Monatsh Chem 139, 299–308 (2008)
DOI 10.1007/s00706-007-0827-7
Printed in The Netherlands

Monatshefte für Chemie
Chemical Monthly

Review

Computational Chemistry techniques: covering orders of magnitude in space, time, and accuracy

Alexander F. Sax

Institut für Chemie, Karl-Franzens-Universität Graz, Graz, Austria

Received 28 June 2007; Accepted 15 October 2007; Published online 14 March 2008

Abstract Computational Chemistry is the branch of chemistry which uses extensively computer models to simulate real systems. Model systems are made to fit the real system in intended respects and with the intended accuracy of the creator or user of the model. The building blocks of model systems and the interactions between them are abstractions of the building blocks of real system and the interactions between them. In computer models a system must be represented by data which are stored in data structures, the interactions between them are represented by mathematical functions or sets of rules. Properties of the model system are calculated with various algorithms, both numerical and non-numerical. A computational chemist must decide which computer model is best suited to represent a real system so that its essential properties can be simulated as intended. Doing computational chemistry means to apply the right computer models to describe a real system or, if necessary, design new computer models.

Keywords Computational Chemistry; Modeling.

Computational Science

"In the broadest sense, science (from the Latin *scire*, to know) refers to any systematic methodology which attempts to collect accurate information about reality and to model this in a way which can be used to make reliable, concrete and quantitative predictions about future events and observations". This is Wikipedia's definition of science [1], which implicitly states that scientific predictions cannot be made without use of models. How this can be done is not said in this definition. There are scientists as well as philosophers who claim that science represents and explains the world (the reality) through the use of theories. Such a statement is, however, useless as long as it is not clear what theories are and how they are used in practical work. After all, science is a human institution run by people with a wide range of interests and motivations. No scientist is simply engaged in the general pursuit of discovery how the world works, but focuses on some particular part of the world. This part is called a system.

Dissecting a system out of the world has both, practical and theoretical reasons: First, it is simply impossible to investigate the whole world, neither are appropriate methods available, nor would the lifetime of a scientist suffice to do such investigations. Second, science wants to find causal factors for a certain phenomenon, and doing that, causal unimportant facts must be eliminated from consideration. That demands that a system should not contain more than necessary for the intended explanation or prediction of a phenomenon. A system is, therefore, man-made to serve a certain scientific goal. Once

---

Correspondence: Alexander F. Sax, Institut für Chemie, Karl-Franzens-Universität Graz, Strassoldogasse 10, A-8010 Graz, Austria. E-mail: alexander.sax@uni-graz.at

more Wikipedia [2]: "System ... is a set of entities, real or abstract, comprising a whole, where each component interacts with or is related to at least one other component, and they all serve a common objective. Any object which has no relation with any other element of the system is not part of that system but rather of the system environment. A subsystem then is a set of elements, which is a system itself, and a part of the whole system. Every division or aggregation of real entities into systems is arbitrary, therefore it is a subjective abstract concept."

The following facts in this description of a system are important:

- The division of the world into a small part, the system, and the large remainder, the environment.
- The fragmentation of the system into small building blocks that are related by some interaction.
- The hierarchical structure of a system due to the possibility that a system consists of subsystems which in turn are composed of building blocks.
- Systems may be material parts of the world made up from material entities (experimental sciences) but mostly systems are made from abstract entities. This is typical for theoretical and computational sciences.

The definition of the structure of a system (setup) is not done by just declaring what the building blocks are and/or which subsystems it has, the description of the interactions between the building blocks as well as between subsystems and between system and environment is an essential part of the definition. A great deal of scientific investigations is related to structure determination and the reason is the well known structure-function paradigm which claims that all properties of a system are a function of its structure. Together with the analytic view of modern science according to which properties of complex systems can be completely understood from a knowledge of the properties of the building blocks and the interactions between them this is the basis for modern science.

As already stated, we cannot investigate the whole world and in general we do not want to do it. To theoretically investigate a part of the world, also called a real system, with respect to the goals of the scientist means that we must use theoretical models or model systems which are systems of abstract entities made so that the structure and interactions between the building blocks show similarity with the real system [3]. To be more precise, a model must be similar to the real system in the intended respects and to the intended degree of accuracy. But the intention is that of the scientist. Abstract entities are the result of an abstraction process by which only the "relevant" properties of the building blocks of the real system are attributed to the building blocks of the model system whereas the irrelevant ones are omitted. The relations and the interactions between abstract entities may also be termed abstract because they are adapted to specific needs. Representation of relations between building blocks can be done with graphs, the interactions in the system which enter the calculation of system properties are mostly described by model potentials or by rules which are part of the definition of the theoretical model.

