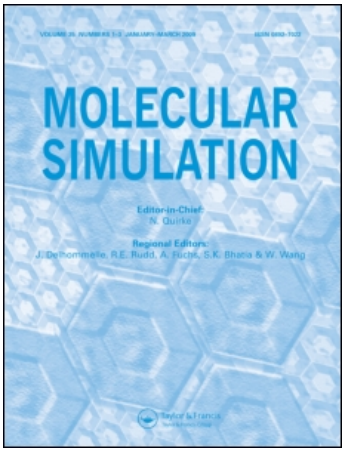


Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Computer Simulation by Molecular Dynamics as a Tool for Modelling of Molecular Systems

W. F. Van Gunsteren[ab]
[a] Department of Physical Chemistry, University of Groningen, Groningen, The Netherlands [b] Department of Physics, Free University, Amsterdam, The Netherlands

To cite this Article Van Gunsteren, W. F.(1989) 'Computer Simulation by Molecular Dynamics as a Tool for Modelling of Molecular Systems', Molecular Simulation, 3: 4, 187 — 200
To link to this Article: DOI: 10.1080/08927028908031372
URL: http://dx.doi.org/10.1080/08927028908031372

Molecular Simulation, 1989, Vol. 3, pp. 187–200
Reprints available directly from the publisher
Photocopying permitted by license only

COMPUTER SIMULATION BY MOLECULAR DYNAMICS AS A TOOL FOR MODELLING OF MOLECULAR SYSTEMS

W.F. VAN GUNSTEREN

Department of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands, and Department of Physics, Free University, P.O. Box 7161, 1007 MC Amsterdam The Netherlands

(*Received October, 1988*)

A survey is given of methods for simulation of molecular systems on a computer. The various assumptions, approximations and limitations are discussed and the possibility of making comparisons with experimental quantities is assessed. Finally, a number of practical applications of molecular dynamics simulation techniques in chemistry are reviewed.

KEY WORDS: Molecular systems, molecular dynamics, chemistry

1 INTRODUCTION

Over the past forty years the computer has achieved an increasingly prominent position in science. This is due to the rapid increase of computer power: every six to seven years the ratio of performance to price has increased by an order of magnitude, and there is no sign of any weakening of this trend. This development has paved the way for simulating in atomic detail a variety of physical processes on a computer. First the degree of reliability of computer simulation procedures and the molecular models and interactomic interaction potentials or force fields used therein had to be assessed. This was and is done by a comparison of simulated with experimental values for those quantities or properties that can be measured experimentally. If a force and simulation technique appear to be sufficiently accurate, computer simulation can be used to predict molecular properties for which experimental determination is too costly, time-consuming or even impossible. This leads to the field of molecular modelling and its use for the design of substances or molecules that possess properties particularly useful in practical applications. Here, one may think of applications in drug or vaccine design, in protein engineering or in material science.

In this paper we will focus our attention to methods, by which fluid-like molecular systems containing N = 100–20000 atoms can be modelled. One may think of a solution of a polypeptide, a protein, an enzyme-ligand complex, a fragment of DNA in any type of solvent, ranging from organic to aqueous (ionic) solutions. A melt of short polymers or a membrane also shows fluid-like behaviour, hence falls in the range of interest too. For these types of systems a variety of properties can be studied. Generally, one is interested in the equilibrium properties: average structure, hydrogen bonding pattern, (free) energy, temperature, density, fluctuations in these quantities

and properties that can be derived from fluctuations, such as the specific heat, compressibility, etc. In modelling one often studies the response of a system to a perturbation: a change in solvent (e.g. polar versus nonpolar), a change in crystal form or from crystal to solution, a change in the composition of the solute (protein engineering) or of the substrate or inhibitor in an enzyme-ligand complex (drug design), a change in temperature or pressure (e.g. in testing materials under adverse (high T,P) conditions). How accurately each property can be simulated will depend on the simulation technique and the characteristics of the force field that is applied. These will be discussed in section 2.

Given a sufficiently reliable force field, the required properties of the N-atom system can be derived using the formulae of statistical mechanics. Due to the complexity of fluid-like many-particle systems, an analytical treatment is of no help. One has to resort to a brute force method like computer simulation to generate the series of molecular configurations that form a statistical ensemble representative of the system under consideration. For a complex molecular system the interaction energy function of atomic coordinates. One of the two basic problems in molecular modelling is how to search the vast phase or configuration space for the global low (free) energy regions, which will be populated by the system in equilibrium.

The common approach to modelling of a molecular system on a computer is a static one. Molecular Mechanics calculations yield one or a few minimum energy conformations of a molecule. On a graphics device molecules are generally studied in terms of fixed conformations. However, a molecular system at room temperature is by no means static. A system of interacting atoms traverses multiple minima of the potential energy surface. In principle one would like to known the multi-dimensional distribution function of all atomic coordinates and the development of this function in time. But, this knowledge can never be complete due to the high dimensionality of the distribution function. In practice, only parts of configuration space can be searched for low (free) energy conformations. The power of a molecular modelling technqiue depends on its ability to find and descibe that part of configuration space that is accessible to the system at the given temperature. The other basic problem is the derivation of a sufficiently accurate interaction energy function for the system of interest. In section 2 both problems are discussed. Section 3 deals with the comparison of simulated with experimental data. Section 4 lists the most important practical applications of dynamic computer simulation in the field of molecular modelling. For other introductions to the field of molecular dynamics computer simulation we refer to [1–7] and references therein.

