Michele Pavone
Dipartimento di Chimica “Paolo Corradini”
Università di Napoli “Federico II”
michele.pavone@unina.it

AB INITIO MODELING OF COMPLEX CHEMICAL SYSTEMS

This Critical Review provides an overview of current computational modeling techniques. Theoretical approaches are discussed in relation to their accuracy and computational feasibility, highlighting successful applications and severe drawbacks.

Modern research in chemical sciences tries to discover by design new molecules and materials with tailored functions and to understand their properties. Such an aim has been stimulating the set-up and tuning of extremely sensitive experimental techniques that are able to investigate the deep and subtle features of the matter. An experimental apparatus records the overall response of the matter to an external perturbation, and, in many cases, the collected data could only be indirectly related to the microscopic structure and dynamics of the system under investigation. Thus, the computer simulations of the undergoing physico-chemical processes provide safe and sound grounds for an unbiased interpretation of ambiguous experimental outputs, and could also suggest new experiments in the most favorable cases [1].

In recent years, the margin of interaction between computational and experimental chemistry has been increasing at a fast pace, thanks also to the development of user-friendly computer programs, implementing a wide range of theoretical methods, and to the availability of high performance computing facilities. Within the framework of molecular modeling techniques, the available theoretical methodologies cover wide time and space length scales. The degree of accuracy of the theoretical model is inversely proportional to the computational feasibility. Thus, the nature of the chemical process under investigation and the level of accuracy required for its understanding can guide the choice of the proper simulation tool: from Quantum Mechanics for nuclear and electronic processes in the nanoscale (e.g., spectroscopy, chemical reactions of molecular systems) to classical approximation, Molecular Mechanics, for atomistic macromolecular simulations (e.g., polymers, proteins) up to the meso-scale, and to simulation based on macroscopic parameters and finite-element methods for the modeling of real size devices, Fig. 1.

