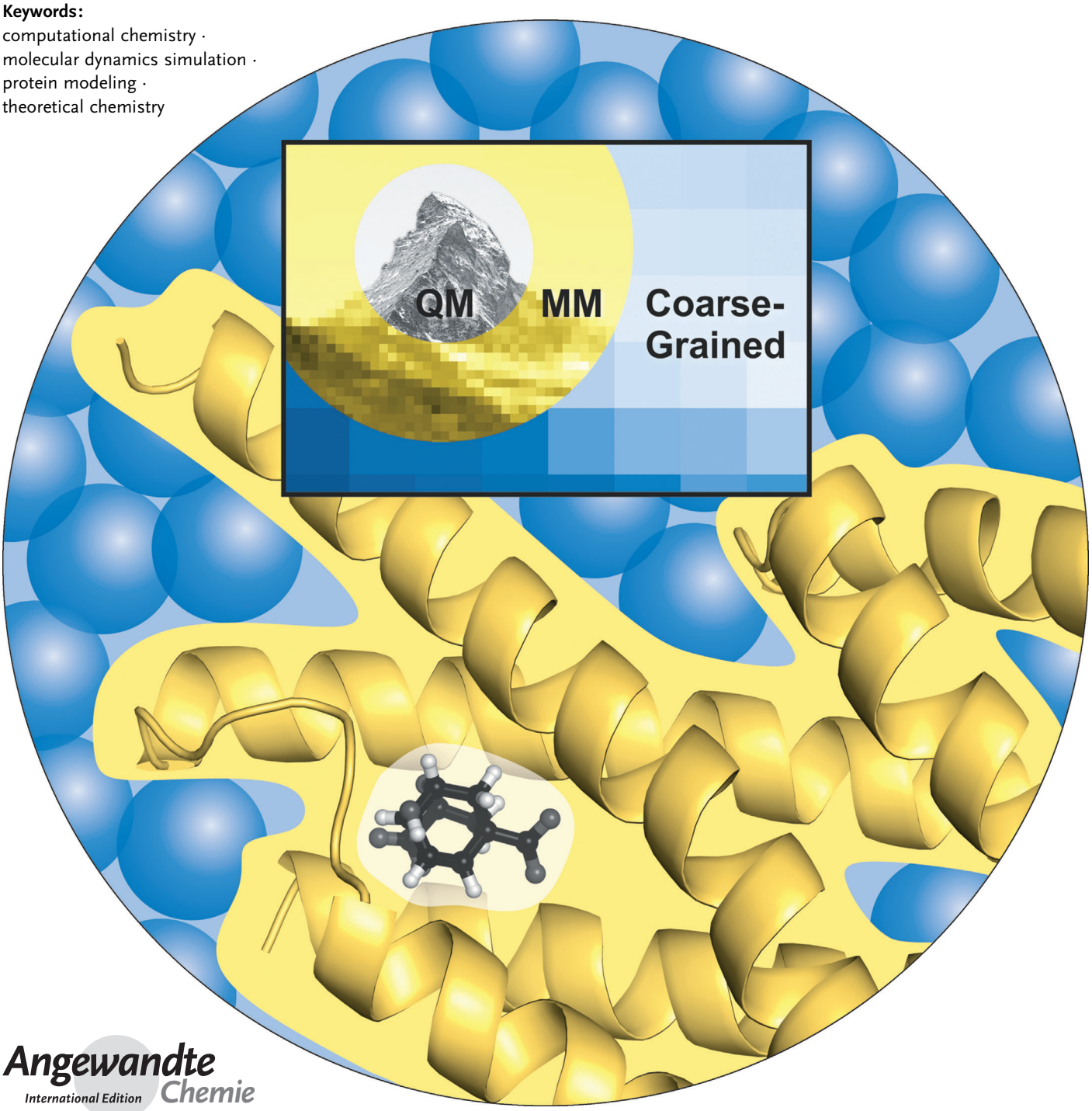
Computational Chemistry

DOI: 10.1002/anie.201205408

Multi-Resolution Simulation of Biomolecular Systems: A Review of Methodological Issues

*Katharina Meier, Alexandra Choutko, Jozica Dolenc, Andreas P. Eichenberger, Sereina Riniker, and Wilfred F. van Gunsteren**

Keywords:
computational chemistry · molecular dynamics simulation · protein modeling · theoretical chemistry



Theoretical-computational modeling with an eye to explaining experimental observations in regard to a particular chemical phenomenon or process requires choices concerning essential degrees of freedom and types of interactions and the generation of a Boltzmann ensemble or trajectories of configurations. Depending on the degrees of freedom that are essential to the process of interest, for example, electronic or nuclear versus atomic, molecular or supramolecular, quantum- or classical-mechanical equations of motion are to be used. In multi-resolution simulation, various levels of resolution, for example, electronic, atomic, supra-atomic or supra-molecular, are combined in one model. This allows an enhancement of the computational efficiency, while maintaining sufficient detail with respect to particular degrees of freedom. The basic challenges and choices with respect to multi-resolution modeling are reviewed and as an illustration the differential catalytic properties of two enzymes with similar folds but different substrates with respect to these substrates are explored using multi-resolution simulation at the electronic, atomic and supra-molecular levels of resolution.

From the Contents

1. Introduction 2821
2. Levels of Resolution 2822
3. Treatment of Quantum Effects in Molecular Simulation 2823
4. Supramolecular Models for Combination with Atomic Models in Molecular Simulation 2826
5. Multi-Resolution Simulation of Four Enzyme–Substrate Complexes 2828
6. Discussion 2832

1. Introduction

Chemistry is more than 125 years old and so is thinking about the reasons for the many observations of chemical processes in daily life. Already in ancient Greece, simple chemical models were proposed, as for example, by Thales in the 6th century B.C. who postulated that all substances would be composed of the simple elements earth, water, air and fire. It took, however, until the 18th century for the real chemical elements to be recognized by the generation of chemists of Lavoisier and Dalton based on a systematic design of experiments to corroborate or repudiate a chemical theory or model. With concepts of atoms and molecules chemistry as it is defined today was born. The 19th century saw a rapid development of the understanding of physical and chemical processes through both experimentation and the concurrent development of theory and models which was reflected in the founding of *Angewandte Chemie* in 1888. Boltzmann laid the foundation of statistical mechanics further developed by Gibbs, and Planck provided the basis for the development of the quantum theory during the first half of the 20th century. The development of computers during the second half of the 20th century induced a transition from analytical solutions of the basic quantum statistical-mechanical equations governing chemical phenomena to numerical ones thereby vastly extending the applicability of chemical theory and modeling. Half a century ago, macromolecular crystallography, quantum chemistry, and classical simulation of substances in the liquid phase could only thrive due to the use of computers by the pioneers in these fields. Nowadays, many chemical processes such as chemical reactions, phase separation, solvation, ligand-protein binding, and protein folding, are investigated using numerical chemical models and state-of-the-art computing devices.

Yet, the accuracy and applicability of theoretical computational models in chemistry is rather limited due to a variety of factors: 1) The degrees of freedom governing chemical processes are electronic, nuclear, atomic and molecular, and the corresponding particles have very different masses and sizes (see Table 1). 2) The interactions between these particles are governed by quantum mechanics, that is, the Dirac or Schrödinger equations of motion, or under particular conditions of not too small mass or low temperature, by classical equations of motion, that is, those of Newton, Lagrange, Hamilton or Langevin. 3) At non-zero temperatures, the behavior of the particles is in addition governed by statistical mechanics, that is, Bose–Einstein, Fermi–Dirac, or Boltzmann ensembles of configurations are to be considered, not single structures. 4) Compared to the strong interaction between nucleons, the Coulomb interaction between nuclei and electrons is spatially rather long-ranged, which induces many-body effects that make an accurate modeling of the resulting forces computationally expensive. 5) The energy or free energy changes of chemical processes can be very small compared to the total energy of the interacting particles involved in the process. 6) The time scale of different chemical processes may easily span 15 orders of magnitude.

These features of chemical processes severely complicate the formulation of accurate predictive models in chemistry. Yet, computational modeling of chemical processes is practised, because it is needed for the interpretation of experimental observations and because it complements the exper-

[*] K. Meier, A. Choutko, J. Dolenc, A. P. Eichenberger, S. Riniker, W. F. van Gunsteren
Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich ETH, 8093 Zürich (Switzerland)
E-mail: wfvgn@igc.phys.chem.ethz.ch

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201205408.

Table 1: Characteristic sizes of particles at different levels of resolution of modeling, scaling of the computational effort as a function of the number of nucleons (N_n), electrons (N_e), atoms (N_a) or beads (N_b), and the reduction of the number of degrees of freedom N_{df} or interactions and the reduction of computational effort that can be achieved by coarse-graining to the next level of resolution.

Level of resolution	I	II	III	IV	V
Particles	nucleons + electrons	nuclei + electrons	atoms	supra-atomic beads	supra-molecular beads
Mass of bead [amu]	10^{-3}–1	10^{-3}–10^{2}	1–10^{2}	10–10^{2}	10–10^{2}
Size of bead [nm]	10^{-6}	10^{-6}–10^{-5}	0.03–0.3	0.5–1.0	0.5–1.0
Interactions	strong Coulomb Pauli principle	Coulomb Pauli	bonded terms, Coulomb repulsive, van der Waals		Coulomb repulsive, van der Waals
Scaling effort	$N_n^{\geq 3}$	$N_e^{\geq 3}$	N_a^{1-2}	N_b^{1-2}	N_b^{1-2}
Reduction number of degrees of freedom	10–100	10–100	2–5	2–10	
Reduction computational effort	$\geq 10^{3}$	$\geq 10^{3}$	2–25	2–100	

imental methodology of investigation. Only a very limited number of properties of a molecular system is actually accessible to measurement by experimental means. In addition, experiments generally provide averages over molecules (space) or time of observable quantities, while the distribution of such quantities over molecules or time remains inaccessible, in particular for microscopic length and time scales. Computational modeling and simulation can provide not only averages of any definable, even non-observable quantity, but also the spatial distribution and time series of such quantities. Another asset of computational modeling over experimentation is that it can be used to investigate mechanistic cause-effect relationships by individually changing model parameters that cannot be changed experimentally without affecting other parameters. The purpose of the present article is to review the key elements of multi-resolution modeling and simulation in chemistry and to illustrate its current status by application of multi-resolution simulation with the aim of interpreting catalytic differences observed between two structurally similar enzymes that are of considerable practical interest.