If a model system and a real system are similar, one can say: "The model fits the world as intended" or simply "The model fits". It is part of scientific work to make claims that a certain model fits the world, meaning that one can use it for making explanations or predictions for a real system. Making such a claim is a theoretical hypothesis which is true if the model fits and wrong otherwise.

Frequently, an explanation in science is expected to give causes for the measured values of system properties. This can be done by calculating properties of the model system and comparing them with the measured ones of the real system: if they agree "to the intended degree of accuracy" then the model fits. The same is true for explanations or predictions of properties of the real system. Doing science theoretically means always that we interact indirectly with the real system because we can only investigate properties of the model system and check how well the model fits the real system. Any kind of representation of a real system by a "similar" model system is called a simulation.

It turns out that most real systems cannot be treated with a single model but need a family of models which differ in the intended respects and the intended degrees of accuracy. There is no model that fits the world in any respect and to any accuracy. And it is not even useful to work with such a model: If, for example, a real system is represented by a model which allows the calculation of properties with an accuracy much higher than the experimental values can be measured, one is wasting resources. It is like cutting butter with a chain saw. Moreover, with such a model one is not able to distinguish between relevant and not relevant properties, but omitting the irrelevant ones is necessary for scientific explanations. On the other hand, using for an investigation a model that never achieves the intended accuracy is like making a brain surgery with a hand axe. Therefore, modeling means doing science with the use of a family of theoretical models from which one picks out the proper one that fits. If no suitable model is available one has to design a new one. (*R. Giere* [3] claims that a scientific theory is a family of models and a set of theoretical hypotheses that select those parts of the world that may fit one of the models in the family.)

In Computational Science the system is mapped on data structures, frequently graphs (lists, trees, or any kind of general graphs). The properties of a building block can be stored in a finite number of data, both numerical and non numerical, which are elements of the nodes of a graph, the relations between the building blocks are represented by the edges. Data structures themselves can be implemented in different ways, the implementation depends on the algorithms applied to the data structure. The choice of a specific data structure, the implementation and the algorithms must serve two purposes: they must properly represent the claimed relations between the building blocks and they must be well suited for the intended computer applications. The potentials of the theoretical model must not only be physically correct, they must, moreover, be computationally suitable with respect to numerical accuracy and computer time. If rules are used, they must be adequately formulated. Therefore, it may be necessary to choose between different numerical algorithms or different model potentials or different rules, according to the intended accuracy. By doing this, one gets a computer model of a real system and Computational Science means doing science with the use of computer models. Computational modeling requires often drastic abstractions in the definition of the properties of the building blocks, the kind of relation between them and the form of the model potentials or the rules. These abstractions are mostly necessary to adapt the structure of the real system to available data structures and/or the sophistication of the algorithms used to calculate the system's properties. Calculation of precise data requires mostly extensive computer resources, so doing computational science demands not only knowledge of a proper scientific model but also which results can be achieved with it and what are the costs to get them.

Any kind of simulation of systems by means of computer models is called a computer experiment or an experiment *in silico*.

Computational Chemistry

Since Computational Chemistry (CC) belongs to the field of Computational Science, it is based solely on computer models. A major issue of CC is the determination of molecular structure and its change, because structure is, according to the structure-function paradigm, responsible for most chemical properties of a substance, and because chemistry is the science dealing with the change of substances.

Another area of CC is the determination of thermodynamic properties or the description of the kinetics of large systems, which are composed of molecular subsystems and whose properties are, according to the structure-function paradigm related to the molecular structures.

Quantitative Structure Activity Relationship (QSAR) or Quantitative Structure Property Relationship (QSPR) is another large class of CC activities where molecular structure and atomic properties are related with molecular properties, like chemical or biological activity.

Structure

Structure in chemistry means mostly molecular structure and it is frequently seen as an intrinsic property of a molecule being determined only by the properties of the building blocks and the interactions between them. If this is done with quantum theoretical methods the important electronic structure has to be considered. All forces acting on the building blocks are called Chemical Forces.

