Linear Response Theory: An alternative to PB and GB methods for the analysis of molecular dynamics trajectories?

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · NOVEMBER 2004

Impact Factor: 2.63 · DOI: 10.1002/prot.20169 · Source: PubMed

6 AUTHORS, INCLUDING:

Xavier de la Cruz
VHIR Vall d'Hebron Research Institute
72 PUBLICATIONS **1,754** CITATIONS
SEE PROFILE

Tim Meyer
Universitätsmedizin Göttingen
25 PUBLICATIONS **387** CITATIONS
SEE PROFILE

Josep L Gelpí
University of Barcelona
101 PUBLICATIONS **2,993** CITATIONS
SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 57:458–467 (2004)

Linear Response Theory: An Alternative to PB and GB Methods for the Analysis of Molecular Dynamics Trajectories?

Antonio Morreale,[1] Xavier de la Cruz,[1,2] Tim Meyer,[1] Josep Lluís Gelpí,[1,3] F. Javier Luque,[4] and Modesto Orozco[1,3]*

[1]*Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Barcelona, Spain*
[2]*Institució Catalana per la Recerca i Estudis Avançats (ICREA), Barcelona, Spain*
[3]*Departament de Bioquímica i Biologia Molecular, Facultat de Química, Universitat de Barcelona, Barcelona, Spain*
[4]*Departament de Fisicoquímica, Facultat de Farmacia, Universitat de Barcelona, Barcelona. Spain*

ABSTRACT **We explore the use of classical Linear Response Theory (LRT) as an alternative strategy to the use of Molecular Mechanics/Poisson-Boltzmann strategies to compute the solvation free energy of macromolecules from molecular dynamics simulations using an explicit representation of solvent. The method reproduces well the free energy of solvation of standard amino acid side chains, small peptides, and proteins. The use of a fully discrete representation of solvent avoids the possible problems of continuum models to represent the solvation of systems containing tightly bound water molecules. Proteins 2004;57:458–467.**

INTRODUCTION

Since its first biochemical application in the mid-1970s,[1–3] molecular dynamics (MD) has become extremely powerful to explore the structure, flexibility, and recognition properties of nucleic acids and proteins[4–15] in environments very close to the physiological ones. Current computers and programs allow the simulation of proteins of medium size (100–200 residues) in the time scale around 10–50 ns. If the increase in the computer power continues in the near future, 1-μs MD simulations might be achieved as routine in a few years, thus widening the range of physiological processes that can be investigated with this technique.

The ability of MD simulations to gain insight into biological systems depends on factors such as (1) the quality of the force-field, (2) the precise definition of the simulation system, (3) the goodness of the simulation protocol, and (4) the length of the trajectory.[5,10,15] Another relevant issue is the quality of the data mining used to extract meaningful information from the huge trajectory file. MD trajectories are typically analyzed to derive average and time-dependent geometrical properties, which provide information about the structure, flexibility, and interaction profiles of the macromolecule. Additional information is obtained from the analysis of the ions and solvent molecules surrounding the macromolecule. Furthermore, processing of MD trajectories yields useful information on the stability of the macromolecule, which is determined by pseudo-free energy determined as shown in Eq. 1.

$$G = E - TS + G_{sol} \quad (1)$$

where E and S are the conformational energy and the entropy of the macromolecular system, and G_{sol} is its solvation free energy.

The conformational energy is the MD-averaged intramolecular energy computed by using the same force-field for the structures sampled along the trajectory. The conformational entropy uses to be calculated using pseudo-harmonic models based on the diagonalization of the mass-weighted covariance matrix.[16–19] Finally, the solvation contribution is determined from the set of MD-collected structures as the addition of steric and electrostatic contributions. The steric term is typically determined from empirical relationships with the solvent accessible surface (SA; 20), while the electrostatic component is determined from continuum solvation models (for discussion, see references[4,5,15]). The two most popular methods based on Eq. 1, denoted MM-PB/SA and MM-GB/SA, are characterized by the use of Poisson-Boltzman (PB)[15,21,22] and Generalized Born (GB)[15,23] models to compute the electrostatic component of the solvation free energy.[4,5,18,24–28]

MM-PB/SA and MM-GB/SA have been successful in the study of many macromolecular systems.[24–28] However, we cannot ignore their conceptual problems, which might limit their applicability and accuracy.[4,5,15] First, small differences in free energy are determined by subtracting large quantities. Second, the entropy analysis is difficult due to the slow convergence of this property. Third, continuum solvation models cannot capture structural macromolecule-water interactions, which might influence the conformational preferences. Many attempts have been

Grant sponsor: Spanish Ministry of Science; Grant numbers: SAF2002-04282 and BIO2003-06848.

*Correspondence to: Modesto Orozco, Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Parc Científic de Barcelona, Josep Samitier 1-5, Barcelona 08028, Spain. E-mail: modesto@mmb.pcb.ub.es

Received 31 December 2003; Accepted 5 March 2004

Published online 22 July 2004 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20169

made to reduce the uncertainties arising from the first two factors. For example, numerical uncertainties are reduced by running independent averages on large trajectories,[4,5,27,28] and time-convergence of entropy can be improved by using extrapolation techniques like that developed by Harris et al.[18] However, to our knowledge no attempts have been made to correct the shortcomings inherent to the use of continuum solvation models (see below).

In this report, we present the use of linear response theory (LRT)[29] for the analysis of the MD trajectories. The suitability of the method is discussed from the comparison with the results obtained from standard MM-PB and MM-GB methods for several protein systems. It is out of the scope of this study to consider empirical approaches used for calculation of the steric term, and we simply assume here that this will depend on the solvent accessible surface and will be the same irrespective of the method used to compute ΔG_{ele}.

THEORETICAL BACKGROUND

The electrostatic component of the free energy of solvation (ΔG_{ele}) is defined as the reversible work necessary to generate the solute charge distribution in a molecular-shaped cavity surrounded by solvent (for reviews see references[15,29–31]). As the solute charge distribution is being created, the solvent is re-organized around the solute, thus generating a reaction field that interacts with the solute charge distribution. Therefore, ΔG_{ele} reflects the balance between the stabilizing contribution due to the solute-solvent interaction, and the work spent in perturbing the solvent. The rigorous calculation of ΔG_{ele} is difficult, but a reasonable approximation can be obtained by assuming linearity in the electrostatic response of the solvent to the generation of the solute charge distribution.[29] In the limit of a non-polarizable solute, linear response theory (LRT) allows ΔG_{ele} to be calculated from Eq. 2.

$$\Delta G_{ele} = \int_0^1 \lambda \sigma V = \frac{1}{2}[\lambda^2]_0^1 \sigma V = \frac{1}{2}\sigma V \qquad (2)$$

where λ is a coupling parameter that varies from 0 (pure solvent) to 1 (fully charged solute), σ is the solute charge distribution, and V is the solvent reaction potential.

