

DFT-based molecular dynamics as a new tool for computational biology: First applications and perspective

by W. Andreoni
A. Curioni
T. Mordasini

Static and molecular dynamics (MD) calculations based on density-functional theory (DFT) are emerging as a valuable means for simulations in the field of biology, especially when coupled with classical simulations. In this contribution, we point out the strengths as well as the limitations of the DFT-MD method, and its possible use to complement existing approaches. Recent applications to systems of biochemical and pharmacological interest are discussed, and an outline is given of steps to be taken regarding future calculations.

Introduction

Density-functional theory (DFT [1]) is currently the method of choice for the study of structural, electronic, and dynamical properties in materials science. The reason

for this lies both in the ability of DFT schemes to treat models of relatively large sizes compared to other first-principles methods, and in the degree of accuracy that can be achieved for diverse systems in a systematic fashion. In particular, molecular dynamics (MD) based on DFT (Car-Parrinello MD (CPMD) [2]) has added flexibility to the standard implementations and has made it possible to study dynamical processes such as structural transformations of covalently bonded materials as well as chemical reactions in condensed phases.

There are several reasons for introducing DFT-based simulations in biology: The principal motivation is a need for an *ab initio* description of the interatomic interactions whenever standard force fields do not yield reasonable results. This is true for problems such as the binding of ligand-protein complexes, the structure of metalloproteins, or the mechanisms driving enzymatic reactions that involve bond redistribution. Thus, DFT-based MD seems in principle to offer an ideal method that combines the

©Copyright 2001 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/01/\$5.00 © 2001 IBM

accuracy of DFT and the power of MD necessary for biological systems which are intrinsically dynamical. However, in practice, the applicability and usefulness of CPMD simulations in the biological field are still unclear.

First of all, it must be emphasized that biological materials systems are much more complex than traditional materials systems, in which the performance of current DFT implementations has been tested or at least has long been under control. In the description of biological materials systems, all of the problems encountered by current DFT schemes, such as accounting correctly for Van der Waals and hydrophobic interactions, are heightened because of the crucial role played by diverse interactions in determining both stability and functionality. Moreover, the relevant energy scales in biochemical materials are at most of the order of kT and thus require a higher accuracy than for traditional materials: This is indeed a difficulty that DFT shares with all other nonempirical methods. Compared to classical MD, DFT-MD also presents practical disadvantages, including limitations in the size range (up to 1000 atoms) and the shorter time scales for simulations (1–10 ps). The size limit has increased rapidly over the past decade (by two orders of magnitude), primarily because of the advent of powerful parallel computing [3]. The time-scale limitation is more difficult to overcome, even by using the algorithms common to classical MD that were devised specifically to increase either the time step or the configuration sampling [4]. Moreover, in principle, activated processes such as enzymatic reactions can be simulated by imposing specific constraints [5], but this possibility remains highly dependent on the specific system and is thus uneven. On the other hand, a “brute-force” DFT-MD simulation is currently impractical for most biological systems and, if confined to oversimplified models and/or irrelevant time scales, may lose its relevance to experiment. However, there are ways in which one can exploit all benefits of this method: either by selecting case studies for which intelligent (but size-limited) models can provide guidance to the physical understanding of more complex systems, or by applying it to the simpler biological models usually investigated in laboratory experiments (in the crystal phase). This track presents useful tests of the computational schemes by extracting important electronic parameters for use in force fields and that are transferable to larger systems, or by combining DFT and classical methods as a special case of coupled quantum-mechanical (QM) and molecular-mechanics (MM) methodologies.

This paper reports on recent attempts to use DFT and DFT-MD as indicated above, with special emphasis on our own projects, both past and current. In particular, details are given about recent calculations, the results of which we present here for the first time. Therefore, in some respects, this paper may appear a little

unbalanced. Indeed, it is not our intention either to present an exhaustive review or to describe a single application in all possible detail. As a basis for our discussion, we have selected a number of papers that well exemplify the state of the art of the calculations. Also, we attempt to provide a critical survey of the field in order to put our own work into its proper context. Finally, future directions for this type of simulation are also outlined.

Structure and dynamics with DFT calculations

Static calculations (local geometry optimizations) on key elements of biomolecules such as DNA base pairs “*in vacuo*” have been performed in a routine manner for some time. Generally, the primary purpose of these calculations was to examine the performance of DFT methods with gradient-corrected exchange-correlation functionals (GGA) versus post-Hartree-Fock methods [6] in accounting for structure, dipole moments, and energetics. Regrettably, calculations of this kind have long suffered from the limited accuracy of the computational approach used (e.g., small basis sets for the electronic wave functions) and thus were unable to provide reliable reference data for useful comparisons. Only very recently has an accurate and systematic DFT study of Watson-Crick base pairs appeared [7] which has allowed one to assess the degree of convergence of different basis sets using localized wave functions and the performance of different (gradient-corrected) exchange-correlation functionals, as well as to make a critical comparison with X-ray crystallographic data. The latter established that the molecular environment present in the crystal, especially the counter ions and water molecules, has an important influence on its structural properties.

A few static DFT calculations [8] were also performed on crystal phases with the gradient-corrected exchange-correlation functionals (GGA) using plane-wave basis sets (e.g., [9–11]). Structural optimization generally begins with the X-ray refinement, as the results for the structure are primarily a test of the computational scheme, but they can also include some complementary features such as the hydrogen pattern. However, the new information provided by calculations is derived from the electronic structure. For instance, in the example reported in Reference [10], a platinum-modified nucleobase pair was studied in the crystal phase. Detailed comparison with calculations *in vacuo* in the absence of the platinum complex allowed simultaneous capture of the effects of metalation on the electronic structure, the chemical bonding, and the geometry of the nucleobase pair in the complex.

Applications of several DFT schemes (GGA and hybrid HF-DFT) to simple models of metal enzymatic centers have recently been reviewed in Reference [12]. Interestingly, comparing these model calculations to experimental data, one could in some cases show that

treating the protein explicitly was not necessary. In fact, its effects could either be indirectly ignored, as in the case of a metal center in the blue-copper protein complex, or represented as a dielectric medium, as in the example of electron transfer in the photosynthetic reaction center of *Rhodobacter sphaeroides*. It would be interesting to understand why and to what extent this is true in other cases as well.