2 MATERIALS AND METHODS

Theoretical description in chemistry involves the physics of complex molecular systems. The relative complexity of (bio)chemical systems limits the ability of physical theory to treat them properly. However, molecular systems are also simple. From a practical chemical point of view they are built up from just two types of particles, nuclei, with negligible size and irrelevant internal structure, and electrons surrounding them. For most practical purposes the Born-Oppenheimer approximation is applicable, separating electronic and nuclear motion. Thus a complex molecular system can be described as a system of point masses or atoms moving in a potential field due to many-body atom–atom interactions. The intramolecular part of the many-body

interaction function generally consists of two-, three- and four-body terms describing various types of interactions such as covalent bond stretching, bond angle bending and torsional angle interactions, and an effective nonbonded pair interaction term. The nonbonded term is generally limited to two-body terms, since three-body terms are too numerous to evaluate. The use of the word effective means that the average effect of the many-body nonbonded terms has been incorporated in the nonbonded pair potential term. For example, a pair potential with an enhanced dipole moment (enhanced atomic charges) may be used as an effective potential that mimics the average effect of polarisability.

2.1 Force fields for molecular systems

A typical molecular force field or effective potential for a system of N atoms with masses m_i ($i = 1,2, \ldots, N$) and cartesian position vectors $\{\boldsymbol{r}_i\}$ looks as follows:

$$V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) = \sum_{\text{bonds}} 1/2\, K_b[b-b_0]^2 \tag{1}$$

$$+ \sum_{\text{angles}} 1/2\, K_\theta[\theta-\theta_0]^2 + \sum_{\text{torsions}} 1/2\, K_\xi[\xi-\xi_0]^2$$

$$+ \sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \delta)] + \sum_{\text{pairs(i,j)}} [C_{12}(i,j)/r_{ij}^{12}$$

$$- C_6(i,j)/r_{ij}^6 + q_i q_j/(4\pi\varepsilon_0 \varepsilon_r r_{ij})].$$

The first term represents the covalent bond stretching interaction along bond b. It is a harmonic potential in which the minimum energy bond length b_0 and the force constant K_b vary with the particular type of bond. The second term describes the bond angle bending (three-body) interaction in a similar form. Two forms are used for the (four-body) dihedral angle interactions: a harmonic term for dihedrals (torsions ξ) that are not allowed to make transitions, e.g. dihedral angles within aromatic rings, and a sinusoidal term for the other dihedrals (ϕ), which may make 360 degree turns. The last term is a sum over all pairs of atoms and represents the effective nonbonded interaction, composed of the van der Waals' and the Coulomb interaction between atoms i and j with charges q_i and q_j at a distance r_{ij}.

There exists a large number of variations on the form (1) [8–25]. Some force fields contain mixed terms like $K_{b\theta}[b-b_0][\theta-\theta_0]$, which directly couple bond-length and bond-angle vibrations [8–10]. Others use more complex dihedral angle interaction terms [11–13]. The choice of the relative dielectric constant ε_r is also a matter of dispute. Values ranging from $\varepsilon_r = 1$ [14] to $\varepsilon_r = 8$ [15] have been used, while others take ε_r proportional to the distance r_{ij} [16–18]. Sometimes the Coulomb term is completely ignored [8,13]. Although hydrogen bonding can be appropriately modelled using (1) [10, 14], in some force fields special hydrogen bonding potential terms are used to ensure proper hydrogen bonding [12, 13, 15–18). Another way to refine expression (1) is to allow for non-atomic interaction centers or virtual sites, that is, interactions between points (e.g. lone pairs) not located on atoms [19]. For solvents, especially water, a variety of molecular models is available [20], of which a few have explicitly been developed for use in mixed solute-water systems [21–24].

When determining the parameters of the interaction function (1) there are essentially two routes to take. The most elegant procedure is to fit them to results of *ab-initio* quantum calculations on small molecular clusters [25]. However, due to

various serious approximations that have to be made in this type of procedure, the resulting force fields are in general not very satisfactory [20]. The alternative is to fit the force field parameters to experimental data (crystal structure, energy and lattice dynamics, infrared, X-ray data on small molecules, liquid properties like density and enthalpy of vaporisation, free energies of solvation, nuclear magnetic resonance data, etc.). In our opinion the best results have been obtained by this semi-empirical method.

The choice of a particular force field should depend on the type of system for which it has been designed. The MM2 force field [8] is based on gas phase structures of small organic compounds. The AMBER [17, 18] and CHARMM [16] force fields are aimed at a description of isolated polypeptides and polynucleotides, in which the absence of a solvent (aqueous) environment is compensated by the use of a distance dependent dielectric constant ε_r. The ECEPP [12] and UNICEPP [11] force fields use $\varepsilon_r = 4$, whereas the GROMOS force field [14] uses $\varepsilon_r = 1$, since it has been set up for simulation of biomolecules in aqueous environment. The quality of the various force fields should be judged from the literature concerning their application to molecular systems.

We note that the choice of a particular force field should depend on the system properties one is interested in. Some applications require more refined force fields than others. Moreover, there should be a balance between the level of accuracy or refinement of different parts of a molecular model. Otherwise the computing effort put into a very detailed and accurate part of the calculation may be easily wasted due to the distorting effects of the crude parts of the model.

2.2 Simulation methods for molecular systems

For applications at normal temperatures and not involving details of the behaviour of hydrogen atoms, it is sound to assume that the laws of classical mechanics are valid. Various techniques can be distinguished by which molecular configurations which are meant to be representative for the system at equilibrium, can be generated.