LRT lies in the foundations of continuum solvation models.[29–33] In these methods, ΔG_{ele} is determined from the interaction between the solute charge distribution and the solvent reaction field generated by an infinite continuum dielectric (the solvent). The calculation of the solvent reaction potential requires solving the Poisson equation (Eq. 3), which can be performed analytically only for simple solute cavities and charge distributions. For more complex solutes, it can be solved numerically by means of finite-difference methods (for discussion, see references[15,21,22]). These methods, nevertheless, are slow and can be grid-dependent, which in practice limits their accuracy.

$$\nabla[\epsilon \nabla V(r)] = -4\pi\rho(r) \qquad (3)$$

where ϵ is the permittivity of the solvent and $\rho(r)$ stands for the charge density.

Poisson equation can be generalized for the introduction of ionic screening, leading to Poisson-Boltzmann equation (PB). In this report, we will use the simplest Poisson equation, but for coherence with standard use of these methods, we will not differentiate between the Poisson and Poisson-Boltzman methods.

The most popular alternative to finite-difference PB methods for the study of macromolecular systems is the Generalized Born (GB) model, where ΔG_{ele} is determined by using Eq. 4, an ad hoc equation that satisfies in the limits both Born's (Eq. 5) and Onsager's (Eq. 6) equations (for discussion, see references[15,23,30,31]).

$$\Delta G_{ele} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_i \sum_j \frac{Q_i Q_j}{\tilde{\delta}_{GB}} \qquad (4)$$

where $\tilde{\delta}_{GB} = (r_{ij}^2 + \alpha_i \alpha_j \exp(r_{ij}^2/4\alpha_i\alpha_j))$, with α being the effective Born radii.

$$\Delta G_{ele} = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\frac{Q^2}{\alpha} \qquad (5)$$

$$\Delta G_{ele} = -\frac{\epsilon - 1}{2\epsilon + 1}\frac{\mu^2}{\alpha^3} \qquad (6)$$

where Q is the solute charge, μ is the dipole, and α is the radius of the spherical cavity containing the solute.

Both PB and GB share many strengths and shortcomings. The two methods provide reasonably accurate estimates of ΔG_{ele}, which is determined including long-range contributions, an effect that is not properly accounted for in discrete simulations. However, as any other continuum method, the contributions to solvation arising from specific solvent molecules are not well described. Moreover, these methods rely on the arbitrary partition of the system into solute and solvent through the definition of a molecular-shaped cavity, whose size is crucial for the magnitude of ΔG_{ele}.[15,34,35] Furthermore, the selection of internal and external dielectric constants is also arbitrary, and no clear rules exist for the definition of internal dielectrics for a large and flexible solute like a protein.[36–38] Finally, since the use of a continuum model implies that the solvent equilibration is much faster than the solute movements, continuum methods are not suitable when strong temporal couplings between solute and solvent motions occur. From a practical point of view, additional problems arise from the numerical errors associated with PB and GB calculations, and the consistency between the force-field parameters used in the MD simulation and the parameters incorporated in PB and GB calculations.

In summary, despite their tremendous power, continuum models present some intrinsic shortcomings that make advisable the development of alternative methods based on discrete approaches. In fact, and quite interestingly, LRT provides a good framework for the use of

discrete models, since Eq. 2 can be rewritten as shown in Eq. 7.[15,39–50]

$$\Delta G_{ele} = \frac{1}{2}\left\langle \sum_{i,j} \frac{Q_i Q_j}{r_{ij}} \right\rangle \quad (7)$$

where the brackets mean that calculation is done using a Boltzmann ensemble of structures, and i and j stand for atoms in the solute and solvent molecules.

Eq. 7 can be easily implemented in MD, thus avoiding the need to carry out a posteriori (and time-consuming) analysis of the trajectory. As LRT is based on a discrete representation of solvent, problems derived from the neglect on specific solute-solvent interactions are avoided. Typically the same force-field used for the MD simulation is used for the calculation of ΔG_{ele}, which guarantees the coherence between solvation calculation and trajectory collection. Additional advantages are that the arbitrariness in the choice of the cavity size or the dielectric constants is avoided and specific solute-solvent interactions are fully considered. Finally, the temporal coupling between solute and solvent distributions is also explicitly accounted for. As a consequence, most of the arbitrary selections implicit in the use of continuum models are avoided. However, we cannot ignore the intrinsic problems of LRT. First, the method works well only in cases of a pure lineal response of the solvent, which cannot be guaranteed for complex solvents. Second, since ΔG_{ele} is computed from an ensemble of solvent and solute configuration instead of from only an ensemble of solute configurations (like in continuum models), the noise in the calculation is larger. Finally, we cannot ignore the problems derived from the use of a discrete non-polarizable force-field for representing both solute and solvent.

The use of LRT, in conjunction with an empirical solvent accessible term to represent steric contributions to solvation, has led to the development of Extended Linear Response methods[45–47] to estimate the solvation free energy of small solutes, to study drug-receptor interactions and even complex electron transfer processes.[39–54] Furthermore, the use of dual sets of charges representing the solute charge distribution in gas phase and solution has lead to the Generalized Linear Response Theory,[55] a method that (at least for small solutes) allows the explicit introduction of solute polarization effects.

Here we will explore the ability of discrete methods based on standard linear response theory to estimate the electrostatic component of hydration free energy of peptides and proteins from discrete MD simulations. Comparison is made with the estimates obtained from continuum classical PB, GB methods and quantum mechanical SCRF MST approaches.

SIMULATION SYSTEMS

The performance of MD-LRT methods to estimate ΔG_{ele} was examined for several test systems. First, we considered a series of small solutes representative of the aminoacidic side chains (ethane, methanol, dimethylthioether, methanethiol, benzene, indole, phenol, acetamide, imidazole, methylammonium and guanidinium cations, and, finally, acetate anion). Second, we determined both absolute and relative hydration free energies for the octa-alanine peptide in four canonical conformations: α, 3_{10}, π-helices, and the β strand. Finally, we examined six small soluble proteins (see Table I) representative of all α-proteins (1CEI, 4ICB), all β-proteins (1ARK, 1SRO), and α + β proteins (2GB1, 3CI2). Starting structures of these proteins were taken from Protein Data Bank (see Table I).

TABLE I. Proteins Used in the Study†

Protein name	PDB entry	No. of residues	No. of ions	Net charge
SH3 domain from human nebulin	1ark	60	0	−4
Colicin E7 immunity protein	1cei	85	0	−9
S1 RNA binding domain	1sro	76	0	1
Protein G immunoglobulin binding domain	2gb1	56	0	−4
Barley serine proteinase inhibitor 2	3ci2	64	0	0
Calcium binding protein	4icb	76	2	−3

†Table includes the PDB references, the number of residues, number of structural ions in the protein (ions tightly bound inside the protein), and net charge of the proteins assuming standard ionization state of aminoacidic side chains.