The important case of the Fe(II)- and Fe(III)-cytochrome c heme-protein complexes involving the nature of the bonding of iron to the protein prosthetic groups has recently been studied using model complexes, namely FeP-Im-DMS (P = porphyrin, Im = imidazole, DMS = dimethylsulfide) [13], in the plane-wave scheme of CPMD [8]. The effect of the environment on the iron-sulfur bond was also investigated by calculating only the electrostatic potential of the protein, using effective point charges obtained from the AMBER parameterization scheme [14]. Surprisingly, this effect was found to be very weak, a result contrary to chemical intuition. No explanation was provided, so that it is still unclear whether this was due to some subtle chemical reason or to an artifact of the modeling.

The possibility of explicitly treating proteins a few hundred atoms in size as well as studying their electronic properties and chemistry exhaustively has recently been realized. In particular, refinement of protein structures, which nowadays relies on *ad hoc* parameterized (empirical) force fields, can be complemented with DFT structural studies. We now describe a case we have investigated in detail in the plane-wave framework of CPMD: a relatively small protein, 1PNH, a scorpion toxin ($\text{PO}_5\text{-NH}_2$) analog with high affinity for apamin-sensitive potassium channels. As shown in **Figure 1**, it consists of 31 amino acid residues (approximately 500 atoms) and has a well-defined globular structure that has been resolved by nuclear magnetic resonance (NMR) [15]. By using a local optimization procedure, the structure of the protein was first refined computationally in the gas phase [16], starting from available NMR measurements [15], and derived with the help of simulations using a classical force field. In this first calculation, the protein was kept in its standard charge state (+6). The calculated root mean square displacement (RMSD) of the computationally refined structure from the experimental structure was only 0.181 Å/atom, demonstrating that the experimentally observed minimum is stable also in our gas-phase model. Moreover, partitioning the total RMSD by atom type, we obtain $\text{RMSD}(\text{C}) = 0.126$ Å/atom, $\text{RMSD}(\text{N}) = 0.116$ Å/atom, $\text{RMSD}(\text{O}) = 0.164$ Å/atom, $\text{RMSD}(\text{H}) = 0.199$ Å/atom, and $\text{RMSD}(\text{S}) = 0.567$ Å/atom. Clearly the two structures are very similar, with the exception of the positions of the sulfur atoms, as is evident from the comparison in **Figure 2**. The reason for the RMSD(S)

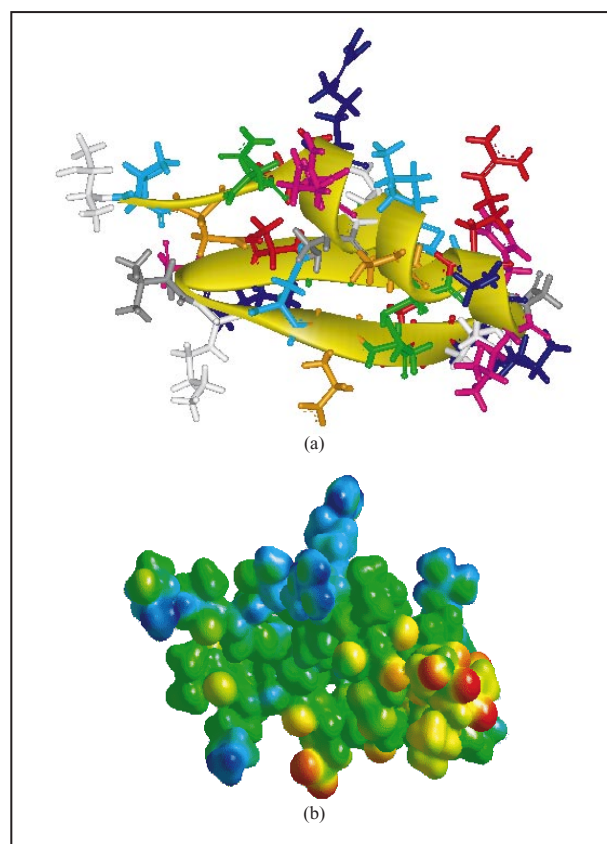


Figure 1

(a) Structure of the 1PNH protein. Different colors are used to represent different amino acids. (b) Calculated electrostatic potential, mapped onto an isodensity surface. The zero value is depicted in green; repulsive regions for an electron are depicted in red and yellow; attractive regions are depicted in blue.

discrepancy is not related to any inadequacy of the computational scheme, but lies in the incorrect values of the original structural refinement [15]. This can be ascribed to a lack of accuracy in the parameterization of the force field regarding the sulfur atoms engaged in -S-S- bridges, which are notoriously difficult to handle in a classical framework. From this result, the usefulness of our *ab initio* procedure for the structural refinement becomes readily apparent as a tool in correcting errors of this general type.

In a second stage, the charge state of the protein was modified; i.e., the global charge was neutralized and its structure reoptimized locally. The calculated RMSD with respect to the original refinement from experiment was 0.141 Å/atom, very close to the one found for the structure in vacuum. This indicates that this (indeed large) change of the charge state of the protein does not destabilize the local energy minimum. Extended to a general protein, this result can be quite important in view

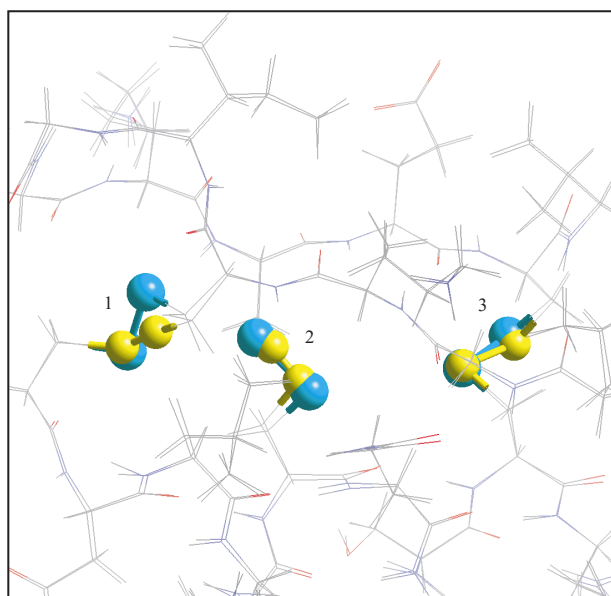


Figure 2

Original structure of the 1PNH protein, obtained from NMR, compared to the result of DFT optimization. The positions of the sulfurs at the $-S-S-$ bridges are depicted as yellow balls in the original structure and as light-blue balls in the computed structure (see text).

of the fact that charge states of proteins cannot easily be determined *a priori*. However, it must be stressed that, although it may play only a minor role in determining the details of a local energy minimum, the specific charge state could still be quite important in determining the global folding pathway.