2.2.1. Systematic search (SS)

If the system contains only a few degrees of freedom, the complete configuration space can be scanned for low energy molecular configurations and an ensemble can be generated using the Boltzmann factor

$$\exp[-V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N)/kT] \tag{2}$$

as a weight function, where k denotes Boltzmann's constant and T refers to the absolute temperature of the system.

2.2.2. Monte Carlo (MC)

If a system contains many degrees of freedom, straightforward scanning of the complete configuration space is impossible. In that case, an ensemble of configurations can be generated by a combination of random sampling and use of the Boltzmann factor (2). Given a starting configuration a new configuration is generated by random displacement of one (or more) atoms. The displacements should be such that in the limit of a large number of successive displacements the available cartesian space of all atoms is uniformly sampled. The newly generated configuration is either accepted or rejected on the basis of an energy criterion involving the change ΔE of the

potential energy (1) with respect to the previous configuration. In the case of rejection, the previous configuration is counted again and used as a starting point for another random displacement. The criterion is the following: accept if $\Delta E \leqslant 0$, or for $\Delta E > 0$, accept if $\exp(-\Delta E/kT) > R$, where R is a random number taken from a uniform distribution over the interval (0,1). In this way each configuration occurs with a probability proportional to its Boltzmann factor (2). In order to obtain high computational efficiency, one would like to combine a large (random) step size with a high acceptance ratio. For complex systems involving many covalently bound atoms, a reasonable acceptance ratio can only be obtained for very small step size. This makes MC much less efficient than MD for (macro) molecular systems.

2.2.3. Molecular Dynamics (MD)

By application of MD a trajectory (configurations as a function of time) of the system is generated by simultaneous integration of Newton's equations of motion

$$d_{\mathbf{r}_i}^2/dt^2 = m_i^{-1} \boldsymbol{F}_i \tag{3}$$

$$\boldsymbol{F}_i = -\partial/\partial \boldsymbol{r}_i V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) \tag{4}$$

for all the atoms in the system. Here, the force on atom i is denoted by $\boldsymbol{F}_i$.

2.2.4. Stochastic Dynamics (SD)

The technique of stochastic dynamics is an extension of MD. A trajectory of the system is generated by integration of the stochastic Langevin equation of motion

$$d^2 \boldsymbol{r}_i/dt^2 = m_i^{-1}[\boldsymbol{F}_i + \boldsymbol{R}_i] - \gamma_i d\boldsymbol{r}_i/dt. \tag{5}$$

Here, a stochastic force $\boldsymbol{R}_i$ and a frictional force proportional to a friction coefficient γ_i are added to eq. (3). An introduction to SD techniques is given in [26, 27].

2.2.5. Energy Minimisation (EM)

Applying EM one searches for a minimum energy configuration by moving along the gradient

$$\Delta \boldsymbol{r}_i \cong -\partial/\partial \boldsymbol{r}_i V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) \tag{6}$$

through configuration space. Since in this way one moves downhill over the energy hypersurface, EM yields only a local minimum energy configuration, which is generally not far from the initial one. In MD the available kinetic energy makes it possible to cross over energy barriers of the order of kT. This feature makes MD a more efficient technique to locate low energy regions of the interaction function (1) than regular EM techniques.

2.2.6. Normal mode analysis (NMA)

Having obtained a minimum energy configuration one may find the eigenvectors of the second derivative matrix

$$\partial^2/\partial \boldsymbol{r}_i \partial \boldsymbol{r}_j V(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_N) \tag{7}$$

at this point in configuration space, from which the harmonic motion of the system around that configuration may be obtained. Although the technique is elegant, its value is rather limited due to the observation that the dynamics of a molecular system

at room temperature cannot satisfactorily be characterised by pure harmonic motion around one molecular configuration.

For simulation of the properties of large molecular systems containing many covalently linked atoms the MD and SD methods are much more efficient than the SS and MC methods, and much more meaningful than EM and NMA. However, due to the enormous size of the configurational space that is to be scanned for low (free) energy configurations, even MD and SD do only search a limited part of configuration space. Therefore, a judicious choice of the initial molecular configuration of a complex molecular system is often required in order to obtain a trajectory that is representative for the system at equilibrium. The power of modelling with MD or SD lies in their ability to search efficiently in regions which are separated by energy barriers of the order of kT. When searching for the global (free) energy minimum, which may be lying behind much larger barriers, their value is very limited. Fortunately there are a number of applications that require only local optimisation. These will be discussed in section 4.

3 COMPARISON OF SIMULATED WITH EXPERIMENTAL DATA

In order to provide a firm basis for the application of computer simulation methods in molecular modelling the results should be compared to experimental data whenever possible.

