Article submitted for:

Molecular Dynamics

Molecular Dynamics

Author: Jarosław Meller

Cornell University, New York USA and Nicholas Copernicus University, Toruń Poland

Glossary

Ab initio methods

Computational methods of physics and chemistry that are based on fundamental physical models and, contrary to empirical methods, do not use experimentally derived parameters except for fundamental physical constants such as speed of light c or Planck constant h.

Docking

Binding of a ligand to a specific binding site of a receptor molecule.

Empirical potential

A certain functional form of the potential energy of a system of interacting atoms with the parameters derived from *ab initio* calculations and experimental data.

Molecular Dynamics

A technique for atomistic simulations of complex systems in which the time evolution of the system is followed using numerical integration of the equations of motion.

Monte Carlo

A simulation technique for conformational sampling and optimization based on a random search for energetically favourable conformations.

Optimization

A procedure in which an extremum of a function is sought. When the relevant extremum is the minimum of a function the optimization procedure is called minimization.

Molecular Dynamics is a technique for computer simulation of complex systems, modelled at the atomic level. The equations of motion are solved numerically to follow the time evolution of the system, allowing the derivation of kinetic and thermodynamic properties of interest by means of "computer experiments". Biologically important macromolecules and their environments are routinely studied using Molecular Dynamics simulations.

Introduction

There is a complex network of chemical entities that evolve dynamically creating life at the molecular level. For example, proteins and nucleic acids fold (adopting specific structure consistent with their function), ions are transported through membranes, enzymes trigger cascades of chemical reactions etc. Because of the complexity of biological systems, computer methods have become increasingly important in the life sciences. With faster and more powerful computers larger and more complex systems may be explored using computer modelling or computer simulations.

Molecular Dynamics (MD) emerged as one of the first simulation methods from the pioneering applications to the dynamics of liquids by Alder and Wainwright and by Rahman in the late 1950's and in the early 1960's. Due to the revolutionary advances in computer technology and algorithmic improvements, MD has subsequently become a valuable tool in many areas of physics and chemistry. Since the 1970's MD has been widely used to study the structure and dynamics of macromolecules, such as proteins or nucleic acids.

There are two main families of MD methods, which can be distinguished according to the model (and the resulting mathematical formalism) chosen to represent a physical system. In the *classical* mechanics approach to MD simulations molecules are treated as classical objects, resembling very much the "ball and stick" model. Atoms correspond to soft balls and elastic sticks correspond to bonds. The laws of classical mechanics define the dynamics of the system. The *quantum* or *first-principles* MD simulations, which started in the 1980's with the seminal work of Car and Parinello, take explicitly into account the quantum nature of the chemical bond. The electron

density functional for the valence electrons that determine bonding in the system is computed using quantum equations, whereas the dynamics of ions (nuclei with their inner electrons) is followed classically.

Quantum MD simulations represent an important improvement over classical approach but they require more computational resources. At present only the classical MD is practical for simulations of biomolecular systems comprising many thousands of atoms over time scales of nanoseconds. From now on, the classical Molecular Dynamics will be simply referred to as Molecular Dynamics.

Computer simulation is a powerful research tool

Experiment plays central role in science. It is the wealth of experimental results that provides a basis for the understanding of the chemical machinery of life. Experimental techniques, such as X-ray diffraction or nuclear magnetic resonance (NMR), allow determining the structure and elucidating the function of large molecules of biological interest. Yet, experiment is possible only in conjunction with models and theories.

Computer simulations have altered the interplay between experiment and theory. The essence of the simulation is the use of the computer to model a physical system. Calculations implied by a mathematical model are carried out by the machine and the results are interpreted in terms of physical properties. Since computer simulation deals with models it may be classified as a theoretical method. On the other hand, physical quantities can be (in a sense) measured on a computer justifying the term "computer experiment".

The crucial advantage of simulations is the ability to expand the horizon of the complexity, which separates "solvable" from "unsolvable". Basic physical theories applicable to biologically important phenomena, such as quantum, classical and statistical mechanics, lead to equations that can not be solved analytically (exactly) except for a few special cases. The quantum Schrödinger equation for any atom but hydrogen (or any molecule) or the classical Newton's equations of motion for a system of more than two point masses can be solved only approximately. This is what physicists call the many-body problem.