Wilfred F. van Gunsteren studied physics and law at the Free University, Amsterdam, receiving his Ph.D. in 1976. After post-doc years at the University of Groningen (1976–1978) and at Harvard University (1978–1980) he stayed till 1990 at the University of Groningen, since 1987 as Professor of Physical Chemistry. From 1987–1992 he served as part-time Professor of Computational Physics at the Free University, Amsterdam. Since 1990 he is Professor of Computer-Aided Chemistry at ETH Zurich. In 1986 he founded a software company, Biomos b.v., of which he still is the director. His main interest is the development of methodology to simulate the behavior of biomolecular systems.

2. Levels of Resolution

In chemistry, one may distinguish different levels of resolution of a theoretical or computational model, for example[1] the ones listed in Table 1. The interactions between particles of levels I and II are governed by quantum mechanics, while the interactions between particles of levels III–V are generally governed by classical statistical mechanics, that is, when quantum effects play a minor role in the process of interest. This is not the case if processes involve a change in electronic structure of the molecular system such as when bond formation or breaking during chemical reactions or photo-excitations are investigated. The number of degrees of freedom, particles or interaction sites will determine, together with the applicable equations of motion, that is, quantum or classical ones, the computational effort required, and thus the reduction of the effort that can be reached by coarse-graining, that is, eliminating particular degrees of freedom in the molecular system (Table 1) that are less relevant to the process of interest. Eliminating degrees of freedom generally leads to a decrease of the applicability of a model. For example, when coarse-graining from level II to level III, relaxation of electronically excited states of molecules is not covered by the model anymore. In addition, coarse-graining may lead to a loss of accuracy of a model, although for particular properties and types of models this need not be true. For example, the properties of liquid water at room temperature and pressure are more accurately described by the SPC model,[2] a level III model, than by a quantum-chemical model based on density-functional theory using the BLYP exchange-correlation functional,[3] a level II model. So, the choice of degrees of freedom to be kept in a model will depend on the properties and phase of interest of the molecular system.

A complete elimination or neglect of particular degrees of freedom of a molecular system is only warranted[1,4] if 1) they are of minor importance to the process or properties of interest, 2) they are very large in number in order to achieve a sufficiently large reduction of the computational effort to compensate the loss of accuracy, 3) the interactions governing

these degrees of freedom are weakly coupled to the interactions governing the other, remaining degrees of freedom of the molecular system, and 4) the interaction along these remaining degrees of freedom can be simply and accurately represented in the coarse-grained model.

The challenge of developing a model that spans two or more different levels of resolution is to find a physically correct balance between the various types of interactions in such a model. For instance, in the enzyme–substrate systems considered as example below, the substrate degrees of freedom are modeled at level II, the enzyme degrees of freedom are modeled at levels III and IV, and the aqueous solution at level V. This requires consistency and compatibility between the quantum, classical, and supra-molecular models that are combined. Combination of level II with levels III–V requires assumptions and approximations in regard to the compatibility of quantum and classical mechanical concepts. Combination of levels III–IV with level V requires assumptions and approximations in regard to the compatibility of different extents of energy and entropy present in the different degrees of freedom of a system in atomic level versus supra-molecular level descriptions.[1] In the following two sections, we briefly review the various combinations of models at the different levels of modeling II–V, and only highlight some conceptual and technical issues to be addressed in any multi-resolution model. In Section 3 we focus on models involving levels II and III–IV and in Section 4 on models involving levels III–IV and V. We do not intend to review the large variety of established models at a particular level of resolution, only indicate the choices to be made when combining different levels of resolution in one model.

3. Treatment of Quantum Effects in Molecular Simulation

Many biomolecular processes can be reasonably well described using purely classical-mechanical (CM) methods. However, when relatively light particles such as electrons or light nuclei such as the proton are involved, quantum effects will come into play, which can be modeled using a variety of techniques. Thereby more approximate techniques can be used for the proton or nuclei that are orders of magnitude heavier than the electron.

3.1. Quantum Effects

In chemistry, mainly quantum effects involving the motion of atoms, electrons and nuclei are of relevance. Which quantum effects are to be accounted for depends on the molecular degrees of freedom that are considered in a simulation. In principle, the energy of a molecular system is quantized, that is, the system can only adopt states (energy levels) with discrete values of energy. The electronic energy levels of a molecule are generally very widely spaced, which means that at physiological temperatures only the ground state will be populated. Thus in atomistic simulation, the atoms and molecules are generally considered to be in their electronic ground state and electronic degrees of freedom are not treated explicitly. Vibrational energy levels involving nuclear motion show much smaller energy differences than the electronic energy levels, in the order of $1600\,\mathrm{cm}^{-1}$ for bond-angle vibrations to $3700\,\mathrm{cm}^{-1}$ for bond-stretching vibrations. However, the thermal energy per degree of freedom in molecular simulation is in the range of $k_\mathrm{B}T$ ($200\,\mathrm{cm}^{-1}$ at 300 K), where k_B denotes the Boltzmann's constant and T the temperature, and is much smaller than the energy needed for the excitation of these vibrational modes. Thus molecular systems in classical simulation are assumed to be in their vibrational ground states with zero-point energy $1/2\,\hbar\omega$, where $\hbar$ is the Planck's constant divided by 2π and ω is 2π times the vibrational frequency. In classical simulation, this energy can be approximated by the energy of a classical harmonic oscillator. In case of bond-stretching vibrations the ground state is simply approximated by a constrained bond length which is a better approximation of the ground state than a classical harmonic oscillator.

For a system of quantum-mechanically described particles, the statistical properties of the distributions of states depend on whether the wave function describing the quantum state is symmetric (Bose–Einstein distribution) or antisymmetric (Fermi–Dirac distribution) with respect to an exchange of the coordinates of two particles occupying the same state (Table 2). If the number of 1-particle states, energy levels, is vastly larger than the number of particles and if the temperature is sufficiently high, these distributions can be closely approximated by a Boltzmann distribution as described by classical molecular simulation.

Table 2: Overview of main differences between quantum and classical mechanics as applied to molecular systems.

Quantum statistical mechanics (QM)	Classical statistical mechanics (CM)
• time-dependent Schrödinger equation: $i\hbar\frac{\partial\Psi(\vec{r}^N,t)}{\partial t}=\hat{H}\Psi(\vec{r}^N,t)$ with wave function Ψ	• Newton's equation: $\vec{f}_i=m_i\vec{a}_i \quad \vec{f}_i=-\frac{\partial V(\vec{r}^N)}{\partial \vec{r}_i}$
• Born-Oppenheimer approximation: nuclei and electrons decoupled	• interaction potential energy function $V(\vec{r}^N)$ is generally pair additive, but polarisation can be included
• probabilistic nature of wave function, $\lvert\Psi(\vec{r}_i,t)\rvert^2$ Uncertainty principle, no trajectory	• phase space trajectory $\vec{r}^N$, $\vec{p}^N$
• Bose-Einstein or Fermi-Dirac statistics (Pauli principle)	• Boltzmann statistics

Another quantum effect that may play a role arises from the Heisenberg uncertainty principle, $\Delta p_x \Delta q_x \geq 1/2\hbar$, where Δp_x is the uncertainty in the momentum p_x in the direction of an x-axis and Δx is the uncertainty in the position x of a particle along this axis. It states that the position and momentum of a particle cannot simultaneously be exactly determined. Thus, the quantum-mechanical wave function $\Psi(x,t)$ of a particle is interpreted in terms of a probability proportional to $|\Psi(x,t)|^2$ of a particle to occupy at time t a particular position along the x-axis. As a consequence defined trajectories of quantum particles do not exist and tunneling through a high energy barrier may occur, for example in the case of proton transfer between molecules.

When electronic degrees of freedom play an essential role in a molecular process, a quantum-mechanical treatment of such degrees of freedom is mandatory. This is the case when modeling changes of chemical bonds or electronic relaxation processes. Different levels of accuracy can be achieved, ranging from high accuracy for computationally highly expensive ab-initio methodology to computationally much cheaper semi-empirical and valence-bond methods (Table 3). If the rearrangement of electronic density in response to a change in the molecular environment is small, a classical treatment using atomic or molecular polarisability or charge transfer between atoms may be appropriate.

3.2. Quantum versus Classical Simulation

In purely quantum-mechanical molecular simulation the fundamental equation describing non-relativistically the motion of the molecular system, that is, the change of the quantum state in time, is the non-relativistic time-dependent Schrödinger equation. Its solution is a wave function $\Psi(\mathbf{r}^N,t)$ that is a function of the coordinates $\mathbf{r}^N$ of N particles (e.g. nuclei and electrons) and of time (see Table 2). Due to their much larger mass, nuclei move considerably slower than the light electrons justifying the Born–Oppenheimer approximation of separating nuclear from electronic motion, that is, writing the wave function Ψ as a product of a nuclear and an electronic wave function and solving the electronic time-independent Schrödinger equation for a given stationary external electric field caused by the nuclei. As mentioned above, exact determination of the position of a quantum particle is restricted by the uncertainty principle and a trajectory cannot be defined, only a probability of occurrence at a given position can be obtained.