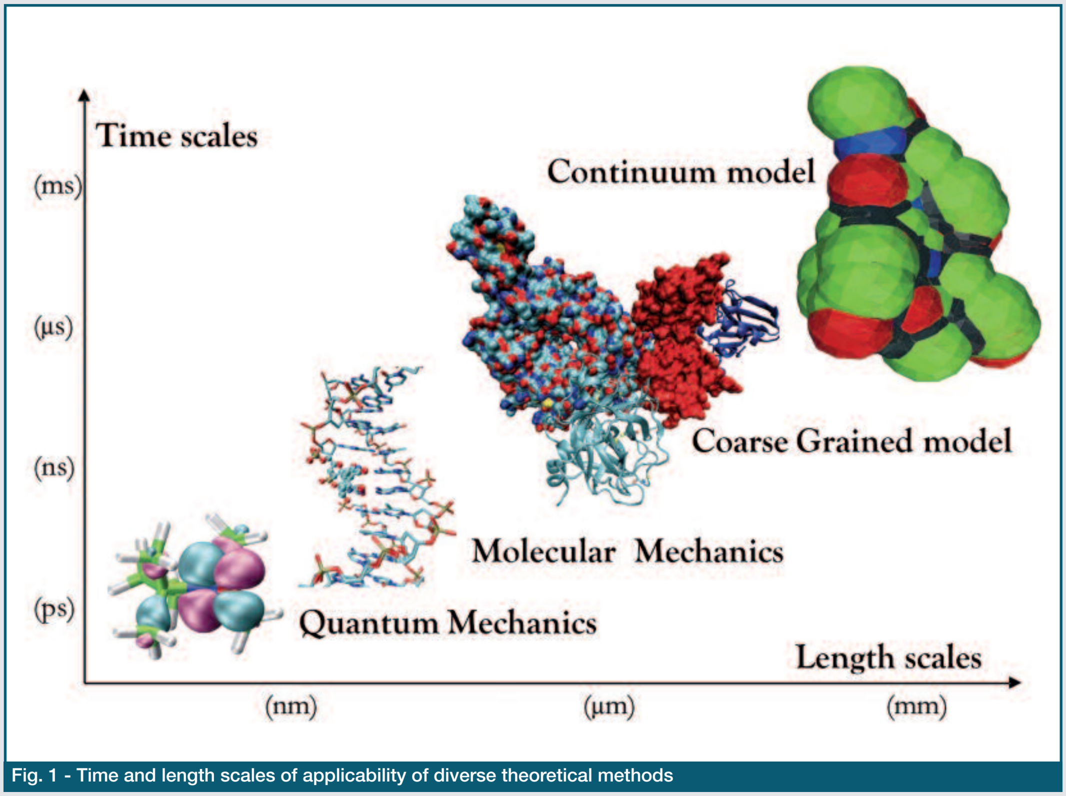

Fig. 1 - Time and length scales of applicability of diverse theoretical methods

From another perspective, the theoretical modeling of complex physico-chemical phenomena is usually feasible, but rarely trivial. A successful computational approach should rely on a proper description of the quantum mechanical behavior of interacting electrons and nuclei, in order to address molecular spectroscopy or chemical reactivity or the design of new nanoscale materials [2]. For example, the accurate first-principles prediction of several spectroscopic parameters has given rise to many remarkable synergic interactions among experimentalists and theoretically oriented chemists. Even within reliable and well tested spectroscopic characterization protocols, the choice among alternative structural hypotheses can be often guided by the agreement between measured and computed molecular parameters. It is quite conceivable that this kind of validation may become routine in a short time, when the computed parameters have disproven a seemingly reliable assignment, at least in those fields of molecular spectroscopy where complications due to flexibility and chemical heterogeneity of the investigated system are not overwhelming [3]. Despite the development of electronic structure methods, especially those based on the Density Functional Theory [4], the computational burden of such approaches imposes obvious limitations on the size of molecular systems amenable to a full QM treatment. To overcome these limits, great theoretical efforts have been devoted to the development of effective hybrid multiscale schemes, which employ full hierarchy of theory levels, from QM to finite-element methods, thus allowing the description of complex chemical processes and paving the route toward the direct and effective comparison between in silico and in vitro experiments.

General background

The aim of predicting molecular properties without any experimental inputs calls for the application of theoretical methods based on Quantum Mechanics (QM) [2]. Unfortunately, no universal QM method exists that is appropriate for all materials and phenomena; today, the most accurate methods are the ab initio post-Hartree-Fock (HF) [5] and Quantum Monte Carlo (QMC) [6] simulations: these methods describe exactly the electron-electron QM interactions of exchange, and very accurately the correlation for the electronic ground and excited states. The most severe drawback of such methods is that they are very expensive. Their computational burden become unaffordable when increasing the number of atoms in the system over few tens, thus preventing their effectively exploitation for real size molecular systems [1].

The most common family of methods for QM modeling uses or builds upon Density Functional Theory (DFT) [4]. DFT is a formally exact ground state theory in which the energy is expressed as a functional of the electron density alone; only two energy terms are usually approximated, because their exact functional expressions remain unknown: the electron kinetic, and the exchange-correlation (XC) energies. Accurate kinetic energies are obtained by introducing orbitals, within the Kohn-Sham (KS) DFT scheme [4]. KS-DFT calculations can be carried out for systems up to many hundreds of atoms. Moreover, computer programs with analytical Gradients and Hessians, and with analytical derivatives of the energy with respect to external fields provided a valuable tool for the characterization of the molecular potential energy surface (PES), for the study of the structure and dynamics of the molecules, and for the prediction of the spectroscopic properties [7]. The limit of KS-DFT in its current application is the approximate XC functional. In contrast to post-HF and QMC methods, no systematic means to improve the XC functional exists. Much effort has been expended to derive approximate XC functionals, either by keeping the functional ab initio (obeying to the known sum rules, bounds, and other physical constraints), or by allowing the functional to become (semi)empirical (fitting parameters in a chosen functional to experimental data) [8]. For most ground state properties, the semilocal generalized gradient approximation (GGA) XC functionals, provide good accuracy. The inclusion of some exact HF exchange, within a slightly more expensive hybrid HF-GGA scheme, improves significantly the accuracy of the method by correcting for the self-interac-