It is intuitively clear that less accurate approximations become inevitable with the growing complexity. We can compute more accurate wave function for the hydrogen molecule than for large molecules as porhyrins, occuring at the active centers of many important biomolecules. It is also much harder to include explicitly the electrons in the model of a protein, rather than representing the atoms as balls and the bonds as springs. The use of the computer makes less drastic approximations feasible. Thus, bridging experiment and theory by means of computer simulations makes possible testing and improving our models using more realistic representation of nature. It may also bring new insights into mechanisms and processes that are not directly accessible through experiment.

On the more practical side, computer experiments can be used to discover and design new molecules. Testing properties of a molecule using computer modelling is faster and less expensive than synthesizing and characterizing it in a real experiment. Therefore drug design is commonly used in the pharmaceutical industry.

Atomic force field model of molecular systems

The atomic force field models describe physical systems as collections of atoms kept together by the interatomic forces. In particular, chemical bond results from specific shape of the interactions between atoms that form a molecule. The interaction law is specified by the potential $U(\mathbf{r}_1, \cdots, \mathbf{r}_N)$, which represents the potential energy of N interacting atoms as a function of their positions $\mathbf{r}_i = (x_i, y_i, z_i)$. Given the potential, the force acting upon i-th atom is determined by the gradient (vector of first derivatives) with respect to atomic displacements:

$$\mathbf{F}_i = -\nabla_{\mathbf{r}_i} U(\mathbf{r}_1, \cdots, \mathbf{r}_N) = -(\frac{\partial U}{\partial x_i}, \frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i})$$

Naturally, the notion of "atoms in molecules" is only an approximation of the quantum-mechanical picture, in which molecules are composed of interacting electrons and nuclei. Electrons are to a certain extent delocalized and "shared" by many nuclei and the resulting electronic cloud determines chemical bonding. It turns out however, that to a very good approximation, known as the adiabatic (or Born-

Oppenheimer) approximation and based on the difference in mass between nuclei and electrons, the electronic and nuclear problems can be separated.

The electron cloud "equilibrates" quickly for each instantaneous (but quasi-static on the time scale of electron motions) configuration of the heavy nuclei. The nuclei, in turn, move in the field of the averaged electron densities. As a consequence, one may introduce a notion of the potential energy surface, which determines the dynamics of the nuclei without explicit account for the electrons. Given the potential energy surface, we may use classical mechanics to follow the dynamics of the nuclei.

Identifying the nuclei with the centers of the atoms and the adiabatic potential energy surface with the implicit interaction law, we obtain rigorous justification of the intuitive representation of a molecule in terms of interacting atoms. The separation of the electronic and nuclear variables implies also that rather than solving the quantum electronic problem (what may be in practice infeasible) we may apply an alternative strategy, in which the effect of the electrons on the nuclei would be expressed by an empirical potential.

The problem of finding a realistic potential that would adequately mimic the true energy surfaces is nontrivial but it leads to tremendous computational simplifications. Atomic force field models and the classical Molecular Dynamics are based on empirical potentials with a specific functional form, resembling physics and chemistry of studied systems. The adjustable parameters are chosen such that the empirical potential represents a good fit to the relevant regions of the *ab initio* Born-Oppenheimer surface or they are based on experimental data. A typical force field, used in the simulations of biosystems, takes the following form:

$$U(\mathbf{r}_1, \cdots, \mathbf{r}_N) = \sum_{bonds} \frac{a_i}{2} (l_i - l_{i0})^2 + \sum_{angles} \frac{b_i}{2} (\theta_i - \theta_{i0})^2 + \sum_{torsions} \frac{c_i}{2} (1 + \cos(n\omega_i - \gamma_i))$$

$$+ \sum_{atom\ pairs} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{atom\ pairs} k \frac{q_i q_j}{r_{ij}}$$

where in the first three terms summation indices run over all the bonds, angles and torsion angles defined by the covalent structure of the system, whereas in the last two

terms summation indices run over all the pairs of atoms (or sites occupied by point charges q_i), separated by distances $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ and not bonded chemically.