Some details of the computations are useful for obtaining a clearer idea of the computational effort they involve. On an IBM RS/6000* SP consisting of 128 200-MHz POWER3 processors, optimization of the electronic wave functions required around 300 seconds per step, while the geometry required 1.5 hours per step, and approximately 40 of these steps were necessary to obtain a converged result for the geometry. These values are reasonable enough to render this technique now feasible and routinely usable in the near future.

Static calculations are far from being exhaustive. Moreover, when testing a method such as DFT vs. experiment, local energy minimization starting from an educated guess is not sufficient and does not guarantee the absence of spurious minima and/or unrealistic conformations. More significantly, the study of the dynamics is vital for understanding the physical and chemical behavior of biomolecules [17] and is certainly essential for a correct description of the influence of water on their conformations and interactions. With respect to liquid water itself, CPMD simulations have shown that it can be treated at a reasonably accurate level within some of the DFT-GGA schemes [18–20]; structural and mobility properties as well as the solvation of diverse ions have been simulated successfully (see for example [21] and [22]). The ability to represent liquid water with a reasonable degree of accuracy is an important achievement, because if use is made of classical force fields, many difficulties are encountered in properly representing the intrinsic properties of water and specific interactions (see for example [23]); it opens up several new possibilities such as determining “structures in solution” by fully *ab initio* methods. This has recently been

Table 1 Sugar-backbone conformation of the cisplatin *dpGpG* adduct (*Start* = starting configuration; *MD* = result of CPMD simulation) compared to that of the $[Pt(NH_3)_2]d(GpG)$ (from NMR) (see text). Notation of the torsion angles as in Figure 3. Reprinted with permission from [22], © 2000 American Chemical Society.

$[Pt(NH_3)_2]-dpGpG$	<i>Start</i>	<i>MD</i>	<i>NMR*</i>
5'-Guanidine			
χ	−143	−147(15)	~ −115
β	165	132(12)	
γ	37	50(10)	52
δ	100	90(10)	87(3)
ϵ	−133	−170(12)	−162(4)
ζ	64	−70(11)	−60
3'-Guanidine			
χ	−127	−147(9)	~ −115
α	−56	−68(9)	−65
β	−159	−170(10)	180(3)
γ	41	60(13)	50–60
δ	143	129(31)	137

*J. H. J. den Hartog, C. Altona, G. C. Chottard, G. P. Girault, J. Y. Lallemant, F. A. A. M. de Leeuw, A. T. M. Marcelis, and J. Reedijk, *Nucl. Acids. Res.* **10**, 4715 (1982).

partially achieved for the monostrand major cisplatin–DNA¹ adduct [22], a system of intense investigation given the well-known antitumoral activity of cisplatin and the interest in the design of less toxic cisplatin derivatives. Here we report only few results from Reference [22], where the interested reader can find more details. As the initial model of the adduct, its experimentally known structure in the crystal phase was considered, and immersed in a water solution. First this system was equilibrated with classical MD simulations, keeping the cisplatin geometry fixed in order to avoid the expected dramatic effects of classical potentials. Subsequently, CPMD simulations [8] were run; these had previously been shown (also in Reference [22]) to reproduce well a number of molecular properties of cisplatin as well as the dynamics of its hydrolysis. Although very short (about 2 ps), the CPMD simulations of the cisplatin–DNA adduct were able to show the “immediate” effects of the aqueous solution, some of which can be argued from the results reported in **Table 1**, where the average sugar-backbone conformation (as depicted in **Figure 3**) found in this way (under the heading *MD*) is compared with the initial conformation (under the heading *Start*). Indeed, the changes are important and compare well with NMR data measured in aqueous solution for another, very similar, cisplatin adduct (see under the heading *NMR* in Table 1) that differs from the model investigated only in its lack of the 5' phosphate moiety. There is a noticeable discrepancy in the χ angles describing the nucleobase/sugar orientation between the experimental NMR-derived (GpG) results and the simulation in water that, apart from the time limitation of the CPMD runs, may also be ascribed to the known conformational flexibility of this molecule and to its sensitivity to the lack of the 5' phosphate moiety in the NMR structure. Also, an apparent effect of the interaction with the solvent is the reduction of the degree of destacking of the two guanines, which could reasonably be ascribed to the loss of hydrophobic interactions between the nucleobases of different complexes present in the crystal phase. In fact, the dihedral angle decreases from 81 to 61(9) degrees, a tendency also confirmed by the NMR value of 53 degrees measured for the analog complex in water.

In the study of the 1PNH protein illustrated above, the effect of the aqueous solution on the local structural minimum was also investigated. Simulating a sufficient number of water molecules to represent a realistic water solvation shell around the protein was computationally too demanding for full DFT calculations. To obtain qualitative information, a much simpler strategy was adopted. First, the protein was fixed at the structure resulting from a DFT refinement and solvated with water treated classically

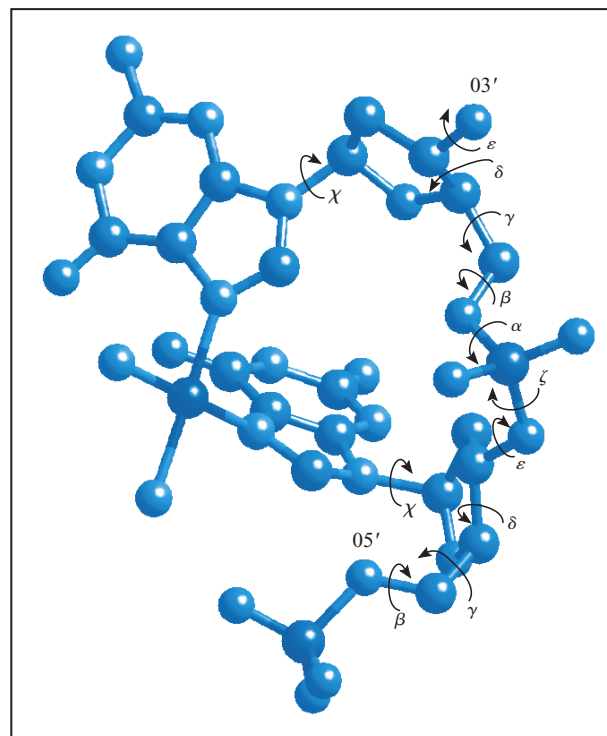


Figure 3

Molecular structure of *cis*-[Pt(NH₃)₂] *d*(pGpG). For clarity, hydrogen atoms are not shown. The torsion angles depicted follow the convention specified in the IUPAC–UB Joint Commission on Biochemical Nomenclature Report, *Eur. J. Biochem.* **131**, 9 (1983). Reprinted with permission from [22]; © 2000 American Chemical Society.