In classical mechanics, using Cartesian coordinates and assuming that the potential energy is independent of the momenta, the motion of the system is fully described by Newton's equations (Table 2). The force on a particular degree of freedom is obtained as the negative gradient of the potential energy function describing the interactions within the system. The non-bonded part of this function is generally pairwise additive, although polarisation can be included in this description. Statistical-mechanically the trajectory of a classical simulation should represent a Boltzmann weighted ensemble.

A second classification of simulation methods is whether a single, for example, minimum-energy, structure or *configuration* of a molecular system is considered, or a particular, for example, canonical, *ensemble* of structures or configurations is investigated, or a dynamic sequence of configurations, that is, *trajectory* is investigated (Table 3). Examples of single configuration results for molecules are solutions of the time-independent Schrödinger equation for electronic degrees of

Table 3: Classification of molecular simulation theory and methods.

Quantum statistical mechanics	Classical statistical mechanics
1. *One single* configuration: Static	
QM: time-independent Schrödinger equation	**CM: energy minimization**
$T=0$	$T=0$
● ab-initio	● (empirical) force field
● density-functional	
● semi-empirical	
● empirical valence bond	
2. *Ensemble* of configurations: Equilibrium statistical mechanics	
QM: density matrices	**CM: Boltzmann sampling**
$T\neq 0$	$T\neq 0$
● path-integral simulation	● force field
	● Monte-Carlo simulation
	● molecular dynamics simulation
	● stochastic dynamics simulation
3. *Trajectories* of configurations: Dynamics	
QD: time-dependent Schrödinger equation	**CD: Newton, Lagrange, Hamilton, Langevin equations**
$T\neq 0$ + dynamics	$T\neq 0$ + dynamics
● wave function propagation	● force field
● density matrix evolution	● molecular dynamics simulation
● real-time path integral	● stochastic dynamics simulation
● Bohmian dynamics	
● surface hopping	

Table 4: Rough historical overview of molecular simulation methods and applications.

Method	Atoms/Nuclei	Electrons
1. Conventional Quantum Chemistry (1960 →)	**Fixed:** $T=0$ K **single configuration**	**QM:** Schrödinger: ab-initio (time-dependent) density-functional theory semi-empirical valence bond **ground + excited states**
2. Conventional Classical Simulation (1970 →)	**CM:** atoms Newton: $T=300$ K **dynamics**	**averaged out** **no reactions**
3. 'Ab-initio' Simulation (1985 →)	**CM:** nuclei Newton: $T=300$ K **dynamics**	**QM:** (as under 1.) **only ground state**
4. Surface Hopping Simulation (1990 →)	**CM:** nuclei Newton: $T=300$ K **dynamics**	**QM:** (as under 1.) **ground + excited states**
5. Hybrid QM/MM Simulation (1990 →)	**CM:** nuclei + atoms Newton: $T=300$ K **dynamics**	**QM:** (as under 1.) for few atoms **only ground state**
6. Path-Integral Simulation (1990 →)	**QM:** atoms **ensemble**	**QM:** (as under 1.) or **averaged out**
7. Hybrid QD/CD Dynamics Simulation (1993 →)	**QM:** few light atoms Schrödinger: **dynamics,** **ground + excited states** **CM:** many atoms Newton: $T=300$ K **dynamics**	**QM:** (as under 1.) for a few atoms or **averaged out**

freedom or a configuration obtained from energy minimization based on an empirical force field or molecular interatomic interaction function. Such static configurations commonly represent a system at temperature $T=0$. Since one is generally interested in a system at non-zero temperature, methods such as Monte-Carlo, molecular or stochastic dynamics simulation are used to generate a Boltzmann ensemble of configurations at a given thermodynamic state point.[5,6] If the degrees of freedom are governed by quantum mechanics, the corresponding simulation method generating an ensemble is path-integral simulation formulated in terms of density matrices.[7–9] If one is interested in dynamic properties of a molecular system, equations of motion are to be integrated, either those of Newton or Langevin for classical particles or those of Schrödinger in one form or the other: wave function propagation,[10–13] density matrix evolution,[14,15] real-time path evolution[16] or Bohmian dynamics.[17,18] The latter two strategies take advantage of the fact that under certain conditions the quantum behavior can be described in terms of classical particle distributions by mathematical equivalence.[6] Transitions to and from excited states can be mimicked using surface hopping[19–21] or multiple spawning techniques.[22–25]

3.3. Hybrid QM/MM Simulation: Methods

Conventional quantum chemistry came of age in the 1960s with the advent of generally programmable computers. Electronic degrees of freedom were treated to obtain the ground and excited states of molecules for fixed positions of the nuclei (Table 4). During the same decade classical dynamics simulation methods were explored for atomic liquids, followed by molecular liquids in the 1970s. The first very simple hybrid quantum-mechanical/molecular-mechanical (QM/MM) simulation in which the nuclei were treated as classical particles while the electronic degrees of freedom were treated quantum-mechanically using the semi-empirical molecular orbital formalism was published in 1976.[26] In addition, the system was spatially divided into a QM region for which the described treatment was applied and a so-called MM region in which the atoms were treated classically. In the 1990s such hybrid QM/MM simulations were further developed[27] and a variety of ways to treat the coupling and boundary between quantum-mechanically and molecular-mechanically treated particles was investigated.[28] So-called "ab-initio" molecular dynamics (MD) simulation was introduced in the 1980-ies: the nuclei of all atoms in the system are treated as classical dynamical (CD) particles while their valence electrons are treated by density-functional theory.[29]

Thus no spatial QM/MM boundary is to be defined. During the 1990s path-integral methods were used to study processes such as electron or proton transfer[30,31] or surface adsorption of light adsorbates[32] through classical simulation. The same decade saw the combination of wave function propagation, density matrix evolution and surface hopping techniques with classical dynamics in mixed quantum-classical simulations;[14,15,33] for reviews see Ref. [21,34,35]. With the increasing popularity of hybrid QM/MM schemes, path integral,[36,37] surface hopping,[38–42] and multiple spawning simulations[43] within a QM/MM framework have been reported.

The classification of methods in Table 3 and 4 enables a systematic description of the literature on hybrid QM/MM simulations: 1) which particles or degrees of freedom such as electrons or nuclei are treated quantum-mechanically and which are treated classical-mechanically, 2) whether a single configuration, an ensemble, or a trajectory is generated, and 3) which quantum or classical method is used to this end. Further distinctions are 4) the way the general coupling between quantum and classical particles is handled and 5) the handling of interactions at the boundary between the quantum-mechanically (QM) and molecular-mechanically (MM) treated regions of a system.

3.4. Hybrid QM/MM Simulation: Coupling and Boundary

The inherent incompatibility of the laws of quantum and classical mechanics at finite temperature complicates a consistent description of mixed quantum-classical systems. Due to their importance in the study of processes in biomolecular systems, especially enzymatic reactions, this section focuses on hybrid QM/MM methodology. Different methods to describe the coupling between the QM and MM regions and the interaction at the QM/MM boundary have been proposed. A straightforward classification scheme of electrostatic coupling models has been presented by Bakowies and Thiel.[28] Accordingly, the simplest model (model A) describes the coupling purely on the basis of classical mechanics where classical point charges are assigned to the atoms of the QM region for evaluation of the QM-MM electrostatic interactions. It lacks polarisation of the QM region and is referred to as mechanical embedding in the literature. The model which is most widely used for QM/MM coupling is called electrostatic embedding (model B) and accounts for the polarisation of the QM electron density by MM point charges through explicitly including them as external point charges in the quantum Hamiltonian. Polarised embedding (models C and D) additionally treat polarisation of the MM region induced by the electric field of the QM part. Although more sophisticated than model B in the sense of a balanced description of polarisation of the QM and MM regions, these methods are still under investigation and not yet standardly used.

A critical issue is the description of long-range QM-MM Coulomb interactions which has only recently been addressed in more detail.[44–53] Among the methods discussed rank charge scaling,[44] variational electrostatic projection,[47,48] generalized solvent boundary potential[46,51–53] and linear-scaling particle-mesh Ewald.[45] In conventional QM/MM schemes, non-bonded van der Waals interactions between the QM and MM regions are treated purely classical-mechanically.

An adequate description of the charge distribution at the QM/MM boundary is yet another issue to be considered. Overpolarisation of the QM electron density by short distances to classical partial charges is to be prevented. Most crucial in that respect is the treatment of covalent bonding across the boundary which leads to unpaired electrons for the corresponding QM atom at the frontier. Available methods to handle these issues range from the popular link-atom approach[54] with different variants to avoid overpolarisation to pseudo-potential methods and the use of frozen localized orbitals. For a detailed characterisation of the various methods see Ref. [50] and references therein. For reviews on QM/MM methodology see Refs. [50,55–57].

4. Supramolecular Models for Combination with Atomic Models in Molecular Simulation

We first consider coarse-graining from level III to level IV, that is, the elimination of intra-molecular degrees of freedom, and then the additional complications when considering coarse-graining from levels III–IV to level V.

4.1. Elimination of Intramolecular Degrees of Freedom of Macromolecules

By treating the aliphatic CH, CH_2, and CH_3 groups as single, so-called united atoms,[58] the number of interaction sites is substantially reduced, 4- to 16-times fewer pairwise non-bonded interactions need to be evaluated between such groups. A speed-up of roughly a factor 10 is obtained at a minor loss of atomic detail.[1]

Another straightforward way to reduce intramolecular degrees of freedom of a macromolecule is to constrain its bond lengths,[4] which does not significantly affect the motion along the other degrees of freedom of the macromolecule[59] and allows a reduction of the computational effort by a factor of 4.[4] Bond-angle degrees of freedom in macromolecules such as proteins cannot be constrained without significantly affecting the torsional-angle and other motions.[59,60] If no torsional-angle degrees of freedom are present in a molecule, for example, as for small solvent molecules, bond-angle degrees of freedom are standardly constrained to obtain a rigid, computationally fast model for the many solvent degrees of freedom.