Physically, the first two terms describe energies of deformations of the bond lengths l_i and bond angles θ_i from their respective equilibrium values l_{i0} and θ_{i0}. Harmonic form of these terms (with force constants a_i and b_i) insures correct chemical structure, but on the other hand it prevents modelling chemical changes such as bond breaking. The third term describes rotations around chemical bond, which are characterized by periodic energy terms (with periodicity determined by n and heights of rotational barriers defined by c_i). The fourth term describes the van der Waals repulsive and attractive (dispersion) interatomic forces in the form of the Lennard-Jones 12-6 potential and the last term is the Coulomb electrostatic potential. Some effects due to specific environments can be accounted for by properly adjusted partial charges q_i (and an effective value of the constant k) as well as the van der Waals parameters ε_{ij} and σ_{ij}.

Molecular Dynamics algorithm

In Molecular Dynamics simulations the time evolution of a set of interacting particles is followed via the solution of Newton's equations of motion:

$$\mathbf{F}_i = m_i \frac{d^2 \mathbf{r}_i(t)}{dt^2}$$

where $\mathbf{r}_i(t) = (x_i(t), y_i(t), z_i(t))$ is the position vector of i-th particle and $\mathbf{F}_i$ is the force acting upon i-th particle at time t and m_i is the mass of the particle. "Particles" usually correspond to atoms although they may represent any distinct entities (e.g. specific chemical groups) that can be conveniently described in terms of a certain interaction law. To integrate the above second order differential equations the instantaneous forces acting on the particles and their initial positions and velocities need to be specified. Due to many-body nature of the problem the equations of motion are discretized and solved numerically. The MD trajectories are defined by both position and velocity vectors and they describe the time evolution of the system in phase space. Accordingly, the positions and velocities are propagated with a finite time

interval using numerical integrators, as for example the Verlet algorithm. The (changing in time) position of each particle in space is defined by $\mathbf{r}_i(t)$, whereas the velocities $\mathbf{v}_i(t)$ determine the kinetic energy and temperature in the system. As the particles "move" at the computer their trajectories may be displayed and analyzed (Figure 1), providing averaged properties. The dynamical events that may influence the functional properties of the system can be directly traced at the atomic level what makes MD especially valuable in molecular biology.

Numerical integration of the equations of motion

The idea of the numerical integration of Newton's equations of motion is to find an expression that defines positions $\mathbf{r}_i(t + \Delta t)$ at time $t + \Delta t$ in terms of the already known positions at time t. Because of its simplicity and stability, the Verlet algorithm is commonly used in MD simulations. The basic formula of this algorithm can be derived from the Taylor expansions for the positions $\mathbf{r}_i(t)$ and it reads:

$$\mathbf{r}_i(t + \Delta t) \cong 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\mathbf{F}_i(t)}{m_i} \Delta t^2$$

The above expression is accurate up to terms of the forth power in Δt. Velocities can be calculated from the positions or propagated explicitly as in alternative *leap-frog* or *velocity* Verlet schemes.

The exact trajectories correspond to the limit of an infinitesimally small integration step. It is however desirable to use possibly large time steps to sample longer trajectories. In practice Δt is determined by fast motions in the system. Bonds involving light atoms (e.g. $O - H$ bond) vibrate with periods of several femtoseconds, implying that Δt should be on a subfemtosecond scale to ensure stability of the integration. Although the fastest and not crucial vibrations can be eliminated imposing constraints on the bond length in the integration algorithm, a time step of more than five femtoseconds can be rarely achieved in simulations of biomolecules.

Force calculation and long-range interactions

Updating the positions and velocities in the stepwise numerical integration procedure requires that the forces acting upon the atoms (which change their relative positions each time frame) have to be recomputed at each step. Biomolecular force fields include long-range electrostatic and dispersion interactions and a summation of order N^2 has to be performed in a straightforward implementation to account for all non-bonded pairs. Therefore the repeated calculation of the forces defines the overall complexity of the MD algorithm and many clever techniques have been developed to deal with the problem of long-range forces.

An aqueous solution is the typical environment for the biological macromolecules and it has to be accounted for in realistic simulations, preferably using the atomically detailed representation of the solvent. Because of the limited computer memory and also to speed up the calculations, only a finite sample of an extended (infinite) system can be represented explicitly in a computer model. The treatment of long-range forces is related to the choice of boundary conditions imposed on a system to deal with its finite size and surface effects. The two common approaches are based on either the periodic boundary conditions (Figure 2) and the Ewald method for lattice summations or spherical boundary conditions and the reaction field method. Recent algorithmic developments, as for example the so-called Particle Meshed Ewald method or Fast Multiple method, allow efficient computation of the long-range interactions without resorting to crude cut-off approximation, in which contributions of sites separated by distance larger than a certain cut-off are neglected.

