Hybrid Usage of Computational Tools in Drug Synthesis

Canan Atilgan1,* and Viktorya Aviyente2

1Faculty of Engineering and Natural Sciences, Sabanci University, Orhanli, 34956, Tuzla, Istanbul, Turkey
2Chemistry Department, Faculty of Art and Sciences, Bogaziçi University, 34342, Bebek, Istanbul, Turkey

Abstract: We describe several computational methodologies used in aiding the chemical synthesis of drugs. We first summarize quantum mechanical approaches that weigh thermodynamical and kinetic factors in selecting the possible pathways during synthesis. The two major problems encountered in computational approaches are the efficient sampling of the conformational space and the incorporation of solvent effect into the system of interest. Thus, conformational search methodologies of small to medium sized molecules, with emphasis on cyclic molecules, are reviewed. Also, the analysis of the solvent effect on the synthesis of drug molecules and yield, using continuum methodologies as well as molecular dynamics, is discussed. How results from these studies are in turn fed back into detailed quantum mechanical calculations with supermolecules of solvent and reaction site are outlined. It is shown that the usage of several computational techniques hand-in-hand provides a plethora of information that may be utilized during the actual synthesis of drug molecules.

Keywords: Conformational analysis, molecular dynamics, quantum chemical computations, rational drug design, solvent effect, macrolide antibiotics.

I. MOTIVATION

Currently there are many alternative therapeutic approaches to conventional drugs. Amongst these, therapies catered to the needs of the individuals are attracting increasing attention. Preventive techniques are also of great interest, as well as using biotechnological approaches to discover if an individual is genetically predisposed to certain disorders. However, the interest in developing new conventional drugs, effective on “all”, is also ever-so-growing. Moreover, new approaches to their discovery are always welcome, considering the competition in the market, in terms of decreasing both the time and money spent on putting a drug into clinical use. In this review we will confine our attention to using computational techniques to aid the chemical synthesis process of such drugs.

There are two general approaches taken in the rational design of new drugs: One either chooses the target (usually a protein) along the biochemical pathways related to the disease/disorder, one then identifies the sites on the protein to which binding of a small molecule is desired, so that the function of that protein may be prevented or changed. This process involves the identification of the target [1], as well as its validation, selection of molecules to be used as leads, and high-throughput screening for a specific target [2]. A knowledge of the three-dimensional structure of the protein, determined by X-ray crystallography or NMR, is essential in this process. Along with other experimental and computational tools, they lead to structure/function/dynamics information useful for understanding biomolecular function [3]. Recent advances in genomics have initiated collaboratories to attack the problem of high-throughput structural genomics [4], working towards obtaining high quality crystals [5], and automated NMR [6] and X-ray [7] structure determination. Identification of the three-dimensional structures of these proteins in complex with ligands in the same biochemical pathway greatly improve the target selection procedure as a means of providing information on their modes of functioning [8]. These efforts have led to fruitful results whereby the modes of binding and inhibition of drug targets have been studied systematically and applied to lead optimization, as exemplified by protein kinases [9]. Nevertheless, it should be kept in mind that any of these structures provide a snapshot from the real process, or rather, an averaged “static” view of the truly dynamical processes involved. Recently, a new methodology that combines NMR spectroscopy techniques and molecular dynamics (MD) simulations to concurrently determine the structure and dynamics of proteins was developed [10]. Such an approach allows the study of small-molecule interactions with multiple, viable conformations of the backbone and side-chains, reducing the estimation errors in binding affinities, thus opening new avenues in drug design.

The other approach to rational drug design involves modification of an existing potent drug, usually isolated as a natural product (a first-generation drug). This is facilitated by the need to increase the potency of the drugs, or to overcome drug resistance [11-14]. In either case, the tactics involve using the scaffold provided by a molecule which is already known to function at a certain point in the biochemical pathway, but modifying positions along that scaffold for better efficacy [15]. There are several novel approaches for the more efficient synthesis of molecules with medicinal value. In particular, advances in methodologies exploiting enzymes have led to the development of many novel bio-active molecules [16-18], in many cases making combinatorial biosynthesis possible [19]. Nevertheless, conventional approaches to molecular synthesis, using sequences of chemical transformations, are also widely applied (for a recent example see [14]).

There are several standing problems during the synthesis of drug molecules, the most obvious being that of yield. Since the synthesis of these molecules usually involve multiple steps, there are several isomers that arise in the final product, only one of which corresponding to the desired product. Thus, new approaches encouraging the synthesis of that isomer is in demand, and many patents target to solve this problem alone.
(see, for example, ref [20]). Methods involving the use of thermodynamical vs. kinetic control over the product, by varying the environmental factors such as temperature and type of solvent are common. Other syntheses involve “mechanical” approaches such as introducing large side-chains so as to create steric effects near regions where reactions may lead to undesired isomers; e.g., see ref. [21].

Depending on the length scale of the problem considered, i.e. the size of the molecule under investigation, and the amount of averaging that may be performed with the currently available computers, computational approaches provide a means of easing some of these problems. This is achieved through saving the time and the money spent by effective screening of the many possible alternatives in silico, before actually performing the wet experiments. For example, the solvent has a large effect on the conformations sampled by the molecules. It is now possible to explore that effect on the synthesis process prior to actually performing the synthesis [22, 23]. Another problem that may be addressed computationally involves which protecting groups would best encourage the targeted sites for reactions such as methylation, protonation, etc. [24]. Moreover, computational tools will further act as a means of explaining why a certain outcome is observed [22, 25].

It is possible to use both quantum mechanical (QM) and classical computational tools in tackling with these problems. In principle, when a chemical reaction is involved, we resort to quantum chemistry, whereas when we investigate which of the plethora of possible conformations occur predominantly in a given environment, we use molecular mechanics (MM). If we are interested in an understanding of the dynamical processes, we resort to molecular dynamics (MD). For an understanding of a given problem from all aspects, it may prove invaluable to apply all of these methodologies hand-in-hand. In this review, we limit ourselves to the above-mentioned, well established and easily accessible computational methodologies, and omit the recent developments in the more elaborate method of quantum dynamics (see reference [26] for the state-of-the-art in this methodology).

We thus deal with the hybrid usage of several computational methodologies to get a better understanding of the drug synthesis process in the following review. In section II, we provide a general description of the methodologies used, i.e., QM, MM, and MD. In section III, we discuss how different pathways are selected during synthesis and describe how computational methodologies may be used in their prediction. In section IV we describe the problem of conformational search (CS) in molecular modeling, and give an outline of the approaches used in the CS of molecules under geometrical restriction, which includes many of the drug molecules. We discuss the effect of solvent in synthesis, and we describe how computational techniques may be used to assess this effect in section V, followed by the concluding remarks of section VI.