using the simple point charge (SPC) water model [24] and the GROMOS96 force field [25]. The water was then thermalized using the GROMOS code.² Finally, only the 85 water molecules contained in a spherical shell having a radius of 2.5 Å around the protein were retained in the model, and the structure of the protein was reoptimized *ab initio*, keeping the water molecules fixed. Preliminary results yielded an RMSD value of 0.120 Å/atom with respect to the gas-phase geometry, confirming again the relative stability of this structure. Note that the CPU time required for such a calculation is nearly double that for the gas-phase protein.

We note that this type of calculation is not meant to be restricted to proteins of affordable size, but can also be applied to protein fragments. An additional interesting result of these calculations was the determination of effective atomic charges from the best fitting of the electrostatic potential (electrostatic potential procedure

¹ The term cisplatin denotes *cis*-diamminedichloroplatinum(II).

² GROMOS Software (C) Biomos bv, University of Groningen, The Netherlands, and ETH Zurich, Switzerland.

(ESP) [26]) on the entire protein. Atomic charges are important because they represent one of the simplest parameters for describing the electrostatic potential generated by a protein—a quantity of fundamental importance in determining the functioning of the protein itself. Often these charges are extracted from calculations on small fragments (e.g., amino acids) or from the Coulomb terms of the force field (with or without charge-equilibration procedures [27]), and assumed to be transferable. In particular, when charge-equilibration procedures are not used, the charge associated with an atom of a certain type, belonging to a certain amino acid, is assumed to be the same irrespective of its position. Our *ab initio* calculations show that this transferability hypothesis is not justified. In fact, the values we find for the atomic charges are such that the charge associated with a carbon atom on the same type of residue—but in a different location in the protein—can vary by more than one charge unit! Therefore, one can reasonably expect relatively large errors in the electrostatic potential calculated using charges that are not derived from an electronic structure calculation on the entire protein (or a sufficiently large protein fragment). It would be very interesting to explore the impact of these findings on the accuracy of the docking models widely used in drug design.

DFT-MD/MM approach

The “modern” approach combining DFT and MM is a special case of QM/MM methods for which the concepts were introduced in the late 1970s in a well-known publication by Warshel and Levitt [28]. Aimed especially at the simulation of enzymatic reactions, the QM/MM approach consists in partitioning the enzyme into two subsystems in order to treat the portion directly involved in the catalytic reaction at the QM level (the QM subsystem); the remainder, where events such as bond breaking and bond formation do not take place, is represented with an MM force field (the MM subsystem). Although the basic philosophy is simple, and practical implementation requires no special computational skills, the practical use of QM/MM methods requires an ingenious treatment of the interface between the two subsystems [29]. Apart from the specific force field used for the MM subsystem and the specific method (theoretical as well as computational) used for the QM subsystem, the model chosen for the description of the interface and the definition of coupling as well as of the information to be exchanged comprise the distinguishing features of the various applications deployed so far [30]. A particularly comprehensive and critical review has recently been published by Monard and Merz [31].

Several attempts have been made to combine the DFT-MD and MM methods as a natural extension of the

CP method, using the CPMD code for the QM part interfaced with classical force-field codes (see for example [32]). An interesting example of applications to biologically relevant systems is that of exploiting the possibility of comparing, within the same framework, the activity of a real enzymatic system treated with the hybrid approach to that of low-molecular-weight analogs (biomimetic compounds) treated fully at the DFT level. In the case of the copper enzyme galactose oxidase, this type of investigation [33] has led to the conclusion that small synthetic analogs can be devised that reproduce the activity of the real system. This was based on the identification of a structural difference between a real synthetic analog that renders it much less active than the real enzyme, and also on the fact that in these calculations the structural and electronic properties of the QM portion (~80 atoms) appear to be insensitive to the behavior of the MM region. Note that the interface was represented by the scaled position method for link atoms, as in [32].

The advantage of using DFT and in particular the plane-wave scheme (as in CPMD) for the QM subsystem, however, comes to light only when much larger systems, i.e., of the order of a few hundred to a thousand atoms, are being treated; as mentioned above, this has become possible only very recently. In fact, increasing the size of the system to be modeled does not imply simply that one can explicitly treat a larger number of atoms around the active region at a higher level of theory, but rather that the problem of “how to treat the interface” becomes much less critical because the MM subsystem is farther away from the active region. Indeed, it should be emphasized that most of the problems encountered in QM/MM applications that are related to detailed modeling of the interface [34], such as the definition of the constraints that must be imposed on the frontier bonds to avoid artifacts in the chemical behavior of the system, arise from the limited size of the QM component, or at least are greatly amplified. Moreover, compared to a QM approach using localized functions, the approach based on plane waves avoids the accuracy problem related to the basis superposition error and allows a better description of the polarization of the electronic wave functions.

Within the DFT-MD/MM scheme we have recently considered BamHI, a prototype of Type II restriction enzymes [35], which are responsible for a particular cleavage of short DNA sequences at both strands, producing 5'-phosphate and 3'-hydroxyl groups. A number of experimental structural studies have elucidated the structure of BamHI-DNA complexes [36]. However, it is unclear how the catalytic reaction evolves, although a few hypotheses about the mechanism have been made [37]. For example, it is known that the catalytic reactions require the presence of divalent metal ions such as Mg, Mn, Co, Zn, or Cd. However, whereas the majority of

them work as catalysts and differ only in the efficiency of their catalytic action, Ca acts as an inhibitor of the enzymatic reaction, and the reason for this is not yet understood. Clearly, the metal ions cannot simply be assumed as point charges, and we risk biasing the result if we make any unsubstantiated hypothesis on the difference of the cations. To answer these questions, we must rely on an *ab initio* approach and thus have decided to carry out a parallel investigation of BamHI with Ca and with Mg using DFT-MD/MM simulations.³

An important feature has emerged from experimental evidence that aided our modeling and simulations, namely that the catalytic reaction drives negligible conformational changes in the BamHI–DNA complex. In fact, the pre-reactive structure measured with the inhibiting calcium ion as well as the post-reactive one (measured in the presence of Mn) are very similar (RMS deviation of only 0.33 Å for the C_α positions). Thus, we chose the pre-reactive state of the complex with Ca as the initial configuration, as indicated in **Figure 4**, and locally optimized the structure within our scheme.