Molecular structure, as it is seen by chemists, is completely described by the spatial coordinates of the atoms in a molecule, therefore it is a classical concept and, thus, in fundamental disagreement with quantum theory which denies that identical particles can be distinguished if they are spatially close, *i.e.*, if their wave functions overlap. However, all models in chemistry contain building blocks with the classical property of a definite spatial position, the models differ in the definition of the building blocks and the relations between them. The attractive and repulsive interactions between the building blocks lead to the equilibrium geometry of a molecule. CC models differ in the way chemical forces are calculated and the building blocks are defined.

Determination of molecular structure is frequently seen as an optimization procedure where structures correspond to local minima of a scalar function which is mostly some kind of energy (total energy, free energy, *etc.*). The mathematical techniques used for structure determination depend on the size of the system: For small systems deterministic methods using gradients or *Hessians* of the energy like Steepest Descent, Conjugate Directions, Quasi-*Newton*, *etc*. can be used. For large systems where the number of degrees of freedom does not allow systematic searches, stochastic methods like Simulated Annealing are often more appropriate.

The interactions between building blocks are mostly described by differentiable or piecewise differentiable potential functions. To such potentials gradient-based optimization methods can be applied. If, however, the interactions show extreme changes in the function values so that the potential is nearly non-continuous, rule-based methods such as some *Monte-Carlo* methods or Cellular Automata are an attractive alternative.

As was stated above, the definition of a system consists not only in the dissection of the system from its environment but also in stating, what are the interactions between system and environment, if there are any. In chemistry, we are primarily not concerned with the vacuum as environment and the interaction of vacuum fluctuations with the molecule, chemically relevant environments are other material parts of the world such as solvent molecules or solids on which molecules may be adsorbed *etc*. The way how the environment is treated, is part of the model definition, and there are at least three ways to do it.

1) Gas phase calculations

One can completely ignore the environment and its interactions with the system, then we talk about "gas phase" calculations which give the intrinsic properties of a system. If all dynamic properties are neglected, one gets only static properties like *e.g.*, the equilibrium structure which correspond to a molecule in gas phase at the absolute zero point. If the dynamics of the molecule is considered, the properties are temperature dependent, but nevertheless, independent of the environment. Most quantum chemical investigations in the past were gas phase calculations at the absolute zero point and, not surprisingly, often criticized by experimental chemists as being irrelevant for experiments in polar solvents and at finite temperatures.

2) Continuum calculations

One can consider the influence of the environment (*e.g.*, the solvent) on the intrinsic molecular properties of a solute, but neglect its atomic structure because it is regarded as unimportant compared to the large number of electric interactions with the solvent molecules. Then the solvent can be treated as a dielectric continuum into which the solute molecule is embedded. Again, both, static and dynamic properties of the embedded molecule can be calculated.

3) Explicit consideration of solvent molecules

If neither of the two mentioned ways are regarded as adequate with respect to the intended similarity of the real system with the model system, the latter has to be enlarged by some solvent molecules which are now treated explicitly as molecular species. Whether deterministic or stochastic methods are used for doing structure investigations depends on the system size.

Electrons and nuclei as building blocks

A molecule built from electrons and nuclei must be described using quantum theory [4]. To avoid the problems of proper treatment of indistinguishable particles (Fermions with state functions which are antisymmetric with respect to an odd number of particle changes, Bosons with state functions which are symmetric with respect to any number of particle changes), only electrons are treated quantum theoretically while nuclei have the classical property of a definite spatial position. This model assumption is known as *Born-Oppenheimer* (*BO*) approximation, it is the basis of quantum chemistry. In the *BO* approximation the nuclei are simply positive charges at definite positions (clamped nuclei) with the electrons moving in this field. There are no analytic functions that can be used to describe the electronic state of a molecule with n electrons in the field of N fixed *Coulomb* potentials, but if there is a complete set of one electron functions (orbitals), still with respect to the positions of the nuclei, one can represent the many-electron function as a series of products of orbitals. If the many electron functions are made to be antisymmetric with respect to the interchange of any two electrons, one ends up with an expansion into a series of completely antisymmetric products of orbitals, called Slater determinants. If the series is truncated after the first term and if the orbitals are optimized for the electronic ground state, one is using the *Hartree-Fock* method to describe the electronic system. That electrons show *Coulomb* repulsion is evident, so they will avoid each other in the best way (*Coulomb* correlation). The antisymmetry of a *Slater* determinant accounts for the *Fermion* character of electrons, therefore, electrons of like spin will avoid each other also because of their spins (*Fermi* correlation). Electrons of different spin, however, feel only the *Coulomb* repulsion, consequently they can come closer than electrons of like spin. But the product form of the multi electron wave function describes a system of electrons which are independent of each other (independent particle model).