II. COMPUTATIONAL METHODOLOGIES

Even a very general account of the computational techniques developed for molecular modeling is beyond our scope. Below we refer to the context in which the various techniques are used for the problems in drug synthesis. Note, however, that these techniques have been around for quite some time, and are so well established that there are many rules and regulations related to how they are used. As such, the modeler need not be very experienced for an effective application of the methodologies to the problems at hand.

Quantum Mechanical Calculations

The complete mathematical description of a molecule includes both quantum mechanical and relativistic effects. Due to the complexity of the problem with small scales, and large velocities, assumptions need to be made. The Born-Oppenheimer approximation takes into account the fact that electrons are several thousand times lighter than the nuclei and therefore move much faster; i.e. the nuclei may be considered stationary.

The fundamental postulate of quantum mechanics is that a wave function exists for any system and that appropriate operators which act upon any system return the observable properties of the system. In the Schroedinger equation,

$$\hat{H}\psi = E\psi \tag{1}$$

the Hamiltonian operator $\hat{H}$ takes into account five contributions to the total energy of a system: the kinetic energies of the electrons and nuclei, the attraction of the electrons to the nuclei, and the interelectronic and internuclear repulsions. These are respectively included in the following mathematical notation of the Hamiltonian

$$H = \sum_{i} \frac{\hbar^2}{2m_i} \nabla_i^2 - \sum_{i<k} \frac{\hbar^2}{2m_{ik}} \nabla_i^2 - \sum_{i<k} \frac{e^2Z_k}{r_{ik}}$$

$$+ \sum_{k<l} \frac{e^2Z_kZ_l}{r_{kl}} \tag{2}$$

where $i$ and $j$ run over the electrons, $k$ and $l$ run over nuclei, $\hbar$ is Planck’s constant divided by $2\pi$, $m_i$ is the mass of the electron, $m_{ik}$ is the mass of the nucleus $k$, $\nabla_i^2$ is the Laplacian operator, $e$ is the charge of the electron, $Z$ is an atomic number, and $r_{ab}$ is the distance between particles $a$ and $b$ [27-30].

The problem of solving equation 1 is simple in concept, but is very complicated due to the techniques that should be developed for its numerical solution for a given geometry of the molecule. The basic problem is the solution of a many body electronic problem for fixed positions of ions, through the utilization of the Born Oppenheimer approximation. This problem can then be solved using a variety of many body techniques of which the most commonly known are Hartree-Fock (HF) and Density Functional Theory (DFT).

Hartree-Fock Theory

The fundamental assumption of the HF theory, that each electron sees all of the others as an average field, allows for tremendous progress to be made in carrying out molecular orbital (MO) calculations. On the other hand, neglect of electron correlation may have chemical consequences in determining accurate wave functions and properties. The development of semi-empirical theories (AM1, PM3) was motivated by the hope that parameterization efforts could compensate for this feature of HF. Another motivation for introducing semi-empirical approximations into HF theory was to facilitate the computation of derivatives so that geometries could be efficiently optimized. If the HF limit is achieved, then the

$$H = \sum_{i} \frac{\hbar^2}{2m_i} \nabla_i^2 - \sum_{i<k} \frac{\hbar^2}{2m_{ik}} \nabla_i^2 - \sum_{i<k} \frac{e^2Z_k}{r_{ik}}$$

$$+ \sum_{k<l} \frac{e^2Z_kZ_l}{r_{kl}} \tag{2}$$
energy associated with the HF approximation for a given system, the electron correlation, can be determined as:

\[ E_{corr} = E - E_{HF} \]  

(3)

where \( E \) is the true energy and \( E_{HF} \) is the system energy in the HF limit. The HF energies were found to be useful for situations where the error associated with ignoring the correlation energy could be made unimportant by comparing two or more systems for which the errors cancel.

The basis set is the set of mathematical functions from which the wave function is constructed. Each MO in HF theory is expressed as a linear combination of basis functions, the coefficients of which are determined from the iterative solution of the HF self-consistent field equations. The full HF wave function is expressed as a Slater determinant formed from the individual occupied MOs. Keeping the total number of basis functions to a minimum is computationally attractive. It is also useful to choose basis set functional forms that permit the various integrals appearing in the HF equations to be evaluated efficiently. Slater functions are not easily implemented in MO calculations, because some of the integrals are difficult to evaluate particularly when the atomic orbitals are centered on different nuclei. It is common in \( \text{ab initio} \) calculations to replace the Slater orbitals by functions based upon Gaussians. A Gaussian function has the form \( \exp(-\alpha r^2) \) and \( \text{ab initio} \) calculations use basis functions comprising integral powers of \( x, y \) and \( z \) multiplied by \( \exp(-\alpha r^2) \):

\[ x^a y^b z^c \exp(-\alpha r^2) \]  

(4)

where \( \alpha \) determines the radial extent (or spread) of a Gaussian function. In this manner, one mimics the orbitals with an alternative that provides a computationally attractive behavior.

The basis functions must also be chosen to have a form that is useful in the chemical sense. That is, the functions should have large amplitude in regions of space where the electron probability density is also large, and small amplitudes where the probability density is small. Basis sets are developed by simultaneous optimization of all these considerations. The 6-31G** basis set is particularly useful where hydrogen acts as a bridging atom. For molecules that have a significant amount of electron density away from the nuclear centres such as anions and molecules containing lone pairs, highly diffuse functions are added to the basis set and the 6-31++G** basis set is used.

