Guimãres Werneck, Cristiano Ruch; Barreiro, Gabriela; Oliveira Fernandes de, César Augusto;
Alencastro Bicca de, Ricardo

On the application of simple explicit water models to the simulations of biomolecules

On the Application of Simple Explicit Water Models to the Simulations of Biomolecules

Cristiano Ruch Werneck Guimarães,‡ Gabriela Barreiro,** César Augusto Fernandes de Oliveira†,
and Ricardo Bicca de Alencastro†

†*Physical Organic Chemistry Group, Departamento de Química Orgânica,*

Instituto de Química, Universidade Federal do Rio de Janeiro, Cidade Universitária,

CT, Bloco A, lab. 609, Rio de Janeiro, RJ 21949-900, Brazil

****Chemistry Department, Wesleyan University, Hall-Atwater Laboratory, Middletown, CT 06459, USA*

‡*Department of Chemistry, Yale University, 225 Prospect St., New Haven, CT 06520, USA*

Computer simulations of biomolecular systems have achieved a significant importance in science as they provide information regarding structure, dynamics, and energetics of biomolecules that are inaccessible to experimental measuring techniques. In this work, some important aspects of the simulation of biomolecular systems are described. An overview of the most popular protein force fields, simple explicit water models for the simulation of liquid water, and different approaches to treat the boundaries of the system is presented. Also, studies conducted in our group illustrating successful simulations of three biomolecules (thrombin, L-type calcium channel and human Cytomegalovirus protease) through the application of simple explicit water models combined with protein force fields are discussed.

1   Introduction

Computer simulations enable us to study biomolecular systems and predict their properties through the use of techniques that consider small replications of the macroscopic system with manageable number of atoms. Molecular dynamics (MD) [1] or Monte Carlo (MC)[2] simulations generate representative configurations of these small replications in such a way that accurate values of structural and thermodynamic properties, such as energy, volume, pressure, enthalpy, entropy, and free energy, can be obtained with feasible amount of computation [3].

In recent decades, computer simulations of biomolecular systems achieved a significant importance in science. This may be attributed to the rapid increase of computer power [1]. For instance, relatively inexpensive Pentium-based computers can now compete with workstations for MD and MC simulations of organic and bioorganic systems [4]. In addition, there are severe limitations of experimental measuring techniques on the characterization of structure, dynamics, and energetics of biomolecules. The X-ray crystallography and NMR spectroscopy are restricted to the structural determination of more rigid molecules. Spectroscopy measuring techniques provide information with respect to dynamics, but are limited to special groups of atoms in a biomolecule [5]. For a review on the application of these techniques in the study of hydration of biomolecules,

see the works of Pal *et al.*[6] and Schoenborn *et al* [7]. As for the energetics, experimental determination of different atomic interaction energies contributing to the total energy of a biomolecular system is almost impossible [8]. In this way, computer simulations and experimental measurements become complementary tools to study biomolecules. The first one provides a detailed atomic picture at a resolution in space, energy or time that is generally inaccessible by experimental means, whereas the second provides the necessary restriction of the configurational space that has to be sufficiently sampled in a simulation.

2   Current force fields for biomolecular systems

Biomolecular systems are very complex to allow for solving the Schrödinger's equation. In other words, the system has too many degrees of freedom, particularly electronic degrees of freedom, to be simulated. Thus, only the degrees of freedom that are essential to a proper representation of the quantity to be calculated or phenomenon to be studied must be explicitly treated in the molecular model. If the quantities or phenomena do not depend explicitly upon the electronic distribution in a molecule, an effective interaction function describing the interactions between the atoms in the molecular model and taking into account the effect of electronic de-

grees of freedom in an average manner, also known as force field, may be applied. A typical force field for biomolecular systems has the following form [9].

$$E_{total} = \sum_{\text{bonds}} K_r(r - r_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2$$

$$+ \sum_{\text{torsions}} \frac{V_n}{2} \left[1 \pm \cos(n\phi - \gamma) \right]$$

$$+ \sum_{\text{non-bonded}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} + \frac{q_i q_j}{r_{ij}} \right] \tag{1}$$

In eq. 1, the total energy of a biomolecule is represented as a sum of energy terms for bonds, angles, torsions, and inter- and intramolecular non-bonded interactions. Bond stretching and angle bending are treated as simple harmonic motions, with force constants, K, and equilibrium bond lengths (r_0) and angles (θ_0). Rotations around bonds are functions of the energy barrier to rotation, V_n, and the dihedral angle, ϕ, and are described by a truncated Fourier series. Interactions between pairs of atoms (i, j) separated by more than two bonds are indicated as a sum of van der Waals and electrostatic energies, given by the Lennard-Jones (parameters A_{ij} and C_{ij}) and Coulombic (atomic partial charges, q) functional forms.

Although most of the force fields for biomolecular systems have a similar functional form, they may differ in the way the parameters are derived. Force field parameters may be derived from *ab initio* quantum-mechanical calculations involving electronic degrees of freedom for small molecular clusters. Alternatively, force field parameters may be fit to experimental data such as crystal structure, energy and lattice dynamics, infrared, X-ray data on small molecules, liquid properties like density and enthalpy of vaporization, free energy of solvation, NMR data, etc.[8] OPLS,[10] CHARMM,[11] GROMOS,[12] AMBER[13] and CVFF[14] are examples of the most popular force fields and have been critically evaluated [15-19].

If the biomolecular system cannot be properly described by a force field because it involves large electron density redistributions or the breaking or formation of chemical bonds, quantum mechanics/molecular mechanics (QM/MM) methods must be applied [20-22]. QM/MM methods treat all the atoms that are directly involved in the chemical process at the QM level while the rest of the system is described by a force field.

3 Force field models for the simulation of liquid water

Water is a highly polarizable molecule. For instance, Canuto and coworkers used sequential MC/QM calculations with different quantum chemical methods to calculate the induced dipole moment of liquid water [23,24]. The authors obtained a statistically converged value of 0.74 ± 0.14 D, which corresponds to a dipole moment of liquid water of 2.60 ± 0.14 D, in excellent agreement with the value derived from the dieletric constant and previous theoretical estimates. In addition, water has a great ability to form hydrogen bonds, both as a hydrogen bond donor and acceptor. For these reasons, water is undoubtedly the most important solvent in biological processes. Due to its key role in several biological processes, water is by far the most studied liquid by computer simulations [25]. Water plays different roles with respect to protein properties. Just to name a few, (i) proteins tend to minimize their apolar surface in aqueous solution (the hydrophobic effect); (ii) water plays a structural role in folded proteins; and (iii) water exerts a dielectric screening on interactions between protein charges. Although continuum dielectric models have been successfully applied in macromolecular simulations,[26] for a complete understanding of the behavior of many biochemical systems, an explicit treatment of water is essential [27]. In a recent review, Orozco and Luque illustrated the different theoretical methods used for the description of solvent effects in biomolecular systems [28].

Water models can be divided into three types. In the simple interaction site models each water molecule is kept in a rigid geometry and the interaction between molecules is described using the non-bonded terms of eq. 1. Flexible models have internal changes in structure, and flexibility is achieved by using, for example, harmonic bond stretching terms for the OH bonds and a harmonic angle bending term for the HOH angle. Finally, models have been developed that explicitly include polarization and many body effects. As computational efficiency with which the energy can be calculated is often an important factor as there may be a very large number of water molecules in the simulation of biomolecular systems, most of the water models use effective pairwise potentials with no explicit polarization and many body effects. For this reason, water models that include polarization [29-32] and many body effects are beyond the scope of this paper [33].