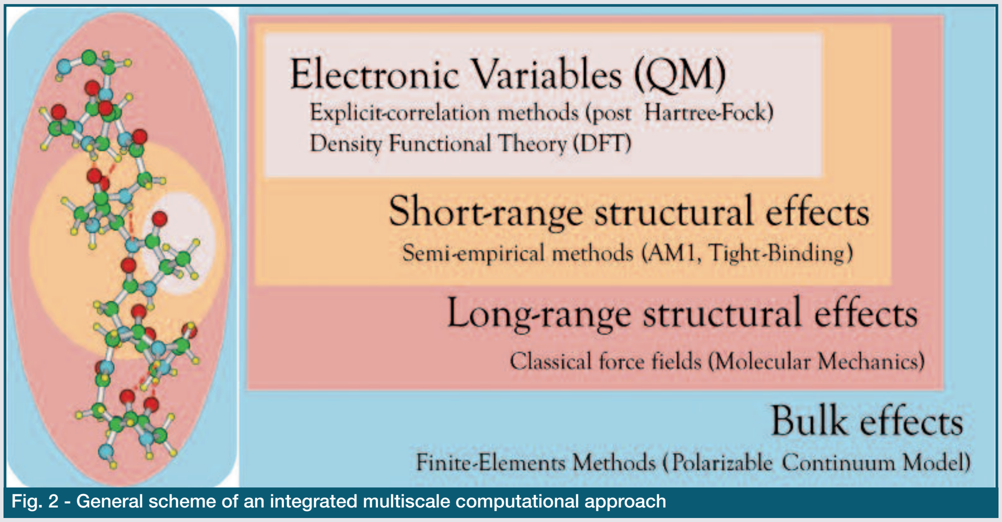

Fig. 2 - General scheme of an integrated multiscale computational approach

tion error [9]. Within this context, exited-state and optical properties can be computed by using the Time-Dependent (TD) DFT, a linear response theory that provides excitation energies, electron energy loss spectra and absorption spectra. The use of TD-DFT with recently proposed long-range corrected XC fuctional provides very promising results, also for difficult cases such as charge-transfer excitations [10]. Another critical issue of DFT is modeling the behavior of molecular, soft, amorphous, or heterogeneous materials. Interactions in soft materials or molecular crystals are determined in large part by van der Waals forces, which are not properly described by standard XC functionals in DFT [11]. However, current developments of tailored schemes purposed to include dispersion within a density functional model are providing very promising results [12].

Simulation of complex heterogeneous molecular systems motivates the recent strategy of multiscale modeling, which aims to bridge length and/or time scales. Many physico-chemical processes are "local". They take place in specific and well defined region of space, and occur in rather fast time, as for example, the chemical reactions in the active site of enzymes, the adsorption of molecules on heterogeneous catalyst surfaces, and the molecular parameters relative to some spectroscopic techniques (NMR chemical shifts, ESR hyperfine coupling constants). Under such circumstances, the computational approach can effectively use an integrated hierarchy of theoretical models, each of which addresses a specific space/time framework [13]. In favorable cases, the system under investigation can be safely divided into concentric layers, the focus being the portion of space where are located the atoms/molecules undergoing the physico-chemical process to be understood. This region represents the core of the system and needs a proper description by QM techniques. The rest of the system could be treated at a lower accuracy, for example by means of Molecular Mechanics (MM) [14], or even losing the atomistic definition, by using a continuum model simulating the macroscopic properties of the system bulk (see Fig. 2) [15]. However, the QM core is not described as an isolated system. It "feels" the surrounding atoms via an effective Hamiltonian that includes the interaction between subsequent layers. For example, in QM/MM calculations it is convenient that the QM core is polarized by the MM partial atomic charges of the embedding layer [16]. Such multiscale partition scheme can be implemented also for more than two layers, the best compromise being reached by the use of methods with decreasing accuracy and increasing efficiency when going outward from the focus of the process.