Table I contains a scheme of the atomic quantities of molecular systems for which comparison between simulated and experimental values is feasible. Three phases are distinguished. The most interesting phase from a practical chemical point of view is

Table I Possible comparison of simulated properties with experimental ones for complex molecules

ATOMIC PROPERTIES	*Experimental method*	*Phase*		
		gas	*solution*	*crystal*
Structure				
– positions	X-ray diffraction			x
	neutron diffraction			x
– distances	NMR		x	
Mobility				
– B-factors	X-ray diffraction			x
	neutron diffraction			x
– occupancy factor	X-ray diffraction			x
	neutron diffraction			x
Dynamic properties				
– vibrational frequencies	infrared spectroscopy	x		
– relaxation rates	various NMR, optical techniques		x	
SYSTEM PROPERTIES				
Thermodynamic properties				
density			x	x
– free energy			x	
– viscosity, conductance			x	

that of a molecule in solution. Although the reliability of the force field of choice should be assessed by a comparison of predicted with measured properties in solution the table shows that in this phase rather thermodynamic properties, like the free energy of solvation, can be tested than atomic properties; atom–atom distances measured by two-dimensional nuclear magnetic resonance (2D-NMR) techniques are generally not very accurate, and atomic relaxation rates, like those obtained by fluorescence depolarisation or quenching measurements for a chromophore are only possible for very specific molecules or groups of atoms in a molecule.

Much more atomic data are available from crystallographic diffraction experiments. High resolution (≈ 1 Å) diffraction experiments yield accurate atomic positions. The information on atomic mobilities is an order of magnitude less accurate, since the crystallographic refinement process is generally based on simplifying assumptions, like the harmonicity and isotropy of the atomic motions [28]. The B-factors and occupancy factors contain all kinds of experimental uncertainty. A test of dynamical properties of a system is generally not possible using data from the crystalline phase. Intramolecular vibrational frequencies can be measured in the gas phase. Unfortunately, the interesting properties of a molecule in solution, like its structure, mobility and free energy of solvation, are not sensitive to the precise distribution of the intramolecular vibrational frequencies. For example, for a biomolecule in aqueous solution these properties will be much more sensitive to an accurate representation of hydrogen bonding in the force field that is used.

We may summarize the possibility of testing a particular force field by a comparison of simulated with experimental data as follows. By simulating molecules in crystalline form the average predicted structure can be tested and in case high resolution data are available, also atomic mobilities (B-factors) and atomic occupancies may be compared. In addition, experimental and simulated densities may be compared. For simulations of molecules in solution a comparison of relative free energies of solvation of different molecules in a variety of solvents forms a sensitive test of the force field. The dynamic properties of simulations of molecular systems are hard to test by comparison with experiment.

4 APPLICATIONS AND PROSPECTS OF MOLECULAR DYNAMICS SIMULATION IN MODELLING

Of the many possible applications of MD simulation techniques in chemistry we briefly discuss a few which have particular practical value.

4.1 Searching configuration space using MD

For a molecular system with more than a few degrees of freedom a systematic search of the configuration space for low (free) energy conformations becomes impossible, since the number of conformations for which the energy is to be evaluated grows exponentially with the number of degrees of freedom.

An alternative is to use MD or SD simulation techniques to generate a series of different conformations for a molecular system. The available kinetic energy may be used to pass over energy barriers that are of the order of kT per degree of freedom. By raising the temperature T a larger part of conformation space can be searched using MD, as has been shown by di Nola *et al.* [29], who generated a series of quite

different conformations of the hormone somatostatin by applying MD at $T = 600$ K and at 1200 K. The system is kept at the required temperature by coupling it to a heat bath [30]. At the elevated temperature the total energy and the potential energy are monitored for conspicuous fluctuations which may signal a possibly significant conformational change. When minima in the total energy occur, the system is cooled down and equilibrated at normal (300 K) temperature. In this way different conformations with comparable free energy were obtained. We note however that the search at elevated temperature favours selection of higher entropy conformations.

Searching conformation space by MD is expected to be efficient for molecules up to a molecular weight of 1000 Dalton. For larger molecules, which may and are likely to show a particular topological fold, MD techniques will not be able to generate major topological rearrangements. Even when the barriers separating two topologically different low energy regions of conformation space are of the order of kT, the time needed for traversing them may be much too long to be covered in a MD simulation of tens of picoseconds.

4.2 Derivation of three-dimensional (3D) molecular structure on the basis of 2D-NMR data by using restrained MD techniques

During the past few years a new type of practical application of MD has emerged in the field of structure determination by NMR. High resolution NMR at frequencies around 500 MHz is able to resolve individual proton resonances of (bio)molecules in solution up to molecular weights of 20000 Dalton. Once the observed resonances have been assigned to individual protons, 2D Nuclear Overhauser Enhancement (NOE) spectra can be used to obtain upper and lower limits or distance constraints to the distances between pairs of protons. The next problem is the derivation of a 3D structure on the basis of the experimentally determined set of distance constraints. Crude molecular structures that approximately satisfy the set of distance constraints, can be obtained from geometrical methods [31–34], or by model building [35]. These crude structures can be subsequently refined by MD using an additional energy term in eq. (1), which represents the set of distance constraints, and which forces the molecule to satisfy the experimentally observed atom–atom distances [35, 36].