Out of a total number of about 6000 atoms, 297 (protein and DNA backbone, plus crystallographic water molecules and metal atoms) were included in the QM subsystem. The complex, maintained at a fixed geometry, was first solvated in water and equilibrated with GROMOS simulations using the SPC model [24]. The effect of the MM subsystem, considered as a geometrically fixed environment and surrounded by an additional ~2300 water molecules, was taken into account via the inclusion of electrostatic and Van der Waals interactions. The electrostatic potential generated by the MM atoms and calculated using the standard GROMOS96 force-field charges [25] was mapped over the tridimensional grid containing the QM atoms. Because there are no conformational changes during the reaction, this calculation could be performed only once. Van der Waals interactions between QM and MM atoms, on the other hand, were calculated at each optimization or MD step. These were represented by the GROMOS96 6–12 Lennard–Jones potential with a cutoff of 10 Å. The open bonds of the QM atoms at the QM/MM interface were saturated with hydrogen atoms [29] and also kept fixed during the entire simulation. This simple modeling of the interface was made possible by two considerations: on the one hand, the large size of the QM system (the atoms of the active region are at least 5 Å away from the interface), and on the other the *a priori* knowledge (mentioned above) that the structure of the MM region is quite insensitive to the reaction. In conclusion, in the above

³ For the QM subsystem, a box of $(22.95 \times 25.69 \times 20.75)$ Å³ was used, and a plane-wave cutoff of 70 Ry. Other details as in [16]. In particular, note that the QM subsystem is treated as an “embedded cluster”; i.e., it is treated as isolated (Hockney method), but the effect of the interaction with the MM system is taken into account.

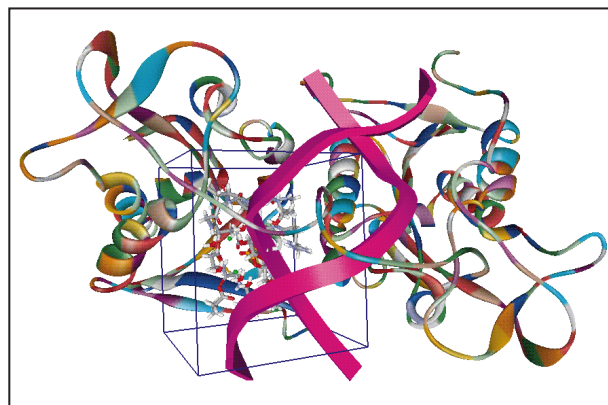


Figure 4

Structure of the BamHI–DNA complex. The “ball-and-stick” model within the box represents the QM subsystem.

scheme, the electronic structure of the QM portion is optimized under the influence of the MM system, represented by the electrostatic potential, and the updating of the atomic coordinates takes into account the Van der Waals contribution of the interaction with the MM atoms. The optimized structure of the pre-reactive complex with Ca does not deviate significantly from the experimental one. The value of the RMSD evaluated over all of the QM atoms (water excluded) is 0.2 Å/atom, while the Ca–Ca distance remains at 4.3 Å and the Ca–O distances vary by about 0.1 Å. When Ca is replaced with Mg, the Mg–O distances, as expected, shrink by 0.3 Å on average, but the overall RMSD value, as calculated above, turns out to be approximately 0.2 Å/atom.

In an attempt to understand why Ca behaves differently from Mg, one might suggest that perhaps the propensity to donate two electrons causes the difference in catalytic activity. From the calculated electronic structure of the pre-reactive geometries, however, it can be clearly demonstrated that this is not the case. For both, two valence electrons are removed, and the localization of the transferred charge on the protein is essentially the same. This striking similarity can be seen by inspecting **Figures 5(a)** and **5(b)**, which illustrate the electrostatic potential mapped on an isodensity hypersurface for the two cases. From a methodological point of view, it is interesting to note the extent and location of the changes of the charge distribution induced in the QM subsystem by the interaction with the MM subsystem. The answer can be found in **Figure 6**, which depicts the difference of the electron density (also mapped on an isodensity surface) of the QM region when it is isolated and when it interacts with the MM region. The difference is indeed sizable,

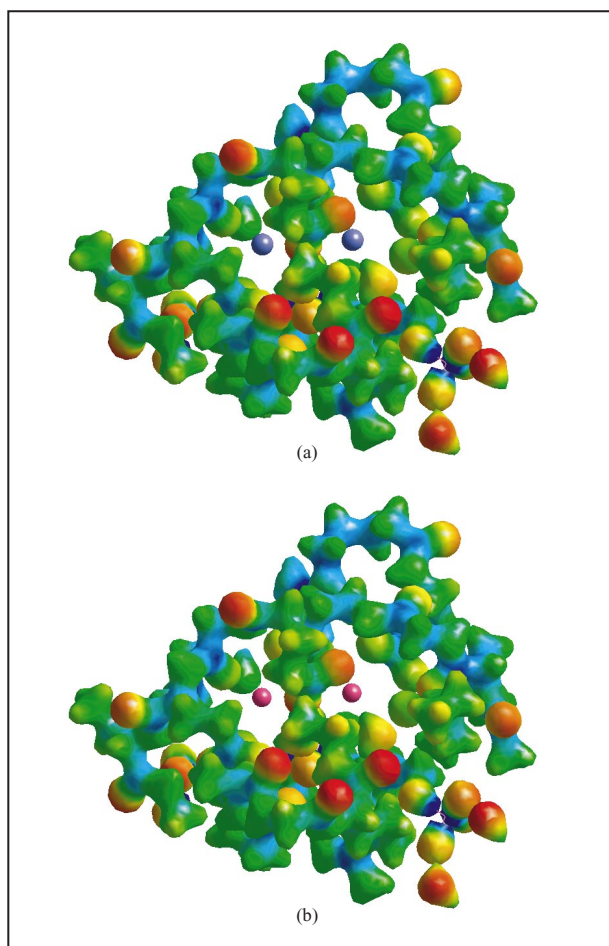


Figure 5

BamHI-DNA complex, depicting the QM subsystem: Electrostatic potential (using the convention of Figure 1 and mapped onto an isodensity surface) for the two choices of metal ion cofactors: (a) Ca; (b) Mg.

especially in the close environment of DNA, where the reaction will eventually take place. Therefore, a purely quantum-mechanical treatment of the ~ 300 -atom subsystem would not be adequate.