This assumption seems at first sight ridiculous when *Coulomb* and *Fermi* correlation are considered, but one can give a rather plausible explanation: Because no electron is independent of the others and because they try to avoid each other, they do not have the properties of isolated electrons that feel the full *Coulomb* interaction with the other electrons and nuclei, instead they behave like particles that feel much less repulsion. Such electrons are not "naked" electrons but "dressed" quasi-electrons which have absorbed a lot of interaction into particle properties. *Hartree-Fock* orbitals always describe "independent" quasielectrons which do not repel each other as strongly as naked electrons because the avoid each other anyway. The concept of Quasiparticles is very fruitful and can be used to simplify systems considerably. The most simple description of the structure of quasielectrons in a molecule is the electron configuration, which gives the number of electrons in a one electron state (orbital). The electron configuration is obtained from applying the aufbau principle to the energetically ordered orbitals. Proper description of a multielectron state demands that the wave function is also an eigenfunctions of the squared total spin operator, such wave functions are termed configuration state functions and they can be rather long linear combinations of Slater determinants. The different ways to truncate the expansion of the many electron wave function in Slater determinants is termed level of theory.

Quantum theory says that the expectation value of every physical quantity can be calculated as the trace of the product of the corresponding operator and the one electron density, which contains the information of the whole system. Wave function based quantum chemical methods (Wave Function Theory, WFT) determine first the wave function and calculate from it the density, whereas in Density Functional Theory (DFT) the calculation of the wave function at the different levels of theory is avoided but a direct approximation of the one electron density is attempted. The discussions about the relative benefits of the two methods have in the meantime become less hostile.

Irrespective of how the total electronic energy of a molecule (plus the nuclear-nuclear repulsion energy) is calculated, it depends always on the nuclear coordinates, and, therefore, it is a potential energy of the nuclei; the gradient of this energy with respect to the nuclear coordinates is proportional to the forces acting on the nuclei, which means it represents the chemical forces. The *Coulomb* interaction of the electrons with each other and with the nuclei are the "glue" holding the atoms in a molecule together. Stable molecular structures correspond to spatial arrangements of the atoms for which no resultant forces exist, and this is the necessary condition for a local minimum on the hypersurface of the potential energy of the nuclei.

It is chemical belief that atomic properties are constituent for molecular properties and that, therefore, the molecular density must be composed of atomic densities as it is claimed in the Atom In Molecule model (AIM). There is no unique definition of boundaries between atoms in a molecule, that means declaring the position of boundaries is part of the definition of the model. One can make even further separations of the density, for example in a core electron density and the density of the valence electrons; such a separation corresponds to the chemical belief that only the valence electrons but not the core electrons are relevant for chemical properties of atoms. This belief is the basis for the development of pseudopotentials or the justification of valence-only-semiempirical methods in Quantum Chemistry.

The equilibrium structure of a molecule can be regarded as the result of the chemical forces acting on the atoms, they are responsible for the existence or absence of bonds between atoms, their relative strengths (bond order). This information constitutes the connectivity of a molecule which is represented by chemical graphs. The metric of a molecule is made up of the numerical values of bond lengths and bond angles, they are a property of the potential energy surface (coordinates of a local minimum). The bond order, however, is the result of an analysis of the wave function (the one electron density) of the molecule.

Although, both, the metric and the connectivity, are properties of the whole molecule, they are attributed to atomic properties, like the hybridization, which is seen as responsible for the local symmetry of an atom in a molecule, and consequently for bond angles and dihedral angles between bonds to adjacent atoms.

Molecular mechanics

Completely different models result when the properties of atoms (hybridization) and the connectivity of the molecule are not the output of the analysis of molecular properties but the input to model calculations on molecules. In such models, called valence force fields, the interactions between bonded atoms like bond stretch, angle bend, torsion about bonds, but also between non bonded atoms (*Coulomb* interaction, *van der Waals* interaction) or the description of hydrogen bonds, are described by potential functions of classical physics (mechanics and electrostatics).