The first water model to enjoy wide use in simulations was the five-site ST2 model of Stillinger and Rahman [34]. Here, the van der Waals interaction between two water molecules is computed using a Lennard-Jones (LJ) function with just a single interaction point per molecule centered on the oxygen atom. Charges are placed on the hydrogen atoms and on two lone pair sites 0.8 Å distant from the oxygen atom. Extensive testing showed that the ST2 model reproduced a wide range of structural, energetic, and dynamic properties of bulk water. To reduce computational time, Jorgensen developed in the early 80s the transferable intermolecular potential surfaces (TIPS) [35,36]. The TIPS model, which has only three interaction sites, comprises a rigid water monomer with charges –q on oxygen and +$1/2q$ on each hydrogen, located at the atomic centers, and a single LJ interaction site centered on the oxygen atom. Us-

ing MC simulations the charges and the LJ parameters were adjusted to reproduce the density and heat of vaporization of liquid water. In this manner, TIPS incorporates many-body forces in an effective way by fitting the parameters of a pairwise-additive function to experimental properties. The SPC model of Berendsen and co-workers was derived by reparameterizing the TIPS model in order to improve the structure beyond the first peak in the oxygen-oxygen radial distribution function (g_{OO}), but with a slight sacrifice in the density [37]. In 1983, the TIP3P model, a reparameterization of TIPS, was performed and demonstrated that the energy and density could be accurately reproduced within the framework of three-site models [38]. However, the TIP3P failed to reproduce the second g_{OO} peak. In 1987, Berendsen and co-workers developed the SPC/E model, an updated version of the SPC model that incorporates a polarization correction [39]. The SPC/E model improves the diffusion coefficient and radial distribution curves.

Four-site water models were also developed. The first one, which is relatively little used nowadays, but is important for historical reasons as it dates from 1933, is the model of Bernal and Fowler [40]. In the Bernal and Fowler (BF) water model, the negative charge is moved off the oxygen and towards the hydrogens at a point 0.15 Å distant from the oxygen atom on the bisector of the HOH angle. The LJ interaction site is still centered on the oxygen atom. Ten distances are required to evaluate the interaction function instead of nine for a three-site model and 17 for a five-site model, such as the ST2 model. Based on the BF water model, Jorgensen and co-workers later developed the TIPS2 and TIP4P potentials [38]. In comparison to experiment, the authors showed that TIP4P reproduced many of the properties of liquid water, including the density (+0.2%), internal energy (+1.5%), enthalpy of vaporization (+1.4%), and heat capacity (+7.3%). The TIP4P model also generated an appropriate profile for the second g_{OO} peak. In a recent work, Hernandes and co-workers used chemometric techniques to investigate the TIP4P potential behavior with respect to perturbations on all intermolecular interaction parameters and their effects on the enthalpy of vaporization, density and some radial distribution functions [41].

Further extension of the TIPnP class of potentials generated the five-site TIP5P model [42]. Similarly to the ST2 model, the van der Waals interaction between two water molecules is computed using a single LJ interaction point per molecule centered on the oxygen atom, and the charges are placed on the hydrogen atoms and on two lone pair sites. However, the cumbersome cubic scaling function to dampen the short-range electrostatic interactions in the ST2 model was discarded in TIP5P. The TIP5P model reproduced the temperature-dependent density and energy with an average error of less than 1% from –37.5 to 62.5^{o}C. Also, the dielectric constant was near 80 and had the correct temperature dependence. However, the TIP5P parameters were obtained using the simple spherical cutoff method for the long-range

electrostatic interactions. Lísal and co-workers[43] showed that when TIP5P is used in conjunction with methods that account for long-range electrostatic interactions, like Ewald summation[44] and reaction field methods,[1] the calculated properties only present a fair agreement with experimental data.

In a recent work, Nada and van der Eerden developed a six-site model for water [45]. In this model, a positive point charge is placed on each hydrogen atom, and a negative charge on each lone-pair site, like in the TIP5P model. A negative charge is also placed on a site located on the bisector of the HOH angle, as in the TIP4P model. A different point from the TIP4P and TIP5P models is that the LJ interaction acts not only on the oxygen atom but also on the hydrogen sites. This six-site water model was developed for the simulation of ice and liquid water close to the melting point. Structural and thermodynamic properties of ice and water close to the melting point were well reproduced.

The use of a rigid model for water, like the models mentioned above, is obviously an approximation and some properties cannot be determined at all. The vibrational spectrum is an example. Flexibility can be incorporated by introducing bond stretching and angle bending terms onto the potential function for a rigid model. For instance, Fergunson developed a flexible water model based on the SPC model by using cubic and harmonic bond stretching terms and a harmonic angle bending term [46]. The charges and LJ parameters were reparameterized and the final values were slightly different from those of the rigid SPC model. Another example of flexible water model is the CVFF water potential, a three-site model with internal geometry corresponding to the gas-phase experimental structure [27]. Flexibility is achieved through the use of harmonic bond stretching and angle bending potentials. With the exception of the hydrogen atom size, the non-bonded parameters are comparable to those of SPC. Because CVFF has deformable internals, a small hydrogen radius was used to avoid catastrophic behavior in dynamics caused by overlapping of the polar hydrogens with negatively charged atoms.

An important aspect of choosing an interaction function for a biomolecular system including explicit water molecules is the consistency between different parts of the interatomic function, for example the protein-protein, protein-solvent, and solvent-solvent terms. When combining a water model with a protein force field, the definition of the protein-solvent interaction requires special attention. In most cases, this interaction is defined using the so-called combination rules [5]. If protein and solvent force fields are of different types, the application of combination rules may lead to an imbalance between protein-protein, protein-solvent, and solvent-solvent interactions.

4 Boundaries

The first biomolecular simulations ignored all solvent molecules due to limited computational resources then available. Vacuum calculations can lead to significant problems as vacuum boundaries tend to minimize the surface area and distort the shape of non-spherical systems. Moreover, the shielding effect of a solvent with high dielectric permitivity, like water, on the electrostatic interactions between charges and dipoles is lacking in vacuum. Application of periodic boundary conditions (PBC) is the best way to avoid distortions due to the presence of boundaries in a finite size system. The atoms of the system to be simulated are put into a periodic cell, which is treated as if it were surrounded by identical translated images of itself. In principle, any cell shape can be used provided it fills all the space by translation operations of the central box in three dimensions. Six shapes satisfy this condition: the cube, the parallelepiped, the hexagonal prism, the truncated octahedron, the rhombic dodecahedron and the elongated dodecahedron. It is often sensible to choose a periodic cell that reflects the underlying geometry of the system. Yet, one should keep in mind that periodicity is an artifact, which may affect the simulated properties, unless the periodic cell is sufficiently large to influence the results. Since the simulation of a biomolecular system in a periodic cell containing many water molecules is computationally expensive, alternative ways of simulating biomolecular systems with explicit water molecules have been proposed.