The most simple form chosen for this distance constraint (dc) term is

$$V_{dc}(r_{ij}, r_{ij}^0) = \begin{cases} 1/2\, K_{dc}[r_{ij} - r_{ij}^0]^2 & \text{if } r_{ij} > r_{ij}^0 \\ 0 & \text{if } r_{ij} \leqslant r_{ij}^0 \end{cases} \tag{8a}$$

for an upper bound r_{ij}^0 to the distance r_{ij} between atoms i and j, and

$$V_{dc}(r_{ij}, r_{ij}^0) = \begin{cases} 0 & \text{if } r_{ij} \geqslant r_{ij}^0 \\ 1/2\, K_{dc}[r_{ij} - r_{ij}^0]^2 & \text{if } r_{ij} < r_{ij}^0 \end{cases} \tag{8b}$$

for a lower bound r_{ij}^0 to the distance r_{ij}. When starting the MD refinement using equations (1) and (8), the temperature T may be raised in order to search conformation space more widely, and the force constant K_{dc} of the distance constraint term should not be chosen too large in order to allow some flexibility to the molecule. The values of T and K_{dc} may be changed during the refinement [37]. We note that it is not necessary to use an all atom force field, since a hydrogen atom can be constructed as a so-called virtual atom: its position is a known function of the coordinates of the

heavy atom it is attached to and its covalently bound neighbours. When assignments of proton resonances are ambiguous due to lack of stereo specific assignments (e.g. CH_2 protons) or because of dynamic effects such as rotation of hydrogens in a methyl group or flipping of aromatic rings, a so-called pseudo atom is constructed at an average position between the non-assigned or rotating protons [36].

From various applications of this technique [37–39] to proteins, polypeptides and and DNA fragments it can be concluded that MD is a very powerful tool to refine structures of biomolecules in solution, in micelles or membrane fragments.

4.3 Use of MD simulation in X-ray crystallographic refinement

MD techniques can also be used to refine molecular structures on the basis of crystallographic data. The method bears close analogy to MD refinement of NMR data. The goal is to find a molecular structure which at the same time has a low energy in terms of molecular force fields like (1), and which yields calculated X-ray diffraction intensities $I_{calc}(h, k, l)$ which are as close as possible to the observed ones $I_{obs}(h, k, l)$. The X-ray reflection intensities $I(h, k, l)$ are proportional to the square of the structure factors $F(h, k, l)$,

$$I(h, k, l) \cong |F(h, k, l)|^2 \tag{9}$$

The structure factors $F(h, k, l)$ are defined as the Fourier transform of the electron density $\varrho(x, y, z)$,

$$F(h, k, l) \cong \iiint \varrho(x, y, z)\, e^{2\pi i(hx+ky+lz)}\, dxdydz. \tag{10}$$

The electron density $\varrho(x, y, z)$ can be directly calculated from one or a series of molecular configurations.

A low resolution X-ray structure can be refined by applying MD using an additional energy term

$$V_{sf}(r) = K_{sf} \sum_{hkl} \{|F_{obs}(h, k, l)| - |F_{calc}(h, k, l)|\}^2 \tag{11}$$

in equation (1). This extra term will try to fit the atoms into an electron density that is compatible with the observed X-ray intensities $I_{obs}(h, k, l)$.

When starting the MD refinement using equation (1) and (11), the temperature may be raised in order to search configuration space more widely [40], and the force constant K_{sf} of the structure factor term should not be chosen too large in order to allow some flexibility to the molecule. It is also good practice [41] to start with only using low resolution intensities in (11) and to increase the resolution during the MD refinement.

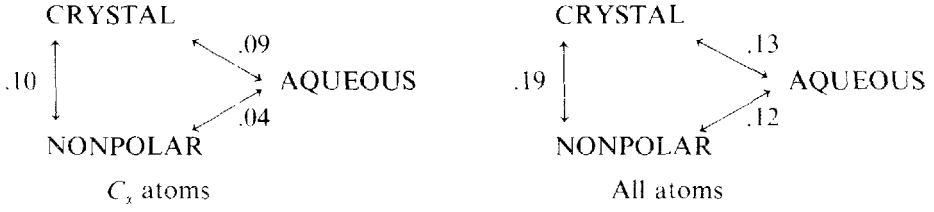
From the few applications of this technique to the proteins crambin [40], α-Amylase inhibitor [40] and phospholipase A2 [41], it can already be concluded that MD simulation is a powerful additional tool in the field of refinement of X-ray crystallographic data.

4.4 Prediction of the dependence of a molecular conformation upon the type of environment

Knowledge about the structure of molecules has been derived primarily from X-ray crystallography. However, it is the behaviour and conformation of a (bio)molecule in

solution that is of more importance than its crystalline form. Therefore, it is of interest to be able to determine the properties and conformation of a molecule as a function of the type of environment, viz. crystalline, nonpolar solution, aqueous solution, etc. Computer simulation by MD can be used to determine the dependence of molecular conformation on the environment by simply simulating the molecule using boundary conditions mimicing the crystal periodicity or an infinitely dilute solution of a chosen type.

An example of this type of analysis is the study of the conformational properties of the immunosuppressive drug cyclosporin A (CPA) by Lautz *et al.* [39]. CPA is a cyclic undecapeptide with amino acid sequence MeBmt, Abu, Sar, MeLeu, Val, MeLeu, Ala, D-Ala, MeLeu, MeLeu, MeVal. The unusual MeBmt residue contains a mainly hydrophobic sidechain with one hydroxyl group. Experimental data were available on its crystal structure from X-ray diffraction and on its conformation in nonpolar solution (chloroform) from NMR measurements. Three MD simulations of CPA were performed: one of CPA in crystalline form, one of CPA dissolved in a nonpolar solvent (carbontetrachloride) and one of CPA dissolved in water. The MD simulation in crystal produced an average structure that deviates from the X-ray structure by 0.03 nm for the C_α atoms (backbone) and by 0.06 nm for all atoms. These values represent the accuracy that was reached by the (GROMOS) force field that was applied. This means that differences in conformation between the simulations in the different environments are only significant, if they are larger than these values. The root mean square positional differences between the different MD time-averaged structures in the different environments (crystalline, aqueous, nonpolar) are given in the following diagram (in nm).