Coarse-graining from level III to level IV, elimination of polar hydrogens or non-hydrogen atoms, is a challenge in particular for inhomogeneously composed macromolecules such as proteins because of their heterogeneity in terms of atom types. The scale invariance that lies at the heart of the renormalisation group approach to coarse-graining of largely homogeneous polymers is not observed for biopolymers such as proteins, polynucleotides and carbohydrates, which are composed of many different, complex structural units that are

connected thorough different types of interactions. When eliminating intra-molecular degrees of freedom, the basic geometry and balance between the various interactions must be preserved in order to avoid losing characteristic features of such molecules.[61] In addition, entropy plays a non-negligible role in molecular rearrangements of systems containing biomolecules in aqueous solution. This means that the loss of entropy due to the elimination of degrees of freedom must be compensated for by a reduction in energy of the remaining degrees of freedom in order to maintain the relevant free energy differences when going from models at level III to models at levels IV or V. As Table 1 shows, the reduction of the computational effort between levels III and IV is rather modest compared to that between levels II and III. This indicates that coarse-graining from level III to level IV, although currently rather popular, does not pay off for macromolecules such as proteins, polynucleotides or carbohydrates: a limited decrease in the number of interaction sites is achieved at the cost of losing the characteristic features of such molecules in terms of intra-molecular interactions, entropy and interactions with the solvent.[6] Only for lipids with their relatively long homogeneous aliphatic carbon tails one may be able to preserve the dominant amphiphilic character of these molecules upon eliminating intra-molecular interaction sites. Because of the abundance of lipids in membranes, the computational gain may off set the loss of atomic detail.

4.2. Elimination of Supramolecular Degrees of Freedom of the Solvent

A much used strategy to reduce the computational effort of simulating a macromolecular solute in solution is to eliminate all solvent degrees of freedom, leading to implicit solvation models,[62] in which the effect of the solvent on the motion along the solute degrees of freedom is modeled as a function of the solute coordinates only. When applied to biomolecular macromolecules in aqueous solution, such implicit solvation models are rather approximate to put it mildly, because of the difficulty of mimicking the essential many-body hydrophobic effect and ionic solvation effects without the water and ionic degrees of freedom.[1] It would therefore be more promising to keep some solvent degrees of freedom. Because of the computational dominance of solvent-solvent and solute-solvent interactions over solute–solute interactions at the atomic level, coarse-graining of solvent degrees of freedom, in particular supra-molecular coarse-graining in which more than one solvent molecule is subsumed into a supra-molecular solvent bead or particle, promises a substantial gain in computational efficiency.

In the case of water, coarse-graining from level III to level V should preserve the major thermodynamic quantities such as heat of vaporisation, density, free energy, entropy and surface tension, the dielectric screening properties reflected in the large static dielectric permittivity of about 78 at room temperature and pressure, and the hydrogen-bonding capacity of liquid water as much as possible.[63] This goal cannot be reached if a supra-molecular water bead is modeled as a Lennard–Jones particle without charge.[64–66] A supra-molecular water bead needs at least two interaction sites with charges, or a single ideal dipole, in order to represent the dielectric screening properties of liquid water. Such a supra-molecular water model representing five water molecules by one water bead containing two interaction sites has been shown to mimic the solvation properties of water rather well,[63] while being one to two orders of magnitude more computationally efficient than a three-site atomistic water model such as the SPC model.

When developing a supra-molecular level V model a few technical issues emerge that play no role when formulating atomic (level III) or supra-atomic (level IV) models. These concern the choice of dielectric permittivity to be used in the expression for the electrostatic energy, and accounting for the number of molecules represented by a single supra-molecular bead when comparing simulated values of particular quantities with experimentally measured ones.[1,63,67]

4.3. Hybrid Atomistic/Supramolecular Simulation: Methods

Generally, a model developed at a particular level of resolution is only used at that resolution. However, as is the case for QM/MM simulations combining models at level II and level III in one simulation, combination of models at levels III, IV and V also enhances the computational efficiency of the simulation of systems with many degrees of freedom. Such a combination of different levels of resolution, called multi-graining, can take different forms.[1]

1. The simulation switches between the levels of resolution in time, which can be done in two ways.
a) The simulation is performed at the supra-atomic level of resolution and particular configurations are mapped back to the atomic level of resolution.[68–76]
b) A coupling parameter λ is introduced that defines a path between the atomic-level model (e.g. $\lambda = 0$) and the supra-atomic or supra-molecular model (e.g. $\lambda = 1$), which allows a smooth switching between different levels of resolution, as for example, in Hamiltonian replica-exchange simulation.[77–79]
2. The system contains a mixture of atomic level and supra-atomic or supra-molecular particles which can be done in two ways.
a) The space occupied by the system is divided into an atomic level resolution region and a supra-atomic or supra-molecular region with a small buffer region in which the particles may change from one resolution level to the other.[80–85]
b) The particles are either atomic or supra-atomic or supra-molecular and may freely mix in space. The resolution of the particles is thus fixed during the simulation.[86–93] This scheme is usually adopted in level II/level III hybrid QM/MM simulations.

4.4. Hybrid Atomic/Supramolecular Simulation: Coupling and Boundary

When developing a model that combines two or more levels of resolution, it is mandatory to define the interactions between the particles of different levels of resolution such that a physically correct distribution of energy and entropy over the different levels of resolution is obtained. For example, the multi-resolution simulations to be discussed in the next section comprise three levels of resolution: level II for the substrates, level III (apart from united atoms) for the enzymes, and level V for the solvent. This means that interactions of the type II/III, II/V and III/V have to be defined such that they are as much as possible thermodynamically, dielectrically and geometrically compatible with the given intra-resolution level interactions of the type II/II, III/III and V/V. The difference in particle size, number of interaction sites and strength of the interaction at the different resolution levels as reflected by the different model parameter values must be accounted for. In the case of atomic/supra-molecular interactions the parameters of the non-bonded interaction can be chosen such that particular properties of mixtures of atomic and supra-molecular particles representing the same liquid are reproduced.[91] For a mixture of atomic level (AL) with supra-molecular level (SM) water particles this was done by using a value, $\varepsilon_{cs}^{AL/SM} = 2.3$, for the dielectric permittivity in the electrostatic interaction in the cut-off sphere between atomic level and supra-molecular level particles in the electrostatic potential energy terms for the supra-molecular water interactions.[91] For the AL-SM Lennard–Jones interactions combination rules applied to AL-AL and SM-SM interaction parameters can be used, and values can be scaled for particular types of combinations of particles if necessary.[92]

If an atomic level solute is solvated in an atomic-level layer of solvent molecules, which is surrounded by supra-molecular solvent particles, the atomic-level solvent can be kept close to the solute by using a weak distance restraining potential energy term in the Hamiltonian.[93]

5. Multi-Resolution Simulation of Four Enzyme–Substrate Complexes

Elucidation of the driving forces and mechanisms of catalysis by enzymes is one of the more intriguing challenges in biochemistry and has practical consequences. Obviously, the observed speed-ups of the catalysed chemical reactions compared to the same reaction in the gas or solution phase are due to the protein environment of the reactant or substrate bound in the active site of the enzyme. The catalytic process can be split into four steps: 1) Binding of the substrate in the active site of the enzyme thereby possibly changing the dominant conformation of the substrate such as to favour the reaction. 2) The reactant mounting the activation barrier towards the transition state of the reaction. The probability of reaching transition-state conformations can be enhanced by positional fluctuations of the enzyme or solvent that induce fluctuations in the forces on the atoms of the substrate. 3) Formation of the product. 4) Release of the product from the binding pocket in the enzyme. Unfortunately, only a few observable quantities characterising the different steps of the catalytic process can be measured experimentally: 1) Overall reaction rate. 2) The free enthalpy of binding of ligands or transition-state analogs, but not of the substrate. 3) Atomic level resolution structure of the apo-enzyme or the enzyme with ligands bound, but not the substrate. These three types of quantities can be measured as function of thermodynamic state point for various mutants of the enzyme and for different ligands. In view of this limited set of experimental data, the detailed driving forces and mechanism of the catalytic process cannot be unambiguously determined by experimental means.

From a theoretical and computational point of view, quantum-statistical mechanics provides the theory by which catalytic processes should be explained. In the computational practice though, one has to resort to severe simplifications and approximations of the quantum-statistical mechanical equations in order to obtain a result for an enzyme–substrate complex, even when using the fastest computers available. For reviews of the state of the art we refer to Ref. [55–57]. The reacting moieties are treated quantum-chemically, that is, at level II resolution, while the enzyme or a larger part of it is treated classical molecular-mechanically, that is, at level III resolution. The aqueous solvent is often represented by a continuum electrostatic reaction field or explicitly by a small spherical region of level III water molecules. If the quantum-chemically treated region is larger and a higher-level quantum-chemical model such as density-functional theory is used, the time span covered by such QM/MM simulations is of the order of tens of picoseconds. This is too short to sample the fluctuations in the forces exerted by the enzyme and solvent on the atoms of the substrate. If one is interested in the effects of the protein and water environment upon the substrate, one may choose to treat the reactant semi-empirically and include a supra-molecular solvent in order to enhance the sampling using a computationally efficient way to represent the solvent degrees of freedom that still accounts for dielectric and entropic effects. This approach is illustrated below through an investigation of the differences in catalytic activity observed for two structurally similar enzymes.

