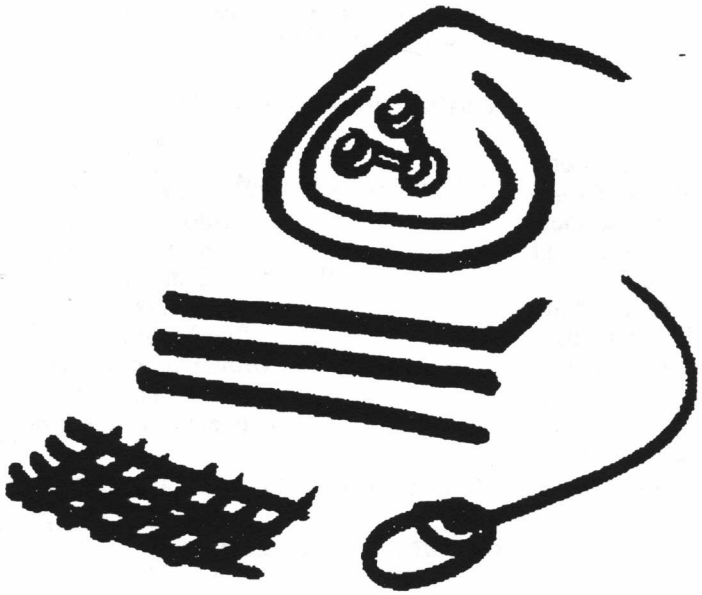
Università degli Studi FEDERICO II Consiglio Nazionale delle Ricerche

GNIC
Gruppo Nazionale di
Informatica Chimica

GICC
Gruppo Interdivisionale
di Chimica Computazionale SCI

3° CONVEGNO NAZIONALE
di
INFORMATICA CHIMICA



27 febbraio - 1 marzo 1997

Università degli Studi FEDERICO II
Facoltà di Ingegneria
Piazzale Tecchio - 80125 Napoli

2-9

MODELING CHEMICAL REACTIONS IN COMPLEX SYSTEMS

Giuliano Alagona, Guido Germano** and Caterina Ghio**

* *Istituto di Chimica Quantistica ed Energetica Molecolare, CNR, Pisa*
** *Dipartimento di Chimica e Chimica Industriale, Università di Pisa*

Background and motivation

The force fields developed in the last 25 years describe with chemical accuracy intra- and intermolecular interactions. Their parameters derive mostly from high quality ab initio calculations performed on smaller molecules, either subparts of the whole system or model compounds, and to a minor extent from experimental data. They have been employed successfully for molecular mechanics (MM) geometry optimizations and for molecular dynamics (MD) and Monte Carlo (MC) simulations of systems with a few tens of thousands of atoms.

However, as their framework is classical mechanics, and the objects they act on are nuclei described as beads connected by springs, they are inadequate to treat phenomena where the electronic degrees of freedom cannot be neglected. The most interesting to the chemist are reactions. Indeed, it is surely appealing to be able to make statements by the mere power of calculus on the structure and on several physical properties of a substance, but what really has characterized chemistry from its beginning has been the study on how to turn one substance into another, i.e. the study of reactivity.

Force fields are not completely useless in this field, as it is possible to slice the reaction path into windows and compute energy and free energy differences between reagents, some selected intermediates, and products, but it is impossible to sample continuously the path joining these separate frames, and to identify with certainty transition states and consequently activation energies.

Of course there have been around for about 50 years methods of molecular quantum mechanics (QM), which are in the meanwhile consolidated and widespread, and do a good job. However they require so much computational resources that only systems with up to hundred atoms are tractable in a reasonable time on state of the art computers. This limitation is ultimately the reason why people came along with force fields, but unfortunately the latter, while gaining in speed, lost important capabilities of the ab initio methods.

Hundred atoms can be enough for some purposes, but what should be done when the inclusion of a larger environment in the model cannot be avoided? This is the case, for instance, in reactions where the solvent plays an important role, and even more in enzyme reactions, towards the understanding of which there has been for long time a considerable interest of both theoretical and practical nature.

We review shortly five methods:

1) Brute force approaches: hardware improvements like faster processors, gigabytes of RAM, parallel architectures - the latter involving a non trivial rethinking of the computing procedure as well. Of course all the following methods benefit from improvements of this kind too.

2) Smarter ways to solve the fundamental equation, so that the number of needed operations scales with a power of the number of atoms N smaller than 3, linear scaling being the ultimate goal. These efforts are pursued both with respect to the Schrödinger equation and the Kohn-Sham equation (Density Functional Theory), and have been given by the authors evocative names like "Divide and conquer"[1] in the first case and "Taking advantage of the near-sightedness of matter"[2] in the second.