In all cases, LRT results were compared with those obtained using PB and GB methods. For the smallest systems, the accurate MST-HF/6-31G(d) SCRF method[15,30,35,56–61] was also considered as an external reference calculation that is not based on the use of un-polarizable classical force-fields.

SIMULATION DETAILS

Model Systems

Model systems were inserted in cubic boxes of water large enough to guarantee that the shortest distance between the solute and the edge of the box was larger than 12 Å. Systems were optimized, heated, and equilibrated for 200 ps and analyzed for solvation during an additional 500 ps run of unrestrained MD simulation. Snapshots used for solvation calculations were taken every 1 ps.

Octa-Alanine Peptides

Peptides were generated using canonical coordinates with the Biopolymer module of InsightII.[62] The N- and C-terminal residues of the peptides were capped with $COCH_3$ and $NHCH_3$ groups. The structures were inserted in boxes of water with the same size criteria used for the model systems. The positions of backbone atoms were fixed using a weak harmonic constraint (10 kcal/mol $Å^2$) to avoid conversion between helices. The optimized system was then heated and equilibrated for 0.5 ns, and analyzed for 1 ns of unrestrained MD simulation. The solvation analysis

was performed by considering all the snapshots collected (1) every 0.5 ps and (2) every 1 ps.

Proteins

The structures were built up using the crystallographic coordinates taken from the Protein Data Bank. Only structural ions (see Table I) were included in the calculations in their crystal positions, but all the structural waters in the crystal were removed. Additional ions needed to achieve electroneutrality were added using the titration module of cMIP[63] near all the charged groups not neutralized by the presence of another charged group in their boundaries. All the systems were then hydrated by using boxes containing from 3,204 to 4,737 water molecules, and then optimized, heated, and equilibrated for 1 ns of unrestrained MD simulation. MD trajectories were continued for 10 ns, and solvation was averaged during the last 8 ns with snapshots collected every 20 ps.

Simulation Protocol

All MD simulations were performed at a constant pressure and temperature (1 atm and 298 K) with an integration time step of 2 fs. SHAKE[64] was used to constrain all the bonds at their equilibrium distances. Periodic boundary conditions and the Particle Mesh Ewald[65] methods were used to treat long-range electrostatic effects. AMBER-95[66] and TIP3P[67] force-fields were used in all cases. For small systems, charges were obtained using the HF-6/31G(d) RESP strategy.[68] All the trajectories were performed using the AMBER 6.0 computer program and associated modules.[69,70]

PB Calculations

The linear form of the Poisson equation was solved numerically using the focusing approach. For the small systems, an outer grid of 0.6 Å was used in conjunction with a finer internal grid of 0.2 Å. For the octa-alanine peptide, calculations were performed with a grid spacing of 1.0 Å for the outer grid, and internal grids with spacings of 1.0, 0.4, and 0.2 Å using a standard focusing procedure. Finally, for the proteins the focusing procedure was used with internal grids of 0.4 and 1.0 Å and an external grid of 1.2 Å. The usual PARSE radii parameters[71] were used to define the solute-solvent boundary, with internal and external dielectric constants of 1 and 80 relative to vacuum, and an exclusion solvent radius of 1.4 Å. Calculations were performed using the MEAD computer program.[72]

GB Calculations

Generalized Born calculations were carried out using Tsui & Case version implemented in AMBER6.1, together with the default parameters for a hydrated trajectory,[73] with external and internal dielectric constants of 80 and 1 relative to vacuum.

LRT Calculations

Eq. 7 was used considering all the water molecules present in the central boxes of each simulation. To account for long-range electrostatic effects Onsager (neutral) or Born (charged) methods (Eqs. 5 and 6) were used with an effective radii α equal to $(0.239Vol)^{1/3}$, with *Vol* being the size of the simulation box at every snapshot. Long-range effects make a small contribution for neutral solutes, but a sizeable contribution for heavily charged proteins like 1CEI.

TABLE II. Electrostatic Component (kcal/mol) of the Hydration Free Energy for the Model Systems Representative of Amino-acidic Side Chains Computed by Using Different Methods

Molecule	PB	GB	LRT	MST/6-31G(d)
CH_3CH_3	0.0	0.0	0.0	−0.1
CH_3OH	−10.8	−9.5	−8.0	−6.1
CH_3SCH_3	−3.9	−3.3	−2.3	−3.0
CH_3SH	−4.0	−3.6	−2.3	−3.2
Benzene	−5.6	−3.5	−3.1	−3.0
Indole	−12.6	−9.2	−7.8	−8.2
Phenol	−13.6	−13.8	−10.0	−7.7
CH_3CONH_2	−17.7	−13.0	−12.6	−11.9
Imidazole	−16.4	−11.5	−10.2	−11.9
$CH_3NH_3^+$	−83.3	−74.8	−58.9	−76.3
CH_3COO^-	−97.4	−84.0	−71.7	−74.3
Guanidinium$^+$	−71.2	−64.1	−52.4	−65.2

MST-HF/6-31G(d) Calculations

For the smallest systems ΔG_{ele} was also determined using a rigorous QM method, which includes specifically contributions to solvation arising from solute polarization under the solvent reaction field.[15,30,35,34,56–61] Calculations were performed using the gas phase optimized geometries (HF/6-31G(d)), and the standard MST parameters for solvation in water.[35]

RESULTS AND DISCUSSION

Model Systems

The electrostatic component (ΔG_{ele}) of the hydration free energy of a series of small molecules, representative of the protein backbone and side chains, was determined using PB, GB, LRT, and MST methods. For these small systems, the MST calculations provide very accurate estimates of the total free energy of hydration (root mean square deviation of 0.7 kcal/mol [0.5 kcal/mol if only neutral compounds are considered], correlation coefficient of 1.00 [0.99 for neutral solutes], and scaling coefficient $\Delta G_{hyd}(exp)/\Delta G_{hyd}$ [MST] deviating only 1% [3% for neutral solutes] from the ideal value). Furthermore, the parametrization of the MST method was done considering solvatochromic and tautomerization data, and was validated by comparison with discrete free energy perturbation calculations, which guarantees a good balance between steric and electrostatic terms. In other words, we can safely use MST values as very good approaches to the "real" electrostatic free energy of hydration.