We are currently performing simulations of the cleavage reaction that should yield useful information on the mechanism of the catalytic reaction and especially identify the reasons why Ca cannot be used as a cofactor.

Conclusions and perspective

There is no doubt that the value and importance of computer simulations in biology are becoming more and more recognized. It is also generally believed, however, that in order to make a real impact, progress will be necessary both in hardware and software. The latter will require not only coding but also (and especially)

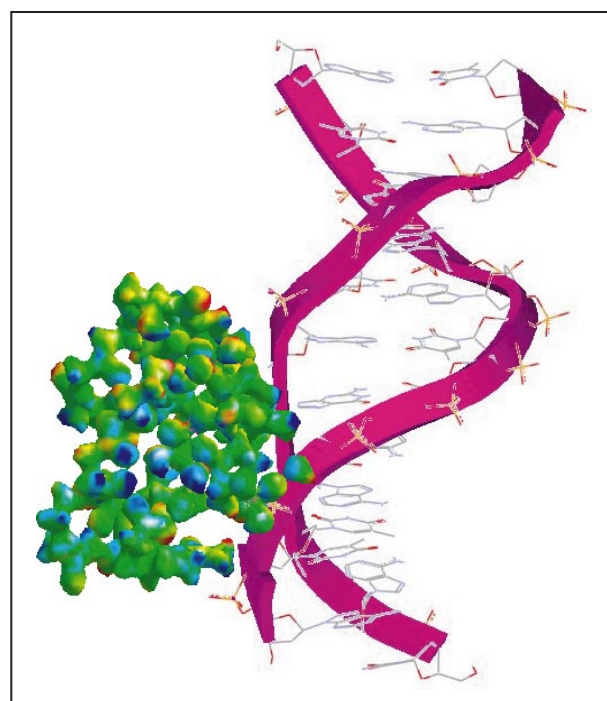


Figure 6

BamHI-DNA complex, depicting the electron density difference between the isolated QM and the QM system coupled with the MM system, as described in the text.

methodological advances. From the scientific point of view, it is clear that computational chemistry methods, and in particular the combination of quantum-mechanical and molecular-dynamics methods, could make a difference by providing the degree of accuracy, reliability, and transferability that force fields generally do not offer. We have briefly surveyed the current state of DFT and DFT-MD calculations, and have reported our own new developments and results, which are a faithful example of what can currently be achieved with CPMD simulations when combined with molecular mechanics. It is probably worth emphasizing that a DFT-based approach allows an accurate calculation of a number of physical and chemical properties, some of which can either be compared directly to experiment or are complementary to it. Beyond the structural characterization of a protein and its electron density distribution reflected in the electrostatic potential (of which we have provided examples), these calculations can in fact provide details on properties such as atomic chemical shifts, measured by nuclear magnetic resonance (see for example [38]), electron spin density distributions (see for example [39]), probed by electron spin resonance, and infrared spectra (see for example [40]). Also, energy barriers for chemical reactions in solution can be inferred

from CPMD simulations (see [3] and references therein), which opens the road to a quantitative assessment of the mechanism of enzymatic reactions.

However, DFT-MD procedures are complex and relatively inefficient when applied to biomolecular systems. The route one has to pursue to render them as efficient as the traditional ones is full of difficulties, but certainly worth the challenge. Regarding an *ab initio* approach to biomolecular systems, there are still a number of fundamental issues to be addressed, in addition to the known problems encountered at the level of chemical accuracy. For example, regarding the DFT approach, the description of dispersion forces still is an unsolved and critical problem of current implementations. This is an obvious consequence of the nature of the approximations (local or semilocal) made thus far for the exchange-correlation density functionals [41], thus requiring novel and clever implementations [42]. Another problem pertains to the correct treatment of hydrogen-bonded systems and of proton-transfer events that may well require that the quantum nature of the nuclei be taken explicitly into account. An extension of DFT-MD that addresses these problems combines it with the path-integral method that has recently been introduced [43, 44]. Coupling classical and quantum approaches is mandatory, given the typical size of biologically relevant systems; however, performing such a coupling in a "correct" and general way is still a challenge in itself. Here, a systematic effort aimed at elucidating the failures and successes of a given scheme and/or at providing a thorough comparison of various schemes would be very useful indeed.

Acknowledgment

We wish to thank an anonymous reviewer for comments and suggestions that we feel have helped to improve our paper.