Therefore, these models are generally termed Molecular Mechanics (MM) [5]. An essential ingredient to valence force fields are atom types, meaning that there are not only different element atoms, but for each element atom there are several "types" which differ in the hybridization and thus the local symmetry; the atom type can also indicate that the atom is used in a certain molecular environment, *e.g.*, an oxygen atom in a carbonyl group or in an alcohol group. Atom types are related to the connectivity, pairs of atoms define the bond order, the force constant of the bond connecting them, and the bond length which is the coordinate of local minimum of the bond stretch potential. Similarly, triples of atoms define a bond angle with a force constant and an equilibrium angle. Summing up, in the definition of a valence force field is stated which interactions are relevant for the description of a molecule, what is the mathematical form of the potentials and what are the numerical parameters like force constants or equilibrium bond lengths, *etc.*, All of that depends on the atom types. The set of the numerical parameters for all atom types is the parameter set of a force field. Most molecular mechanics models use training sets of molecules for the fitting of the parameters to experimental data or data from quantum theoretical calculations. Non bonded interactions are defined for atom pairs, the potential parameters are derived from atomic contributions as arithmetic or geometric means or by even more complicated expressions. All well known force fields like MM4, AMBER, and CHARMM, *etc.* are force fields of that type.

A disadvantage of training sets is that all parameters of the force field depend on each other because they were obtained in the same fit procedure. Taking parameters from other force fields destroys generally the quality of the results. If, however, all interatomic parameters are obtained from atomic properties (not only those for non bonded interactions), one can easily extend the number and kind of interaction terms and include new element types. This concept was for example used to define the Universal Force Field (UFF).

To increase computational efficiency the large number of non bonded interactions must be reduced. This can be done by making different model assumptions. One can calculate the interactions only out to a given distance (cutoff radius) or one can change the mathematical form of an interaction, *e.g.*, using a modified *Coulomb* interaction so that it becomes more short ranged. Such a change of range of *Coulomb* interaction can be explained as the result of a distance dependent dielectric constant.

A third model assumption is, however, much more important: The number of interactions becomes smaller, when the number of building blocks is reduced, for example by condensing atom groups into structureless larger abstract atoms with appropriately changed atomic properties. This technique is called in MM united atom approach and is heavily used for the calculation of large molecules like proteins or polymers. Creating united atoms is an abstraction process where the number of building blocks and the structural differentiation is reduced and the symmetry of the abstract object is increased, as in case of a CH_3 group which is replaced by a spherical united atom. Concomitantly, the interaction with such new objects becomes much simpler, but they must be defined.

This kind of abstraction can easily be extended to residues in the biosciences (united residue force field) or even to whole molecules in the modeling of condensed phase systems. In the latter case the abstract building blocks may have a form faraway from the original one. Frequently ellipsoids, spheres, rods but also disks are used, in the latter case a three dimensional system may even become two dimensional.

Continuum models

Whenever the electric interaction of a molecule with its environment cannot be neglected and only the structure of the latter is regarded as irrelevant, it can be treated as a continuum into which the molecule is embedded. The continuum models [6] differ in the way how both, the solute and the solvent, are described.

A very simple model represents the charge distribution of a solute by a multipole expansion which induces multipoles in the continuum, the interaction energy between these multipoles gives the interaction energy of the solvent and the solute. More sophisticated are those models where the solute occupies a cavity in a polarizable dielectric continuum (Polarizable Continuum Model, PCM). The molecule, which is described by quantum methods, is in contact with the continuum at the surface of the cavity which may be defined by a union of spheres centered on each atom, the radii of the spheres are related to the van der Waals radii of the atoms. The charge of the embedded molecule polarizes the continuum and generates an apparent surface charge on the cavity's surface which feeds back on the solute. The interaction of the electrostatic potential of the embedded molecule with the surface charge, but also the gradient of this energy, can be calculated when the surface is divided into small tesserae.

In COSMO (conductor-like screening model) the environment is not treated as a dielectric continuum but as a conductor, which allows to establish certain boundary conditions.

Change of structure

Change of the system's structure covers all processes which cause changes in the relations between its building blocks or subsystems. Two aspects of structure change are important: 1) The dynamic aspect, here one asks what is responsible for the change, 2) the kinetic aspect, where one wants to know how fast the structure change is.

Change of electronic structure

Changes in the electronic structure, called transitions, are described by a change in the electron configuration and in the spin multiplicity. Most transitions are caused by interactions of molecules with the electric component of the electromagnetic field. In the treatment of the change of electronic structure the dynamic aspect is generally described as a perturbation, mostly by the electromagnetic field, but also by collisions with other molecules as in chemical reactions. The theoretical description of transitions in spectroscopy and during chemical reactions is traditionally done in quantum chemistry.