Molecular Dynamics is a statistical mechanics method

According to statistical mechanics, physical quantities are represented by averages over microscopic states (configurations) of the system, distributed in accord with a certain statistical ensemble. One important example is the microcanonical ensemble, in which only the different states corresponding to a specific energy E have non-zero probability of occurring. Another example is the canonical ensemble, in which the temperature T is constant whereas the energy can be exchanged with the surroundings (thermal bath) and the distribution of states is given by the Boltzmann function.

Newtonian dynamics implies the conservation of energy and MD trajectories provide a set of configurations distributed according to the microcanonical ensemble. Therefore, measuring a physical quantity by MD simulation is equivalent to taking an arithmetic average over instantaneous values of that quantity obtained from the trajectories. In the limit of infinite simulation time such averages converge to the true value of the measured thermodynamical properties. In practice the quality of sampling and the accuracy of the interatomic potentials used in simulations are always limited. In fact the quality of sampling may be very poor, especially for processes of time scale larger than typical MD simulations, and caution should be exerted when drawing conclusions from such computer experiments.

Generalizations of MD can be used to sample other statistical ensembles. For example, introducing a coupling to the thermal bath, we may obtain a set of trajectories representing the canonical ensemble. As a matter of fact, the energy is never strictly conserved even in the microcanonical simulation because of the accumulating errors in the numerical integration (estimating the fluctuation of energy is a good test for the quality of MD simulations!). Tricks of the trade like occasional rescaling of velocities may need to be used to adjust the energy, making a thorough equilibration of the system an important issue.

Perceived from the point of view of statistical mechanics, MD is merely a method of conformational sampling that yields average structural and thermodynamical properties. However, sampling many possible configurations can be also used as an optimization method. The (quasi) equilibrium configurations correspond to local minima in the potential energy U (and not the total energy E) and the most stable configuration corresponds to a global minimum of U. With a sufficiently long MD run (and some additional tricks) we may have enough luck to overcome numerous energy barriers and reach the global minimum on the complicated energy surface of a protein.

Limitations of Molecular Dynamics

It is important to realize what are the limitations of MD in order to make a reasonable use of it. Besides, addressing current problems gives a flavour of new frontiers and developments that will define the future of MD simulations.

Quantum effects

Oxygen binding to hemoglobin, catalytic cleavage of the peptide bond by chymotrypsin or the light induced charge transfer in the photosynthetic reaction center are well known examples of biologically important processes. These dynamical events involve quantum effects such as change in chemical bonding, important non-covalent intermediates, tunneling of proton or electron and dynamics on electronically excited states. Straightforward atomic force field simulations cannot be used to model such phenomena.

The change in bonding and the existence of intermediates characteristic for the enzyme reactions can be accounted for using first-principles Molecular Dynamics. However, quantum (or *ab initio*) MD simulations for all valence electrons are yet impractical for large systems. Besides, first-principles MD techniques such as the Car-Parinello method are based on the ground state Density Functional Theory (DFT) and they are at present restricted to the dynamics on the ground state adiabatic surfaces. Therefore, various schemes of quantum-mechanical calculations for an *embedded* quantum subsystem (with the rest of the degrees of freedom treated classically) have been proposed as an alternative approach.

Reliability of the interatomic potentials

The atomic force field defines the physical model of the simulated system. The results of simulations will be realistic only if the potential energy function mimics the forces experienced by the "real" atoms. On the other hand potential should have a simple functional form to speed up the evaluation of the forces. Ideally empirical potentials should be also transferable and applicable to possibly many systems under different conditions. Designing a good force field is a challenging task!

The standard approach combines experimental data and the results of *ab initio* calculations on model systems that can be used as building blocks for a macromolecule. For example, structural parameters as the equilibrium bond lengths can be taken from crystallographic studies and the initial guess of the partial atomic charges can be based on *ab initio* electronic densities. Further refinement of the parameters is possible through computer simulations of well characterized systems and comparison of calculated and experimental results. The available force fields,

such as AMBER (Cornell et al., 1995), CHARMM (Brooks et al., 1983) or GROMOS (van Gunsteren et al., 1996) proved to be sufficiently accurate in terms of kinetic and thermodynamic properties derived by means of MD simulations for proteins or nucleic acids. Yet, it is clear that there is a vast room for improvements. The coming era of long time simulations will pose new challenges in terms of stability and accuracy of MD and systematic techniques of improving the quality of the atomic force fields are highly desirable.