The simplest way of doing this is to surround the solute with a "skin" of solvent molecules. The number of solvent molecules is usually significantly smaller than would be required in the analogous PBC. Boundary effects are then transferred from the solute-vacuum interface to the solvent-vacuum interface, resulting in a more realistic treatment of the solute. However, this treatment still suffers from surface tension distortions of the solvent layer and lack of dielectric screening due to the vacuum outside the solvation shell. In the extended wall region approach, the distorting effect of the vacuum outside the biomolecular system is reduced by harmonically restraining or even fixing a layer of atoms of the system. Alternatively, their motion may be coupled to a heat bath by applying, for example, stochastic boundary conditions to account for exchange of energy with the surroundings [47]. In any case, the extended wall region forms a buffer between the fully unrestrained part of the system and the vacuum surrounding it. Simulations using a spherical droplet (cap) of water have also been proposed. In this method, the least sophisticated approach for preventing system evaporation and for treating the neglected interactions with molecules beyond the boundary is to use a half-harmonic boundary potential [48]. Using a more complicated approach, the surface-constrained all-atom solvent method applies harmonic restraints that are functions of the water's distance from the boundary and the orientation of the molecular dipole moment vector. Long-range electrostatic interactions are treated using a reaction field term [49].

5 Studies of biomolecules using explicit water molecules

In this section, three studies of our group demonstrate successful simulations of biomolecules through the application of simple explicit three-site water models, flexible and rigid, combined with protein force fields.

Thermodynamic Analysis of Thrombin Inhibition by Novel Benzamidine Derivatives via Free-Energy Perturbations

Clot formation results from a complex sequence of biochemical events that comprise the coagulation cascade, which involves the interaction of specific blood proteins followed by platelet aggregation [50-52]. Although this process occurs in blood vessels to repair minor internal injuries, exaggerated clot formation leads to several cardiovascular disorders such as venous and arterial thrombosis, atrial fibrillation, stroke, and myocardial infarction [53]. Thrombin is a serine protease that plays a central role in the coagulation cascade through the conversion of fibrinogen to fibrin and platelet activation [54,55]. The p-amidinophenylpyruvate (APPA) ($K_i = 620$ nM)[56,57] and benzamidine (Bz) ($K_i = 300$ nM)[56,58,59] compounds are examples of thrombin inhibitors (Fig. 1). The first one is an electrophilic inhibitor, whose carbonyl group is attacked by the Ser195 residue of the catalytic triad, whereas the second is a noncovalent inhibitor. In order to improve the binding to thrombin of Bz and APPA, we have designed four new benzamidine derivatives [60-62]. In Fig. 1, PMBz, PEBz, PPBz, and POPBz are p-methylbenzamidine, p-ethylbenzamidine, p-(1-propyl)benzamidine, and p-(2-oxo-1-propyl)benzamidine, respectively.
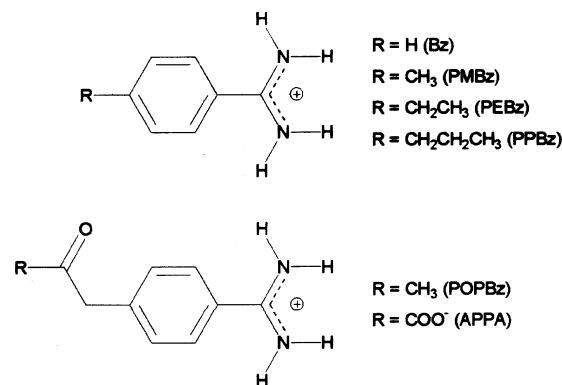

Figure 1. Benzamidine derivatives studied.

APPA (g) $\xrightarrow{\Delta A_{\mathrm{hyd}}\ \mathrm{APPA}}$ APPA (sol) + Enz (sol) $\xrightarrow{\Delta A_{\mathrm{bind(ncov)}}\ \mathrm{APPA}}$ APPA.Enz (sol) $\xrightarrow{\Delta A_{\mathrm{reaction}}\ \mathrm{APPA}}$ APPA-Enz (sol)

$\Delta A\mathrm{tr}_{(g)}$ APPA-POPBz $\downarrow$ ... $\Delta A\mathrm{tr}_{(sol)}$ APPA-POPBz ... $\Delta A_{\mathrm{bind(ncov)}}$ POPBz ... $\Delta A\mathrm{tr}_{(comp(ncov))}$ APPA-POPBz ... $\Delta A_{\mathrm{reaction}}$ POPBz ... $\Delta A\mathrm{tr}_{(comp(cov))}$ APPA-POPBz

POPBz (g) $\xrightarrow{\Delta A_{\mathrm{hyd}}\ \mathrm{POPBz}}$ POPBz (sol) + Enz (sol) $\longrightarrow$ POPBz.Enz (sol) $\longrightarrow$ POPBz-Enz (sol)

$\Delta A\mathrm{tr}_{(g)}$ POPBz-PPBz $\downarrow$ ... $\Delta A\mathrm{tr}_{(sol)}$ POPBz-PPBz ... $\Delta A_{\mathrm{bind(ncov)}}$ PPBz ... $\Delta A\mathrm{tr}_{(comp(ncov))}$ POPBz-PPBz

PPBz (g) $\xrightarrow{\Delta A_{\mathrm{hyd}}\ \mathrm{PPBz}}$ PPBz (sol) + Enz (sol) $\longrightarrow$ PPBz.Enz (sol)

$\Delta A\mathrm{tr}_{(g)}$ PPBz-PEBz $\downarrow$ ... $\Delta A\mathrm{tr}_{(sol)}$ PPBz-PEBz ... $\Delta A_{\mathrm{bind(ncov)}}$ PEBz ... $\Delta A\mathrm{tr}_{(comp(ncov))}$ PPBz-PEBz

PEBz (g) $\xrightarrow{\Delta A_{\mathrm{hyd}}\ \mathrm{PEBz}}$ PEBz (sol) + Enz (sol) $\longrightarrow$ PEBz.Enz (sol)

$\Delta A\mathrm{tr}_{(g)}$ PEBz-PMBz $\downarrow$ ... $\Delta A\mathrm{tr}_{(sol)}$ PEBz-PMBz ... $\Delta A_{\mathrm{bind(ncov)}}$ PMBz ... $\Delta A\mathrm{tr}_{(comp(ncov))}$ PEBz-PMBz

PMBz (g) $\xrightarrow{\Delta A_{\mathrm{hyd}}\ \mathrm{PMBz}}$ PMBz (sol) + Enz (sol) $\longrightarrow$ PMBz.Enz (sol)

$\Delta A\mathrm{tr}_{(g)}$ PMBz-Bz $\downarrow$ ... $\Delta A\mathrm{tr}_{(sol)}$ PMBz-Bz ... $\Delta A_{\mathrm{bind(ncov)}}$ Bz ... $\Delta A\mathrm{tr}_{(comp(ncov))}$ PMBz-Bz