3) First principles MD (Car-Parrinello)[3]. This kind of MD is attractive not only because it considers the electronic degrees of freedom too, but also because, as the name suggests, it does not need previously a tiring development of new force field parameters for every newly encountered chemical environment of an atom, and is as completely general as ab initio methods. Many leading scientists in the field share the opinion that the MD of the future will be

ab initio MD[4], but as several practical and conceptual problems remain to be solved, nobody can tell when this future will come. One well known practical problem arises in aperiodic systems because the plain waves development of the density functional (actually the best working implementation, though Gaussians have been experimented too) requires a huge amount of memory and computing resources. A more fundamental problem is that the trajectory is bound to the ground state Born-Oppenheimer energy surface.

4) Methods for reactions in solution. The methods seen so far allow to increase the size of the studied system. This method and the following one take a different approach. They conserve a traditional ab initio description of a small number of atoms around the reaction center, where bonds are broken and formed, and try to incorporate in some way the effect of a much larger environment treated at a lower level of accuracy. For instance, a homogeneous solvent can be described as a continuous dielectric medium with a cavity containing the solute molecules. The charge distribution of the solute induces polarization charges on the cavity surface; a proper generation of these and their inclusion in the molecular hamiltonian provides a good and cheap description of solution properties[5]. However this approach is not suited to more complex environments like e.g. enzymes.

5) Combined QM/MM. The system is divided into two regions, one treated by QM and one by MM. The former region is centered around the reaction center and is rather small, while the latter can encompass as many as some tens of thousands of atoms. The construction of a mixed QM/MM potential is not difficult[6], even if an optimal solution on how to treat covalent bonds crossing the barrier has not been devised yet. But in spite of the triumphalistic tones of some literature, according to which everything has been already solved in this field, there remain relevant problems of a practical nature: we are not aware of any publicly available program that implements a fully combined QM/MM approach along the published theoretical guidelines. Even the apparently simple task of ab initio geometry optimizations in the field of fixed external electrostatic charges, an essential prerequisite for combined QM/MM, is either not supplemented at all or supplemented only badly within the most popular software packages. Gaussian 94[7] for instance does not have this feature.

A simple simulation protocol

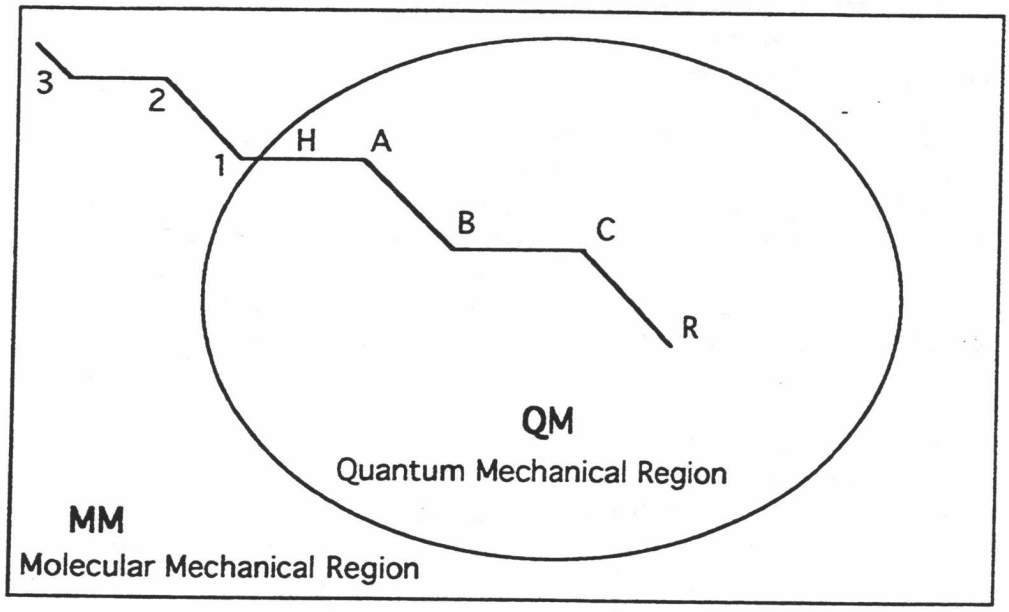
The panorama outlined above seems depressing. There are good starting points, but they all end sooner or later against an obstacle, not of a fundamental nature, yet big enough to apparently delay for years the possibility to go along that way up to the goal. In the spirit of a quick and dirty implementation of a combined QM/MM method in the field of enzyme catalysis taking advantage of a popular QM program[7] and a popular MM program[8], we propose the following simulation protocol:

1) Saturate with hydrogen atoms the open valences left in the QM region by covalent bonds crossing the barrier (see the figure at the end). Ignore the electrostatic charges on the first and second MM neighbors beyond the barrier, as they are too close to the last QM center (the "link" hydrogen) and produce spurious effects on the wavefunction. Resorting to a link hydrogen is not the most sophisticated solution, but clear and definitive arguments in favor of more involved schemes are still missing. Moreover it is easy to implement, and it has been the classic choice since the first attempts[9] of combined QM/MM.
2) Extend Gaussian 94 to allow efficient ab initio geometry optimizations in the field of the electrostatic charges placed on the MM centers. To this purpose, add the contribution of the external charges to that of the nuclei in the monoelectronic hamiltonian, and add the contribution of the MM centers to the energy gradient felt by the nuclei when they move. The interaction between the MM centers and the nuclei must be of both electrostatic and Van der Waals type. Otherwise, in absence of a repulsive part of the potential, a collapse of the nuclei into the MM centers is at risk. The first, second and third nuclei along the covalent bond chain interrupted by the QM/MM boundary must instead be subjected to bond, angle, dihedral and 1-4 nonbonded

MM potentials, or be fixed in space, to prevent them to drift away from their MM partners beyond the barrier, leading thus to irrealistically huge MM potential energies in step 4.
3) Optimize the coordinates of the QM part keeping the MM centers fixed.
4) Optimize the coordinates of the MM centers using the new geometry of the QM part obtained in the previous step kept fixed in their turn.
5) Repeat steps 3 and 4 until no more significant variations are detected.
6) Move forward the leading parameters along the reaction coordinate in the QM part, and return to step 3.

This scheme reproduces both the constraints imposed on the position of the reacting moieties by the environment and the modification of the energy surface along which the reaction develops. It does not account for dynamic aspects, however, i.e. how the environment provides the reacting species with the kinetic energy necessary to cross the energy barriers that separate them from the products. Instead, the environment is pulled behind by the QM part, which is on its turn pulled uphill by the simulator. Of course a fully integrated motion of both the QM and the MM part together, in place of the two succesive steps 3 and 4 repeated until self-consistency, would be more satisfactory. Nevertheless, considering the moderate programming cost of this approach, it seems worthy to try.

We are applying this protocol to the study of the reaction in Mandelate Racemase. This communication is intended to focus on the method; preliminary results have been communicated elsewhere[10].



R is the reaction center, which requires a QM treatment. When a covalent bond is split across the boundary between the two regions, the free valence of QM atom A is filled in by a so-called link hydrogen H placed at the appropriate distance from atom A along the line between atom A and the last MM atom 1. The interaction between the electrons in the QM zone and the charges on the MM groups containing 1 and 2 is neglected, because these groups are too close to H. QM nucleus A is connected by a harmonic MM bond potential to MM atom 1, by an angle potential to atoms 1 and 2, by a torsional potential to atoms 1, 2 and 3, and by a 1-4 nonbonded potential to atom 3. QM nucleus B is subject only to an angle potential with A and 1, a torsional potential with A, 1 and 2, and a 1-4 nonbonded interaction with 2. Last, QM nucleus C feels a torsional interaction with B, A and 1, and a 1-4 nonbonded interaction with 1. Alternatively, atoms H, A and B can be fixed in space.

References

1 S.L. Dixon, K.M. Merz, *J. Chem. Phys.* **104**, 6643-6649 (1996)
2 W. Kohn, to be published
3 R. Car., M. Parrinello, *Phys. Rev. Lett.* **55**, 2471-2474 (1985)
4 Proceedings of the Euroconference on Computer Simulation in Condensed Matter Physics and Chemistry "Monte Carlo and Molecular Dynamics of Condensed Matter Systems", Como, July 3 - 28, 1995, K. Binder, G. Ciccotti Editors, Vol. 49, SIF Bologna 1996
5 J. Tomasi, M. Persico, *Chem. Rev.* **94**, 2027-2094 (1994)
6 J. Gao, *Reviews in Computational Chemistry* Vol. 7 Chapter 3, K.B. Lipkowitz, D.B. Boyd Editors, VCH Publishers, New York 1996
7 M.J. Frisch et. al., Gaussian 94, Rev. D.4, Gaussian, Inc., Pittsburgh PA, 1996
8 D.A. Pearlman et. al., Amber 4.1, University of California, San Francisco, 1995
9 U. Chandra Singh, P.A. Kollman, *J. Comp. Chem.* **7**, 718-730 (1986)
10 G. Alagona, G. Germano, C. Ghio, P.A. Kollman, poster communication to the National Congress of the Physical Chemistry Division of the Italian Chemical Society, Pisa, February 10 - 14, 1997