The conformation in nonpolar solution agrees with the NMR data. The conformation of the backbone (C_α atoms) differs in solution from that in crystalline form. The sidechains differ even more and also show a difference between aqueous and nonpolar solution. In nonpolar solution the MeBmt hydrophobic sidechain is in extended form, allowing the hydroxyl group to form a hydrogen bond to the MeBmt backbone C=O group. When CPA in this conformation is dissolved in aqueous solution, the hydrophobic sidechain folds over the molecule and the hydroxyl group exchanges its hydrogen bond with the carbonyl oxygen for one with a water molecule.

This application illustrates the power of MD simulation techniques when studying conformational properties of flexible molecules.

4.5 Calculation of relative binding constants by evaluating free energy differences between various molecular complexes

A new and promising development in the field of computer simulation of molecular systems is the so-called thermodynamic cycle integration technique, which combines well known results from statistical thermodynamics with powerful computer simula-

tion methods. The free energy difference between two states A and B of a system can be determined from a MD simulation in which the potential energy interaction function $V(\lambda)$ is slowly changed such that the system slowly changes from state A to state B. In principle the free energy difference $\Delta F_{BA} = F(B) - F(A)$ is determined as the work necessary to change the system from A to B over a reversible path.

The basis on which the thermodynamic cycle approach rests is the fact that the (Helmholtz) free energy F is a thermodynamic state function. This means that as long as a system is changed in a reversible way the change in free energy ΔF will be independent of the path. Therefore, along a closed path or cycle one has $\Delta F = 0$. This result implies that there are two possibilities of obtaining ΔF for a specific process; one may calculate it directly using the technique sketched above along a path corresponding to the process, or, one may design a cycle of which the specific process is only a part and calculate the ΔF of the remaining part of the cycle. The power of this thermodynamic cycle technique lies in the fact that on the computer also non-chemical processes such as the conversion of one type of atom into another type may be performed.

In order to visualize the method we consider the relative binding of two inhibitors I_A and I_B to an enzyme E. The appropriate thermodynamic cycle for obtaining the relative binding constant is

$$\begin{array}{ccc} E + I_A & \xrightarrow{1\,(\mathrm{exp})} & (E{:}I_A) \\ {\scriptstyle 3\,(\mathrm{sim.})}\downarrow & & \downarrow{\scriptstyle 4\,(\mathrm{sim.})} \\ E + I_B & \xrightarrow[2\,(\mathrm{exp.})]{} & (E{:}I_B) \end{array} \qquad (12)$$

where the symbol : denotes complex formation. The relative binding constant equals

$$K_2/K_1 = \mathrm{e}^{-(\Delta F_2 - \Delta F_1)/RT}. \qquad (13)$$

However, simulation of processes 1 and 2 is virtually impossible, since it would involve the removal of many solvent molecules from the active site of the enzyme to be substituted by the inhibitor. But, since (12) is a cycle we have

$$\Delta F_2 - \Delta F_1 = \Delta F_4 - \Delta F_3 \qquad (14)$$

and, if the composition of inhibitor I_B is not too different from that of I_A, the desired result can be obtained by simulating the non-chemical processes 3 and 4. A short review of the method is presented in [6]. From a number of recent applications [42–48] it is clear that the method of thermodynamic cycle integration by MD computer simulation is a very promising and widely applicable tool in the study of molecular processes at the atomic level.

4.6. Prediction of energetic and structural changes caused by modification of amino acids in enzymes or of base pairs in DNA

The thermodynamic integration technique which was discussed in section 4.5 can also be used to study the effect of amino acid substitution in a protein or of base pair substitution in DNA fragments upon the stability of protein-substrate or DNA-repressor complexes. In the former application the symbols in scheme (12) have the following meaning: E = substrate, I_A = protein, I_B = protein with substituted amino acid residue. In the latter application one has: E = repressor protein,

I_A = DNA fragment, I_B = DNA fragment with substituted base pair. An example of this type of study can be found in [47].

4.7. Prediction of properties of materials under extreme conditions of temperature and pressure

In computer simulation one may investigate details of atomic motion which are inaccessible to experimental probes or study a system under extreme conditions which are inaccessible to experimental probes or study a system under extreme conditions which are too costly to attain experimentally. Here, one may think of the behaviour of materials under high pressure or at high temperatures. The availability of MD simulation methods [30], in which the system is coupled to a heat bath of a specified temperature or to a pressure bath of a specified pressure allows for a study of material properties under extreme conditions of temperature and pressure.

Acknowledgments

The author is much indebted to Herman Berendsen for a stimulating collaboration over the last ten years, and to Masao Fujinaga and Joerg Lautz for their work on the material presented in sections 4.3 and 4.4. This work was supported in part by the Foundation for Chemical Research (SON) and in part by a NATO Science Fellowship both under the auspices of the Netherlands Organisation for the Advancement of Pure Research (ZWO). The hospitality of the Department of Pharmaceutical Chemistry of the University of California at San Francisco is gratefully acknowledged.

The GROMOS (Groningen Molecular Simulation) program library is available at nominal cost from the author.

References

[1] R.W. Hockney and J.W. Eastwood, *Computer Simulation using Particles*, McGraw-Hill, New York (1981).

[2] W.F. van Gunsteren and H.J.C. Berendsen, "Molecular Dynamics Simulations: Techniques and Applications to Proteins," in *Molecular Dynamics and Protein Structure*, J. Hermans ed., Polycrystal Book Service, P.O. Box 27, Western Springs, Ill 60558, USA 5–14 (1985).