There is a general agreement (see Table II) between the different estimates, which confirm that despite their theoretical differences all methods are providing a similar picture of hydration for the selected series. Thus, we found good correlations ($r = 0.99$ for all compounds; $r = 0.98$ only

TABLE III. Electrostatic Component (kcal/mol) of the Hydration Free Energy Computed for the Ala_8 Peptide in Four Canonical Conformations Using Different Solvation Models†

	PB(0.2)	PB(0.4)	PB(1.0)	GB	LRT
α-helix	−62(0.2)	−65(0.2)	−94(0.2)	−51(0.1)	−53(0.2)
3_{10}-helix	−56(0.1)	−59(0.1)	−89(0.2)	−44(0.1)	−47(0.2)
π-helix	−65(0.1)	−67(0.1)	−94(0.3)	−55(0.2)	−58(0.3)
β-strand	−87(0.2)	−93(0.2)	−141(0.2)	−63(0.1)	−72(0.2)

†PB calculations were performed using three-grid spacing (0.2, 0.4, and 1.0 Å). Standard errors in the averages (kcal/mol) are displayed in parentheses.

for neutral molecules) between PB and MST results. However, the scaling factors $c = \Delta G_{ele}^{PB}/\Delta G_{ele}^{MST}$ (1.14 for all the compounds; 1.45 only for neutral molecules) are far from unity, especially for neutral solutes. Note that no such systematic deviation is found for GB and LRT calculations, as noted in the respective scaling factors: $\Delta G_{ele}^{GB}/\Delta G_{ele}^{MST}$(1.09) and $\Delta G_{ele}^{LRT}/\Delta G_{ele}^{MST}$ (1.00). This finding suggests that current PB calculations might overestimate the solvent's electrostatic response (at least for small solutes), due probably to the use of a too small solute cavity. Therefore, we think that caution is required when PB calculations are used as the "best" continuum approach to solvation, since the ΔG_{ele} values might be too large (in absolute values) with the cavity definitions used here.

Finally, we should emphasize for the purposes of this report the good correlation between MST and LRT calculations ($r = 0.99$ for the entire set and $r = 0.96$ for the neutral compounds, with scaling factor equal to 1.0) is remarkable, and supports the quality of the LRT to reproduce the ΔG_{ele}, in agreement with previous studies by other authors.[39–52]

Peptide Models

The ΔG_{ele} values of the Ala_8 peptide in four canonical conformations determined from MD simulations using the standard PB approach (with three different grid spacings; see Methods), GB, and MD/LRT methods are given in Table III. Time profiles (data not shown) indicate that the ΔG_{ele} values are well converged along the production part of the MD trajectory. Furthermore, no significant difference was found in the results determined from the snapshots taken either at every 1 ps or every 0.5 ps. Therefore, Table III gives reliable averages of the ΔG_{ele} values for each conformation of the octapeptide.

The results in Table III show the strong dependence of PB results on the grid spacing. Clearly, the numerical errors due to the use of finite grids are reasonably small if the grid spacing is very narrow. Thus, when grid spacings of 0.4 and 0.2 Å are compared, results appear reasonably converged, with changes in the absolute ΔG_{ele} values of 3–6 kcal/mol (i.e., an error of 5–7%), no error in the relative ΔG_{ele} values of the helical conformations, and a difference around 3 kcal/mol in the relative ΔG_{ele} values of α and β structures. However, when a wider grid spacing of 1 Å is used, the deviation in the absolute ΔG_{ele} values from the converged values can be as large as 50 kcal/mol. Moreover, the relative ΔG_{ele} between α and β structures differ by around 100% compared to the values obtained with the finer grid. Even the relative ΔG_{ele} values for the helical structures are poorly reproduced by the larger grid spacing. Certainly, part of the grid-dependence might be corrected by more efficient smoothing algorithms,[74] but much caution is necessary unless a very fine grid is used.

The GB results correlate reasonably well ($r = 0.94$) with the best PB estimates. The relative ΔG_{ele} values between helical structures computed by GB and PB(0.2) calculations are very similar. A larger difference is, nevertheless, found in the relative ΔG_{ele} of the α and β structures: 24 (PB(0.2)) versus 11 (GB) kcal/mol. In absolute terms, the GB model gives ΔG_{ele} values around 30% smaller than the PB(0.2) results. However, and quite remarkably, both in absolute and in relative terms GB results are closer to the "best" PB continuum results (see Table III and Fig. 1) than PB values computed with a grid spacing of 1 Å.

LRT calculations correlate surprisingly well with the best PB results ($r = 0.99$), as well as with GB estimates ($r = 0.98$). The three methods provide the same relative order of preferential electrostatic hydration: β-strand>π-helix>α-helix>3_{10} helix. In fact, the discrepancy in the hydration of α and β structures computed from PB(0.2) and LRT methods is only 6 kcal/mol (Table III). The LRT ΔG_{ele} values are systematically smaller (around 19% in absolute values) than the PB(0.2) ones, but the differences are lower than those found between PB and GB results (see Table III and Fig. 1). In summary, there is a general good agreement between the "best" PB estimates and discrete LRT values. It is worth noting that the fact that LRT underestimates the magnitude of the solvent reaction field with respect to PB calculations is not an undesirable effect, since MST calculations (see above) suggest that PB (with current cavity definitions) overestimates $|\Delta G_{ele}|$.

Proteins

The MD trajectories of different proteins (1ARK, 1CEI, 1SRO, 2GB1, 4CI2, and 3ICB; see Methods) were stable and sampled structures very close to the experimental ones (data not shown). After the first ns, most energy terms were reasonably converged. However, the profiles for ΔG_{ele} (see Fig. 2) showed the occurrence of slow relaxation processes along the first 2 ns for some cases. Particularly, it is worth noting that in the first ns of the trajectories for 1ARK and 1SRO continuum methods detect a decay (in absolute value) in ΔG_{ele}, but the reverse trend is detected by LRT methods. This different behavior reflects the incomplete equilibration of the solvent around the protein at the beginning of the simulation, due probably to fast and local arrangements in the protein structure, which argues against the assumption of complete solvent equilibration implicit to continuum methods. Accordingly, for our purposes here only the last 8 ns of the trajectory, where all the methods give provide stable and similar time-dependent solvation profiles (Fig. 2), were used for further analysis.