Change of molecular structure

Change of molecular structure means always changes in the metric and maybe in the connectivity. In the latter case the changes are due to chemical reactions and are frequently accompanied with changes in the electronic structure of subsystems or the whole system. Change of molecular structure causes displacements of some building blocks, the dynamics of such a process relates the displacement to forces which can be derived from the interaction potentials. Without interaction with the environment, these forces will yield structures with lower potential energy and, eventually, an equilibrium structure. This is done during a conventional geometry optimization using deterministic techniques.

If energy is supplied from the environment, *e.g.*, in form of kinetic energy (finite temperatures), it may be used to reach molecular structures of higher potential energy or to surmount energy barriers; this is used in techniques like simulated annealing or quenched dynamics which are frequently the best choice to find different local minima and, hopefully, the global minimum of large systems.

Frequently, one is interested not only in an equilibrium structure but also in the complete way from an initial structure to it, the succession of structures lying in between is called the trajectory. Trajectories can be obtained by brute force integration of *Newton*'s Equations of Motion (*N*EM), for which only the forces acting on the atoms must be known. The method for describing such a dynamic development of the system is called Molecular Dynamics (MD) [7]. For the numerical solution of the differential equations one needs discrete time steps, the lengths of which depend on the potential: for slowly varying potentials the time steps can be large, for rapidly varying potentials they must be small, if the trajectory shall indeed describe the systems dynamics. In quenched dynamics a trajectory is calculated only to generate starting points for conventional geometry optimizations, and, therefore, need not have any physical relevance. This is, however demanded from trajectories which shall simulate processes like the folding of a protein. In such cases, not only the final structure must be correctly predicted, but also the complete time development of the system, *i.e.*, the rate of the structure change. Only then it is possible to classify processes as fast or slow.

The solutions of the *N*EM contain parameters which can be used to define initial values of the velocities. Since only *Maxwell*'s demon might know them for a certain process, they are set by using random numbers, which classifies MD as a stochastic method. The application of MD is not limited to the description of changes of molecular structure, a main application field is the calculation of physical quantities of large systems, the values of which are mean values which can only be calculated with statistical methods. If random forces, representing interactions with solvent molecules, are included in the *N*EM the method is called *Langevin* or *Brownian* dynamics, with these techniques one can *e.g.*, simulate the influence of the solvent on the solute's structure.

Description of ensembles

Condensed matter systems show properties which depend essentially on the interactions between the building blocks. Therefore, these interaction must be adequately described. Building blocks in chemically relevant systems may be atoms, molecules or abstract objects like hard spheres, disks, cells, *etc.*

The interactions between them are defined in the model. Any description of bulk properties of crystalline solids is simple compared to the description of amorphous solids or liquids or adsorbed molecules on surfaces. Such systems can grow, they can undergo phase transitions, one can observe segregation and mixing, and an important mechanism in all these processes is diffusion, that means transport of parts of the system in space and time.

Due to the large size of the systems their states are highly degenerate. The density of states is the function whose values are the number of states having a certain energy. The density of states is related to the partition function which is the working horse of statistical physics [8]. All macroscopic properties can be calculated from it. To get the number of degenerate states of a certain potential energy, one has to find all possible spatial arrangement of building blocks having this energy. These different structures of the system are represented by a point in a high-dimensional configuration space. If also dynamic aspects are accounted for a system, it is represented as a point in the phase space. In either case, calculation of mean values of quantities like the energy means to calculate the energy corresponding to different points in configuration or phase space and adding them up.

Averaging in configuration space is done with *Monte-Carlo* (*MC*) [7] methods, averages in phase space are calculated with MD, but according to the ergodic hypothesis, these two averages are equal. In MD the forces acting at a certain time on the building blocks decide on the length and direction of the spatial displacements; in *MC* displacements are, however, randomly chosen and, therefore, may lead to structures of high energy, which should not contribute too much to the average values. To avoid the contributions of too many energetically unfavored structures, rules are needed which decide on the fraction of high energy structures to be accepted. It is an advantage of the the *MC* method that one can tailor models to specific needs by defining the displacements and the acceptance rules.