Bz (g) $\xrightarrow{\Delta A_{\mathrm{hyd}}\ \mathrm{Bz}}$ Bz (sol) + Enz (sol) $\longrightarrow$ Bz.Enz (sol)
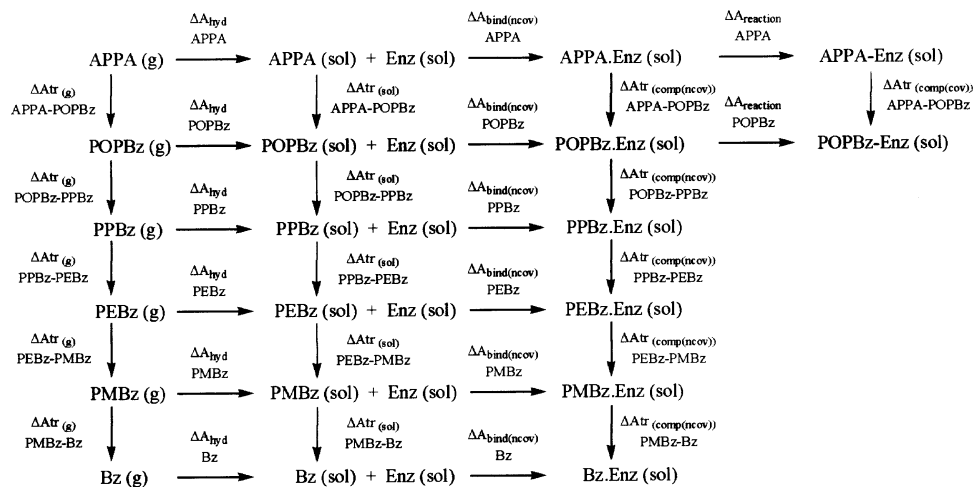
Figure 2. Thermodynamic cycles used for the calculation of relative free energies.

Relative free energies of hydration ($\Delta\Delta A_{\mathrm{hyd}}$) and binding to thrombin ($\Delta\Delta A_{\mathrm{bind(ncov)}}$ for the noncovalent compounds and $\Delta\Delta A_{\mathrm{bind(cov)}}$ for the electrophilic molecules, like APPA and POPBz) for all candidates were computed through MD simulations in conjunction with the finite difference thermodynamic integration (FDTI) algorithm,[63] a free energy perturbation method [64-67].

$$\Delta A = -kT\sum_{i=1}^{n}\frac{\ln\left\langle e^{-[U(\lambda i+\delta\lambda)-U(\lambda i)]/kT}\right\rangle_i}{\delta\lambda}\cdot\Delta\lambda i \quad (2)$$

In eq. 2, ΔA is the free energy variation, $\delta\lambda$ is the increment used to compute the numerical derivatives, $U(\lambda i)$ and $U(\lambda i+\delta\lambda)$ are the potential energy functions for the states λi and $\lambda i+\delta\lambda$, respectively, $<>_i$ refers to an average over the ensemble of configurations generated using $U(\lambda i)$, n is the number of quadrature points, and $\Delta\lambda i$ is a parameter that depends on the numerical integration scheme. When calculating relative free energy differences, a thermodynamic cycle is generally applied [68]. We have used the thermodynamic cycles exhibited in Fig. 2 [60-62]. Because the quantities ΔA_{hyd}, $\Delta A_{\mathrm{bind(ncov)}}$, and $\Delta A_{\mathrm{reaction}}$ (free energy of reaction with thrombin for the electrophilic molecules, APPA and POPBz) are impracticable to compute due to the time scale and the enormous atom reorganization involved in these thermodynamic processes, the relative free energies are obtained through the simulation of non-physical paths by transforming one molecule into another in different environments. To obtain $\Delta\Delta A_{\mathrm{hyd}}$, the transformations in aqueous phase ($\Delta A\mathrm{tr}_{(sol)}$) were performed in a 20 Å cubic box with approximately 250 water molecules using PBC. Due to computational limitations to calculate $\Delta\Delta A_{\mathrm{bind(ncov)}}$ and $\Delta\Delta A_{\mathrm{reaction}}$, the transformations inside the noncovalent ($\Delta A\mathrm{tr}_{(comp(ncov))}$) and covalent ($\Delta A\mathrm{tr}_{(comp(cov))}$) complex models were performed using spherical cap of water with 19 Å of radius. To prevent the evaporation of water molecules, a half-harmonic restraining potential of 0.5 kcal $\cdot$ mol$^{-1}\cdot$Å^{-2} was used when the distance between a water oxygen atom and the center of the solvation model exceeded 19 Å. As shown by Essex and Jorgensen, [48] the use of a spherical cap of water rather than the conventional PBC affects the calculated free energies of hydration in simple systems. Therefore, to cancel the errors introduced by the application of a spherical cap of water in the solvated noncovalent complex model, we employed an identical approximation for the calculation of $\Delta A\mathrm{tr}_{(sol)}$. It is interesting to note that $\Delta\Delta A_{\mathrm{reaction}}$ added to $\Delta\Delta A_{\mathrm{bind(ncov)}}$ between POPBz and APPA gives the relative free energy of covalent binding to thrombin ($\Delta\Delta A_{\mathrm{bind(cov)}}$) for these two compounds. The all-atom CVFF force field [14] combined with the CVFF flexible water potential [27] was employed in all calculations. By employing protein and solvent force fields of the same type, we guaranteed a perfect balance between protein-protein, protein-solvent, and solvent-solvent interactions.

The agreement between the experimental[60] and the calculated $\Delta\Delta A_{\mathrm{bind(ncov)}}$ between PMBz and Bz (experimental is 0.68 kcal/mol; calculated is 0.65 kcal/mol) is excellent. This supports the quality of our simulations and suggests that the calculated binding order (POPBz > PEBz > PPBz > PMBz > Bz > APPA) is correct. It also suggests that errors introduced due to the application of a spherical cap of water are cancelled when a thermodynamic cycle is used. It should be noted that all candidates proposed in this work have a higher affinity for the enzyme than APPA and Bz, compounds originally selected from the literature. Analysis of the relative hydration revealed two possible hydration orders for the p-substituted benzamidine derivatives: APPA > POPBz > PMBz > PEBz > PPBz > Bz and APPA > POPBz > PMBz > PEBz > Bz > PPBz [61,62].

To
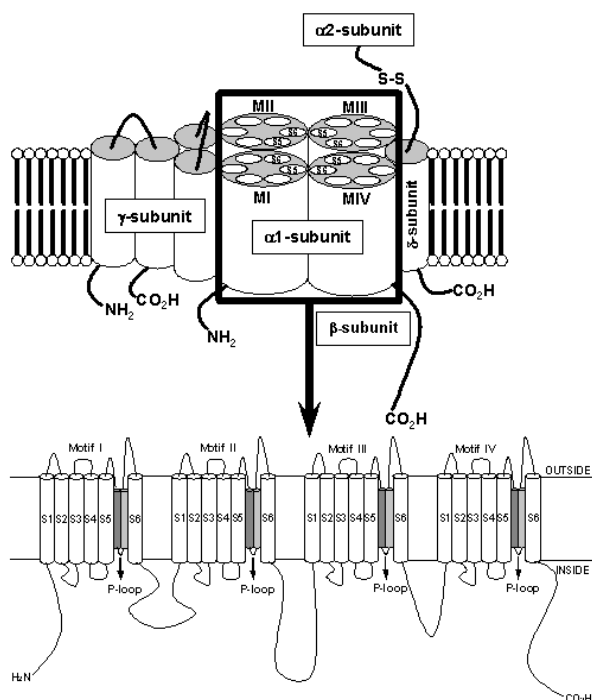

Figure 3. Molecular structure of the VGCC complex with zoom of the proposed transmembrane arrangement of the VGCC α 1-subunit. The dark and light gray cylinders are the SS1 and SS2 segments, respectively.

conclude, we showed in this study that hydrogen bonds with the N-H group of Gly193 in the protein active site and hydrophobic interactions with the wall of the cavity formed by His57, Ala190, Glu192, Gly193, and Val213 increase the binding affinity for thrombin. Moreover, we showed that it is wise to avoid the design of candidates with a negatively charged group because of electrostatic repulsions with negative residues in the active site, and because of a large desolvation penalty.