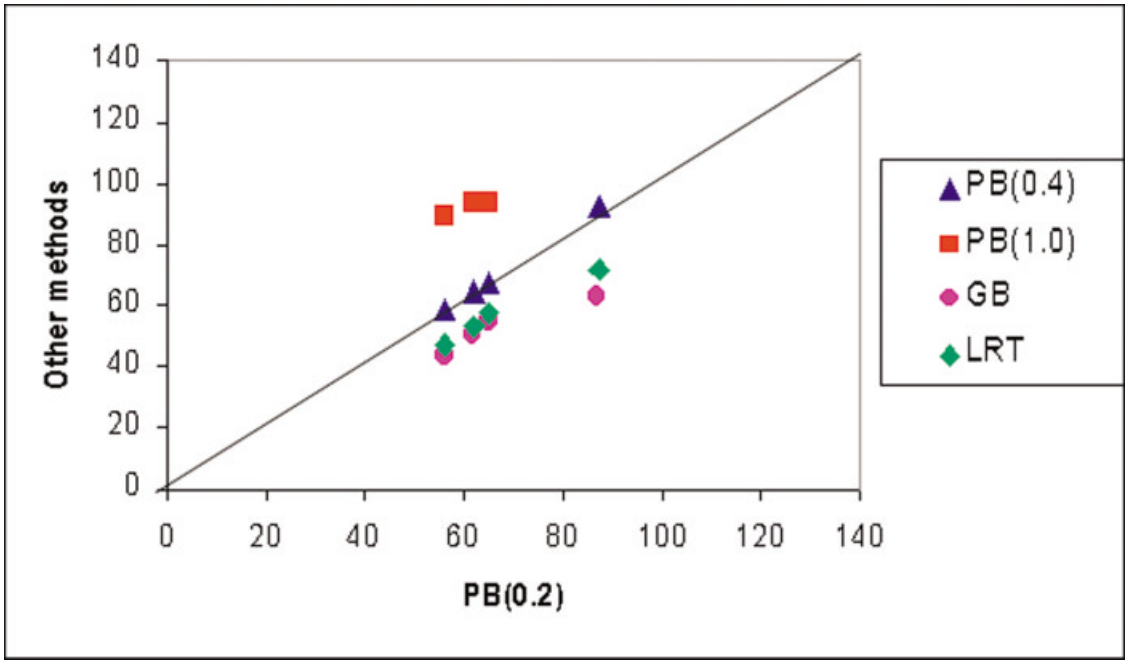

Fig. 1. Representation of MD-averaged ($-\Delta G_{ele}$) for different helical conformations (see text) computed with Poisson-Boltzmann (grid spacings of 1.0 and 0.4 Å), Generalized Born, and Linear Response theories with respect to the reference calculation (Poisson-Boltzmann with a grid spacing of 0.2 Å).

Inspection of Figure 2 also shows that the LRT profile exhibits larger-amplitude fluctuations compared to the continuum ones, leading to standard deviations in the average ΔG_{hyd} around 35% larger for those derived from continuum models. Again, this trend can be attributed to the simultaneous equilibration between solute and solvent, which is explicitly included in the LRT method. We must emphasize that this feature should not be considered an undesirable characteristic of the LRT calculations, but a meaningful physical property arising from the solvent relaxation time, which does not affect the accuracy in the determination of the average ΔG_{ele}.

The profiles in Figure 2 were created by taking snapshots every 10 ps in each trajectory, as is typically done in simulations of macromolecules, where PB calculations are rather expensive. To investigate the statistical quality of the averaged ΔG_{ele} (or ΔG_{hyd}) values, LRT calculations were repeated for the snapshots collected every 1 ps, but no significant differences were observed in the averaged values (and standard deviations), which vary 0–4 kcal/mol for the two different samplings (see Table IV).There is also agreement between the average values determined for 10 groups of 500 snapshots randomly selected along the trajectory (the maximum difference between sub-averages was only around 6 kcal/mol for total ΔG_{ele} 1,000–2,000 kcal/mol). Finally, standard errors in the averages are small even for the sets obtained using 10-ps spaced samplings, giving support to the statistical quality of the averages. For computational reasons, we cannot repeat these calculations for the more expensive continuum methods, but a similar behavior can be expected.

Comparison of the average ΔG_{ele} values computed from PB calculations (see Fig. 3) shows a marked dependence with the grid spacing, as already noted for the Ala_8 system. Thus, the change from 1.0 to 0.4 Å in the grid spacing changes the average ΔG_{ele} by around 400–600 kcal/mol (around 20–40%), a value similar to that obtained for the Ala_8 peptide. Based on the results shown in Table II, we can guess that the PB estimates for a grid spacing of 0.4 Å are only around 5% overestimated (in absolute terms) with respect to those obtained with a very fine (0.2 Å) grid, which is presumably close to the limit of accuracy of PB calculations. Therefore, the PB(0.4) values can be reasonably used as good estimates of the ΔG_{ele}.

GB and PB(0.4) correlate very well ($r > 0.99$), showing GB and PB provide very similar relative values (see Fig. 3). The GB method gives ΔG_{ele} values systematically smaller (in absolute terms) than the PB(0.4) calculations. The difference is small (around 15%), thus demonstrating a satisfactory agreement between continuum models. Our experience with other macromolecular systems confirms that in general (but unfortunately not always) PB and GB methods give similar solvation free energies,[28] a result that has been now extensively proved by Feig et al.[74] This general agreement, combined with the computational efficiency of GB method (see Table V and Fig. 3), gives support to its general use in the analysis of solvation in MD trajectories.

It is worth noting that, despite the differences in the computational approaches and the size and complexity of the systems studied here, LRT calculations provide very similar results to those derived by continuum models (see Fig. 3). Thus, LRT and GB ΔG_{ele} estimates are very close (deviation <2%) and well correlated ($r > 0.99$). LRT and PB(0.4) correlates also very well ($r > 0.99$), though the former method gives ΔG_{ele} around 16% smaller (in abso-