Furthermore, the issue of structural flexibility may well represent the main bottleneck in the formulation of a general computational protocol. In favorable cases, grid searches and relaxed potential energy scans can be used, but the involved computational effort rules out these simple approaches when the complexity of the problem increases, and other conformational search techniques must be brought into play. For these screening phases one wants to employ lower levels of theory than, e.g., for the more demanding computation of spectroscopic parameters. In particular, exploration of the potential energy surface by means of molecular dynamics (MD) simulations would be much easier by the use of reliable force fields based on MM [14]. Unfortunately, in some cases these are still not available. The development of a tailored force-field is a possibility, albeit not the most attractive one from the viewpoint of a non-specialist, probably. Conceptually, solvent effects can also be related to structural flexibility. In both cases the central issue is the difficulty of adequately sampling a proper phase space. Approaches in which the solvent degrees of freedom are accounted for in an average way, e.g., in terms of a reaction field, have proven to be particularly effective. The polarizable continuum model (PCM) [15] represents the best known example. This model consists of embedding the solute molecule within a cavity generated by the envelope of atom-centred spheres, whose radii have been parameterized for reproducing experimental solvation free energies; inside the cavity, the relative dielectric constant has the same value as in a vacuum ($\varepsilon = 1$), and it steeply goes to the solvent bulk value outside (e.g., $\varepsilon = 78.4$ for water). The presence of a solute perturbs the polarizable medium, which undergoes new equilibration acting back onto the solute molecular parameters by means of an effective reaction field. Such phenomenon is described by the PCM in terms of a pattern of effective charges on the cavity surface. The simple introduction of the PCM often provides substantial improvements in energies, geometries and spectroscopic parameters computed for molecular systems in a variety of solvent media. However, the PCM may also display limitations in some cases, e.g., when highly specific interactions, like hydrogen bonds, come into play. Cluster approaches can provide a straightforward route to describe localized phenomena, including the spectroscopic transitions of molecules; once again, the PCM can be brought into play to account for the "bulk" solvent, and to reduce the number of solvent molecules to be described explicitly [17]. In other words, such a discrete-continuum scheme takes into account both the quantum mechanical nature of specific solute-solvent interaction, i.e. hydrogen bonds, and the classical Coulomb long-range effects tuned by the dielectric properties of the solution. The resulting cluster-PCM description represents a very versatile tool that can be adapted to different structural and spectroscopic situations. Which and how many solvent molecules need to be explicitly described is a question that has to be defined case by case, depending on the nature, shape and size of the molecule. Moreover, a single structure of the solute-solvent cluster or supra-molecular frame could not be representative of a very flexible situation [18]. This means that the structures of the solute and its closest solvent molecules must be statistically averaged among all the energetically accessible configurations. Thus, the cluster-PCM approach is also attractive for the computation of averaging effects brought about by dynamics. Physically, the measured molecular properties reflect the contribution of different populated configurations. From a computational viewpoint (and assuming ergodicity holds), when a representative MD simulation has been generated, the statistical distribution of the parameter values in the system can be reproduced by extracting from the trajectory a sufficient number of uncorrelated frames, and repeating a cluster-PCM calculation on each of them. Thus, averaging the computed values provides an estimate of the observable.