**Density Functional Theory**

The basis for DFT is provided by the proof of Hohenberg and Kohn, that the ground-state electronic energy is determined completely by the electron density \( \rho \) [31]. In this approach there exists a one-to-one correspondence between the electron density of a system and the energy. A wave-function for an \( N \)-electron system contains \( 3N \) coordinates, three for each electron. The electron density is the square of the wave function, integrated over \( N-1 \) electron coordinates, this only depends on three coordinates, independent of the number of electrons. While the complexity of a wave function increases with the number of electrons, the electron density has the same number of variables independent of the system size. The goal of DFT methods is to design functionals connecting the electron density with the energy [32]. The energy functional may be divided into three parts, kinetic energy, \( T[\rho] \), attraction between the nuclei and electrons, \( E_{\text{ne}} [\rho] \), and electron-electron repulsion, \( E_{\text{ee}} [\rho] \). The foundation for the use of DFT methods in computational chemistry was the introduction of orbitals by Kohn and Sham [33]. The key to Kohn-Sham theory is the calculation of the kinetic energy under the assumption of non-interacting electrons. A general DFT energy expression can be written as:

\[ E_{\text{DFT}} [\rho] = T[\rho] + E_{\text{ne}} [\rho] + J[\rho] + E_{\text{ee}} [\rho] \]  

(5)

where \( E_{\text{ne}} \) is the exchange correlation functional and \( J[\rho] \) is the electrostatic repulsion term. The difference between DFT methods is the choice of the functional form of the exchange-correlation energy. Functional forms are designed to have certain limiting behaviour and fitting parameters to known accurate data. Which functional is the better will have to be settled by comparing the performance with experiments at high-level wave-mechanics calculations. In the Local Density Approximation (LDA), it is assumed that the density locally can be treated as a uniform electron gas. In the more general case, where the spin densities are not equal, LDA is replaced by the Local Spin Density Approximation (LSDA). LSDA methods are found to provide results with an accuracy similar to that obtained by wave mechanics HF methods. Improvements over the LSDA approach have to consider a non-uniform electron gas. Gradient Corrected or Generalized Gradient Approximation (GGA) methods make the exchange and correlation energies dependent not only on the electron density, but also on the derivatives of the density. Gradient corrected methods usually perform much better than LSDA. Hybrid methods perform even better. For systems containing multi-reference character, DFT methods are found to generate results of a quality comparable to that obtained with coupled cluster methods [34]. Another significant advantage is that, DFT methods based on unrestricted determinants for open-shell systems are not very prone to “spin contamination”. On the other hand, weak interactions due to dispersion are poorly described by the current functionals. Furthermore, DFT methods are at present not well suited for excited states of the same symmetry as the ground state.

**Molecular Mechanics**

When chemical reactions do not take place in the system, so that the electronic structure is not much altered due to conformational changes, it is possible to lump the behavior of all the subatomic particles in an atom to a single point that signifies the whole atom. This brings us from the QM description to the classical limit where one needs an accurate description of the interactions between the atoms. MM is based upon a simple model of interactions within a system with contributions from processes such as the stretching of bonds, the bending of angles and the rotations about single bonds. The transferability of a force field enables a set of parameters to be developed and tested on a relatively small number of cases, and to be applied to a much wider range of problems. Force fields can be interpreted in terms of a simple four-component picture of the intra- and inter-molecular forces within the system. The force field contains a term that describes the energy associated with the deviation of bonds and angles away from their equilibrium values, there is a function that describes how the energy changes as bonds are rotated, and finally the force field contains terms that describe the interaction between non-bonded parts of the system. A functional form for such a force...
field that can be used to model single molecules or assemblies of atoms and/or molecules is:

\[ V(r^N) = \sum_{\text{bonds}} \frac{k_i}{2} (l_i - l_{i,0})^2 + \sum_{\text{angles}} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 + \]

\[ \sum_{\text{torsions}} \frac{V_i}{2} \left[ 1 + \cos(n\varphi - \gamma) \right] + \]

\[ \sum_{i=1}^{N} \sum_{j=i+1}^{N} 4\epsilon_ij \left[ \gamma_{ij} \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} \right] + \frac{q_i q_j}{4\pi \varepsilon_0 r_{ij}} \]  

Here, \( V(r^N) \) denotes the potential energy, which is a function of the positions \( r \) of \( N \) particles. The first term models the interaction between pairs of bonded atoms by a harmonic potential that gives the increase in energy as the bond length deviates from the reference value \( l_{i,0} \). It maintains the connectivity of the polymeric molecule. The second term is a summation over all valence angles in the molecule. The third term is a torsional potential that models how the energy changes as a bond rotates. The fourth contribution is the non-bonded term. This is calculated between all pairs of atoms \( (i \) and \( j) \) that are in different molecules or that are in the same molecule but separated by at least four bonds. In a simple force field, the non-bonded term is usually modeled using a Coulomb potential term for electrostatic interactions, a Lennard-Jones potential for van der Waals interactions [27, 28, 30]. In the former, the interaction is defined between the explicit charges of the atoms, as well as the parameterized partial charges that represent the polarity of the bonds. The latter accounts for the excluded volume interactions in the systems being considered.

There is a wide variety of force fields developed for the study of different molecular systems. A first generation force field is derived by fitting experimental data sets; MM2 [35] and CVFF [36] are well used examples of these. They have been shown to handle a wide range of organic systems; due to their extensive use, they have also been widely tested. They were primarily intended for studies of structures and binding energies, although they predict vibrational frequencies and conformational energies reasonably well. CFF91 [37] and PCFF [38-40] are second generation force fields, derived based on \textit{ab initio} models. CFF91 was parameterized against a wide range of experimental observables for organic compounds containing H, C, N, O, S, P, halogen atoms and ions, alkali metal cations, and several biochemically important divalent metal cations. PCFF is based on CFF91, extended so as to have a broad coverage of organic polymers, metals, and zeolites. Common force fields used in the description of biomolecules include CHARMM [41], AMBER [42], OPLS [43], GROMOS [44]. These differ in the source of the data used for parameterization; e.g. from various experimental techniques or \textit{ab initio} calculations. New generation force fields such as AMOEBA also treat the polarizability of atoms [45] by including an explicit term in the potential energy function for the polarization of charge distribution by the environment. Various publications comparing the force fields to each other and to experimental findings for different molecular systems can be found in the literature; e.g. see [46-48] and references cited therein.

In the general MM methodology, starting from an initial “guess” structure for the system, the energy computed \textit{via} a suitable force field is minimized using standardized numerical minimization schemes such as steepest descents, conjugate gradients, Newton techniques, to name a few. The efficiency of the minimization depends on the nature of the system as well as the selected minimization scheme [49]. The procedure is then repeated many times for different starting configurations to get many local minima. The low energy structures are assumed to be also structurally relevant; the basis of such assumptions will be discussed in section III.

MM ignores the time evolution of the system and focuses on finding particular geometries and their associated energies or other static properties. Many experimental properties such as vibrational frequencies, sublimation energies, and crystal structures can be reproduced with a classical forcefield, not because the systems inherently behave classically, but because the forcefield is fit to reproduce relevant observables and therefore includes most of the quantum effects empirically.