[3] G. Ciccotti and W.G. Hoover eds., *Molecular-Dynamics Simulation of Statistical-Mechanical Systems*, Proceedings of the International School of Physics "Enrico Fermi," course 97, North-Holland, Amsterdam 1–610 (1986).

[4] J.A. McCammon and S.C. Harvey, *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, London 1–234 (1987).

[5] H.J.C. Berendsen, W.F. van Gunsteren, E. Egberts and J.de Vlieg, "Dynamic Simulation of Complex Molecular Systems," Amer. Chem. Soc. Symposium Series 353, Supercomputer Research in Chemistry and Chemical Engineering, K.F. Jensen and D.G. Truhlar, eds., Amer. Chem. Soc., Washington D.C., USA 106–122 (1987).

[6] W.F. van Gunsteren and H.J.C. Berendsen, "Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry," *J. Comput.-Aided Molec. Design*, **1**, 171–176 (1987).

[7] M.P. Allen and D.J. Tildesley, "Computer Simulation of Liquids," Clarendon, Oxford, (1987).

[8] N.L. Allinger, "Conformational Analysis. 130. MM2. A hydrocarbon Force Field Utilizing V_1 and V_2 Torsional Terms," *J. Amer. Chem. Soc.*, **99**, 8127–8134 (1977).

[9] U. Burkert and N.L. Allinger, *Molecular Mechanics*, American Chemical Society, Washingtion D.C. 1–339 (1982).

[10] A.T. Hagler, P.S. Stern, S. Lifson and S. Ariel, "Urey-Bradley Force Field, Valence Force Field, and

ab Initio Study of Intramolecular Forces in Tri-tert-butylmethane and Isobutane," *J. Amer. Chem. Soc.* **101,** 813–819 (1979).

[11] L.G. Dunfield, A.W. Burgess and H.A. Scheraga, "Energy Parameters in Polypeptides. 8. Empirical Potential Energy Algorithm for the Conformational Analysis of Large Molecules," *J. Phys. Chem.*, **82,** 2609–2616 (1978).

[12] F.A. Mommany, R.F. McGuire, A.W. Burgess and H.A. Scheraga, "Energy Parameters in Polypeptides. VII Geometric Parameters, Partial Atomic Changes, Nonbonded Interactions, Hydrogen Bond Interactions, and Intrinsic Torsional Potentials for the Naturally Occurring Amino Acids," *J. Phys. Chem.* **79,** 2361–2381 (1975).

[13] M. Levitt, "Molecular Dynamics of Native Protein. I. Computer Simulation of Trajectories," *J. Molec. Biol.*, **168,** 595–620 (1983).

[14] W.F. van Gunsteren and H.J.C. Berendsen, *Groningen Molecular Simulation* (GROMOS) *Library Manual,* Biomos, Nijenborgh 16, Groningen, The Netherlands, 1–229 (1987).

[15] Z.I. Hodes, G. Nemethy and H.A. Scheraga, "Model for the Conformational Analysis of Hydrated Peptides. Effect of Hydration on the Conformational Stability of the Terminally Blocked Residues of the 20 Naturally Occurring Amino Acids," *Biopolymers,* **18,** 1565–1610 (1979).

[16] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan and M. Karplus, "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations," *J. Comput. Chem.*, **4,** 187–217 (1983).

[17] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta and P. Weiner, " A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins," *J. Amer. Chem. Soc.*, **106,** 765–784 (1984).

[18] S.J. Weiner, P.A. Kollman, D.T. Nguyen and D. Case, "An All Atom Force Field for Simulations of Proteins and Nucleic Acids," *J. Comput. Chem.*, **7,** 230–252 (1986).

[19] E. Platt and B. Robson, "Ab-initio Refinement of an Orbital-centred Force Field for Biomolecules. Test Cases Including Peptides, a Sulphonamide and Modelling of DNA Helices," *J. Theor. Biol.*, **96,** 381–399 (1982).

[20] D.L. Beveridge, M. Mezei, P.K. Mehrotra, F.T. Marchese, G. Ravi-Shanker, T.R. Vasu and S. Swaminathan, "Monte Carlo Computer Simulation Studies of the Equilibrium Properties and Structure of Liquid Water," in *Molecular-Based Study of Fluids,* J.M. Haile and G.A. Mansoori eds., Advances in Chemistry Series Vol. 204, Amer. Chem. Soc., Washington D.C. 297–352 (1983).

[21] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren and J. Hermans, "Interaction models for water in relation to protein hydration," in *Intermolecular Forces,* B. Pullman ed., Reidel, Dordrecht, The Netherlands 331–342 (1981).

[22] J. Hermans, H.J.C. Berendsen, W.F. van Gunsteren and J.P.M. Postma, "A Consistent Empirical Potential for Water-Protein Interactions," *Biopolymers,* **23,** 1513–1518 (1984).

[23] W.L. Jorgensen, "Revised TIPS for simulations of liquid water and aqueous solutions," *J. Chem. Phys.*, **77,** 4156–4163 (1982).

[24] W.L. Jorgensen, D.J. Madura and C.J. Swenson, "Optimized Intermolecular Potential Functions for Liquid Hydrocarbons," *J. Amer. Chem. Soc.*, **106** 6638–6646 (1984).

[25] E. Clementi, "Ab Initio Computational Chemistry," *J. Phys. Chem.*, **89,** 4426–4436 (1985).