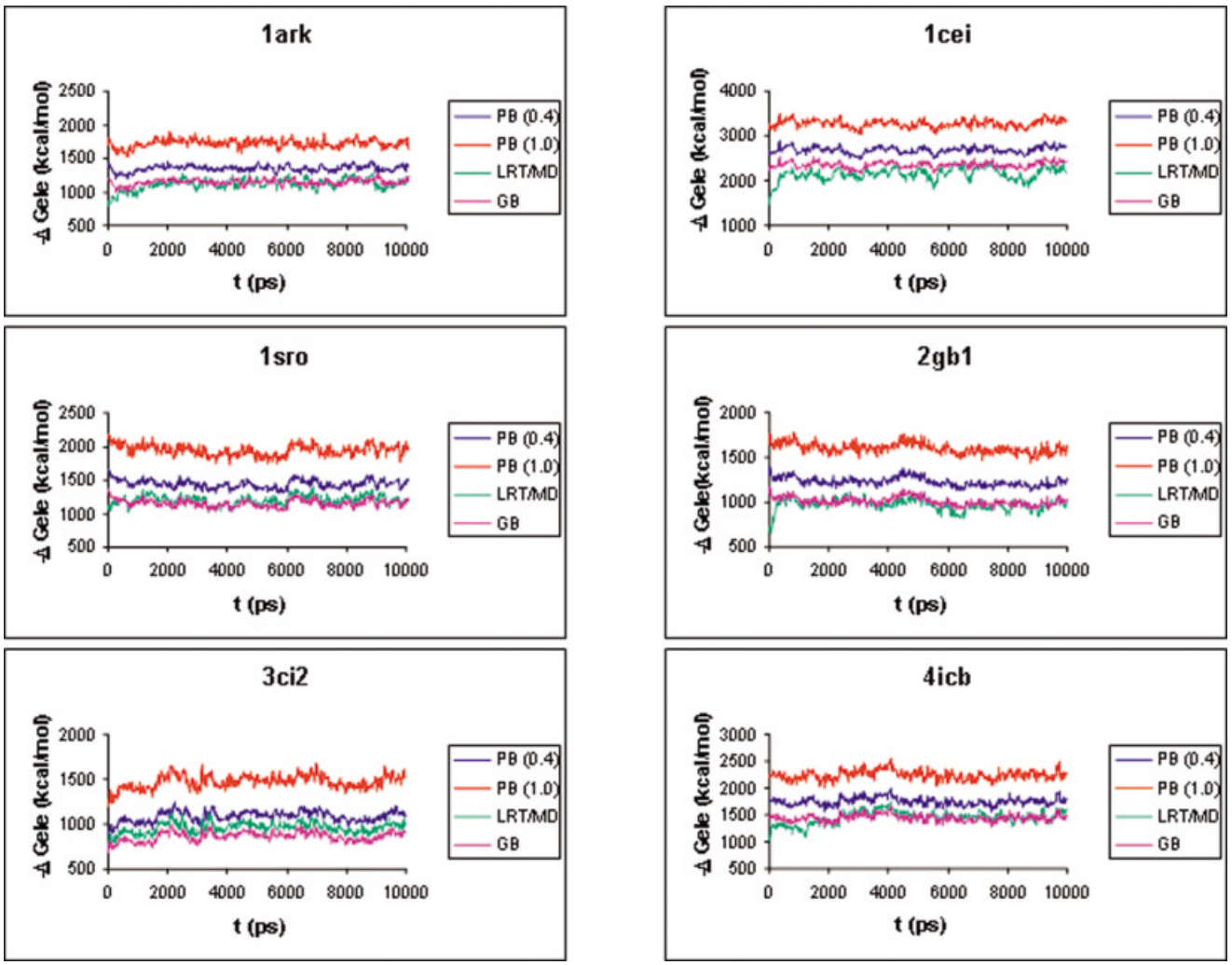

Figure 2.

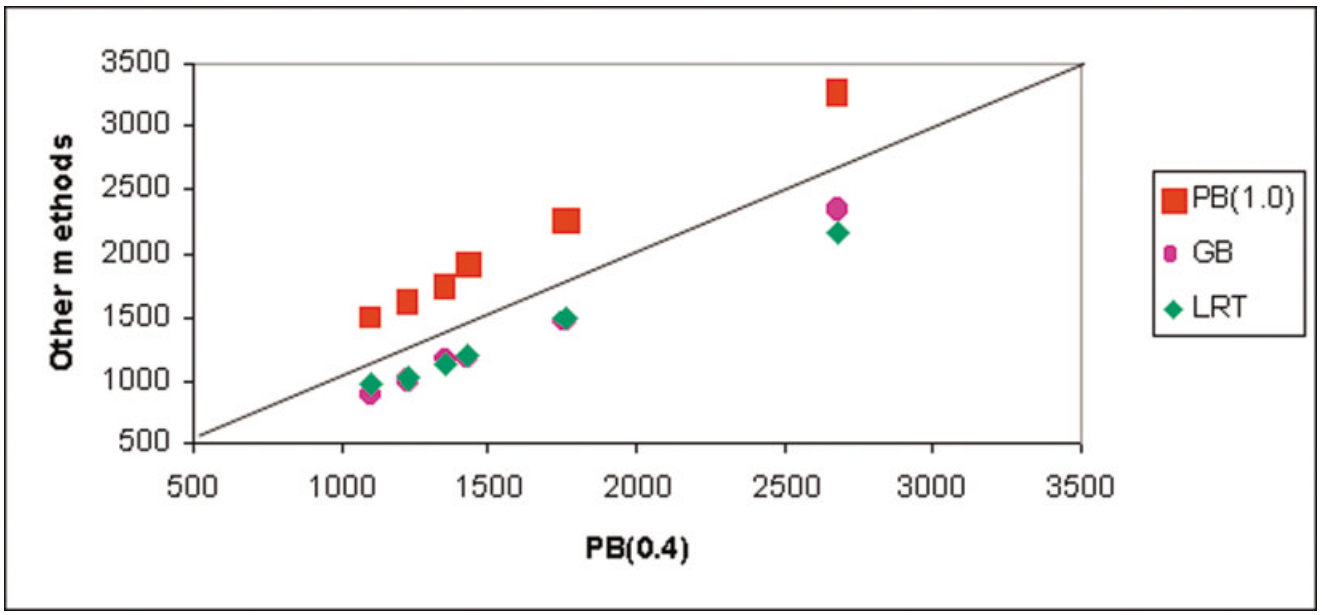

Figure 3.

TABLE IV. Electrostatic Component (kcal/mol) of the Hydration Free Energy Computed for 6 Representative Proteins Using Different Solvation Algorithm†

	PB(0.4)	PB(1.0)	GB	LRT
1ARK	−1,373(2)	−1,751(3)	−1,170(2)	−1,158(3)
				−1,160(1)
1CEI	−2,710(3)	−3,294(4)	−2,380(3)	−2,199(6)
				−2,195(1)
1SRO	−1,456(3)	−1,955(3)	−1,171(2)	−1,224(3)
				−1,225(1)
2GB1	−1,240(2)	−1,640(3)	−1,023(2)	−1,034(3)
				−1,034(1)
3CI2	−1,123(2)	−1,513(2)	−896(2)	−997(2)
				−998(1)
4ICB	−1,782(3)	−2,294(4)	−1,472(3)	−1,514(4)
				−1,515(1)

†PB calculations were performed using two grid spacings: 0.4 and 1.0 Å. Standard errors in the averages (kcal/mol) are displayed in parentheses. Numbers in italics correspond to averages obtained by collecting data every 1 ps instead of the usual time-step of 10 ps.

TABLE V. CPU Times (in seconds of an AMD 1.7 GHz computer) for the Calculation (with different methods) of the Electrostatic Component of the Solvation Free Energy for 100 Snapshots of the Six Proteins Considered Here†

	PB(0.4)/SA	PB(1.0)/SA	GB/SA
1ARK	17,800	12,700	145
1CEI	24,800	19,100	215
1SRO	22,300	17,600	201
2GB1	16,300	12,800	130
3CI2	6,400	4,100	98
4ICB	14,600	9,900	111

†Note that in all the cases, the cost of LRT calculations is negligible since the data is directly determined along the MD simulation.

lute terms) than the PB(0.4) ones. Let us note that this implies around 10% difference with respect to the "optimum" PB estimates (see above). In summary, discrete the LRT method provides close values to those obtained using the "best" continuum model, but in principle at no computational cost, since solute-solvent interaction energies can be collected during the trajectory. Finally, we should remember that, as noted above, the fact that LRT provides ΔG_{ele} values slightly smaller (in absolute terms) than those determined by PB calculations cannot be considered an undesirable effect.