### Hybrid Quantal/Classical Models

Frequently, we wish to make use of the tools of QM to accurately model an electronic structure problem, while in the surrounding region the explicit representation of the supersystem is important. The level of model applied can be reduced in complexity owing to the more simply understood influence of the outer region on the process as a whole. Hybrid methods have been designed for modeling such cases, where the active site is calculated by electronic structure methods, while the backbone is calculated by a force field method [50]. When the level applied to the outer system is MM, the complete Hamiltonian for the system must be some kind of hybrid QM/MM methodologies.

\[ H_{\text{complete}} = H_{\text{QM}} + H_{\text{MM}} + H_{\text{QM/MM}} \]  

where \( H_{\text{QM}} \) accounts for the full interaction energy between the quantum mechanical particles, \( H_{\text{MM}} \) accounts for the full interaction energy between the classical particles, and \( H_{\text{QM/MM}} \) accounts for the energy of all interactions between pairs of quantum mechanical and classical particles [27]. The main challenge in QM/MM schemes is deciding how the two parts should be connected. Partial charges on the MM atoms can be incorporated into the electronic HF equations, analogously to nuclear charges, and the QM atoms thus feel the electric potential due to all the MM atoms. van der Waals potentials from the MM atoms are also added to the QM part to prevent QM and MM atoms from bumping into each other. In many cases, the MM and QM parts belong to the same molecule, and the connection between the two bonds must be made by cutting a molecular bond. The QM part is terminated by adding “link” atoms such as hydrogens to each of the dangling bonds. The forces from the force field must be added to the atoms treated by the electronic structure method and vice versa [51]. Obviously, there is no unique way of deciding which part should be treated by a force field and which by QM. The concept has been generalized in the ONIOM method to include several layers, for example using high level \textit{ab initio} in the central part, lower-level electronic structure theory in an intermediate layer and a force field to treat the outer layer [52].

### Molecular Dynamics

MM leads to static structures which are relevant in the context of real-life function if the starting positions of the
atoms are close to the functional three-dimensional shape. However, quite often, scientific interest lies in an understanding of the dynamics of the systems. In MD, successive configurations of the system in time are generated by integrating Newton’s laws of motion. The result is a trajectory that specifies how the positions and velocities of the particles in the system vary with time. The trajectory is obtained by solving the differential equations embodied in Newton’s second law (F = ma):

\[ \frac{d^2 x_i}{dt^2} = \frac{F_{xi}}{m_i} \] (8)

This equation describes the motion of a particle of mass \( m_i \) along one coordinate (\( x_i \)) with \( F_{xi} \) being the force on the particle in that direction. In MD, first it is necessary to establish an initial configuration of the system, which may be obtained from experimental data, from a theoretical model or from a combination of the two. Initial velocities are assigned to the atoms by using a Maxwell-Boltzmann distribution at the temperature of interest. The initial velocities are often adjusted so that the total linear and angular momentum of the system is zero. At each step, the force on each atom must be calculated by differentiating the selected potential function (equation 6). The first stage of an MD simulation is the equilibration phase, which allows to bring the system to equilibrium from the starting configuration. During equilibration, the kinetic, potential, total energies, velocities, pressure and temperature are monitored. During the production phase that follows, various properties are calculated and stored for subsequent analysis and processing. It is also usual to store the positions, energies and velocities of configurations at regular intervals, from which other properties can be determined once the simulation is finished. MD generates configurations of the system that are connected in time. It can provide information about the conformational properties of molecular systems and the way in which the conformation changes with time. In fact, using a realistic trajectory and the theory that connects the atomic positions and velocities to the properties of interest, in principle, it is possible to obtain all properties of the system. In practice, for systems of current size with number of atoms on the order of \( 10^4 \), we are limited by the accessible computational times, which is on the order of tens of nanoseconds on conventional computers of our day [27, 28, 53]. It should be noted that microsecond simulations on systems with size of current interest were first introduced by Duan and Kollman, on the villin headpiece protein with 36 residues (295 atoms) and 3000 water molecules by utilizing the supercomputer powers of 256 CPU Cray T3D and T3E-600 machines [54]. More recently, through the folding@home project (http://folding.stanford.edu/) which utilizes over 1,000,000 CPUs worldwide, it has been possible to study the dynamics of several similar systems surpassing the microsecond time scale [55]. However, sub-microsecond simulations remain to be typical for most applications.

**QM/MD**

The Car-Parrinello method [56] was developed as a unified approach to classical MD and DFT as the electronic structure method of choice. Every MD step involves taking a phase point, computing the energy and the gradients for that point given the nuclear positions, and propagating a short time step prior to repeating this process. This formalism is time consuming. Car and Parrinello showed that one does not need to fully converge the Kohn-Sham wave function at every step; instead the Kohn-Sham MO coefficients are treated as dynamical variables. To further increase speed, the method usually uses a plane-wave basis set, which is suited to the periodic boundary conditions and allows fast Fourier transform methods to facilitate solution of the SCF equations. The method allows for a full representation of the solvent with the advantage that solvent participation in reactions is handled naturally. With improved DFT functionals and increasing computer speeds, this method is very promising for the future, although it is still sufficiently time-consuming that present day applications are limited [57, 58].

**III. SELECTING PATHWAYS DURING SYNTHESIS**

**Kinetic vs. Thermodynamic Control in Chemical Reactions**

We need to use QM approaches in determining the preferred pathways during synthesis, since we must describe bond breaking and formation. When the product composition of a reaction is governed by the equilibrium thermodynamics of the system, i.e. if formation of the thermodynamically more stable product is the driving-force of the reaction, the reaction is said to be thermodynamically controlled. At high temperatures, the reaction is under thermodynamic control (equilibrium, reversible conditions) and the major product is the more stable system. However, if the product composition is governed by competing rates of formation of products, i.e. if the product with the lowest activation energy is formed, the reaction is said to be under kinetic control. At low temperatures, the reaction is under kinetic control (irreversible conditions) and the major product is that from the fastest reaction (see Fig. (1) for some examples).

**Examples from Erythromycin A Derivatives**

Macrolides are a class of macrocyclic lactone molecules to which are attached one or more deoxysugar groups. The first macrolide to be introduced for clinical use in humans was erythromycin A (Fig. (2)). It is effective against most gram-positive bacteria, and some gram-negative bacteria. However, it undergoes decomposition within the acidic medium of the stomach which results in undesired gastrointestinal effects [59-61]. To overcome the side effects, the acid stability of the macrolide is improved through modification of functional groups involved in the degradation process, which engages interactions between the hydroxyl groups at C-6 and C-12 with the carbonyl at C-9. Therefore, chemical derivatives of erythromycin A have been well studied, and several have been approved for clinical use [13].