As far as the nature of the simulation is involved, classical MD provides many advantages for the cheapness of the MM approach, but also increases the difficulties for the unavailability of reliable force fields. This represents a concrete limitation. Alternative approaches are raising to provide dynamical trajectories which are long enough to allow for reasonable averaging of spectroscopic parameters. They include classical simulations employing energies and gradients computed by QM methods, i.e., Born-Oppenheimer dynamics and, especially, extended-Langrangian approaches based on the Car-Parrinello scheme [19]. It is worth noting that the quality of the potential-energy landscape underlying Car-Parrinello simulations, which presently rely on conventional DFT, may well turn out to be insufficient for a direct use in the computation of structurally sensitive parameters. On the other hand, the aforementioned multiscale approach has also effectively been implemented in a time-dependent scheme for the finite-temperature dynamical simulations of complex liquids and solutions, in close resemblance with laboratory experiments [20].

The role of molecular dynamics and chemical environment in tuning spectroscopic observables

The accurate prediction of spectroscopic properties of complex molecular systems represents a challenging task with a remarkable impact on chemical studies in both the fields of Life and Materials sciences.

The theoretical machinery for the calculation of spectroscopic parameters is well established [13]. However, the standard QM approach, routinely applied by non-specialists, is still based on the description of the molecular system under investigation in the gas-phase, and at its equilibrium minimum-energy structure. Such a static picture could hardly be related to real-life experiments on complex systems performed at finite-temperature and/or in condensed phases. Therefore, the improvement of the accuracy in the simulations of spectroscopic properties requires to account for both dynamics (i.e. molecular vibrations) and environmental (at least solvent librations) effects.

In particular, intra-molecular dynamics effects related to molecular vibrations affect several spectroscopic properties [21], and should be taken into account in the calculation of all spectroscopic parameters, not only for IR and Raman. In the case of electronic transitions, the vibrations are responsible for the vibronic structure of UV-Vis absorption, fluorescence emission or electronic circular dichroism spectra via the Franck-Condon selection rules (direct effect) [22]. On the other hand, the energies governing the transitions of interest in magnetic resonance spectroscopy are much lower than vibrational quanta, so that the long-time dynamics of the molecule, e.g. the diffusion and rotational motions, determine the spectral line-shapes. However, also in this case the temperature has the non-negligible effect of averaging out the magnetic parameters over the structural fluctuations of the molecule, i.e. the short-time dynamics due to molecular vibrations (indirect effect) [23].

In particular, Fig. 3 shows the Nitrogen hyperfine coupling constant (hcc_N) values with respect to the out of plane angle of a nitroxide spin probe (PROXYL) computed along the gas phase MD trajectory [24]. It clearly appears a direct structure-property relationship, which can be rationalized by an analysis of the electronic structure of the nitroxide