5.1. Chorismate Mutase from Escherichia coli and Isochorismate Pyruvate Lyase from Pseudomonas aeruginosa

Chorismate mutase (CM) from *Escherichia coli* (EcCM) catalyses the transformation of chorismate (chr) to prephenate (Figure 1).[94] Isochorismate-pyruvate lyase (IPL) from *Pseudomonas aeruginosa* (PchB) catalyses the transformation of isochorismate (ichr) to salicylate and pyruvate (Figure 1).[95] Crystal structures of both enzymes have been derived from X-ray diffraction patterns[96,97] and show that both enzymes have the same fold (Figure 2), despite their low sequence homology of 21% (Figure 3).[95,98] PchB also shows some catalytic activity with respect to chorismate, but with a lower efficiency than that of EcCM. The chorismate mutase activity of natural PchB could be improved by amino-acid

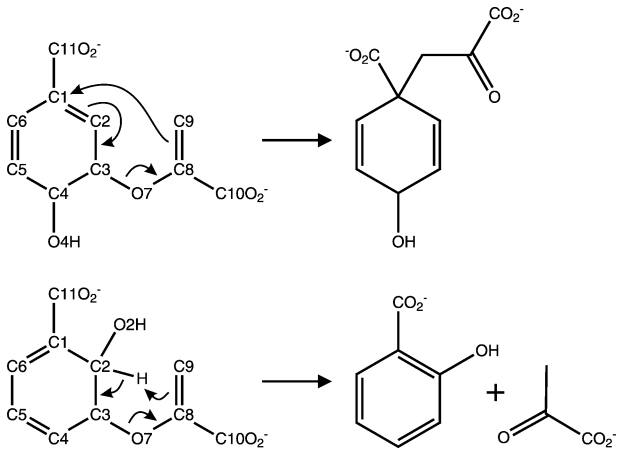


Figure 1. Transformations of chorismate to prephenate (CM-catalyzed reaction 1) and isochorismate to salicylate and pyruvate (IPL-catalyzed reaction 2).

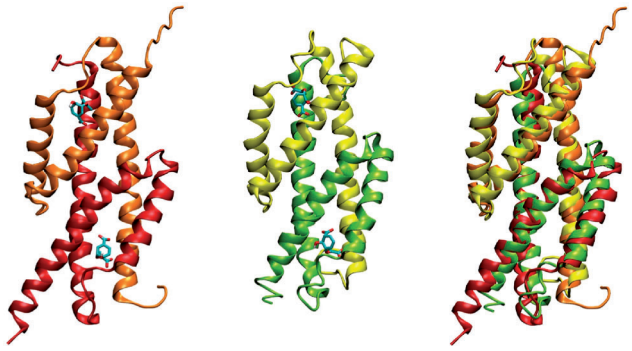


Figure 2. The simulated enzymes. Configurations taken after energy minimization of the representative X-ray structures for the EcCM-chr (left) and PchB-ichr (middle) systems together with the substrates. On the right the CM and IPL enzymes are superimposed. EcCM is shown in orange (monomer 1) and red (monomer 2) and PchB in yellow (monomer 1) and green (monomer 2). For the two substrates, carbon atoms are depicted in cyan, oxygen atoms in red, and hydrogen atoms in white.

mutagenesis.[98] This successful engineering of an enzyme's catalytic activity led to similar attempts to induce IPL activity in EcCM by mutagenesis. However, up till now no trace of IPL activity was found in EcCM mutants engineered to obtain a more PchB-like active site (Don Hilvert, personal communication). These experimental observations raised the question whether a model calculation would be able to detect the reasons for the differential catalytic activity of these two enzymes with regard to their respective substrates despite their similar spatial folds. The key must lie in the different side chains surrounding the active sites, leading to different forces on the atoms of the substrates, or in different motional patterns of the enzymes. So four different enzyme–substrate complexes in aqueous solution were simulated for 5 ns using multi-resolution molecular dynamics (MD) (Table 5). One of the substrates was represented quantum-chemically (level II), the other substrate and the enzymes molecular-mechanically (level III), and the solvent by a supra-molecular water model (level V). The obtained trajectories were analyzed and compared in terms of substrate conformations and forces by the enzymes on substrate atoms.

```
          1(1)        11(1)       21(1)       31(1)       41(1)
EcCM: MTSENPLLAL RekISALDEK LLALLAERRE LAVEVGKAkL LSHRPVRDID

          51(1)       61(1)       71(1)       81(1)       91(1)
      RERDLLERLI TLGKAHHLDA HYITRLFQLI IEDSVLTQQA LLQQHLNKINP

          1(2)        11(2)       21(2)       31(2)       41(2)
      MTSENPLLAL REKISALDEK LLALLAERRE LAVEVGKAKL LSHRPVRDID

          51(2)       61(2)       71(2)       81(2)       91(2)
      RERDLLERLI TLGKAHHLDA HYITRLFQLI IEDSVLTQQA LLQQHLNKINP

          1(1)        11(1)       21(1)       31(1)       41(1)
PchB: MKTPEDCTGL ADIREAIDRI DLDIVQALGR RMDYVKAASR FKAS.EAAIPA

          51(1)       61(1)       71(1)       81(1)       91(1)
      PERVAAMLPE RARWAEENGL DAPFVEGLFA QIIHW.YIAEQ IKYWRQTRGAA

          1(2)        11(2)       21(2)       31(2)       41(2)
      MKTPEDCTGL ADIREAIDRI DLDIVQALGR RMDYVKAASR FKAS.EAAIPA

          51(2)       61(2)       71(2)       81(2)       91(2)
      PERVAAMLPE RARWAEENGL DAPFVEGLFA QIIHW.YIAEQ IKYWRQTRGAA
```

Figure 3. Sequences of the dimeric enzymes EcCM and PchB with residue sequence numbers and alignment. The residues surrounding the first active site are indicated in bold, those surrounding the second active site in bold italics. The monomer sequence number is given in parentheses.

Table 5: Overview of MD simulations.

Simulation label	Enzyme	Substrate	Number of water beads	Box volume [nm^3]
EcCM-chr	*E. coli* CM	Chorismate	8467	1328
EcCM-ichr	*E. coli* CM	Isochorismate	8469	1329
PchB-chr	*P. aeruginosa* IPL	Chorismate	9444	1479
PchB-ichr	*P. aeruginosa* IPL	Isochorismate	9444	1480

We do not intend to explain catalytic activity in terms of differences in (free) energy difference of the transition state and the reactant state between the different environments of enzyme, aqueous solution, or gas phase, for which one would have to apply umbrella sampling along a hypothetical transition pathway[99,100] leading to a hypothetical transition state.[101,102] Neither do we intend to explain catalytic activity in terms of near attack configurations of the substrate.[103–105] We only investigate the conformational ensembles of the four enzyme–substrate complexes to extract indications of the effect of the different protein environments on the substrates. Our investigation differs thus in a number of aspects from other computational studies of chorismate mutase catalytic activity: 1) We consider EcCM and PchB, while most computational studies concern CM from *Bacillus subtilis*.[99,101,102,104–108] 2) We investigate substrate binding to EcCM and PchB, while most other studies aim at explaining CM or IPL catalysis.[100] 3) We do not calculate a potential of mean force along a reaction coordinate, but analyse enzyme–substrate configurational ensembles.

5.2. Molecular Model and Computational Procedure

Both EcCM and PchB are active as dimers and thus have two active sites or substrate binding sites which are surrounded by residues of both monomers (Figure 2). The detailed atomic structures of the two active sites of a dimer show differences according to the X-ray studies.[96,97] In our multi-resolution model, one of the substrates (first active site) was treated quantum-chemically using the semi-empirical OM3 Hamiltonian[109] as implemented in the MNDO program.[110] The other substrate and the enzymes were modeled using the GROMOS force field 45A3[111] for biomolecular systems. The force-field parameters for the substrate in the second binding pocket are specified in Ref. [112]. The amino acid sequences and their alignment are shown in Figure 3. The ionisable groups were set to their protonated or deprotonated state according to the standard pK_a values of amino acids and a pH of 6.8. The histidine residues were protonated at N_δ for histidines, 66(2), 67(1), 71(1,2) and 95(1) and at N_ε for histidines 43(1,2), 66(1), 67(2) and 95(2) of EcCM and at N_ε for histidine 84(1,2) of PchB. The solvent was modeled using a CG water model[63] in which a supra-molecular bead represents five water molecules. The interactions between the QM substrate and its MM environment consisting of protein and supra-molecular water was modeled as described in Ref. [113]. A cut-off radius $R_{QM/MM} = 1.4$ nm around each atom of the QM substrate determined the partial charges of the enzyme and water beads to be included in the QM Hamiltonian at every MD step. The interactions between the supra-molecular solvent and the substrates and protein were modeled as described in Refs. [91–93]. For the electrostatic interactions within the cut-off sphere for non-bonded interactions a relative dielectric permittivity of $\varepsilon_{cs} = 1$ was used for interactions within the substrate and enzyme, a value $\varepsilon_{cs}{}^{mix} = 2.3$ for interactions between the supra-molecular water and the enzyme and substrate atoms, and a value $\varepsilon_{cs} = 2.5$ for supra-molecular water–water interactions.

The simulations were carried out using the GROMOS biomolecular simulation software,[114–116] which can be coupled[113] to executables of the quantum-chemical software packages MNDO[110] and TURBOMOLE.[117]

The respective X-ray structures of dimers of EcCM (PDB entry 1ECM)[96] and PchB (PDB entry 3HGX)[97] were taken as starting structures. The ligands present in the active sites were replaced by the substrates chorismate or isochorismate to obtain the four enzyme–substrate systems of Table 5. The substrates were energy minimized while keeping the protein atoms harmonically restrained with a force constant of 2.5×10^4 kJ mol^{-1} nm^{-2}.

The solute was solvated with supra-molecular water beads in a cubic box with a minimum distance between the non-hydrogen protein atoms and the box wall of 1.6 nm (Table 5). Using periodic boundary conditions the water was energy-minimized and equilibrated while keeping the solute atoms positionally restrained, as described in Ref. [92].