Time and size limitations

The time limitation is the most severe problem in MD simulations. Relevant time scales for biologically important processes extend over many orders of magnitude. For example, the nitric oxide (NO) rebinding to myoglobin takes tens of picoseconds, R to T conformational transition in hemoglobin takes tens of microseconds whereas protein folding may take minutes. However, the presence of significant fast motions limits the time step in numerical integration to about one femtosecond. Thus, following the allosteric transition in hemoglobin requires tens of billions of steps for a system of about ten thousand atoms! While this may become feasible in near future, nanosecond time scale for biosystems comprising several tens of thousands of atoms is the current domain of standard MD simulations. The desired length of simulations also limits increasing the size of the problem to tackle relatively strong interactions of macromolecules with their water and lipid environments. Computationally more demanding evaluation of the forces for large systems implies that each integration step takes longer time. Many alternative strategies and extensions of MD are being explored to study slow conformational changes and activated processes.

Molecular Modeller Kit

Molecular Dynamics is only one of computer methods available to molecular modellers. Global optmization techniques, free energy methods and alternative approaches to conformational analysis enable applications to much broader scope of problems.

Monte Carlo method

Thermodynamic and structural equilibrium properties are static averages independent of the dynamics of the system. Hence, they can be calculated by any (efficient and correct) sampling method. One such method is the widely used Monte Carlo (MC) technique which was developed even before MD. The core of MC algorithm is a heuristic prescription for a plausible pattern of changes in the configurations assumed by the system. Such an elementary "move" depends on the type of the problem. In the realm of protein structure it may be for instance a rotation around a randomly chosen backbone bond. A long series of random moves is generated with only some of them considered as "good" moves. In the standard Metropolis MC a move is accepted unconditionally if the new configuration results in a better (lower) potential energy. Otherwise it is accepted with a probability given by the Boltzmann factor:

$$P_{\mathbf{r} \to \mathbf{r}'} \propto \exp(-\frac{\Delta U}{k_B T})$$

where $\Delta U = U(\mathbf{r}') - U(\mathbf{r})$ denotes the change in the potential energy associated with a move $\mathbf{r} \to \mathbf{r}'$ and k_B is the Boltzmann constant. As a result average properties obtained from the accepted configurations are consistent with the canonical ensemble at temperature *T*. The advantage of MC method is its generality and a relatively weak dependence on the dimensionality of the system. However, finding a "move" which would ensure efficient sampling may be a non-trivial problem.

Except for conformational sampling, MC can be also used as a global optimization method in which we seek a global minimum of the potential energy. Global optimization is an important and difficult problem in protein structure determination. The MC optimization protocol is particularly efficient in conjunction with the so-called *simulated annealing* method based on the sequential "heating" and "cooling" of the system. Indeed, chances of accepting energetically unfavourable moves (or in other words chances to overcome barriers) can be increased by rising the effective temperature.

Free energy methods

The Helmholtz (canonical ensemble) free energy is defined as $F = E - TS$ where E and S are energy and entropy of a system, respectively. Relative free energies determine relative stability of different states of a system whereas the free energy barriers determine the rates of transitions between different states. For example, the relative affinities of ligand binding to receptor molecules result from the differences in free energy, in accord with the relation $\log K \propto -\Delta F / k_B T$ where K is the equilibrium constant for association (defined as relative concentration of liganded and free species under equilibrium conditions). Therefore the atomic simulations allowing estimation of the relative free energies are of great importance. The standard techniques of calculation free energy differences include the so-called *thermodynamic integration* and *thermodynamic perturbation* methods. Both approaches usually imply a series of MD or MC simulations (for intermediate states) with an extensive sampling and are computationally very intensive. Development of better free energy methods is an active research field fuelled by the interest of the pharmaceutical industry.