As described in previous works,[15] and contrary to common beliefs, continuum and discrete methods share a common physical framework, and a general agreement in the results obtained from the two types of approaches is not surprising. However, we must admit that we did not expect such a good agreement for huge and heavily charged macromolecules (for example, the net charge of 1CEI is −9!). The agreement between PB, GB, and LRT provides a combination of techniques to determine solvation free energies from MD samplings. A routine procedure can be to repeat the three types of analysis, averaging values to reduce the global uncertainty. We should note, however, that while in many cases the three methods would provide results of similar quality, especially in relative terms, there are cases where discrete LRT-type calculations seems to be the best choice. A typical example is when a few water molecules are strongly bound to the macromolecules or trapped inside cavities. Another example is when the macromolecule is not equilibrated, but changes its conformation along the simulation (for example, in folding simulations). In both cases, there is a breakdown on the basic hypothesis on which continuum models are founded, and the use of continuum methods might not be appropriate.

Fig. 2. Variation of ($-\Delta G_{ele}$) along the collection part of the trajectory of the six proteins considered in the study. Calculation is performed with Poisson Boltzmann (grid spacings of 1.0 and 0.4 Å), Generalized Born, and Linear Response theories.

Fig. 3. Representation of MD-averaged ($-\Delta G_{ele}$) for the six proteins computed with Poisson-Boltzmann (grid spacings of 1.0 Å), Generalized Born, and Linear Response theories with respect to the reference Poisson-Boltzmann (grid spacings of 0.4 Å) calculation.

ACKNOWLEDGMENTS

We are indebted to D. Bashford for a copy of MEAD and help in its practical use. We acknowledge Prof. Tomasi for the original version of his MST code, which was later modified in our laboratory. We thank the CIRI-CEPBA Consortium for computational facilities. A.M. is a fellow from the Spanish Ministry of Science and Technology.

REFERENCES

1. Warshel A. Bicycle-pedal model for the first step in the vision process. Nature 1976;260:679–683.
2. McCammon JA. Molecular Dynamics Study of the Bovine Pancreatic Trypsin Inhibitor, in "Report of the 1976 Workshop, Models for Protein Dynamics" Berendsen HJC, editor. Models for protein dynamics. CECAM, Université de Paris IX Orsay, France; 1976. p 137.
3. McCammon JA, Gelin BR, Karplus M. Dynamics of folded proteins. Nature 1977;267:585–590.
4. Orozco M, Rueda M, Blas JR, Cubero E, Luque FJ, Laughton CA. Molecular dynamic simulations of nucleic acids. In: Encyclopedia of Computational Chemistry. New York: Wiley (in press).
5. Orozco M, Pérez A, Noy A, Luque FJ. Theoretical Methods for the Simulation of Nucleic Acids. Chem Soc Rev 2003; 32:350–364
6. McCammon JA, Harvey SC. Dynamics of proteins and nucleic acids. New York: Cambridge University Press; 1987.
7. Wang W, Donini O, Reyes CM, Kollman PA. Biomolecular simulations: recent developments in force-fields simulations of enzyme catalysis protein-ligand protein-protein and protein-nucleic acid noncovalent interactions. Annu Rev Biophys Biomol Struct 2001; 30:211–243.
8. Kollman PA. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. Acc Chem Res 1996;29:461–469.
9. Cheatham TE, Kollman PA. Molecular dynamics simulation of nucleic acids. Ann Rev Phys Chem 2000;51:435–471.
10. Van Gunsteren WF, Mark AE. On the interpretation of biochemical data by molecular dynamics computer simulation. Eur J Biochem 1992;204:947–961.
11. Van Gunsteren WF, Berendsen HJ. Computer simulation of molecular dynamics: methodology applications and perspectives in chemistry. Angew Che Int Ed Engl 1990;29:992–1023.
12. Karplus M, Petsko GA. Molecular dynamics simulations in biology. Nature 1990;347:631–639.
13. Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature Struct Biol 2002;9:646–652.
14. Warshel A. Computer modeling of chemical reactions in enzymes and solutions. New York: Wiley; 1991.
15. Orozco M, Luque FJ. Theoretical methods for the description of