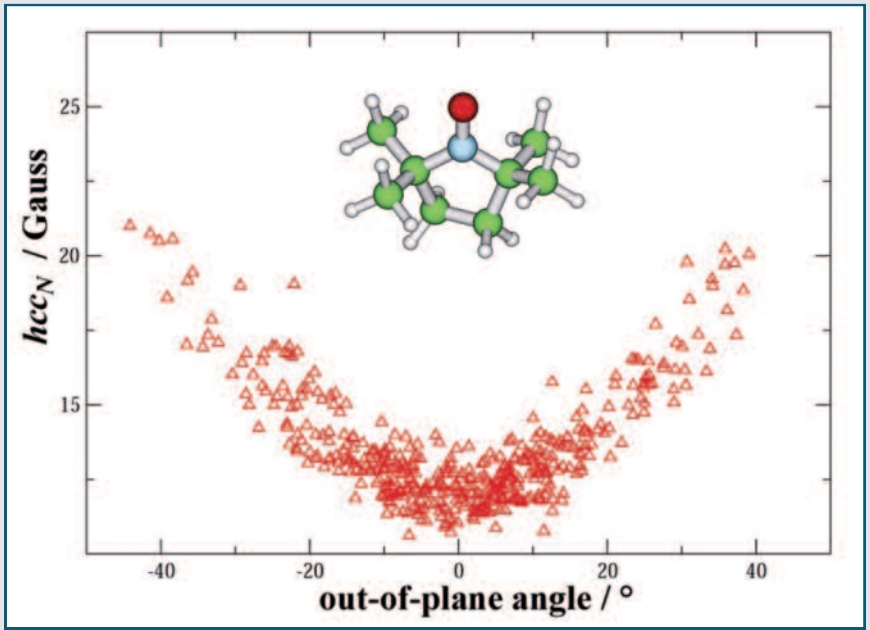

Fig. 3 - Nitrogen hyperfine coupling constant (hcc_N) versus out-of-plane angle of the PROXYL spin-probe as computed along an ab initio MD trajectory

Tab. 1 - Solvent shifts of spectroscopic observables when going from gas phase to aqueous solution. Values inter parentheses are percentage errors

Acetone UV ($n \rightarrow \varepsilon^*$) (eV)			Acetone ^{17}O NMR (ppm)		
OPT	<AIMD>	Exp.	OPT	<AIMD>	<Exp.
0.39 (95%)	0.21 (5%)	0.2	101 (33%)	83 (9%)	76
Proxyl hcc_N (Gauss)			**Proxyl Δg_{iso} (ppm)**		
OPT	<AIMD>	Exp.	OPT	<AIMD>	Exp.
2.44 (4%)	2.5 (6%)	2.35	-341 (21%)	-428 (0.5%)	-430

probe. In the planar conformation of the nitroxide (which corresponds to minimum on the potential energy surface for the PROXYL), the only contribution to hcc_N is provided by spin polarization. Pyramidalization at the nitrogen center allows for the direct involvement of the nitrogen s orbital in the singly occupied molecular orbital (SOMO), and this results in a large contribution to the Fermi contact term, which overcomes the spin polarization contribution and determines the observed increase in hcc_N. Therefore, a proper treatment of such a dynamic effect led to an average value of 14 Gauss, quantitatively in agreement with the reference experimental value of 14.02, and quite far from the minimum-energy structure counterpart of 11.8 Gauss [24]. Noteworthy, the classical MD average is very closely in agreement with the value of 14.4 Gauss obtained by describing the molecular dynamics of Proxyl with a QM time-independent approach based on the solution of the vibrational Hamiltonian, which accounts for anharmonicity via a feasible and effective second order perturbation scheme [25].

Regarding the chemical environment, the PCM approach was able to effectively reproduce the influence of the solvent on the spectroscopic parameters. In some instances, the explicit introduction of some first-shell solvent molecules [21] also proved to be necessary. The resulting discrete/continuum (cluster/PCM) model is very powerful, when a correct sampling of the solute-solvent energetically accessible configurations can be obtained. However, one could expect that the use of geometry optimized solute-solvent clusters for the computation of spectroscopic properties could lead to an overestimation of the solvent effects, since the thermal fluctuations of the system are being essentially neglected [17]. When these subtle influences are of interest, MD simulations represent the method of choice for exploring the time evolution of liquid phase systems at finite temperatures. A detailed analysis of the many features and advantages of different MD approaches is clearly beyond the aim of this Review [14]. In this section, we just intended to outline the importance of a dynamical description also for solute-solvent systems, when the spectroscopic computation aims at an accuracy quantitatively comparable with experimental data. For obvious reasons, the following examples concern aqueous solutions.

The effect on the UV ($n \rightarrow \pi^*$ transition) and NMR (^{17}O chemical shifts) parameters of a carbonyl group will be described in the following in the case of the acetone molecule. Instead, dynamical solvation effects on ESR observables [Nitrogen hyperfine coupling constants (hcc_N) and the isotropic g-tensor shift (Δg_{iso})] will be illustrated for the nitroxyl group of a prototypical spin-probe nitroxide, namely the PROXYL. Details of the Car-Parrinello (CP) and classical MD simulations and of the spectroscopic calculations are available elsewhere [18, 26]. Here, we wish to compare the different predictions of solvent effects that are provided by the optimized-cluster approach, and by averaging of the computed parameters along the CPMD trajectories. Tab. 1 summarizes the computed solvent shifts with respect to the corresponding experimental values, for each of the spectral observables listed above.