Molecular Dynamics and Potential of Mean Force Calculations on an L-type Calcium Channel Model

The family of voltage-gated calcium channels (VGCCs) is responsible for the major influx pathway for Ca^{2+} in different types of cells. Ca^{2+} ions play important roles in a wide variety of processes such as muscle contraction, cell proliferation, gene expression, neurotransmitter release and other secretion processes, and cell death [69-72]. One low (T type) and five high VGCC types (L, N, P, Q, R) have been studied through pharmacological and electrophysiological studies. These studies revealed that the L-type calcium channel complex is a heteropentamer consisting of $\alpha 1$, β, $\alpha 2/\delta$, and γ subunits [69,70,73]. Ca^{2+} ions reach the cytoplasm by passing selectively through the pore of the $\alpha 1$-subunit,[69-73] formed by four repeating motifs (MI-MIV), each comprising six hydrophobic segments (S1-S6). A highly conserved segment connecting the S5 and S6 trans-

membrane domains in each motif, termed the P loop or "SS1-SS2" region, is responsible for calcium selectivity in the pore region (Fig. 3) [69,72]. More specifically, four closely aligned glutamate residues, the EEEE locus, localized in the SS2 segment of each P loop segment, molecularly express the calcium selectivity of VGCCs.

We recently proposed a molecular model for the L-type calcium channel pore from the human cardiac $\alpha 1$ subunit [74,75]. This is not the first molecular model for the L-type calcium channel,[76] but it was the first that did not use the crystal structure of the KcsA K^+ channel as a template [77]. The reasons for that are discussed in details in refs. 74,75. In a recent review, Sather and McCleskey described the attempts by theoretical groups, including our molecular model, to explain the combination of high selectivity and high flux that characterizes calcium channels [78].

To assure perfect balance between protein-protein, protein-solvent, and solvent-solvent interactions, the all-atom CVFF force field[14] combined with the CVFF flexible water potential[27] was employed in all calculations. The proposed α-helix structure for the SS1 segment of each motif was studied separately through molecular dynamics simulations in aqueous-phase under PBC. The analysis of dynamical behavior of the $i - i + 4$ helical hydrogen bond (hHB) between the amino acid residues revealed for all SS1 segments that a few hHB interactions were broken during

the time spanned by the simulation and had the consequent water insertion into the broken hHB [74]. Nevertheless, as shown by Ramachandran plots for the average structures, the integrity of the α-helix structure was maintained for all SS1 segments because only a reduced number of hHB was disrupted during the simulation.
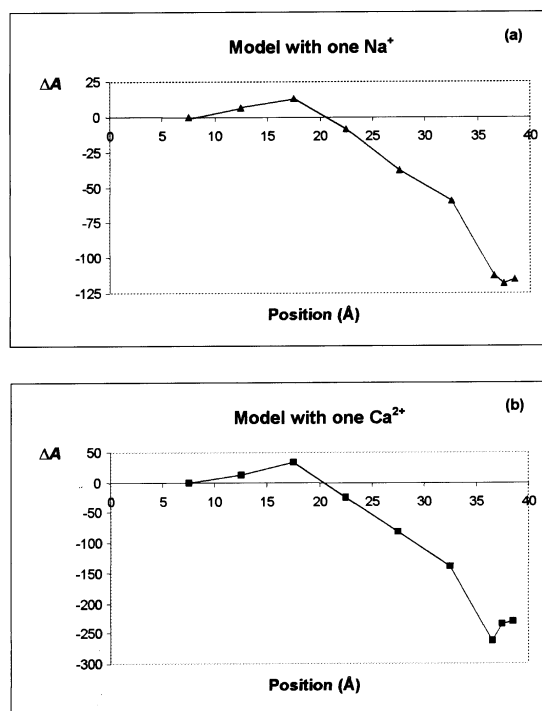
Figure 4. Potential of mean force curves (a) for the position change of a Na^+ ion inside the channel, and (b) for the position change of a Ca^{2+} ion inside the channel.

The model for the L-type calcium channel pore, constructed by docking the four motif models (each motif model formed by the respective SS1 segment and six residues of the SS2 segment) and following experimental studies regarding Ca^{2+} channels, was energy minimized and placed in a parallelepiped with $\sim$1550 water molecules, with no PBC applied to the calculations. To avoid evaporation, a 5 Å layer of water molecules surrounding the whole system was kept fixed during all simulations. To understand the mechanisms of Na^+/Li^+ permeation at submicromolar Ca^{2+} concentrations, Na^+/Li^+ blocking at higher Ca^{2+} concentrations (10^{-6}-10^{-4} M), and Ca^{2+} permeation at milimolar Ca^{2+} concentrations,[79,80] eq. 2 was applied to derive potential of mean force (pmf) curves[65] for the position change of one Na^+ and one Ca^{2+} ion inside the channel. In addition, MD simulations were performed to investigate the behavior of the channel model in the presence of one and two Ca^{2+} and also of one Ca^{2+} and one Na^+ at the binding site. The pmf curves for the change in position of one Na^+ ion and one Ca^{2+} ion in the L-type VGCC pore model are shown in Figs. 4a and 4b. As expected, the pmf curves point out the

EEEE locus as the region of coordination; more precisely the free energy minimum was located in x coordinate values of 37.5 Å and 36.5 Å for Na^+ and Ca^{2+}, respectively. The pmf curves also show that the interaction of one Ca^{2+} ion with the binding site is more favorable than the interaction of one Na^+ by approximately 140 kcal/mol, with respect to the entrance of the channel (x coordinate value of 7.5 Å). This result shows that our model reproduces the high selectivity of the channel for Ca^{2+}, and suggests that this high selectivity derives from strong electrostatic interactions between one ion with charge 2+ and the EEEE locus with charge 4-. The MD simulations showed that the EEEE locus must form a single high-affinity binding site, in agreement with the mutation studies of Yang *et al.* (Fig. 5a) [79]. We were able to confirm this result through the analysis of an MD simulation of the pore model in presence of two Ca^{2+} ions at the binding site. We observed that at the end of the MD run the first Ca^{2+} ion is displaced by the second one from the EEEE locus and that the coordination of the second ion with the single high affinity-binding site starts to dominate the process (Fig. 5b). Differently from what we observed with two calcium ions, Ca^{2+} was not displaced from its binding site by the Na^+ (Fig. 5c). This last result is in agreement with the experimentally observed blocking of Na^+ flux at micromolar Ca^{2+} concentrations.