Studies on conformational changes in proteins

The intrinsic flexibility of proteins manifests in folding, which proceeds through large conformational changes, and in many functionally relevant motions in the neighbourhood of the native state. For example, significant domain motions in a host protein may assist binding of a ligand whereas the thermal motions, such as side chain rotations occurring at equilibrium, are often the precondition for functional activity. Some conformational changes (the fast ones) can be followed by MD simulations. Slow processes can be studied computing the relative free energies of different states. High temperature simulations, simplified models of the system (e.g. with a continuum representation of the solvent) or enhanced sampling of important regions of the free energy profile along a plausible reaction path are often used to explore conformational changes occurring on nanosecond (and longer) time-scales.

One of the problems with comprehensive data is the series of events that myoglobin undergoes upon binding of oxygen or other ligands such as CO and NO. Picosecond relaxation of the protein, the kinetics of ligand rebinding and the paths of ligand

diffusion (Figure 1) have been studied, adding to our understanding of the discrimination mechanism that prevents (dysfunctional) binding of alternative ligands. The thermodynamics of the cooperativity of oxygen binding to hemoglobin units has been studied and changes in the free energy of cooperativity upon crucial point mutations have been quantitatively reproduced (for a recent review see Kuczera, 1996).

Folding of peptides and small proteins (Figure 2) have been also investigated extensively. The structure and kinetics of folding of many peptides have been studied using MD simulations in conjunction with free energy and global optimization methods. Unfolding MD simulations, often performed at high temperatures to enhance the rate of unfolding, provide insights into forces that stabilize proteins. Important kinetic intermediates can be identified and subsequently studied in refolding computer experiments (for a recent review see Brooks, 1998). Initial stages in folding of 36-residue protein (with explicit representation of water) have been directly observed in a 1-microsecond MD simulation (Duan and Kollman, 1998). Other important applications of MD to conformational changes in proteins include studies of ion channel proteins such as gramicidin and other transmembrane proteins (for a recent review see Sansom, 1998) and studies on the ligand-induced or solvent-induced conformational changes such as "hinge bending" observed in many enzymes.

Studies on substrate/inhibitor binding to proteins

Enzyme reactions, diverse control functions, molecular recognition and signalling occur as a result of intermolecular interactions between receptor molecules (typically proteins) and ligands (typically small molecules or peptides). Binding of a ligand to a specific binding site of the receptor is known as the docking problem. Finding the geometry and strength of ligand binding (ligand affinity) is the main goal of computational studies of docking. This goal may be achieved using global optimization and free energy methods coupled with MD or MC sampling. Docking techniques are based either on the simple *key-and-lock* model, in which binding partners are regarded as rigid, or on the more demanding computationally *induced-fit* model, which takes into account the flexibility of ligands and ligand-induced

conformational changes in host proteins. The concept of the thermodynamical cycle is often used to study relative affinities of ligand binding (Figure 3).

Binding of an inhibitor influences the host molecule (by suppressing hinge motions for instance) and reduces its ability to bind primary substrates. Docking studies of the substrate/inhibitor binding to proteins supplement experimental efforts in the drug design. One important example is the discovery and development of several potent and selective inhibitors of the human immunodeficiency virus (HIV) protease as drugs against HIV infection. Free energies of binding, structure and enzyme dynamics of various enzyme-substrate and enzyme-inhibitor complexes of HIV protease have been studied using simple energy minimization, MD and MC coupled with free energy methods and combination of quantum and classical MD simulations (for a recent review see Wlodawer and Vondrasek, 1998). Other important examples include docking studies of peptide antigens binding to specific major histocompatibilty complex (MHC) receptors (for a recent review see Rosenfeld et al., 1995). Predicting which peptides bind to a certain MHC is important for the understanding of the immune response and the design of vaccines.

References:

Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S and Karplus M (1983) CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**: 187-217.

Brooks CL 3rd (1998) Simulations of protein folding and unfolding. *Current Opinion in Structural Biology* **8**: 222-226.

Chakrabar A and Baldwin RL (1995) Stability of α–helices. *Advances in Protein Chemistry* **46**: 141-176.

Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM Jr, Fergusson DM, Spellmeyer DC, Fox T, Caldwell JW and Kollman PA (1995) A second generation force field for the simulation of proteins and nucleic acids. *Journal of American Chemical Society* **117**: 5179-5197.

Duan Y and Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **282**: 740-744.