The MD simulations were carried out at a temperature of 310 K and a pressure of 1 atm. Solute and solvent were separately coupled to the heat bath using the weak coupling method[118] and a temperature coupling time $\tau_T = 0.1$ ps, a pressure coupling time $\tau_p = 0.5$ ps and an isothermal compressibility $\kappa_p = 4.575 \times 10^{-4}$ (kJ mol^{-1} nm^{-3})$^{-1}$. Newton's equations of motion were integrated using the leap-frog scheme[119] with a time step of 0.5 fs. The bond lengths in the enzyme were constrained to the ideal values applying the SHAKE algorithm[120] using a relative geometric precision of 10^{-4}. Evaluating the non-bonded interactions, a triple-range scheme was used with a short-range cut-off radius of 1.4 nm, an intermediate-range cut-off radius of 2.0 nm and an update frequency of 20 time steps for the short-range pairlist and intermediate-range interactions. A reaction field long-range force[121] was applied to the protein atoms and water beads using the experimental relative dielectric permittivity $\varepsilon_{rf} = 78.5$. The systems were equilibrated for 25 ps. A picture of one of the simulated systems is presented in Figure 4.

The analyses were performed on the ensembles of system configurations extracted at 0.25 ps time intervals from the simulations.

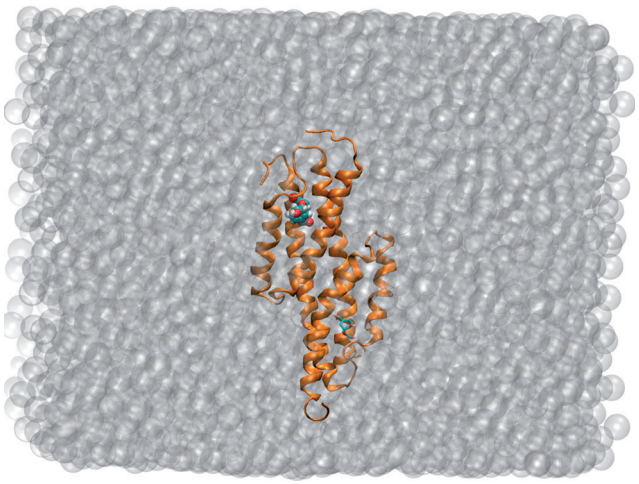


Figure 4. Picture of one of the four simulated enzyme–substrate combinations in supra-molecular water. QM substrate in bold cyan, MM substrate in licorice cyan, atomic-level enzyme in orange, and supra-molecular water beads in transparent gray.

The projection f_{ij} of the force $\vec{f}_i$ exerted by the enzymes on an atom i of a substrate along the direction to atom j of the substrate (Figure 6), was calculated as

$$f_{ij} = \frac{\vec{f}_i \cdot \vec{r}_{ij}}{|\vec{r}_{ij}|}, \qquad (1)$$

where $\vec{r}_{ij} = \vec{r}_i - \vec{r}_j$. The forces on the atoms of the QM substrate exerted by the enzymes were recalculated from the MD trajectories by applying the averaged OM3 net charges (Table S1, Supporting Information) for the QM substrate and the van der Waals parameters of the GROMOS MM force field (45A3) in the GROMOS expression for the Coulomb and van der Waals non-bonded forces.

5.3. Conformational Analysis of the Substrates using Atom–Atom Distances

For the EcCM catalyzed reaction 1 to occur, the C3–O7 bond of chorismate has to be broken and a bond has to be formed between carbon atoms C1 and C9. For the PchB catalyzed reaction 2 to occur, the C3–O7 bond of isochorismate has to be broken as well and a hydrogen at position C2 has to be abstracted by carbon C9 (Figure 1). The distributions of distances C1–C9 and C2–C9 during the course of the MD simulations are shown in Figure 5. The C1–C9 distances in chorismate (first column), which should be smaller for reaction 1 to occur, are for the QM substrate (first row) and for the MM substrate (second row) in the EcCM simulation indeed smaller than in the PchB simulation (third row for the QM substrate and fourth row for the MM substrate). Even a short time period during which the C1 and C9 atoms come close to each other would be enough for electron transfer to occur and for a new bond to be formed. Consequently, it is also of interest to evaluate the smallest distance between the atoms of interest in the simulation. The smallest C1–C9 distance measured in chorismate is 0.32 nm for the EcCM simulation and 0.45 nm for the PchB simulation. Thus EcCM brings the substrate chorismate in a conformation where its transformation to the product prephenate is more likely to occur, while PchB does not.

The C2–C9 distances in isochorismate (fourth column), which should be smaller for the hydrogen on carbon C2 to be abstracted and for reaction 2 to occur, are indeed smaller in the PchB simulation with mean values of 0.41 nm and 0.38 nm for the QM and MM substrates respectively than in the EcCM simulation with mean values of 0.42 nm and 0.41 nm. Thus PchB brings the substrate isochorismate in a conformation where its transformation to the products salicylate and pyruvate is more likely to occur, while EcCM does not.

The distance analysis shows that EcCM and PchB both fail to bring the "non natural" substrate, isochorismate for EcCM (upper two panels, fourth column) and chorismate for PchB (lower two panels, first column) in a conformation where the "non natural" reaction is more likely to occur.

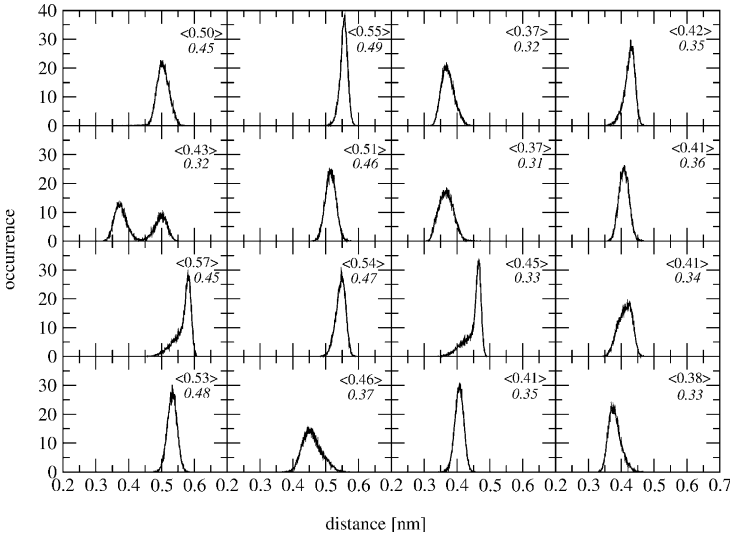


Figure 5. Distribution of the distances between atoms C1–C9 (first two columns) and C2–C9 (last two columns) for the substrates chorismate (first and third columns) and isochorismate (second and fourth columns) in the four simulations. For each enzyme, two sets of distances are given, as there are two substrates, one in each catalytic pocket of the dimer, from top to bottom the two active sites of EcCM (first row QM substrate, second row MM substrate) and of PchB (third row QM substrate, fourth row MM substrate). The average values, $\langle ... \rangle$, and the minimal distances (in nm) are given in the top right corners.

5.4. Forces Exerted on the Substrates by the Enzymes

The forces on atoms C1 and C9 of chorismate and atoms C2 and C9 of isochorismate projected along the lines connecting these atoms and exerted by the enzymes are shown in Figure 6. According to Equation (1), a negative

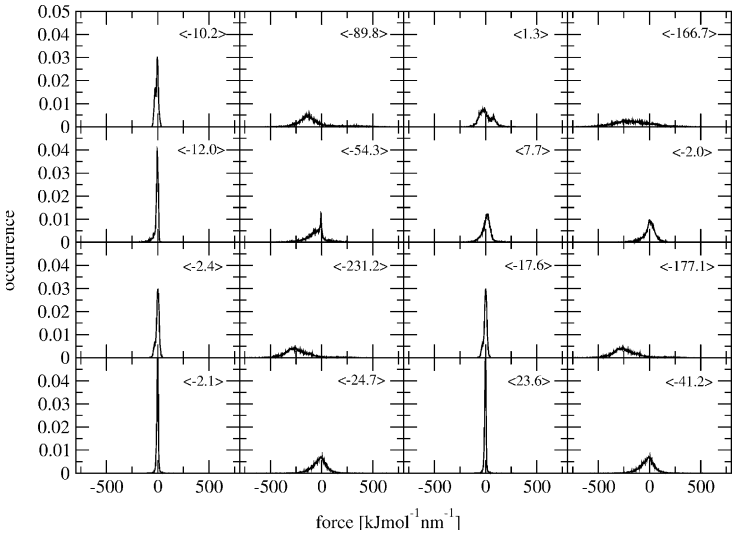


Figure 6. Distribution of the projected forces f_{C1-C9} (first column), f_{C9-C1} (second column), f_{C2-C9} (third column) and f_{C9-C2} (fourth column) exerted by the enzyme for the substrates chorismate (first and second columns) and isochorismate (third and fourth columns) in the four simulations. For each enzyme, two sets of forces are given, as there are two substrates, one in each catalytic pocket of the dimer, from top to bottom for the two subunits of EcCM (first row QM substrate, second row MM substrate) and of PchB (third row QM substrate, fourth row MM substrate). The average values are given (in $kJ\,mol^{-1}\,nm^{-1}$) in the top right corners.

force f_{ij} indicates that it pushes atom i towards atom j. The more negative f_{ij}, the more the substrate is forced towards a configuration favourable for the reaction. The forces on chorismate (left columns), are for the C1 atom (left column) more negative for EcCM (upper rows) than for PchB (lower rows). For the C9 atom (second column) of the QM substrate (first and third row) this observation does not hold, which may be related to the approximations made when using the average OM3 net charges from the MD trajectories. The forces on isochorismate (right columns) are for PchB (lower rows) more negative than for EcCM (upper rows) except for those on the C2 atom of the MM substrate. Overall, there is a slight preference of each substrate for its own enzyme. More detailed analyses of the forces on the substrate atoms exerted by the different residues of the enzymes may hold the key to the origin of the differential catalysis of EcCM and PchB, but ly beyond the scope of the present article.