In our previous experience in modeling the selective methylation in the synthesis of clarithromycin, various steps have been modeled with QM tools [24]. First, the selective protection of the hydroxyl groups by a protecting group such as chlorortrimethylsilane (\( \text{CH}_3\text{SiCl} \)) has been considered by examining four-membered transition structures. The protection preference among the hydroxyl groups has been assessed by considering the barrier heights along this reaction. The next step was the methylation of the non-protected hydroxyl groups. This reaction involved the formation of the anions, followed by methylation. The formation of anions by abstraction of hydrogen in basic medium is a very rapid reaction and does not
Not For Distribution

Equilibrates to the thermodynamically more stable immediately cyclizes to give epidirithromycin. The latter for the formation of the Schiff base of epidirithromycin which cyclization yielding oxazine. There is an experimental evidence formation of the Schiff base controls the facial selectivity of the face of the Schiff base intermediate gave dirithromycin whereas epidirithromycin is known to occur equilibrium ratio in favor of dirithromycin is reached in several hours in acetonitrile. Analysis of the reaction mixture erythromycylamine with 2-(2-methoxyethoxy)-acetaldehyde for we studied using QM tools. It is synthesized by treating 9(2

Temperature adjustment will not change the ratio of A to B (mostly B). This reaction is under thermodynamic control. Kinetic or thermodynamic control Under carefully controlled conditions the reaction can be stopped at A (the kinetic product). Under more forcing conditions, i.e. higher temperatures, A can be converted into B, the thermodynamic product. In Case 3, TSA has a much higher energy barrier than TSB. When R is converted into A it will equilibrate rapidly to give B. Temperature adjustment will not change the ratio of A to B (mostly B). This reaction is under kinetic control.

Contribute to the rate-determining step of the overall reaction. As mentioned within the context of short, strong hydrogen bonds, these types of reactions have low activation barriers [62]. In this case, the stability of the anions was examined, and then the barrier for the methylation by CH$_3$I through an S$_N$2 reaction was modeled. Experimentally, the synthesis of clarithromycin from erythromycin A is carried out at room temperature [21] suggesting kinetically controlled steps in line with our findings [24].

Fig. (1). Reaction coordinates depicting thermodynamic and kinetic control. In Case 1, both products A and B are energetically significantly more stable than reactant R. The reverse reaction from A or B to R is energetically disfavored and the reaction is effectively irreversible (regardless of what the ratio of A to B is). Whether A or B are formed will depend on the activation energy to reach TSA or TSB. This is an example of kinetic control. In Case 2, the activation energies for the formation of A and B from R are very similar. Formation of A is only slightly more favourable than formation of B. This reaction can be under kinetic or thermodynamic control. Under carefully controlled conditions the reaction can be stopped at A (the kinetic product). Under more forcing conditions, i.e. higher temperatures, A can be converted into B, the thermodynamic product. In Case 3, TSA has a much higher energy barrier than TSB. When R is converted into A it will equilibrate rapidly to give B. Temperature adjustment will not change the ratio of A to B (mostly B). This reaction is under thermodynamic control.

Fig. (2). Structure of erythromycin A.

Dirithromycin is another derivative of erythromycin A which we studied using QM tools. It is synthesized by treating 9(S)-erythromycylamine with 2-(2-methoxyethoxy)-acetaldehyde for several hours in acetonitrile. Analysis of the reaction mixture revealed that epidirithromycin is very rapidly formed as the initial product and it subsequently epimerizes until an 85:15 equilibrium ratio in favor of dirithromycin is reached in solution [63]. The formation and the epimerization of epidirithromycin is known to occur via a Schiff base intermediate. The cyclization of the C-11 hydroxyl to the Re face of the Schiff base intermediate gave dirithromycin whereas the Se face of the intermediate gave epidirithromycin. The formation of the Schiff base controls the facial selectivity of the cyclization yielding oxazine. There is an experimental evidence for the formation of the Schiff base of epidirithromycin which immediately cyclizes to give epidirithromycin. The latter equilibrates to the thermodynamically more stable dirithromycin in the end of the reaction [63]. By a detailed QM study of the possible steps during the synthesis of epidirithromycin and dirithromycin via the condensation of the 9(S)-erythromycylamine with 2-(2-methoxyethoxy)-acetaldehyde, the experimental observations were rationalized [25]. The pathways for the formation of the Schiff bases for dirithromycin and epidirithromycin were modeled. Interconversion of epidirithromycin to dirithromycin via epimerization of the Schiff bases was also taken into account. Consideration of thermodynamic vs. kinetic products at various stages of the syntheses made it possible to explain the large difference in the equilibrium ratio of dirithromycin vs. epidirithromycin.

Hardness and Softness as a Measure of Selectivity

For a qualitative explanation of selectivity, an analysis has been carried out using the reactivity descriptors such as the global hardness($\eta$) and softness ($S$) calculated from

$$\eta = \frac{1}{2} (IE - EA); S = \frac{1}{\eta}$$

(9)

where IE and EA are the first vertical ionization energy and electron affinity of the molecule, respectively [64]. The analysis thus pertains to the kinetics of the reaction. There is a conceptual relationship between the properties called nucleophilicity and basicity, the most useful qualitative approach for making predictions being the hard and soft acids and bases (HSAB) concept. Hard nucleophiles prefer hard electrophiles, while soft nucleophiles prefer soft electrophiles. Although the HSAB principle cannot be used as a tool for mechanistic explanations of the processes, it may provide clues to likely targets and the loci of action. In S$_N$2 reactions, the results involving the HSAB principle have been interpreted by correlating the energy difference between the two ion-molecule complexes ($X...CH_3Y$ and $Y...CH_3X$), with the group hardness difference between X and Y [65]. Their correlation has been interpreted such that increasing the hardness of Y also hardens the neighboring C atom of the CH$_3$ group, favoring the attack of a harder nucleophile [66]. The HSAB principle in this form provided an explanation to the selectivity of the various anions in the selective methylation of azithromycin [22].
IV. SAMPLING THE CONFORMATIONAL SPACE

The Problem of Conformational Search

An initial task prior to making any QM or MD calculation is to find a meaningful starting structure for the system of interest. “Meaningful” here refers to a structure that is relevant under the conditions where the function of interest is served. The conformational space available to molecules grows exponentially with the number of atoms in the system. For example, with the MM2 force field, cycloundecane is known to have 40 energy minima within the range of ca. 34 kcal/mol above the global energy minimum (GEM); of these 11 are within the 3 kcal/mol range [67]. In comparison, for cycloheptadecane, which has often been used as a benchmark molecule for testing the efficiency of CS algorithms, there are 262 minima in the 3 kcal/mol range of the GEM with the same force field [68-71]. Thus, a 1.5 fold increase in the number of atoms leads to a 24-fold increase in the number of low-lying energy minima. Moreover, these are examples from cyclic molecules whose conformational space is much more restricted compared to those without geometrical restrictions.