Elber R, Roitberg A, Simmerling C, Goldstein R, Li H, Verkhivker G, Keasar C, Zhang J and Ulitsky A (1994) MOIL: a program for simulation of macromolecules. *Computer Physics Communication* **91**: 159-189.

Kuczera K (1996) Dynamics and thermodynamics of globins. In: Elber R (ed*) Recent developments in theoretical studies of proteins*, pp. 1-64. Singapore: World Scientific.

Rosenfeld R, Vajda S and DeLisi C (1995) Flexible docking and design. *Annual Review of Biophysics and Biomolecular Structure* **24**: 677-700.

Sansom MS (1998) Models and simulations of ion channels and related membrane proteins. *Current Opinion in Structural Biology* **8**: 237-244.

van Gunsteren WF, Billeter SR, Eising AA, Hünenberger PH, Krüger P, Mark AE, Scott WRP and Tironi IG (1996) Biomolecular Simulation: The GROMOS96 Manual and User Guide. Zürich: Vdf Hochschulverlag.

Simmerling C, Elber R and Zhang J (1995) MOIL-View – a program for visualization of structure and dynamics of biomolecules and STO – a program for computation of stochastic paths. In: Pullman A (ed) *The Proceeding of the Jerusalem Symposium on Theoretical Biochemistry*. Netherlands: Kulwer Academic Publishers.

Wlodawer A and Vondrasek J (1998) Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annual Review of Biophysics and Biomolecular Structure* **27**: 249-284.

Further reading:

Allen MP and Tildesley DJ (1987) *Computer Simulation of Liquids*. Oxford: Clarendon.

Auffinger P and Westhof E (1998) Simulations of the molecular dynamics of nucleic acids. *Current Opinion in Structural Biology* **8**: 227-236.

Brooks CL 3rd, Karplus M and Pettitt BM (1988) *A Theoretical Perspective of Dynamics, Structure and Thermodynamics*. New York: Wiley Interscience.

Creighton TE (ed) (1992) *Protein Folding*. New York: Freeman and Company.

Frenkel D and Smit B (1996*) Understanding Molecular Simulation. From Algorithms to Applications*. San Diego: Academic Press.

Heile JM (1992) *Molecular Dynamics Simulations: Elementary Methods*. New York: Wiley.

McCammon JA and Harvey SC (1987) *Dynamics of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press

Sagui C and Darden TA (1999) Molecular dynamics simulations of biomolecules: longe-range electrostatic effects. *Annual Review of Biophysics and Biomolecular Structure* **28**: 155-179.

Straatsma TP (1996) Free energy methods by molecular simulation. In: Lipkowitz KB and Boyd DB (eds) *Reviews in Computational Chemistry*, Volume 9, pp. 81-127. New York: VCH Publishers.

Tieleman DP, Marrink SJ and Berendsen HJ (1997) A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochimica Biophysica Acta* **1331**: 235-270

Figure legends.

Figure 1 Ligand diffusion pathway through myoglobin mutant (Phe29) as observed in MD simulation (Meller and Elber, unpublished result). The subsequent positions of

the carbonmonoxy ligand with respect to the protein, as the ligand escapes from the heme (marked in red) to the external medium, are recorded and overlapped in order to obtain a suggestive view of the trajectory (CO is represented by spheres in the figure). Several alternative diffusion pathways have been reported in MD simulations of myoglobin. The MOIL package (Elber et al., 1994) was used to perform the simulation and the figure was generated using MOIL-View program (Simmerling et al., 1995).

Figure 2 Simulation model of C peptide (Chakrabar and Baldwin, 1995) in aqueous solution. The peptide (in the native conformation) is put into a box with about 1400 water molecules and periodic boundary conditions are employed to define the implicit lattice copies of the simulation box. MD folding simulations of peptides and small proteins are becoming feasible.

Figure 3 Schematic presentation of the thermodynamical cycle which can be used to facilitate computation of relative affinities of ligand binding to a host protein. Alternative ligands are represented by empty and filled rectangles, respectively. The difference in relative free energies $\Delta F_1 - \Delta F_3$ indicates which ligands bind more strongly. Instead of computationally demanding calculation of ΔF_1 and ΔF_3 one may use equation $\Delta F_1 - \Delta F_3 = \Delta F_2 - \Delta F_4$ and compute ΔF_2 and ΔF_4, which is easier since smaller spatial changes are involved.