Molecular Dynamics Studies on the Human Cytomegalovirus Protease Homodimer

Human cytomegalovirus (HCMV) is a highly species-specific DNA virus infecting up to 80% of the general population [81]. Except for a mononucleosis-like illness in some persons, infection with HCMV rarely causes disease in immunocompetent individuals [82]. However, HCMV is responsible for severe morbidity and mortality in immunocompromised individuals including organ transplant recipients and HIV-infected persons [83]. Clinical manifestations comprise disseminated disease, pneumonitis, retinitis, and gastrointestinal infections [81]. The viral genome contains the open reading frame UL80, which encodes the full-length 80 kDa HCMV serine protease and its substrate [84]. Full-length HCMV protease is composed of an N-terminal 256-amino-acid proteolytic domain, called assemblin, a linker region, and a C-terminal structural domain, the assembly protein precursor [85]. In HCMV, the assembly protein precursor interacts with the major capsid protein and acts as a scaffold in the nucleus around which the capsid assembles [85,86]. Thus, the inhibition of the HCMV protease has been considered as an alternative target for antiviral chemotherapy [84].

X-ray crystallographic studies [87-89] revealed that the protein fold of assemblin, comparatively soluble and active as the full-length HCMV protease, is very different from that of any other known serine protease. In addition, HCMV protease has a unique catalytic triad in which the third member is a histidine residue (Ser132, His63, and His157) and exists
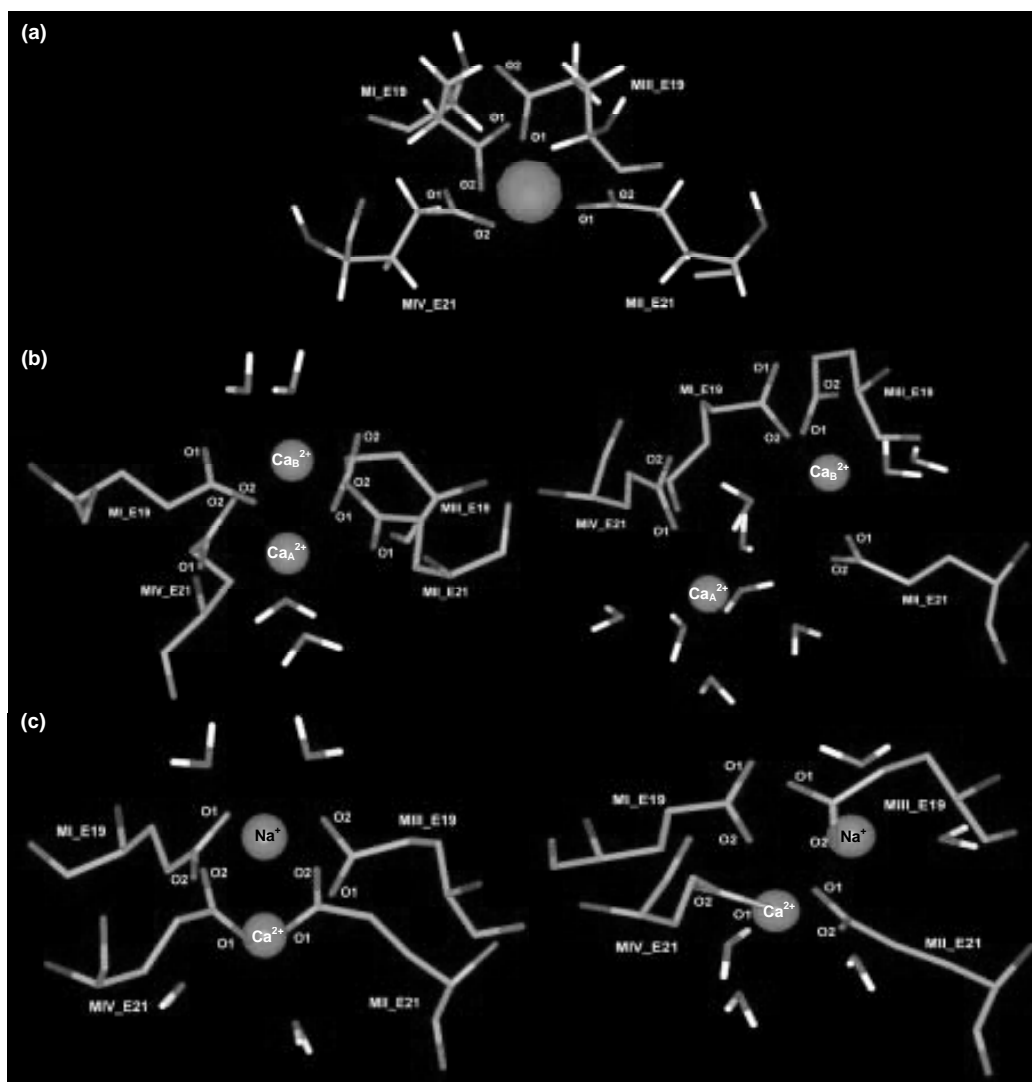
Figure 5. (a) The arrangement of the EEEE locus in two close but non-equivalent planes occupying *trans* positions. (b) Two Ca^{2+} ions, one of them is displaced from the binding site. (c) The Ca^{2+} ion is not displaced by the Na^+ ion. On the left in (b) and (c), the starting positions, on the right, the final positions of the MD simulations.

as a dimer, believed to be the sole active form [90,91]. In a recent work, we reported MD simulations on the HCMV protease using PBC.[92] The AMBER force field[13] combined with the TIP3P rigid water potential[38] was employed in all calculations.

The isotropic temperature factor (B-factor) from atom-positional root-mean-square fluctuations of Cα atoms were calculated from the MD simulations. The B-factors, plotted as a function of residue number, were obtained from eq. 3 [93]

$$\left\langle \Delta r_i^2 \right\rangle = \frac{3B_i}{8\pi^2} \qquad (3)$$

where $\left\langle \Delta r_i^2 \right\rangle$ is the mean-square fluctuation for the Cα atom of residue i. Fig. 6a compares the calculated B-factors for the dimeric form of assemblin complexed noncovalently to a

peptidyl-activated carbonyl inhibitor,[94] very similar structurally to BILC 821, with the experimental B-factors for the covalent complex between BILC 821 and the dimer of assemblin [95]. Figs. 6b compares the calculated and the experimental B-factors for the dimeric form of unbound assemblin [87]. The results exhibited in Figs. 6a and b show very good agreement between the simulation and experimental values in that peaks and valleys are coincident. Regions that show flexibility in the crystal structure were well reproduced by the MD trajectories. For example, the missing loops in the crystal structures had the largest calculated B-factors. As discussed by Halle,[96] this is probably derived from the solvent exposure and non-regular secondary structures of these regions. It should be noted, however, that the atomic fluctuations in the simulations are greater than
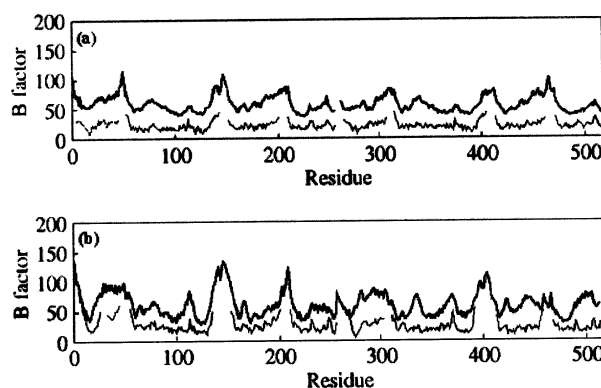
Figure 6. Comparison of B-factors on a *per* residue basis from atom-positional root-mean-square fluctuations of Cα atoms. (a) Calculated B-factors for the dimeric form of assemblin complexed noncovalently to a peptidyl-activated carbonyl inhibitor, very similar structurally to BILC 821 (thick line) vs. experimental B-factors for the covalent complex between BILC 821 and the dimer of assemblin (thin line, PDB code: 2WPO). (b) Calculated (thick line) and experimental B-factors (thin line, PDB code: 1WPO) for the dimeric form of unbound assemblin.

the experimental values. This result may be attributed to the increased mobility of the protein atoms in solution when compared to a more rigid structure in the crystal, as discussed by Stocker and coworkers [97].