References and notes

1. P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Phys. Rev.* **136**, B864–B887 (1964); W. Kohn and L. J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Phys. Rev.* **140**, A1133–A1138 (1965); R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford Science Publications, New York, 1989; R. M. Dreizler and E. K. U. Gross, *Density-Functional-Theory*, Springer-Verlag, Berlin, 1990.
2. R. Car and M. Parrinello, "Unified Density-Functional-Theory and Molecular Dynamics," *Phys. Rev. Lett.* **55**, 2471–2474 (1985).
3. W. Andreoni and A. Curioni, "New Advances in Chemistry and Materials Science with CPMD and Parallel Computing," *Parallel Computing* **26**, 819–842 (2000).
4. W. F. van Gunsteren, "Molecular Dynamics and Stochastic Dynamics Simulations: A Primer," *Computer Simulation of Biomolecular Systems*, Vol. 2, W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, Eds., ESCOM, Leiden, The Netherlands, 1993, pp. 3–36.
5. E. A. Carter, G. Ciccotti, and J. T. Hynes, "Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events," *Chem. Phys. Lett.* **156**, 472–477 (1989); G. Ciccotti, M. Ferrario, and J. T. Hynes, "Constrained Molecular Dynamics and the Mean Potential for an Ion Pair in a Polar Solvent," *Chem. Phys.* **129**, 241–251 (1989).
6. See for example J. Sponer, J. Leszczynski, and P. Hobza, "Structures and Energies of Hydrogen-Bonded DNA Base Pairs. A Nonempirical Study with Inclusion of Electron Correlation," *J. Phys. Chem.* **100**, 1965–1974 (1996).
7. C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, and E. J. Baerends, "Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment," *J. Amer. Chem. Soc.* **122**, 4117–4128 (2000).
8. Calculations were performed with the CPMD code written by Jürg Hutter, © IBM Corporation, 1990–1997; Max-Planck-Institut für Festkörperlphysik, 1997.
9. J. Hutter, P. Carloni, and M. Parrinello, "Non-Empirical Calculations of a Hydrated RNA Duplex," *J. Amer. Chem. Soc.* **118**, 8710–8712 (1996).
10. P. Carloni and W. Andreoni, "Platinum-Modified Nucleobase Pairs in the Solid State: A Theoretical Study," *J. Phys. Chem.* **100**, 17797–17800 (1996).
11. P. Carloni, W. Andreoni, and M. Parrinello, "Self-Assembled Peptide Nanotubes from First Principles," *Phys. Rev. Lett.* **79**, 761–764 (1997).
12. P. E. M. Siegbahn and M. R. A. Blomberg, "Density Functional Theory of Biologically Relevant Metal Centers," *Ann. Rev. Phys. Chem.* **50**, 221–249 (1999).
13. C. Rovira, P. Carloni, and M. Parrinello, "The Iron-Sulfur Bond in Cytochrome c," *J. Phys. Chem.* **103**, 7031–7035 (1999).
14. W. D. Cornell, P. Cieplak, C. I. Bayley, I. R. Gould, K. M. Merz, Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. E. Kollman, "A Second Generation Force-Field for the Simulation of Protein and Nucleic Acids," *J. Amer. Chem. Soc.* **117**, 5179–5197 (1995).
15. S. Meunier, J.-M. Bernassau, J.-M. Sabatier, M.-F. Martin-Eauclaire, J. van Rietschoten, C. Cambillau, and H. Darbon, "Solution Structure of $\text{PO}_5\text{-NH}_2$, a Scorpion Toxin Analog with High Affinity for the Apamin-Sensitive Potassium Channel," *Biochemistry* **32**, 11969–11976 (1993).
16. Computational details: an orthorhombic box of size $(38.63 \times 28.05 \times 28.05) \text{ \AA}^3$; a 55 Ry energy cutoff for the plane-wave expansion; gradient-correction exchange-correlation functional BLYP [A. D. Becke, "Density-Functional-Exchange-Energy Approximation with Correct Asymptotic Behavior," *Phys. Rev. A* **38**, 3098–3100 (1988); C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density," *Phys. Rev. B* **37**, 785–789 (1988)]; Hockney method [as described in R. N. Barnett and U. Landman, "Born-Oppenheimer Molecular-Dynamics Simulations of Finite Systems: Structure and Dynamics of $(\text{H}_2\text{O})_n$," *Phys. Rev. B* **48**, 2081–2097 (1993)] to treat isolated systems.
17. M. Karplus and G. A. Petsko, "Molecular Dynamics Simulations in Biology," *Nature* **347**, 631–639 (1990).
18. M. Sprik, J. Hutter, and M. Parrinello, "Ab Initio Molecular Dynamics Simulation of Liquid Water: Comparison of Three Gradient-Corrected Density Functionals," *J. Chem. Phys.* **105**, 1142–1152 (1996).
19. E. Schwegler, G. Galli, and F. Gygi, "Water Under Pressure," *Phys. Rev. Lett.* **84**, 2429–2432 (2000).
20. M. Boero, K. Terakura, T. Ikeshoji, C. C. Lew, and M. Parrinello, "Hydrogen Bonding and Dipole Moment of Water at Supercritical Conditions: A First Principles Molecular Dynamics Study," *Phys. Rev. Lett.* **85**, 3425–3428 (2000).