6. Discussion

Any theoretical-computational modeling endeavour to interpret experimental observations involves a choice of 1) essential degrees of freedom, of interactions, that is, functional form and parametrization, governing the motion along these degrees of freedom, 2) of a method to generate a Boltzmann ensemble or dynamic trajectory of configurations of the degrees of freedom, and 3) of the way the interactions with the outside world are represented. In chemistry, essential degrees of freedom can be electronic, nuclear, atomic, molecular and supra-molecular. At each of these levels of resolution different assumptions and approximations are involved to render a model computationally feasible. However, for the six reasons mentioned in the Introduction, the accuracy and predictive potential of computational modeling has its limitations, in particular when investigating the behavior of complex many-body molecular systems which is determined by electronic as well as molecular degrees of freedom with sizable electrostatic and entropic contributions characterising the different states of the systems. In such cases, the development of a model that comprises different levels of resolution, in space or in time, is a viable but rather challenging approach to enhance computational efficiency without loosing too much accuracy. In the example given here, the use of a supra-molecular water model for the solvent instead of an atomic one resulted in a speed-up of a factor 6. This made the three-level resolution simulation three times faster than the single-level, fully atomic-level resolution one.[112]

Any computational model will yield results, but these only make sense when the physico-chemical basis of the result, the modeled mechanism, corresponds to reality. This will not be the case when a correct mechanism is clouded by using too crude assumptions and approximations. In such a case only pretty pictures remain.

The present article aims at presenting an overview of the challenges posed by the development of multi-resolution models to represent chemical processes. Although at each separate level of resolution, quantum-chemical, classical-mechanical, atomic, supra-atomic, or supra-molecular, more or less robust models have been developed over the past decades, their integration into a multi-resolution model is a more difficult task because of the different extents of energy and entropy and particle sizes involved at the different levels of resolution. In addition, quantum mechanics has to be married to classical mechanics which liaison still involves unresolved issues. Here, an overview was given of the choices regarding assumptions and approximations in regard to degrees of freedom and interactions essential for a particular chemical phenomenon or process of interest. The example of the investigation of the differential catalytic activity of two structurally similar enzymes with respect to their respective mutual substrates using multi-resolution molecular dynamics simulation only serves to demonstrate the usefulness of molecular modeling, despite its rather approximate quality, with respect to questions that are of practical biochemical interest but cannot be answered by experimental means.

We did not intend to give an overview of 125 years of theoretical-computational chemistry, a clearly impossible task, but to offer a birthday present to *Angewandte Chemie* who has significantly stimulated the development of chemistry, in the form of an illustration of the state of the art in theoretical-computational chemistry and a description of some of the many challenges in this field of research that still have to be taken on.

This work was financially supported by the National Center of Competence in Research (NCCR) in Structural Biology, the Swiss National Science Foundation (grant 200020-137827), and the European Research Council (grant 228076).

Received: July 9, 2012
Revised: September 12, 2012
Published online: February 18, 2013

[1] S. Riniker, J. R. Allison, W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423 – 12430.
[2] A. Glättli, X. Daura, W. F. van Gunsteren, *J. Chem. Phys.* **2002**, *116*, 9811 – 9828.
[3] B. Guillot, *J. Mol. Liq.* **2002**, *101*, 219 – 260.
[4] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Phys.* **1977**, *34*, 1311 – 1327.
[5] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Clarendon, Oxford, **1987**.
[6] H. C. J. Berendsen, *Simulating the Physical World: Hierachical Modeling from Quantum Mechanics to Fluid Dynamics*, Cambridge University Press, Cambridge, **2007**.
[7] R. P. Feynman, A. R. Hibbs, *Quantum Mechanics and Path Integrals*, McGraw-Hill, New York, **1965**.
[8] D. Chandler, P. G. Wolynes, *J. Chem. Phys.* **1981**, *74*, 4078 – 4095.
[9] B. J. Berne, D. Thirumalai, *Annu. Rev. Phys. Chem.* **1986**, *37*, 401 – 424.
[10] G. G. Balint-Kurti, R. N. Dixon, C. C. Marston, *Int. Rev. Phys. Chem.* **1992**, *11*, 317 – 344.
[11] R. Kosloff, *Annu. Rev. Phys. Chem.* **1994**, *45*, 145 – 178.
[12] N. Balakrishnan, C. Kalyanaraman, N. Sathyamurthy, *Phys. Rep.* **1997**, *280*, 79 – 144.
[13] H. D. Meyer, G. A. Worth, *Theor. Chem. Acc.* **2003**, *109*, 251 – 267.
[14] J. Mavri, H. J. C. Berendsen, W. F. van Gunsteren, *J. Phys. Chem.* **1993**, *97*, 13469 – 13476.
[15] H. J. C. Berendsen, J. Mavri, *J. Phys. Chem.* **1993**, *97*, 13464 – 13468.
[16] N. Makri, *Comput. Phys. Commun.* **1991**, *63*, 389 – 414.
[17] D. Bohm, *Phys. Rev.* **1952**, *85*, 166 – 179.
[18] D. Bohm, *Phys. Rev.* **1952**, *85*, 180 – 193.
[19] J. C. Tully, R. K. Preston, *J. Chem. Phys.* **1971**, *55*, 562.
[20] J. C. Tully, *J. Chem. Phys.* **1990**, *93*, 1061 – 1071.
[21] M. Barbatti, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 620 – 633.
[22] T. J. Martinez, M. Ben-Nun, R. D. Levine, *J. Phys. Chem.* **1996**, *100*, 7884 – 7895.
[23] T. J. Martinez, M. Ben-Nun, G. Ashkenazi, *J. Chem. Phys.* **1996**, *104*, 2847 – 2856.
[24] T. J. Martinez, R. D. Levine, *Chem. Phys. Lett.* **1996**, *259*, 252 – 260.
[25] A. Toniolo, C. Ciminelli, M. Persico, T. J. Martinez, *J. Chem. Phys.* **2005**, *123*, 234308.
[26] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227 – 249.
[27] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700 – 733.
[28] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580 – 10594.
[29] R. Car, M. Parrinello, *Phys. Rev. Lett.* **1985**, *55*, 2471 – 2474.
[30] R. Egger, C. H. Mak, *J. Chem. Phys.* **1993**, *99*, 2541 – 2549.
[31] D. H. Li, G. A. Voth, *J. Phys. Chem.* **1991**, *95*, 10425 – 10431.
[32] C. Chakravarty, *J. Phys. Chem. B* **1997**, *101*, 1878 – 1883.
[33] S. Klein, M. J. Bearpark, B. R. Smith, M. A. Robb, M. Olivucci, F. Bernardi, *Chem. Phys. Lett.* **1998**, *292*, 259 – 266.
[34] N. Makri, *Annu. Rev. Phys. Chem.* **1999**, *50*, 167 – 191.
[35] G. Pierdominici-Sottile, S. F. Alberti, J. Palma, *Adv. Quantum Chem.* **2010**, *59*, 247 – 282.
[36] D. T. Major, M. Garcia-Viloca, J. L. Gao, *J. Chem. Theory Comput.* **2006**, *2*, 236 – 245.
[37] J. Gao, K.-Y. Wong, D. T. Major, *J. Comput. Chem.* **2008**, *29*, 514 – 522.
[38] L. V. Schäfer, G. Groenhof, M. Boggio-Pasqua, M. A. Robb, H. Grubmüller, *PLoS Comput. Biol.* **2008**, *4*, 1 – 14.
[39] C. Ciminelli, G. Granucci, M. Persico, *Chem. Phys.* **2008**, *349*, 325 – 333.
[40] M. Ruckenbauer, M. Barbatti, B. Sellner, T. Muller, H. Lischka, *J. Phys. Chem. A* **2010**, *114*, 12585 – 12590.
[41] O. Weingart, P. Altoe, M. Stenta, A. Bottoni, G. Orlandi, M. Garavelli, *Phys. Chem. Chem. Phys.* **2011**, *13*, 3645 – 3648.
[42] Z. G. Lan, Y. Lu, E. Fabiano, W. Thiel, *ChemPhysChem* **2011**, *12*, 1989 – 1998.
[43] A. M. Virshup, C. Punwong, T. V. Pogorelov, B. A. Lindquist, C. Ko, T. J. Martinez, *J. Phys. Chem. B* **2009**, *113*, 3280 – 3291.
[44] A. R. Dinner, X. Lopez, M. Karplus, *Theor. Chem. Acc.* **2003**, *109*, 118 – 124.
[45] K. Nam, J. L. Gao, D. M. York, *J. Chem. Theory Comput.* **2005**, *1*, 2 – 13.
[46] P. Schäfer, D. Riccardi, Q. Cui, *J. Chem. Phys.* **2005**, *123*, 014905.
[47] B. A. Gregersen, D. M. York, *J. Phys. Chem. B* **2005**, *109*, 536 – 556.
[48] B. A. Gregersen, D. M. York, *J. Comput. Chem.* **2006**, *27*, 103 – 115.
[49] D. Riccardi, P. Schaefer, Y. Yang, H. B. Yu, N. Ghosh, X. Prat-Resina, P. Konig, G. H. Li, D. G. Xu, H. Guo, M. Elstner, Q. Cui, *J. Phys. Chem. B* **2006**, *110*, 6458 – 6469.
[50] H. M. Senn, W. Thiel in *Topics in Current Chemistry*, *Vol. 268*, *QM/MM Methods for Biological Systems*, Springer, Berlin, **2007**, Chapter "Atomistic Approaches in Modern Biology: From Quantum Chemistry to Molecular Simulations", pp. 173 – 290.
[51] T. Benighaus, W. Thiel, *J. Chem. Theory Comput.* **2008**, *4*, 1600 – 1609.
[52] T. Benighaus, W. Thiel, *J. Chem. Theory Comput.* **2009**, *5*, 3114 – 3128.
[53] T. Benighaus, W. Thiel, *J. Chem. Theory Comput.* **2011**, *7*, 238 – 249.
[54] U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1986**, *7*, 718 – 730.
[55] H. M. Senn, W. Thiel, *Angew. Chem.* **2009**, *121*, 1220 – 1254; *Angew. Chem. Int. Ed.* **2009**, *48*, 1198 – 1229.
[56] O. Acevedo, W. L. Jorgensen, *Acc. Chem. Res.* **2010**, *43*, 142 – 151.
[57] K. E. Ranaghan, A. J. Mulholland, *Int. Rev. Phys. Chem.* **2010**, *29*, 65 – 133.
[58] M. Levitt, S. Lifson, *J. Mol. Biol.* **1969**, *46*, 269 – 279.
[59] W. F. van Gunsteren, M. Karplus, *Macromolecules* **1982**, *15*, 1528 – 1544.
[60] D. Juchli, U. Stocker, W. F. van Gunsteren, *Mol. Simul.* **2003**, *29*, 123 – 138.
[61] M. Müller, K. Katsov, M. Schick, *Phys. Rep.* **2006**, *434*, 113 – 176.
[62] W. F. van Gunsteren, F. J. Luque, D. Timms, A. E. Torda, *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847 – 863.
[63] S. Riniker, W. F. van Gunsteren, *J. Chem. Phys.* **2011**, *134*, 084110.
[64] S. J. Marrink, A. H. de Vries, A. E. Mark, *J. Phys. Chem. B* **2004**, *108*, 750 – 760.
[65] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J. Phys. Chem. B* **2007**, *111*, 7812 – 7824.
[66] W. Shinoda, R. Devane, M. L. Klein, *Mol. Simul.* **2007**, *33*, 27 – 36.