The acetone UV $n \rightarrow \pi^*$ transition energy directly involves a change in the electronic structure of the molecule. The dielectric properties of the solvent affect the relative positions of the ground and the excited states, but in addition to this indirect effect, when the solute is dissolved in a protic solvent like water, the presence of hydrogen bonded molecules allows for a partial delocalization of the acetone electron density toward the H-bonding partners. The computed values of the UV transition energies are thus strongly sensitive to any change in the position of the surrounding water molecules. As a consequence, the solvent shift computed on the supra-molecular minimum-energy structure is overestimated with respect to the experimental value. Averaging over the CPMD trajectory provides a very satisfactory agreement with experiment. Purposely developed classical MM force field for the acetone molecule also provided great accuracy, ensuring at the same time a proper statistical sampling of the liquid solution phase space [27].

NMR chemical shifts represent in a sense more "local" properties. However, parts of the molecule that are directly exposed to the solvent medium may display a marked dependence on specific solute–solvent interactions, e.g., hydrogen bonding. A case in point is the carbonyl oxygen atom of acetone in aqueous solution. Again, solvent dynamics exerts a clear effect on the computed solvent shift, as listed in Tab. 1. Within the accuracy of the method employed to compute the property, averaging over the MD trajectories reproduces rather well the experimental figure, while the static cluster-PCM approach clearly overestimates solvent effects.

Going now to the ESR parameters of the nitroxide spin probe (Proxyl) in aqueous solution, and focusing on the most commonly available data, i.e., nitrogen hyperfine coupling constant (hcc_N) and isotropic shift of the electronic g-tensor (Δg_{iso}), Tab. 1 points out a substantial convergence of the static minimum-energy approach with the results obtained by averaging over the MD trajectory in the case of hcc_N. The nitrogen atom is not greatly affected by the dynamics of the solvent molecules, since it is not directly exposed to the aqueous medium. Thus, the solvent is well described by a mean field model, which is essentially unchanged in the optimized cluster and in the dynamical average. However, the picture changes when Δg_{iso} is considered. In this case (much like for the UV transition energy), the spectroscopic value depends strongly on the subtle features of the solute electronic wave-function, and thus on the specific geometry of the H-bonding network. The difference between the minimum energy and the CPMD averaged values, listed in Tab. 1, is quite large, the second being quantitatively in agreement with experiment [26].

Overall, the cases here discussed highlight the importance of taking into proper account the effects of solvent dynamics when computing spectroscopic properties for system in the liquid phase. The quality of the dynamics and the extent of the sampling (time scale of the MD simulations) are also crucial to the accuracy of the final results.

Conclusions and perspectives

The present review provides an outline of the theoretical methods to day available for an effective modeling of complex chemical systems.

The above examples are related to the prediction of spectroscopic properties of molecular probes in liquid solutions, pointing out the relevance of a proper accounting, within the computational protocol, of two key factors, namely the thermal averaging effects and the role of the chemical surrounding. Noteworthy, the effectiveness of the described multiscale integrated approach based on QM/MM/PCM methods, together with Car-Parrinello or Classical MD simulations, has been highlighted for the very good agreement with experimental references.

The development of new density-functional models for the QM description of complex phenomena and further integration of the multiscale approach to longer space and time scales by using, respectively, coarse grained techniques and stochastic methods are promising to bridge the gap between the molecular and the macroscopic worlds. Indeed, by pursuing the aim of a direct comparison between experiments and simulations, theoretical developments will deliver a valuable tool for supporting interpretation of ambiguous laboratory outputs and for driving design and synthesis of new materials toward unexplored directions.