the solvent effect in biomolecular systems. Chem Rev 2000;100: 4187–4225.
16. Andricioaei I, Karplus M. On the calculation of entropy from covariance matrices of the atomic fluctuations. J Chem Phys 2001;115:6289–6292.
17. Schlitter J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. Chem Phys Let 1993;215:617–621.
18. Harris SA, Gavathiotis E, Searle MS, Orozco M, Laughton CA. Cooperativity in drug-DNA recognition: a molecular dynamics study. J Am Chem Soc 2001;123:12658–12663.
19. Schafer H, Mark AE, van Gunsteren WF. Absolute entropies from molecular dynamics simulation trajectories. J Chem Phys 2000; 113:7809–7817.
20. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combination of pairwise overlaps (LCPO). J Comp Chem 1999;20:217–230.
21. Gilson MK, Honig B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies binding energies and conformational analysis. Proteins 1988;4:7–18.
22. Davis ME, McCammon JA. Electrostatics in biomolecular structure and dynamics. Chem Rev 1990;90:509–521.
23. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990; 112:6127–6129.
24. Srinivasan J, Cheatham TE, Kollman PA, Case DA. Continuum solvent studies of the stability of DNA RNA and phosphoramidate-DNA helices. J Am Chem Soc 1998;120:9401–9409.
25. Lee MR, Duan Y, Kollman PA. Use of MM/PB/SA in estimating the free energies of proteins: application to native intermediates and unfolded vilin headpiece. Proteins 2000;39:309–316.
26. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, Lee M, Lee T, Duan Y, Wang W, Donini O, Cieplak P, Srinivasan J, Case DA, Cheatham TE. Calculating the structure and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 2000;33:889–897.
27. Cubero E, Luque FJ, Orozco M. Theoretical studies of d(A:T)-based parallel-stranded DNA duplexes. J Am Chem Soc 2001;123: 12018–12025.
28. Cubero E, Abrescia NGA, Subirana JA, Luque FJ, Orozco M. Theoretical study of a new DNA structure: the antiparallel Hoogsteen duplex. J Am Chem Soc 2003;125:14603–14612.
29. Böttcher CJF. Theory of electrostatic polarisation. Amsterdam: Elsevier; 1952.
30. Tomasi J, Persico M. Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. Chem Rev 1994;94: 2027–2094.
31. Cramer CJ, Truhlar DF. Implicit solvation models: equilibria structure spectra and dynamics. Chem Rev 1999;99:2161–2200.
32. Luque FJ, Curutchet C, Muñoz-Muriedas J, Bidon-Chanal A, Soteras I, Morreale A , Gelpí JL, Orozco M. Continuum solvation models: Dissecting the free energy of solvation. Phys Chem Chem Phys 2003;5:3827–3836.
33. Rivail JL, Rinaldi D. In: Liquid-State Quantum Chemistry: Computational Applications of the Polarizable Continuum Models, Leszczynski J, editor. Computational chemistry reviews of current trends. Singapore: World Scientific; 1995. p 139.
34. Curutchet C, Cramer CJ, Truhlar DG, Ruiz-López MF, Rinaldi D, Orozco M, Luque FJ. Electrostatic component of solvation comparison of SCRF continuum models. J Comp Chem 2003;24:284–297.
35. Curutchet C, Orozco M, Luque FJ. Solvation in octanol: parametrization of the continuum MST model. J Comp Chem 2001;22:1180–1193.
36. Warshel J, Aqvist J. Electrostatic energy and macromolecular function. Annu Rev Biophys Biophys Chem 1991;20:267–298.
37. Schutz CN, Warshel A. What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins 2001; 44:400–417.
38. Warshel A, Russell ST, Churg AK. Macroscopic models for studies of electrostatic interactions in proteins: limitations and applicability. Proc Natl Acad Sci USA 1984;81:4785–4789.
39. Lee FS, Chu ZT, Warshel A. Microscopic and semimicroscopic calculations of electrostatic energies in proteins by the POLARIS and ENZYMIX programs. J Comp Chem 1993;14:161–185.
40. Lee FS, Chu ZT, Bolger MB, Warshel A. Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. Prot Eng 1992;5:215–228.
41. Warshel A, Rusell ST. Calculations of electrostatic interactions in biological systems and in solutions. Q Rev Biophys 1984;17:283–422.
42. Roux B, Yu HA, Karplus M. Molecular basis for the Born model of ion solvation. J Phys Chem 1990;94:4683–4688.
43. King U, Warshel A. Investigation of the free-energy functions for electron-transfer reactions. J Chem Phys 1990;93:8682–8692.
44. Kong YS, Warshel A. Linear free energy relationships with quantum mechanical corrections: classical and quantum mechanical rate constants for hydride transfer between NAD+ analogs in solutions. J Am Chem Soc 1995;117:6234–6242.
45. Aqvist J. Ion water interaction potentials derived from free-energy perturbation simulations. J Phys Chem 1990;94:8021–8024.
46. Aqvist J, Medina C, Samuelsson JE. A new method for predicting binding affinity in computer-aided drug design. Prot Eng 1994;7: 385–391.
47. Aqvist J, Hansson T. Validity of electrostatic linear response in polar solvents. J Phys Chem 1996;100:9512–9521.
48. Orozco M, Luque FJ, Habibolahzadeh D, Gao J. The polarization contribution to the free energy of hydration. J Chem Phys 1995;102:6145–6152.
49. Carlson HA, Jorgensen WL. An extended linear response method for determining free energies of hydration. J Phys Chem 1995;99: 10667–10673.
50. Pierce AC, Jorgensen WL. Estimation of binding affinities for selective thrombin inhibitors via Monte Carlo simulations. J Med Chem 2001;44:1043–1050.
51. Wesolowski SS, Jorgensen WL. Estimation of binding affinities for celecoxib analogues with COX-2 via Monte Carlo-extended linear response. Bioorg Med Chem Lett 2002;12:267–270.
52. Levy R.M, Belhadj M, Kitchen, D.B. Gaussian fluctuation formula for electrostatic free-energy changes in solution. J Chem Phys 1991;3627–3633.
53. Kuharsi R.A, Bader J.S, Chandler D, Sprik M, Klein M, Impey R.W. Molecular model for aqueous ferrous-ferric electron transfer. J Chem Phys 1991;3627–3633.
54. Ulstruo, J. Charge transfer processes in condensed media. New York: Springer; 1979.
55. Orozco M, Luque FJ. Generalized Linear Response Approximation in Discrete Methods. Chem Phys Lett 1997;265:473–480
56. Miertus S, Tomasi J. Approximate evaluations of the electrostatic free energy and internal energy changes in solution processes. Chem Phys 1982;65:239–245
57. Miertus S, Scrocco E, Tomasi J. Electrostatic interaction of a solute with a continuum. A direct utilization of ab initio molecular potentials for the prevision of solvent effects. Chem Phys 1981;55: 117–129
58. Bachs M, Luque FJ, Orozco M. Optimization of solute cavities and van der Waals parameters in ab initio MST-SCRF calculations of neutral molecules. J Comput Chem 1994;15:446–454
59. Orozco M, Bachs M, Luque FJ. Development of optimized MST/SCRF methods for semiempirical calculations: the MNDO and PM3 Hamiltonians. J Comput Chem 1995;16:563–575
60. Luque FJ, Zhang Y, Aleman C, Bachs M, Gao J, Orozco M. Solvent Effects in Chloroform Solution: parametrization of the MST/SCRF Continuum Model. J Phys Chem 1996;100:4269–4276
61. Luque FJ, Alemán C, Bachs M, Orozco M. Extension of MST/SCRF method to organic solvents: ab initio and semiempirical parametrization for neutral solutes in CCl4. J Comput Chem 1996;17:806–820
62. Insight II Computer Program Accelrys Co., San Diego, CA. 2003
63. Gelpi JL, Kalko SG, Barril X, Cirera J, de la Cruz X, Luque FJ, Orozco M. Classical molecular interaction potentials: improved setup procedure in molecular dynamics simulations of proteins. Proteins 2001;45:428–437
64. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comput Phys 1977;23:327–341
65. Darden TA, York DM, Pedersen LG. Particle mesh Ewald: an N×log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089–10092
66. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Fergurson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second

generation force field for the simulation of proteins nucleic acids and organic molecules. J Am Chem Soc 1995;117:5179–5197
67. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926–935
68. Bayly CE, Cieplak P, Cornell WD, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. J Phys Chem 1993;97: 10269–10280
69. Case DA, Pearlman DA, Caldwell JW, Cheatham TE III, Ross WS, Simmerling CL, Darden TL, Marz KM, Stanton RV, Cheng AL, Vincent JJ, Crowley M, Tsui V, Radmer RJ, Duan Y, Pitera J, Massova I, Seibel GL, Singh UC, Weiner PK, Kollman PA. AMBER6. University of California, San Francisco. 1999
70. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S, Fergurson D, Seibel G, Kollman PA. AMBER a package of computer programs for applying molecular mechanics normal mode analysis molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. Computer Physics Communications 1995;91:1–41
71. Weise J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). J Comp Chem 1999;20:217–230
72. Bashford D, Gerwert K. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. J Mol Biol 1992; 224:473–496
73. Tsui V, Case D. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. J Am Chem Soc 2000;122: 2489–2498.
74. Feig M, Onufriev A, Lee MS, Im W, Case DA, Brooks CL III. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. J Comput Chem 2004;25:265–284.