[67] W. F. van Gunsteren, J. Dolenc, A. E. Mark, *Curr. Opin. Struct. Biol.* **2008**, *18*, 149 – 153.
[68] W. Tschöp, K. Kremer, O. Hahn, J. Batoulis, T. Bürger, *Acta Polym.* **1998**, *49*, 75 – 79.
[69] G. Milano, F. Müller-Plathe, *J. Phys. Chem. B* **2005**, *109*, 18609 – 18619.
[70] S. Izvekov, G. A. Voth, *J. Phys. Chem. B* **2005**, *109*, 2463 – 2473.
[71] B. Hess, S. Leon, N. F. A. van der Vegt, K. Kremer, *Soft Matter* **2006**, *2*, 409 – 414.
[72] A. Y. Shih, P. L. Freddolino, S. G. Sligar, K. Schulten, *Nano Lett.* **2007**, *7*, 1692 – 1696.
[73] A. P. Heath, L. E. Kavraki, C. Clementi, *Proteins Struct. Funct. Bioinf.* **2007**, *68*, 646 – 661.
[74] T. Carpenter, P. J. Bond, S. Khalid, M. S. P. Sansom, *Biophys. J.* **2008**, *95*, 3790 – 3801.
[75] A. J. Rzepiela, L. V. Schäfer, N. Goga, H. J. Risselada, A. H. de Vries, S. J. Marrink, *J. Comput. Chem.* **2010**, *31*, 1333 – 1343.
[76] A. Samiotakis, D. Homouz, M. S. Cheung, *J. Chem. Phys.* **2010**, *132*, 175101.
[77] M. Christen, W. F. van Gunsteren, *J. Chem. Phys.* **2006**, *124*, 154106.
[78] E. Lyman, F. M. Ytreberg, D. M. Zuckerman, *Phys. Rev. Lett.* **2006**, *96*, 28105.
[79] P. Liu, G. A. Voth, *J. Chem. Phys.* **2007**, *126*, 045106.
[80] M. Praprotnik, L. Delle Site, K. Kremer, *J. Chem. Phys.* **2005**, *123*, 224106.
[81] B. Ensing, S. O. Nielsen, P. B. Moore, M. L. Klein, M. Parrinello, *J. Chem. Theory Comput.* **2007**, *3*, 1100 – 1105.
[82] A. Heyden, D. G. Truhlar, *J. Chem. Theory Comput.* **2008**, *4*, 217 – 221.
[83] J. H. Park, A. Heyden, *Mol. Simul.* **2009**, *35*, 962 – 973.
[84] S. Poblete, M. Praprotnik, K. Kremer, L. Delle Site, *J. Chem. Phys.* **2010**, *132*, 114101.
[85] C. Junghans, S. Poblete, *Comput. Phys. Commun.* **2010**, *181*, 1449 – 1454.
[86] M. Neri, C. Anselmi, M. Cascella, A. Maritan, P. Carloni, *Phys. Rev. Lett.* **2005**, *95*, 218102.
[87] Q. Shi, S. Izvekov, G. A. Voth, *J. Phys. Chem. B* **2006**, *110*, 15045 – 15048.
[88] J. Michel, M. Orsi, J. W. Essex, *J. Phys. Chem. B* **2008**, *112*, 657 – 660.
[89] M. Masella, D. Borgis, P. Cuniasse, *J. Comput. Chem.* **2008**, *29*, 1707 – 1724.
[90] A. J. Rzepiela, M. Louhivuori, C. Peter, S. J. Marrink, *Phys. Chem. Chem. Phys.* **2011**, *13*, 10437 – 10448.
[91] S. Riniker, W. F. van Gunsteren, *J. Chem. Phys.* **2012**, *137*, 044120.
[92] S. Riniker, A. P. Eichenberger, W. F. van Gunsteren, *Eur. Biophys. J.* **2012**, *41*, 647 – 661.
[93] S. Riniker, A. P. Eichenberger, W. F. van Gunsteren, *J. Phys. Chem. B* **2012**, *116*, 8873 – 8879.
[94] E. Haslam, *Shikimic Acid: Metabolism and Metabolites*, Wiley, New York, **1993**.
[95] C. Gaille, P. Kast, D. Haas, *J. Biol. Chem.* **2002**, *277*, 21768 – 21775.
[96] H. Goerisch, *Biochemistry* **1978**, *17*, 3700 – 3705.
[97] J. Zaitseva, J. Lu, K. L. Olechoski, A. L. Lamb, *J. Biol. Chem.* **2006**, *281*, 33441 – 33449.
[98] D. Kunzler, S. Sasso, M. Gamper, D. Hilvert, P. Kast, *J. Biol. Chem.* **2005**, *280*, 32827 – 32834.
[99] S. Martí, J. Andres, V. Moliner, E. Silla, I. Tunon, J. Bertran, *Chem. Eur. J.* **2003**, *9*, 984 – 991.
[100] S. Martí, J. Andres, V. Moliner, E. Silla, I. Tunon, J. Bertran, *J. Am. Chem. Soc.* **2009**, *131*, 16156 – 16161.
[101] F. Claeyssens, K. E. Ranaghan, F. R. Manby, J. N. Harvey, A. J. Mulholland, *Chem. Commun.* **2005**, 5065 – 5070.
[102] A. Crespo, M. A. Marti, D. A. Estrin, A. E. Roitberg, *J. Am. Chem. Soc.* **2005**, *127*, 6940 – 6941.
[103] T. C. Bruice, *Acc. Chem. Res.* **2002**, *35*, 139 – 148.
[104] M. Štrajbl, A. Shurki, M. Kato, A. Warshel, *J. Am. Chem. Soc.* **2003**, *125*, 10228 – 10237.
[105] K. E. Ranaghan, A. J. Mulholland, *Chem. Commun.* **2004**, 1238 – 1239.
[106] H. Guo, Q. Cui, W. N. Lipscomb, M. Karplus, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9032 – 9037.
[107] T. Ishida, *J. Am. Chem. Soc.* **2010**, *132*, 7104 – 7118.
[108] F. Claeyssens, K. E. Ranaghan, N. Lawan, S. J. Macrae, F. R. Manby, J. N. Harvey, A. J. Mulholland, *Org. Biomol. Chem.* **2011**, *9*, 1578 – 1585.
[109] "Semiempirische Verfahren mit Orthogonalisierungskorrekturen: Die OM3 Methode": M. Scholten, PhD thesis, Heinrich-Heine-Universität Düsseldorf, **2003**.
[110] W. Thiel, MNDO99 v. 6.1 ed., Max-Planck-Institut für Kohlenforschung, Mülheim an der Ruhr, Germany, **2003**.
[111] L. Schuler, X. Daura, W. F. van Gunsteren, *J. Comput. Chem.* **2001**, *22*, 1205 – 1218.
[112] A. Choutko, W. F. van Gunsteren, *Prot. Sci.* **2012**, submitted.
[113] K. Meier, N. Schmid, W. F. van Gunsteren, *J. Comput. Chem.* **2012**, *33*, 2108 – 2117.
[114] N. Schmid, C. D. Christ, M. Christen, A. P. Eichenberger, W. F. van Gunsteren, *Comput. Phys. Commun.* **2011**, *33*, 340 – 353.
[115] A. P. E. Kunz, J. R. Allison, D. P. Geerke, B. A. C. Horta, P. H. Hünenberger, S. Riniker, N. Schmid, W. F. van Gunsteren, *J. Comput. Chem.* **2012**, *33*, 340 – 353.
[116] W. F. van Gunsteren et al. http://www.gromos.net.
[117] R. Ahlrichs et al. http://www.cosmologic.de/turbomole.html.
[118] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* **1984**, *81*, 3684 – 3690.
[119] R. W. Hockney, J. W. Eastwood, *Computer Simulation using Particles*, McGraw-Hill, **1981**.
[120] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327 – 341.
[121] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 5451 – 5459.