21. E. J. Meijer and M. Sprik, "A Density-Functional Study of the Addition of Water to SO_3 in the Gas Phase and in Aqueous Solution," *J. Phys. Chem. A* **102**, 2893–2898 (1998).
22. P. Carloni, M. Sprik, and W. Andreoni, "Key Steps of the *cis*-Platin–DNA Interaction: Density Functional Theory-Based Molecular Dynamics Simulations," *J. Phys. Chem. B* **104**, 823–835 (2000).
23. D. L. Beveridge, S. Swaminathan, G. Ravishanker, J. M. Withka, J. Srinivasan, C. Prevost, S. Louise-May, D. R. Langley, F. M. DiCapua, and P. H. Bolton, "Molecular Dynamics Simulations on the Hydration, Structure and Motions of DNA Oligomers," *Water and Biological Macromolecules*, E. Westhof, Ed., MacMillan, London, 1993, pp. 165–225.
24. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans, "Interaction Models for Water in Relation to Protein Hydration," *Intermolecular Forces*, B. Pullman, Ed., D. Reidel Publishing Co., Dordrecht, Netherlands, 1981, pp. 331–342.
25. W. R. P. Scott, P. H. Huenenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krueger, and W. F. van Gunsteren, "The GROMOS Biomolecular Simulation Program Package," *J. Phys. Chem. A* **103**, 3596–3607 (1999).
26. K. M. Merz, Jr., "Analysis of a Large Database of Electrostatic Potential Derived Atomic Charges," *J. Comput. Chem.* **13**, 749–771 (1992).
27. A. K. Rappé and W. A. Goddard III, "Charge Equilibration for Molecular Dynamics Simulations," *J. Phys. Chem.* **95**, 3358–3363 (1991).
28. A. Warshel and M. Levitt, "Theoretical Studies of Enzymatic Reactions: Dielectric, Electrostatic and Steric Stabilization of Carbonium Ion in the Reaction of Lysozyme," *J. Mol. Biol.* **103**, 227–249 (1976).
29. The most widely used treatment of atoms at the QM/MM frontier is introduced in the following two papers: U. C. Singh and P. A. Kollman, "A Combined Ab Initio Quantum Mechanical and Molecular Mechanical Method for Carrying out Simulations on Complex Molecular Systems: Applications to the $\text{CH}_3\text{Cl}^+\text{Cl}^-$ Exchange Reaction and Gas-Phase Protonation of Polyethers," *J. Comput. Chem.* **7**, 718–730 (1986); M. J. Field, P. A. Bash, and M. Karplus, "A Combined Quantum Mechanical and Molecular Mechanical Potential for Molecular Dynamics Simulations," *J. Comput. Chem.* **11**, 700–733 (1990).
30. For recent applications, see for example J. Bentzien, R. P. Muller, J. Florián, and A. Warshel, "Hybrid Ab Initio Quantum Mechanics/Molecular Mechanics Calculations of Free Energy Surfaces for Enzymatic Reactions: The Nucleophilic Attack in Subtilisin," *J. Phys. Chem. B* **102**, 2293–2301 (1998); M. Peräkylä and P. A. Kollman, "Ab Initio Quantum Mechanical and Molecular Dynamical Study of Intra- and Intermolecular Anhydride Formation," *J. Phys. Chem. A* **103**, 8067–8074 (1999); B. Kuhn and P. A. Kollman, "QM–FE and Molecular Dynamics Calculations on Catechol *O*-Methyltransferase: Free-Energy of Activation in the Enzyme and in Aqueous Solution and Regioselectivity of the Enzyme-Catalyzed Reaction," *J. Amer. Chem. Soc.* **122**, 2586–2596 (2000); P. D. Lyne, M. Hodoscek, and M. Karplus, "A Hybrid QM–MM Potential Employing Hartree–Fock or Density Functional Methods in the Quantum Region," *J. Phys. Chem. A* **103**, 3462–3471 (1999); Q. Cui and M. Karplus, "Molecular Properties and Combined QM/MM Methods. 2. Chemical Shifts in Large Molecules," *J. Phys. Chem. B* **104**, 3721–3743 (2000).
31. G. Monard and K. M. Merz, Jr., "Combined Quantum Mechanical/Molecular Mechanical Methodologies Applied to Biomolecular Systems," *Acc. Chem. Res.* **32**, 904–911 (1999).
32. M. Eichinger, P. Tavan, J. Hutter, and M. Parrinello, "A Hybrid Method for Solutes in Complex Solvents: Density Functional Theory Combined with Empirical Force Fields," *J. Chem. Phys.* **110**, 10452–10467 (1999).
33. U. Röthlisberger, P. Carloni, K. Doclo, and M. Parrinello, "A Comparative Study of Galactose and Active Site Analogs Based on QM/MM Car–Parrinello Simulations," *J. Bioinorg. Chem.* **5**, 236–250 (2000).
34. N. Reuters, A. Dejaegere, B. Maigret, and M. Karplus, "Frontier Bonds in QM/MM Methods: A Comparison of Different Approaches," *J. Chem. Phys. A* **104**, 1720–1735 (2000).
35. A. K. Aggarwal, "Structure and Function of Restriction Endonucleases," *Curr. Opin. Struct. Biol.* **5**, 11–19 (1995).
36. H. Viadiu and A. K. Aggarwal, "The Role of Metals in Catalysis by the Restriction Endonuclease BamHI," *Nat. Struct. Biol.* **5**, 910–916 (1998).
37. A. Pingoud and A. Jeltsch, "Recognition and Cleavage of DNA by Type-II Restriction Endonucleases," *Eur. J. Biochem.* **246**, 1–22 (1997).
38. Y.-G. Yoon, B. G. Pfrommer, F. Mauri, and S. G. Louie, "NMR Chemical Shifts in Hard Carbon Nitride Compounds," *Phys. Rev. Lett.* **80**, 3388–3391 (1998).
39. A. Curioni, M. Boero, and W. Andreoni, "Alq₃: Ab Initio Calculations of Its Structural and Electronic Properties in Neutral and Charged States," *Chem. Phys. Lett.* **294**, 263–271 (1998).
40. W. Andreoni and A. Curioni, "Ab Initio Approach to the Structure and Dynamics of Metallofullerenes," *Appl. Phys. A* **66**, 299–306 (1998).
41. S. Kristyán and P. Pulay, "Can (Semi)local Density Functional Theory Account for the London Dispersion Forces?" *Chem. Phys. Lett.* **229**, 175–180 (1994).
42. See for example E. K. U. Gross, J. F. Dobson, and M. Petersilka, "Density Functional Theory of Time-Dependent Phenomena," *Density Functional Theory*, R. F. Nalewajski, Ed., *Topics in Current Chemistry*, Vol. 181, Springer, Heidelberg, 1996, pp. 81–172.
43. D. D. Marx and M. Parrinello, "Ab-Initio Path Integral Molecular Dynamics: Basic Ideas," *J. Chem. Phys.* **104**, 4077–4082 (1996).
44. M. E. Tuckerman, D. Marx, M. L. Klein, and M. Parrinello, "On the Quantum Nature of the Shared Proton in Hydrogen Bonds," *Science* **275**, 817–819 (1997).

Received September 20, 2000; accepted for publication January 24, 2001

Wanda Andreoni *IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland (and@zurich.ibm.com).* After receiving a Ph.D. degree in physics from the University of Rome, Italy, Dr. Andreoni continued her scientific career as a physicist and expert for computer simulations in physics and chemistry at academic and industrial institutions in Italy, France, the U.S.A., and Switzerland. She joined the Zurich Research Laboratory as a Research Staff Member in 1986. She became Project Leader of the Computational Chemistry and Physics group in 1994, and has been Manager of Computational Materials Science since 1995. Elected a member of the IBM Academy of Technology in 1999, Dr. Andreoni also serves as a member of the Deep Computing Council of the IBM Deep Computing Institute. Since she joined IBM, her research has focused on diverse frontier areas of chemistry and physics: metal and semiconductor clusters and surfaces, fullerene-based materials, organic light-emitting devices, metal oxides, and, more recently, biologically relevant systems.

Alessandro Curioni *IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland (cur@zurich.ibm.com).* After receiving a Ph.D. degree in chemistry from the Scuola Normale Superiore, Pisa, Italy, in 1996, Dr. Curioni joined the Zurich Research Laboratory in 1998 as a Research Staff Member. Starting with his thesis work on the application of theoretical chemistry methods to the study of organic conducting materials, his research has been devoted mainly to the development of *ab initio* computational methods and to their application to diverse phenomena in the chemistry of materials such as fullerenes, aluminum oxide, and organic materials used in light-emitting devices. Dr. Curioni has recently extended his research to the field of biology.

Tiziana Mordasini *IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland (tiz@zurich.ibm.com).* Dr. Mordasini joined the Zurich Research Laboratory as a Research Staff Member in January 2000. She received a Ph.D. degree in chemistry in 1996 from the Eidgenössische Technische Hochschule, Zurich, Switzerland, and held postdoctoral appointments at the University of Zurich, Switzerland, and at the University of California, San Diego. Dr. Mordasini has been using several different computational methods to study chemical phenomena (such as the free energy of organic host-guest complexes and the electrochemical properties of fullerenes) and more recently to study biochemical phenomena, focusing on the simulation of enzymatic reactions.