Cellular Automata (CA) [9] demand a dramatic abstraction from the properties of a real system. Here, a system consists of a finite number of cells, arranged as a grid, a finite set of ingredients which may occupy the cells, and a finite set of rules governing the behavior of the ingredients. What makes CA so special is, that it is completely discretized.

The cells may have various shapes according to the structural properties of the system to be modeled. A square cell is *e.g.*, well suited to represent a water molecule, which has four possibilities to form hydrogen bonds to other water molecules which are in cells in the neighborhood of a given cell. The cell itself can be empty or occupied by a water molecule. A water molecule can move from cell to cell, it can stick to a molecule in an adjacent cell or break such a relation. The rules governing these processes are based on probabilities, which give the degree to which a certain process may occur.

The size and form of the grid, the shape of the cells, and the states in which a cell can be together with the rules, offer much freedom for the modeling of a system. The applications of CA in chemistry cover the investigation all kind of properties of condensed matter systems like the system structure, growth processes, adsorption, mixing and demixing of two-phase liquids, percolation and much more.

Some more modeling techniques

If any of the modeling techniques discussed above alone does not allow to investigate systems with intended accuracy, this may be achieved when several techniques are combined ("combining the best of several worlds"). Hybrid [10] or QM/MM methods [11] are an example for such a combination, they are used for large systems which cannot be treated with QM methods alone, and for which MM is not sufficient in all respects. In such cases, only a small subsystem is treated with quantum theoretical methods and the complementary subsystem with MM. Hybrid methods can not only be used to determine equilibrium structures, but also to investigate dynamic or statistical properties. The separation of a system is not limited to just two subsystems, in ONIOM the separation into three subsystems is allowed, each one can be treated with a different method ("level of theory").

In QSAR [12, 13], biologically important properties of molecules are correlated with the molecular structure and the underlying atomic properties. Important structural properties are *e.g.*, the molecular volume and surface, charge distributions, electrostatic potentials, all of which are calculated from atomic properties. Although QSAR seems to be rather simple compared with the other techniques already discussed, it is nevertheless a demanding modeling issue. That one can get molecular properties from a linear combinations of atomic properties is a modeling assumption, which can only be fulfilled when intelligent definitions of atomic properties and simple algorithms to calculate them are available. QSAR is very important in drug design and in synthesis planning.

If the physical properties of the building blocks and the correct physical description of the interactions between them are less important than the fact, that a certain building block is related to others, graph algorithms become very important. The application of graphs to chemical systems ranges from a simple *Hückel* treatment of π-systems to applications in Bioinformatics, where trees and strings are used to describe evolution processes or to store biological properties in proteins or *DNA* [14]. In these applications, the building blocks (nucleic acids or amino acids) are regarded as nodes building large strings of different kind (linear lists and trees), and appropriate algorithms are necessary to get intended information from these large data structures (data mining, sequence analysis, protein structure prediction, *etc.*)

Another kind of system modeling uses extrapolation techniques to simulate completeness of description, when mathematical completeness is impossible. Every basis set used in quantum chemical calculations is mathematically incomplete, therefore, any calculation of molecular properties like total energies, zero point energies, *etc.* is approximate with respect to basis set completeness for each level of theory. Basis set extrapolation to the complete basis is done with a sequence of property values which are calculated with basis sets with of increasing size, from these data one gets with an extrapolation scheme the value for the basis set limit [15].

As last example of system modeling I mention the extrapolation schemes like the *Gaussian* (*Gn*) theories [16] or the *Weizmann* (*Wn*) theories [17] to get thermochemical data like bond energies and barrier heights with higher than chemical accuracy. As basis ingredients thermochemical data like atomization enthalpies of molecules in a training set are needed together with the prescription how the energies of a molecule obtained at different levels of theory can be extrapolated.

Applications

An example for an extremely fast reaction is the spin flip reaction during the photodissociation of the HF molecule in solid argon [18]. The system consists of 256 *Ar* atoms in a face centered cubic grid with a HF molecule in a monosubstitutional site. The system's geometry was optimized with simulated annealing at 20 K. Within 1 fs after vertical excitation of the HF molecule into the $^1\Pi$ state the $^3\Pi$ state populates, although there is only weak spin-orbit coupling in HF. The dynamics is described classically, the nuclei propagate on the $^1\Pi$ energy hypersurface until the spin flip (surface hopping) occurs.