Acknowledgments: The author (M. P.) feels deeply in debt with Prof. Vincenzo Barone for the many lessons learned from him. All the people at the Laboratory of Structure and Dynamics of Molecules (http://lsdm.campusgrid.unina.it) are kindly acknowledged.

Since November 2008 Michele Pavone is Researcher at the Department of Chemistry of University of Naples Federico II. In 2004, he graduated cum laude in Chemistry at the University of Naples Federico II and at the same University he received the Ph.D. in Chemical Science in 2007. Since 2002 he has been a member of the research group directed by Prof. Vincenzo Barone, working on the development and the application of computational approaches for an effective description of physico-chemical processes.

References

[1] a) V. Barone, A. Polimeno, *Chem. Soc. Rev.*, 2007, **36**, 1724;
b) E.A. Carter, *Science*, 2008, **321**, 800.
[2] R. McWeeny, Methods of Molecular Quantum Mechanics, Academic Press, London, 1992.
[3] V. Barone *et al., J. Mol. Struct.-Theochem,* 2007, **811**, 323.
[4] W. Koch, M.C. Holthausen, A Chemist's Guide to Density Functional Theory, Wiley-VCH, Weinheim, 2002.
[5] A. Szabo, N.S. Ostlund, Modern Quantum Chemistry, Dover, New York, 1989.
[6] W.M.C. Foulkes *et al., Rev. Mod. Phys.,* 2001, **73**, 33.
[7] J.B. Foresman, A.E. Frish, Exploring Chemistry With Electronic Structure Methods, Gaussian Inc., Wallingford, 1996.
[8] S. Kurth *et al., Int. J. Quantum Chem.,* 1999, **75**, 889.
[9] C. Adamo, V. Barone, *J. Chem. Phys.,* 1999, **110**, 6158.
[10] D. Jacquemin *et al., J. Chem. Theory Comput.,* 2008, **4**, 123.
[11] S. Kristyán, P. Pulay, *Chem. Phys. Lett.,* 1994, **229**, 175.
[12] M. Pavone *et al., Chem. Phys. Lett.,* 2008, **452**, 333.
[13] V. Barone *et al., Acc. Chem. Res.,* 2008, **41**, 605.
[14] A.R. Leach, Molecular Modelling: Principles and Applications, Prentice Hall, London, 2001.
[15] J. Tomasi *et al., Chem. Rev.,* 2005, **105**, 2999.
[16] T. Vreven,*et al., J. Chem. Theory Comput.*, 2006, **2**, 815.
[17] M. Pavone *et al., J. Am. Chem. Soc.,* 2006, **128**, 4338.
[18] O. Crescenzi *et al., J. Phys. Chem. B,* 2005, **109**, 445.
[19] R. Car, M. Parrinello, *Phys. Rev. Lett.,* 1985, **55**, 2471.
[20] N. Rega *et al., Chem. Phys. Lett.,* 2006, **422**, 367.
[21] R. Improta, V. Barone, *Chem. Rev.,* 2004, **104**, 1231.
[22] F. Santoro, V. Barone, *Int. J. Quantum Chem.,* 2010, **110**, 476.
[23] M. Zerbetto *et al., Comp. Phys. Commun.,* 2009, **180**, 2680.
[24] M. Pavone *et al., J. Phys. Chem. B,* 2010, **114**, 11509.
[25] V. Barone, *J. Chem. Phys.,* 2005, **122**, 014108.
[26] M. Pavone *et al., J. Phys. Chem. B,* 2007, **111**, 8928.
[27] M. Pavone *et al., ChemPhysChem,* 2006, **7**, 148.

RIASSUNTO

Modeling ab initio di sistemi chimici complessi

Questa Critical Review offre una panoramica delle attuali metodologie in modellistica computazionale: gli approcci teorici allo stato dell'arte sono discussi in relazione alla loro accuratezza e al relativo costo computazionale, mettendone in luce alcune applicazioni di successo e i limiti.