CS is the repetitive application of a MM step for many different initial structures. To provide a visual-aid on the problem of CS, please refer to Fig. (3). The potential energy surface of even a moderately sized molecule is decorated with a plethora of energy minima, sometimes termed as localized microstates (LM); e.g. all the minima in Fig. (3). Some of these LM will have similar energy values (e.g. minima i and ii in Fig. (3)), whereas some others will be structurally similar, although they may have very different energy values (e.g. minima iii and iv). The minima for which the structures are not too different from each other may be clustered into groups, referred to as wide microstates (WM). WM will be located close to each other on the potential energy hyper-surface; such WM are exemplified by regions I, II and III in Fig. (3). Clustering of structures into WM may be performed through some defined metrics such as having values below a threshold root mean square value, or having backbone dihedral angles with close values, i.e. backbone motifs [72].

![Fig. (3). Hypothetical demonstration of the potential energy wells of a molecule. GEM corresponds to the global energy minimum, whereas I, II, and III refer to wide microstates (WM) within which the structures corresponding to each minimum are similar to each other, whereas their energetic values may be quite different. During the conformational search of a medium-sized molecule, such as those of the drugs considered in this review, it is usually enough to find one low-lying representative of each WM for further analysis.](image)

The relative population $p_i$ for a chosen WM $i$ is calculated from $\Delta F_{i,m}$, the free energy difference between a reference WM, $m$ and WM, $i$ [73]:

$$p_i = \exp(-\Delta F_{i,m} / k_B T) / \sum_j \exp(-\Delta F_{j,m} / k_B T)$$ (10)

To complicate matters even further, the true probability of observing an available state is given not by the interaction energy, but rather by the free energy, as displayed in eq. 10. Yet, if one assumes that the curvatures of the various low lying potential energy wells are similar so that the entropy, $S$, due to different WM have similar values, the entropic term contributes an additive constant to the free energy (since $F = E - TS$ at constant $T$). Under such circumstances, the probability of a state $i$ will be the same both when its population is calculated from the free energy or only the interaction energy of eq. 6. This is possible since the probabilities are computed through the differences in energy rather than the energy itself. The assumption is necessitated by the fact that it is usually extremely hard to compute the entropy and hence the free energy of a molecular system [74].

The force field gives an approximation to the real potential energy surface of the molecule, as its derivation involves many assumptions. Thus, different force fields will lead to different energy surfaces. Note, however, that “good” forcefields will lead to similar low energy structures. Although the absolute value of the calculated energies will be different, since the reference states for different terms in the potential originate from different resources, the energy differences should have similar values.

Conformational Search of Small-To-Medium Sized Molecules Under Geometrical Restrictions

Most drug molecules are small organics with geometrical restrictions. In many cases, the mobility of the backbone is restricted via cyclization, as in macrolide antibiotics [13]. In many others, rigid backbones are ensured with the presence of aromatics and other ring structures, as in non-nucleoside reverse transcriptase inhibitors [11]. Geometrical restrictions
limit the conformational space available to the drug by orders of magnitude, providing control over their three-dimensional shape so that in the environment where they are functional, the ligands are presented to the binding site in the conformation that they are effective.

The most guaranteed way for a thorough CS that will locate all the energy minima is a systematic search where the internal degrees of freedom of the molecule are varied at regular intervals. However, a fine-grained choice of the interval necessitates an explosive number of conformations to be sampled, whereas a coarse interval will miss some of the important structures (see reference [75] for an efficient variant of this approach). Similarly, a CS method that is based on random selection of conformations followed by energy minimization will mostly lead to the highly populated high-energy structures, therefore demanding a very large number of initial structures to be chosen. Therefore, an efficient method should give a strong preference for generating low-energy structures. In the CS of molecules under geometrical restriction, there is the added difficulty of satisfying geometrical constraints such as loop closure or cyclization.

One approach includes the pioneering work of Go and Scheraga, which provides a solution for the ring-closure problem of a peptide of \( N \) backbone dihedral angles [76]. The method was later extended to flexible geometry [77]. A related procedure was suggested where the conformation of the first \( N/2 \) bonds is determined at random, while that of the last \( N/2 \) bonds is obtained under restrictions that guarantee the closure of the ring [70]. With another method, a random conformation of the linear molecule is first generated, and the dihedral angles are “tweaked” to close the ring [78]. The common feature of these methods is that a large set of ring conformations without severe atomic overlaps are generated first, and their energies are minimized at the next stage.

In another family of approaches, low-energy conformations are obtained with the help of a stochastic process. Thus, at each step, an initial constrained structure is first deformed and energy minimized. Then, a decision is made between accepting the new structure or another previously minimized structure with the help of a “selection procedure” that gives preference for accepting low-energy structures (see ref. [79] and the references cited therein). This structure is then deformed and energy-minimized, and the selection/minimization process is continued for a sufficiently large number of times. The methods which pertain to this category can be distinguished by their different deformation techniques and selection procedures. An early version of the method performed small random “kicks” to the Cartesian coordinates of the atoms [67]. The great advantage of the latter method stems from its simplicity and the possibility that it is suitable for handling dense systems. However, it is relatively inefficient, since a lot of effort is spent on deforming stiff degrees of freedom. More efficient versions of this approach apply the deformations on the soft degrees of freedom of the torsional angles [80, 81].

There is a plethora of other methods for CS, which makes use of simulated annealing, distance geometry, genetic algorithms, as well as more novel approaches which have been recently reviewed [82]. A more inferior alternative, as demonstrated as a result of a scientific bet [69], to all the above-listed methodologies is to use MD for CS purposes. This approach is nevertheless attractive, since it is already implemented in many of the commercially available software that also supply the force fields used in any of the above-mentioned methodologies. It thus saves the time to code the CS algorithms, which at times can get rather complicated and may also be problem specific, consuming the computer time that is liberated due to their efficiency. Running MD at unrealistically high temperatures ensures that high energy barriers are surmounted and a large portion of the conformational space is sampled. Structures recorded at regular intervals during the MD runs are later energy minimized and ranked to get a pool of conformers. Using high temperature MD, it was possible to find 97 significantly different structures for clarithromycin [24], and 82 for azithromycin [22], including the X-ray structures, in reasonable computational times (overnight with the currently available computers with Intel Pentium 4 processor technology).