[26] W.F. van Gunsteren, H.J.C. Berendsen and J.A.C. Rullmann, "Stochastic dynamics for molecules with constraints: Brownian dynamics of n-alkanes," *Molec. Phys.*, **44,** 69–95 (1981).

[27] W.F. van Gunsteren and H.J.C. Berendsen, "A leap-frog algorithm for stochastic dynamics," *Molecular Simulation,* **1,** 173–185 (1988).

[28] J. Kuriyan, G.A. Petsko, R.M. Levy and M. Karplus, "Effect of Anisotropy and Anharmonicity on Protein Crystallographic Refinement," *J. Molec. Biol.*, **190,** 227–254 (1986).

[29] A. DiNola, H.J.C. Berendsen and O. Edholm, "Free energy determination of polypeptide conformations generated by molecular dynamics," *Macromolecules,* **17,** 2044–2050 (1984).

[30] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola and J.R. Haak, "Molecular dynamics with coupling to an external bath," *J. Chem. Phys.*, **81,** 3684–3690 (1984).

[31] T. Havel, I.D. Kuntz and G.M. Crippen, "The theory and practice of distance geometry," *Bull. Math. Biol.*, **45,** 665–720 (1983).

[32] G.M. Crippen, "Distance Geometry and Conformational Calculations," in *Chemometrics Research Studies Series,* D. Bawden ed., Wiley, New York 1–58 (1983).

[33] T. Havel and K. Wuthrich, "A distance geometry program for determining structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular H-H proximities in soluton," *Bull. Math. Biol.*, **46,** 673–698 (1984).

[34] W. Braun and N. Go, "Calculation of Protein Conformations by Proton–Proton Distance Constraints, A New Efficient Algorithm," *J. Molec. Biol.*, **186,** 611–626 (1985).

[35] R. Kaptein, E.R.P. Zuiderweg, R.M. Scheek, R. Boelens and W.F. van Gunsteren, "A Protein Structure from Nuclear Magnetic Resonance Data. Lac Repressor headpiece," *J. Molec. Biol.*, **182,** 179–182 (1985).

[36] W.F. van Gunsteren, R. Boelens, R. Kaptein, R.M. Scheek and E.R.P. Zuiderweg, "An Improved Restrained Molecular Dynamics Technique to Obtain Protein Tertiary Structure from Nuclear Magnetic Resonance Data," in *Molecular Dynamics and Protein Structure*, J. Hermans ed., Polycrystal Book Service, P.O. Box 27, Western Springs, Ill 60558, USA 92–99 (1985).

[37] J. de Vlieg, R. Boelens, R.M. Scheek, R. Kaptein and W.F. van Gunsteren, "Restrained Molecular Dynamics Procedure for Protein Tertiary Structure Determination from NMR Data: A Lac Repressor Headpiece Structure Based on Information on J-coupling and from Presence and Absence of NOE's, *Israel J. Chem.*, **27,** 181–188 (1986).

[38] L. Nilsson, G.M. Clore, A.M. Gronenborn, A.T. Brunger and M. Karplus, "Structure Refinement of Oligonucleotides by Molecular Dynamics with Nuclear Overhauser Effect Interproton Distance Restraints: Application to 5' $d(C\text{-}G\text{-}T\text{-}A\text{-}C\text{-}G)_2$," *J. Molec. Biol.*, **188,** 455–475 (1986).

[39] J. Lautz, H. Kessler, R. Kaptein and W.F. van Gunsteren, "Molecular Dynamics Simulations of the Crystal and Dynamic Modelling of the Solution Structure of Cyclosporin A based on NMR-Data," in *Computational Methods in Chemical Design*, J. Stezowski ed., Oxford University Press, Oxford (1988) in press.

[40] A.T. Brunger, J. Kuriyan and M. Karplus, "Crystallographic R Factor Refinement by Molecular Dynamics," *Science*, **235,** 458–460 (1987).

[41] M. Fujinaga, P. Gros and W.F. van Gunsteren, "Testing the Method of Crystallographic Refinement with Molecular Dynamics," *J. Applied Crystallography*, (1988) in press.

[42] W.F. van Gunsteren and H.J.C. Berendsen, "The Power of Dynamic Modelling of Molecular Systems," in *Computational Methods in Chemical Design*, J. Stezowski ed., Oxford University Press, Oxford (1988) in press.

[43] H.J.C. Berendsen, "Biophysical Applications of Molecular Dynamics," *Comput. Phys. Comm.*, **44,** 233–242 (1987).

[44] T.P. Lybrand, J.A. McCammon and G. Wipff, "Theoretical calculation of relative binding affinity in host-guest systems," *Proc. Natl. Acad. Sci. USA*, **83,** 833–835 (1986).

[45] C.F. Wong and J.A. McCammon, "Computer Simulation and the Design of New Biological Molecules," *Israel J. Chem.*, **27,** 211–215 (1986).

[46] P.A. Bash, U.C. Singh, F.K. Brown, R. Langridge and P.A. Kollman, "Calculation of the Relative Change in Binding Free Energy of a Protein-Inhibitor Complex," *Science*, **235,** 574–576 (1987).

[47] S.N. Rao, U.C. Singh, P.A. Bash and P.A. Kollman, "Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin," *Nature*, **328,** 551–554 (1987).

[48] W.F. van Gunsteren and P.K. Weiner, eds., *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, ESCOM Science Publishers, Leiden, The Netherlands (1988) in press