Two other interesting results were obtained from our MD simulations [92]. The conformational changes observed for the dimer-BILC 821 covalent complex were reproduced by the MD simulations of the noncovalent complex between the dimer and the peptidyl-activated carbonyl inhibitor selected from the literature,[94] suggesting that the HCMV protease is a novel example of serine protease that operates by an induced-fit mechanism. The MD simulations also showed that the dimeric form is necessary to activate the enzyme because of an induced stabilization of the oxyanion hole, in agreement with mutagenesis studies [98]. Finally, the quality of our results indicate that the combination of the AMBER force field[13] with the TIP3P water potential[38] is well balanced, even though both methods were developed separately by different groups.

Acknowledgments

References

[1] W.F. van Gunsteren, H.J.C. Berendsen, Angew. Chem. Int. Ed. Engl. **29**, 992 (1990).

[2] M.P. Allen, D.J. Tildesley, *In Computer Simulation of Liquids*; Clarenden Press, Oxford, 1987, pp. 110.

[3] A.R. Leach, *In Molecular Modelling: Principles and Applications;* Longman, Essex, 1996.

[4] J. Tirado-Rives, W.L. Jorgensen, J. Comput. Chem. **117**, 1385 (1996).

[5] W.F. van Gunsteren, F.J. Luque, D. Timms, and A.E. Torda, Annu. Rev. Biophys. Biomol. Struct. **23**, 847 (1994).

[6] S.K. Pal, J. Peon, B. Bagchi, and A.H. Zewail, J. Phys. Chem. B **106**, 12376 (2002).

[7] B.P. Schoenborn, A. Garcia, and R. Knott, Prog. Biophys. Molec. Biol. **64**, 105 (1995).

[8] W.F. van Gunsteren, A.E. Mark, Eur. J. Biochem. **204**, 947 (1992).

[9] M.L. Lamb, W.L. Jorgensen, Curr. Opin. Chem. Biol. **1**, 449 (1997).

[10] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, and W.L. Jorgensen, J. Phys. Chem. B **105**, 6474 (2001).

[11] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).

[12] J. Hermans, H.J.C. Berendsen, W.F. van Gunsteren, and J.P.M. Postma, Biopolymers **23**, 1513 (1984).

[13] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman, J. Am. Chem. Soc. **117**, 5179 (1995).

[14] P. Dauber-Osguthorpe, V.A. Roberts, D.J. Osguthorpe, J. Wolff, M. Genest, and A.T. Hagler, Proteins **4**, 31 (1988).

[15] D.J. Price, C.L. Brooks, J. Comput. Chem. **23**, 1045 (2002).

[16] P.A. Kollman, Acc. Chem. Res. **29**, 461 (1996).

[17] W.L. Jorgensen, D.S. Maxwell, and J. Tirado-Rives, J. Am. Chem. Soc. **118**, 11225 (1996).

[18] G.A. Kaminski, W.L. Jorgensen, J. Phys. Chem. **100**, 18010 (1996).

[19] V. Helms, R.C. Wade, J. Comput. Chem. **18**, 449 (1997).

[20] A. Warshel, M. Levitt, J. Mol. Biol. **103**, 227 (1976).

[21] U.C. Singh, P.A. Kollman, J. Comput. Chem. **7**, 718 (1986).

[22] P.A. Bash, M. Field, and M. Karplus, J. Am. Chem. Soc. **109**, 8092 (1987).

[23] W.R. Rocha, K. Coutinho, W.B. Almeida, and S. Canuto, Chem. Phys. Lett. **335**, 127 (2001).

[24] K. Coutinho, R.C. Guedes, B.J.C. Cabral, and S. Canuto, Chem. Phys. Lett. **369**, 345 (2003).

[25] B. Guillot, Y. Guissani, J. Chem. Phys. **114**, 6720 (2001).

[26] D. Bashford, D. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).

[27] K.F. Lau, H.E. Alper, T.S. Thacher, and T.R. Stouch, J. Phys. Chem. **98**, 8785 (1994).

[28] M. Orozco, F.J. Luque, Chem Rev. **100**, 4187 (2000).

[29] P. Barnes, J.L. Finney, J.D. Nicholas, and J.E. Quinn, Nature **282**, 459 (1979).

[30] M.W. Mahoney, W.L. Jorgensen, J. Chem. Phys. **115**, 10758 (2001).

[31] H.A. Stern, F. Rittner, B.J. Berne, and R.A. Friesner, J. Chem. Phys. **115**, 2237 (2001).

[32] P. Ren, J.W. Ponder, J. Phys. Chem. B **107**, 5933 (2003).

[33] F.N. Keutsch, J.D. Cruzan, and R.J. Saykally, Chem. Rev. **103**, 2533 (2003).

[34] F.H. Stillinger, A. Rahman, J. Chem. Phys. **60**, 1545 (1974).

[35] W.L. Jorgensen, J. Am. Chem. Soc. **103**, 335 (1981).

[36] W.L. Jorgensen, J. Chem. Phys. **77**, 4156 (1982).

[37] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, and J. Hermans, In *Intermolecular Forces*, edited by B. Pullman (Reidel, Dordrecht, 1981), p. 331.

[38] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, and M.L. Klein, J. Chem. Phys. **79**, 926 (1983).

[39] H.J.C. Berendsen, J.R. Grigera, and T.P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).

[40] J.D. Bernal, R.H. Fowler, J. Chem. Phys. **1**, 515 (1933).

[41] M.Z. Hernandes, J.B.P. da Silva, and R.L. Longo, J. Comput. Chem. **24**, 973 (2003).

[42] M.W. Mahoney, W.L. Jorgensen, J. Chem. Phys. **112**, 8910 (2000).

[43] M. Lísal, J. Kolafa, and I. Nezbeda, J. Chem. Phys. **117**, 8892 (2002).

[44] T.A. Darden, D.M. York, and L.G. Pedersen, J. Chem. Phys. **98**, 10089 (1993).

[45] H. Nada, J.P.J.M. van der Eerden, J. Chem. Phys. **118**, 7401 (2003).

[46] D.M. Fergunson, J. Comput. Chem. **16**, 501 (1995).

[47] M. Berkowitz, J.A. McCammon, Chem. Phys. Lett. **90**, 215 (1982).

[48] J.W. Essex, W.L. Jorgensen, J. Comput. Chem. **16**, 951 (1995).

[49] G. King, A. Warshel, J. Chem. Phys. **91**, 3647 (1989).

[50] C.A. Coburn, D.M. Rush, P.D. Williams, C. Homnick, E.A. Lyle, S.D. Lewis, B.J. Lucas Jr., J.M. Di Muzio-Mower, M. Juliano, J.A. Krueger, K. Vastag, I.-W. Chen, and J.P. Vacca, Bioorg. Med. Chem. Lett. **10**, 1069 (2000).