V. INCORPORATION OF SOLVENT EFFECT

While a lot of molecular modeling studies are carried out in vacuo, real molecules are usually present in a solvent environment. The incorporation of the solvent effect into the simulations, while computationally expensive, is at times a must due to their effect on the conformations sampled. Interactions between the solvent and the solute modify the energy surface of the solute, so that the low energy structures might change, giving totally different properties to the system. On the other hand, to correctly incorporate the solvent effect, one must extend the system so as to include hundreds of solvent molecules per solute. Since the simulation time varies roughly with the square of the number of particles present in the system, we are faced with a huge computational burden. We are left with the dilemma of having to spend a very large computational time to simulate the behavior of many molecules not of direct interest!

In modeling the properties of drug molecules and the intermediates that form during their synthesis, the local environment around the reaction site is extremely important. The exact complex that will form around a given location appears due to a sensitive interplay between the steric effects operative in the immediate environment, the bulk effect of the solvent (forming a frictional environment as well as randomly colliding with the solute atoms), and specific interactions between the solute and the solvent molecules. The latter is found to be especially important in these systems, shaped by the complex web of interactions that propagate throughout the solvent [22, 23]. Note that these solvent effects also control the reaction kinetics by adjusting the barrier heights, hence affecting the reaction rate.

Since solute-solvent interactions can have drastic effects on molecular energies, structures and properties of most chemical systems, in recent years many different models describing solute-solvent interactions implicitly have been proposed. One way around the problem is to treat the solvent as a continuum. These models have reached a broad success thanks to their efficiency. Below we review some of their key aspects [83].

Continuum Models

Continuum solvent models are convenient for their limited computational burden; they also include an average over the different solvent distributions. In these models, the bulk of the solvent is represented as a structureless polarizable medium
characterized mainly by a dielectric constant $\varepsilon$, the solute is hosted in a suitable cavity built within the solvent and the total free energy of the solute ($G_{sol}$) is computed as the sum of two terms [84]:

$$G_{sol} = G_{el} + G_{non-el}$$  \hspace{1cm} (11)

The methods used to treat $G_{el}$, which takes into account the solute-solvent long-range electrostatic contributions, can be classified as (i) apparent surface charge methods (ASC), (ii) multipole expansion methods (MPE), (iii) generalized Born approximation methods (GB), (iv) image charge methods (IC), (v) finite difference methods (FD), (vi) finite elements methods (FE). The methods today more in use in computational chemistry belong to the ASC, MPE, GB and FD families. Several of those methods may introduce, via the effective Hamiltionian or with less formal procedures, solute-solvent interaction effects of non-electrostatic origin. The SCRF set of methods due to Rivail and coworkers belong to the MPE family [85], the AMSOL methods due to Cramer and Truhlar belong to the GB family [86, 87].

$G_{non-el}$ refers to all non-electrostatic contributions, the solute-solvent van der Waals interactions (dispersion + repulsion) and the cavitation energy. Two different general strategies have been devised to compute $G_{non-el}$. A single set of parameters can be used to compute all non-electrostatic terms, as a function of a solute property, such as the solvent exposed surface. van der Waals and cavitation contributions can be calculated separately as is the case of the polarizable continuum model.

The acronym PCM (polarizable continuous model) indicates a set of methods addressed to the study of solvation problems at the quantum mechanical level with the use of continuum solvent distributions [88]. In PCM the attention is focused on the ‘solute’ while the ‘solvent’ is treated at a lower level of accuracy. It is a quantum mechanical method in which use is made of an effective Hamiltonian for the solute $M$, and the Schrodinger equation is generally treated at the ab-initio level. PCM formulates the basic electrostatic problem with the aid of already defined ASC on the surface of a cavity in the solvent where $M$ is accomodated. The PCM solute cavity is built as the envelope of spheres centered on the solute atoms, supplemented by some additional spheres used to smooth the cavity surface. The PCM provides the free energy of the molecular solute in the liquid environment, as the sum of different contributions:

$$G_{sol} = G_{el} + G_{cav} + G_{dis} + G_{rep}$$  \hspace{1cm} (12)

$G_{el}$ is computed by modeling the solvent as an infinite polarizable dielectric, whose reaction field is represented by an apparent solvation charge appearing on the cavity walls. $G_{cav}$ is defined as the work needed to form a suitable cavity in the solvent. $G_{dis}$ and $G_{rep}$ (dispersion and repulsion contributions) are related to short-range electronic solute-solvent interactions. The cavitation free energy is always positive, disfavoring the solvation process. This term is usually computed by the hard sphere model where the solvent is represented as a collection of non-interacting hard spheres with the proper macroscopic density. It is computed as the logarithm of the probability that in such a fluid a spherical cavity is found large enough to accommodate the solute. In order to avoid the limitation to spheres, cavities formed from a collection of spheres have been used.

In most used PCM implementations, only $G_{el}$ is included in the solute Hamiltonian. The electrostatic contribution affects the solute structure and electronic properties, and is the most important term in polar solvents. There are several cases, however, when the solvation free energy depends on non-electrostatic terms. All the processes involving non-polar solvents or apolar solutes are affected by non-electrostatic interactions. Furthermore, for large size solutes the cavitation contribution can be larger than the electrostatic one.

Another successful solvation model is the conductor-like polarizable continuum model, CPCM [89]. CPCM and PCM define the cavities as envelopes of spheres centered on atoms or atomic groups: a number of cavity models have been suggested. Inside the cavity the dielectric constant is the same as in vacuo, outside it takes the value of the desired solvent. Once the cavity has been defined, the surface is smoothly mapped by small regions, called tesserae. Each tessera is characterized by the position of its center, its area and the electrostatic vector normal to the surface passing through the center. The tessellation of the spheres making up the surface was originally based on parallels and meridians and later on the inscribed pentacosahedrons. Recently, the CPCM method has been improved and extended in Gaussian03 so that the cavity can be chosen in a number of ways. In CPCM and PCM the choice of the cavities is important because the computed energies and properties depend on the cavity size. UA0 cavity is built up using the united atom topological model (UATM) applied on atomic radii of the universal force field (UFF). The UAHF and UAKS cavities use UATM with radii optimized for the H/6-31G(d) and PBE0/6-31G(d) levels of theory respectively. For the Pauling and Bondi cavities, each solute atom and group is assigned van der Waals values obtained from Pauling or Bondi atomic radii. Benchmarks of different variations of CPCM for the computation of solvation energies of neutral and ionic organic species have been performed and compared to other work in the literature. The CPCM-UAKS method provides the aqueous solvation free energies in agreement with experimental data, the mean absolute deviations being around 2.6 kcal/mol. The largest solvation energy errors are obtained for anions and cations [89]. The 6-31G, 6-31G(d), 6-31+G(d), 6-31+G(d,p), 6-311+G(2d,p) basis sets were utilized to investigate the dependence on basis sets for aqueous solvation free energies. Even when basis sets are enlarged up to 6-311+G(2d,p), the MAD values from the experiment are very similar indicating that diffuse and polarization functions on basis sets hardly change the aqueous solvation effects.

