) 7556.
[24] J. Caldwell, L.X. Dang and P.A. Kollman, J. Am. Chem. Soc. 112 (1990) 9144.
[25] T.P. Straatsma and J.A. McCammon, Chem. Phys. Letters 177 (1991) 433.

[26] L.X. Dang, J.E. Rice, J. Caldwell and P.A. Kollman, J. Am. Chem. Soc. 113 (1991) 2481.
[27] J.W. Halley, J.R. Rustad and A. Rahman, J. Chem. Phys. 98 (1993) 4110.
[28] J. Cao and B.J. Berne, preprint.
[29] D.N. Bernardo, Y. Ding, K. Krogh-Jespersen and R.M. Levy, J. Phys. Chem. 98 (1994) 4180.
[30] K. Sharp, J.-C. Arald and B. Honig, J. Phys. Chem. 96 (1992) 3822.
[31] A.A. Rashin, J.R. Rabinowitz and J.R. Banfelder, J. Am. Chem. Soc. 112 (1990) 4133.
[32] D. Bashford and M. Karplus, Biochemistry 29 (1990) 10219.
[33] A.-S. Yang, M.R. Gunner, R. Sampogna, K. Sharp and B. Honig, Proteins 15 (1993) 252.
[34] M. Maroncelli and G.R. Fleming, J. Chem. Phys. 89 (1988) 5044.
[35] A. Warshel, F. Sussman and G. King, Biochemistry 25 (1986) 8368.
[36] A.-S. Yang and B. Honig, Curr. Opin. Struct. Biol. 2 (1992) 40.
[37] M.M. Teeter, Ann. Rev. Biophys. Chem. 20 (1991) 577.
[38] S.L. Chan and C. Lim, J. Phys. Chem. 98 (1994) 692.
[39] G.S. Del Buono, F.E. Figueirido and R.M. Levy, Intrinsic p$K_a$s of Ionizable Residues in Proteins: An Explicit Solvent Calculation for Lysozyme, Proteins, in press.
[40] H. Schreiber and O. Steinhauser, Chem. Phys. 168 (1992) 75.
[41] S.W. de Leeuw, J.W. Perram and E.R. Smith, Proc. Roy. Soc. A (1980) 27.
[42] S.W. de Leeuw, J.W. Perram and E.R. Smith, Proc. Roy. Soc. London A (1980) 57.
[43] D.J. Adams, Chem. Phys. Letters 62 (1979) 329.
[44] A.W. Appel, SIAM J. Sci. Stat. Comput. 6 (1985) 85.
[45] J. Barnes and P. Hut, Nature 324 (1986) 446.
[46] L. Greengard, The rapid evaluation of potential fields in particle systems (MIT Press, Cambridge, MA, 1988).
[47] J.A. Board, J.W. Causey, J.F.J. Leathrum, A. Windemuth and K. Schulten, Chem. Phys. Letters 198 (1992) 89.
[48] H.-Q. Ding, N. Karasawa and W.A. Goddard III, J. Chem. Phys. 97 (1992) 4309.
[49] F.S. Lee and A. Warshel, J. Chem. Phys. 97 (1992) 3100.
[50] F.E. Figueirido and R.M. Levy, J. Comput. Phys., submitted for publication.
[51] A. Beguelin, J. Dongarra, A. Geist, R. Mancheck and V. Sunderam, Technical Report No. ORNL/TM-11826, Mathematical Sciences Section, Oak Ridge National Laboratory, Oak Ridge, TN.