A major issue of the simulation of crystal growth of urea from solution [19] is the interaction solvent–solute, surface–solvent, and surface–solute and the time evolution of the system. MD allows only a description of a system with a few thousand atoms for a few nanoseconds. A realistic system is of the order of micrometer on the length scale and milliseconds on the time scale. The authors assumed that the global growth process consists of independent processes on the molecular level, which can be described with MD and which can be used in the second part of the simulation, where the much slower global growth is simulated. In the MD runs the dissolution of urea from crystal surfaces in solvent is simulated, until the saturated solution concentration is reached. Here, local crystallization and dissolution processes are in equilibrium. The calculated thermodynamic properties like the dissolution enthalpy but also the reaction rates of crystallization and dissolution were calculated. The influence of extended defects at crystal surfaces and the extended dimension of the crystal on the growth process was simulated with a *MC* variant, called kinetic *Monte-Carlo* in two and three dimensions. With this method the length and the time scale could be extended to micrometer and milliseconds, respectively.

As an example for CC calculation with very precise results I discuss a QC study of the $O_3 \cdots OH$ complex at 0 K in the gas phase [20]. Using large basis sets together with the complete active space SCF method with large active spaces, the agreement with experimental values for ozone in the bond distances is $0.24\,\acute{\mathring{A}}$ and in the bond angle 0.7 degrees, the maximum deviation of the calculated frequencies form the experimental ones was 8 percent. The ground state of the hydrogen bonded complex is nearly degenerate, the $^2A''$ lies only 1.7 kJ/mol above the $^2A'$. The activation of breaking and forming a new hydrogen bond was estimated to be 2.5 kJ/mol in the $^2A'$ ground state and 4.2 kJ/mol in the $^2A''$ state.

References

1. Wikipedia contributors, Science – Wikipedia, The Free Encyclopedia (2007) http://en.wikipedia.org/w/index.php?title=Science&oldid=137260069
2. Wikipedia contributors, System – Wikipedia, The Free Encyclopedia (2007) http://en.wikipedia.org/w/index.php?title=System&oldid=134493943
3. Giere RN (1997) Understanding Scientific Reasoning, 4th edn. Many phrases in Chapter 1 are taken from this book without explicit citation. Wadsworth, Belmont, CA
4. Yarkony DR (1995) Modern Electronic structure Theory, Part I and Part II, World Scientific, Singapore
5. Rappé AK, Casewit CJ (1997) Molecular Mechanics across Chemistry, University Sience Books, Sausalito
6. Klamt A (1998) COSMO and COSMO-RS. In: Rague-Schleyer Pv (ed) Encyclopedia of Computational Chemistry. John Wiley, Chicester, p 604
7. Frenkel D, Smit B (2002) Understanding Molecular Simulation. Academic Press, San Diego
8. Chandler D (1987) Introduction to Modern Statistical Mechanics. Oxford University Press, New York
9. Kier LB, Seybold PG, Cheng, CK (2005) Modeling Chemical Systems using Cellular Automata. Springer, Dordrecht
10. Froese RDJ, Morokuma K (1998) Hybrid methods. In: Rague-Schleyer Pv (ed) Encyclopedia of Computational Chemistry. John Wiley, Chicester, p 1244
11. Gao J (1998) Hybrid quantum mechanical/molecular mechanical (QM/MM) methods. In: Rague-Schleyer Pv (ed) Encyclopedia of Computational Chemistry. John Wiley, Chicester, p 1257
12. Kubinyi H (1998) Quantitative structure-activity relationships in drug design. In: Rague-Schleyer Pv (ed) Encyclopedia of Computational Chemistry. John Wiley, Chicester, p 2309
13. Jurs J (1998) Quantitative structure-property relationships (QSPR). In: Rague-Schleyer Pv (ed) Encyclopedia of Computational Chemistry. John Wiley, Chicester, p 2320
14. Trinajstič N (1992) Chemical Graph Theory. CRC Press, Boca Raton, p 37
15. Petersson GA, Frisch MJ (2000) J Phys Chem A 104:2183
16. Curtiss LA, Redfern PC, Raghavachari K (2007) J Chem Phys 126:084108
17. Karton A, Rabinovich E, Martin JML, Ruscic B (2006) J Chem Phys 125:2183
18. Cohen A, Gerber RB (2007) Chem Phys Lett 441:48
19. Piana S, Gale JD (2006) J Crystal Growth 294:46
20. Mansergas A, Anglada JM (2006) Chem Phys Chem 7:1488