[51] V. Pavone, G. De Simone, F. Nastri, S. Galdiero, N. Staiano, A. Lombardi, and C. Pedone, J. Biol. Chem. **379**, 987 (1998).

[52] K. Lee, Korean J. Med. Chem. **7**, 127 (1997).

[53] S.D. Kimbal, Curr. Pharm. Des. **1**, 441 (1995).

[54] W.C. Ripka, Curr. Opin. Chem. Biol. **1**, 242 (1997).

[55] D.K. Jones-Hertzog, W.L. Jorgensen, J. Med. Chem. **40**, 1539 (1997).

[56] R.E. Babine, S.L. Bender, Chem. Rev. **97**, 1359 (1997).

[57] Z. Chen, Y. Li, A.M. Mulichak, S.D. Lewis, and J.A. Shafer, Arch. Biochem. Biophys. **322**, 198 (1995).

[58] D. Banner, P. Hadváry, J. Biol. Chem. **266**, 20085 (1991).

[59] A. Scozzafava, F. Briganti, and C.T. Supuran, Eur. J. Med. Chem. **34**, 939 (1999).

[60] C.R.W. Guimarães, R. Bicca de Alencastro, J. Phys. Chem. B **106**, 466 (2002).

[61] C.R.W. Guimarães, R. Bicca de Alencastro, Int. J. Quantum Chem. **85**, 713 (2001).

[62] C.R.W. Guimarães, R. Bicca de Alencastro, J. Med. Chem. **45**, 4995 (2002).

[63] M. Mezei, J. Chem. Phys. **86**, 7084 (1987).

[64] R. Zwanzig, J. Chem. Phys. **22**, 1420 (1954).

[65] D.L. Beveridge, F.M. DiCapua, Annu. Rev. Biophys. Biophys. Biochem. **18**, 431 (1989).

[66] W.L. Jorgensen, Acc. Chem. Res. **22**, 184 (1989).

[67] P.A. Kollman, Chem. Rev. **93**, 2395 (1993).

[68] C.F. Wong, J.A. McCammon, J. Am. Chem. Soc. **108**, 3830 (1986).

[69] G. Varadi, M. Strobeck, S. Koch, L. Caglioti, C. Zucchi, and G. Palyi, Crit. Rev. Biochem. Molec. Biol. **34**, 181 (1999).

[70] G. Varadi, Y. Mori, G. Mikala, and A. Schwartz, Trends Pharmacol. Sci. **16**, 43 (1995).

[71] K. Imoto, FEBS Lett. **325**, 100 (1993).

[72] W.A. Catterall, Science **242**, 50 (1988).

[73] A. Randall, C.D. Benham, Mol. Cell. Neurosci. **14**, 255 (1999).

[74] G. Barreiro, C.R.W. Guimarães, and R. Bicca de Alencastro, Protein Eng. **15**, 109 (2002).

[75] G. Barreiro, C.R.W. Guimarães, and R. Bicca de Alencastro, Protein Eng. **16**, 209 (2003).

[76] G.M. Lipkind, H.A. Fozzard, Biochemistry **40**, 6786 (2001).

[77] D.A. Doyle, J. Morais Cabral, R.A. Pfuetzner, A. Kuo, J.M. Gulbis, S.L. Cohen, B.T. Chait, R. MacKinnon, Science **280**, 69 (1998).

[78] W.A. Sather, E.W. McCleskey, Annu. Rev. Physiol. **65**, 133 (2003).

[79] J. Yang, P.T. Ellinor, W.A. Sather, J.-F. Zhang, and R.W. Tsien, Nature **366**, 158 (1993).

[80] E.W. McCleskey, E. W. J. Gen. Physiol. **113**, 765 (1999).

[81] W. Ogilvie, M. Bailey, M.-A. Poupart, A. Abraham, A. Bhavsar, P. Bonneau, J. Bordeleau, Y. Bousquet, C. Chabot, J.-S. Duceppe, G. Fazal, S. Goulet, C. Grand-Maître, I. Guse, T. Halmos, P. Lavallée, M. Leach, E. Malenfant, J. O'Meara, R. Plante, C. Plouffe, M. Poirier, F. Soucy, C. Yoakim, and R. Déziel, J. Med. Chem. **40**, 4113 (1997).

[82] M.D. Jong, G.J. Galasso, B. Gazzard, P.D. Griffiths, D.A. Jabs, E.R. Kern, and S.A. Spector, Antiviral Res. **39**, 141 (1998).

[83] B.C. Holwerda, Antiviral Res. **35**, 1 (1997).

[84] S.A. Margosiak, D.L. Vanderpool, W. Sisson, C. Pinko, and C.-C. Kan, Biochemistry **35**, 5300 (1996).

[85] W. Gibson, Nat. Struct. Biol. **8**, 739 (2001).

[86] R.L. LaFemina, K. Bakshi, W.J. Long, B. Pramanik, C.A. Veloski, B.S. Wolanski, A.I. Marcy, and D.J. Hazuda, J. Virol. **70**, 4819 (1996).

[87] L. Tong, C. Qian, M.-J. Massariol, P. Bonneau, M. Cordingley, and L. Lagacé, Nature **383**, 272 (1996).

[88] X. Qiu, J.S. Culp, A.G. DiLella, B. Hellmig, S.S. Hoog, C.A. Janson, W.W. Smith, and S.S. Abdel-Meguid, Nature **383**, 275 (1996).

[89] H.S. Shieh, R.G. Kurumbail, A.M. Stevens, R.A. Stegeman, E.J. Sturman, J.Y. Pak, A.J. Wittwer, M.O. Palmier, R.C. Wiegand, B.C. Holwerda, and W.C. Stallings, Nature **383**, 279 (1996).

[90] P.L. Darke, J.L. Cole, L. Waxman, D.L. Hall, M.K. Sardana, and L.C. Kuo, J. Biol. Chem. **271**, 7445 (1996).

[91] S.R. LaPlante, P.R. Bonneau, N. Aubry, D.R. Cameron, R. Déziel, C. Grand-Maître, C. Plouffe, L. Tong, and S.H. Kawai, Biochemistry **121**, 2974 (1999).

[92] C.A.F. Oliveira, C.R.W. Guimarães, G. Barreiro, and R. Bicca de Alencastro, Proteins **52**, 483 (2003).

[93] J.A. McCammon, S.C. Harvey, In Dynamics of Proteins and Nucleic Acids. Cambridge: Cambridge University Press; 1987.

[94] P.R. Bonneau, C. Grand-Maître, D.J. Greenwood, L. Lagacé, S.R. LaPlante, M.-J. Massariol, W.W. Ogilvie, J.A. O'Meara, and S.H. Kawai, Biochemistry **36**, 12644 (1997).

[95] L. Tong, C. Qian, M.-J. Massariol, R. Déziel, C. Yoakim, and L. Lagacé, Nat. Struct. Biol. **5**, 819 (1998).

[96] B. Halle, Proc. Natl. Acad. Sci. USA **99**, 1274 (2002).

[97] U. Stocker, K. Spiegel, and W.F. van Gunsteren, J. Biol. NMR **18**, 1 (2000).

[98] P.-H. Liang, K.A. Brun, J.A. Field, K. O'Donnel, M.L. Doyle, S.M. Green, A.E. Baker, M.N. Blackburn, and S.S. Abdel-Meguid, Biochemistry **37**, 5923 (1998).