**Supermolecule-Polarizable Continuum Approach**

Pure self-consistent reaction field (SCRF) calculations based on dielectric continuum models completely ignore the solvent structure and might not account for some important effects caused by specific solute-solvent interactions, especially for chemical reactions assisted directly by solvent molecules. The pure reaction field calculation can be improved by coupling with a supermolecule model that includes solute and a few solvent molecules that interact with the solute. The hybrid supermolecule-polarizable continuum approach is based on a supermolecule model in which solvent water molecules having hydrogen bonds with solutes are explicitly included in the reaction system. This methodology has been successfully used by Zhan and others [90-93].
Solvent Structure in MD Simulations with Explicit Solvent

A computationally more expensive approach to the description of solute – solvent interactions, where the reaction is known to be assisted by the solvent molecules is to model the dynamics of the system. Although MD will not show how the reaction actually takes place, a correct model will nevertheless show the preferred orientations of the solute molecules around the reaction site so as to provide clues on the propensities.

The local structure is assessed by the radial distribution functions (RDFs), where the probability distribution of finding a given solvent atom around a reaction site is computed. The RDF, $g_{ij}(r)$, is given by $g_{ij}(r) = \rho_{ij}(r) / <\rho_j>$, where $\rho_{ij}(r)$ is the number density of atom $j$ at a distance $r$ from an atom $i$, and $<\rho_j>$ is the average number density of atom $j$. The $g_{ij}(r)$ are averaged over a statistically significant number of snapshots recorded from the MD trajectories. RDFs are normalized so that $g(r)$ → 1 when the solvent density approaches the bulk value far from the solute. At short distances, the coordination between solute and solvent atoms appear as peaks in the curves.

Using this approach, we have been able to explain the effect of different solvents on the yields in the methylation process at different positions of 2',4'-(O-bis(TMS) erythromycin A 9-O-(dimethylhexylsilyl)oxime). To understand the effect of the solvent on the methylation process, we have performed detailed MD simulations in pure DMSO, pure THF and DMSO:THF(1:1) mixture by using the anions at the C-6, C-11 and C-12 positions of the oxime under the assumption that the anions are stable on the sub-nanosecond time scale [23]. The slight differences in the solvent populations near the reaction site were in excellent accord with the experimentally observed yields. To gain further insight into the results of RDFs of the anions with the solvent molecules, the reactions of the anion with the solvent molecules, DMSO and THF, were explored for small model systems using the supermolecule-polarizable continuum approach described in the previous subsection. These complexes – supermolecules – were found to prevail throughout the MD trajectories, further corroborating the findings.

VI. CONCLUDING REMARKS: A MODULAR APPROACH OF USING COMPUTATIONAL TOOLS IN DRUG SYNTHESIS

Drug synthesis through step-wise reactions involves processes for which phenomena occurring at different time and length scales need be considered. The chemical reaction itself occurs locally, but that local environment is the result of extensive interactions between the involved species, whose conformations are largely affected by the solvent medium and temperature. Computational techniques provide a means of explaining the molecular basis of the experimental observations. Moreover, molecular modeling has predictive power on how the synthesis of a certain organic molecule will proceed under prescribed environmental conditions, using the well established modeling tools of QM and MD hand-in-hand.

Our prescription for treating these problems involves several steps: First, a detailed CS is performed at the classical limit to achieve meaningful conformers to model the reactions. Next, the alternative chemical reactions are studied in vacuo using QM tools, the level of theory depending on the number and the size of the systems of interest. A general trend of the alternative pathways will emerge at this point, and it will be relatively easy to discriminate some of the alternatives if the barrier heights are too high (kinetic control) or the energy differences are too different (thermodynamic control). On the other hand, if energy differences are on the order of only several kcal/mol, then the environment plays a discriminating role on the different alternatives.

It is computationally burdensome to study the dynamic behavior of the solute in the explicit solvent environment using quantum trajectories by solving the time-dependent Schrodinger equation. Although it is possible to study the reaction site using QM with implicit solvent models, many of the solvent molecule types have not been parameterized or tested extensively. For this reason, we study the local environment of the most important reaction intermediates in the solvent of interest using MD simulations. Although this approach will not provide direct knowledge on how the reaction proceeds, it will give information on the dynamic environment around the reaction site, provided by the local structural solvent molecules and the mean field effect of the bulk solvent.

If the solvent has a strong effect on the solute, it will cause significant conformational changes that will be observed in the MD trajectories, necessitating further QM calculations to be conducted on the new conformation of the solute. If the solvent has a weak effect, it will cluster around preferred regions without initiating significant conformational change. This situation is diagnosed by the use of radial distribution functions. Thus, MD trajectories are analyzed in detail to identify the interaction geometries of solvent-solute atoms. These are compared with the geometries obtained from small supermolecules modeled at the QM level. The rule of thumb used at this stage is that, solvent molecules that are held tighter by the solute will be harder to displace by the reacting agent.

When working with a certain class of molecules, applying a general approach to molecular modeling will provide valuable information about their synthesis. Drug molecules are medium-sized organic molecules under geometrical restriction. Our experience with such molecules, also typified by small cyclic peptides [73, 94, 95], show that they have rather stable properties, making their electronic structure portrayable by similar parameters across the whole family. Thus, once the force fields (at the classical limit) and basis sets (in the quantum chemical applications), that well describe interactions are verified using findings on a few representatives of the family, they can be imported to their variants with relative comfort. This property gives us the flexibility to study many aspects of these systems, such as their reaction pathways at the QM level, solvent effect on their dynamics at the classical limit by detailed MD simulations, and their conformational properties by both approaches.

REFERENCES
