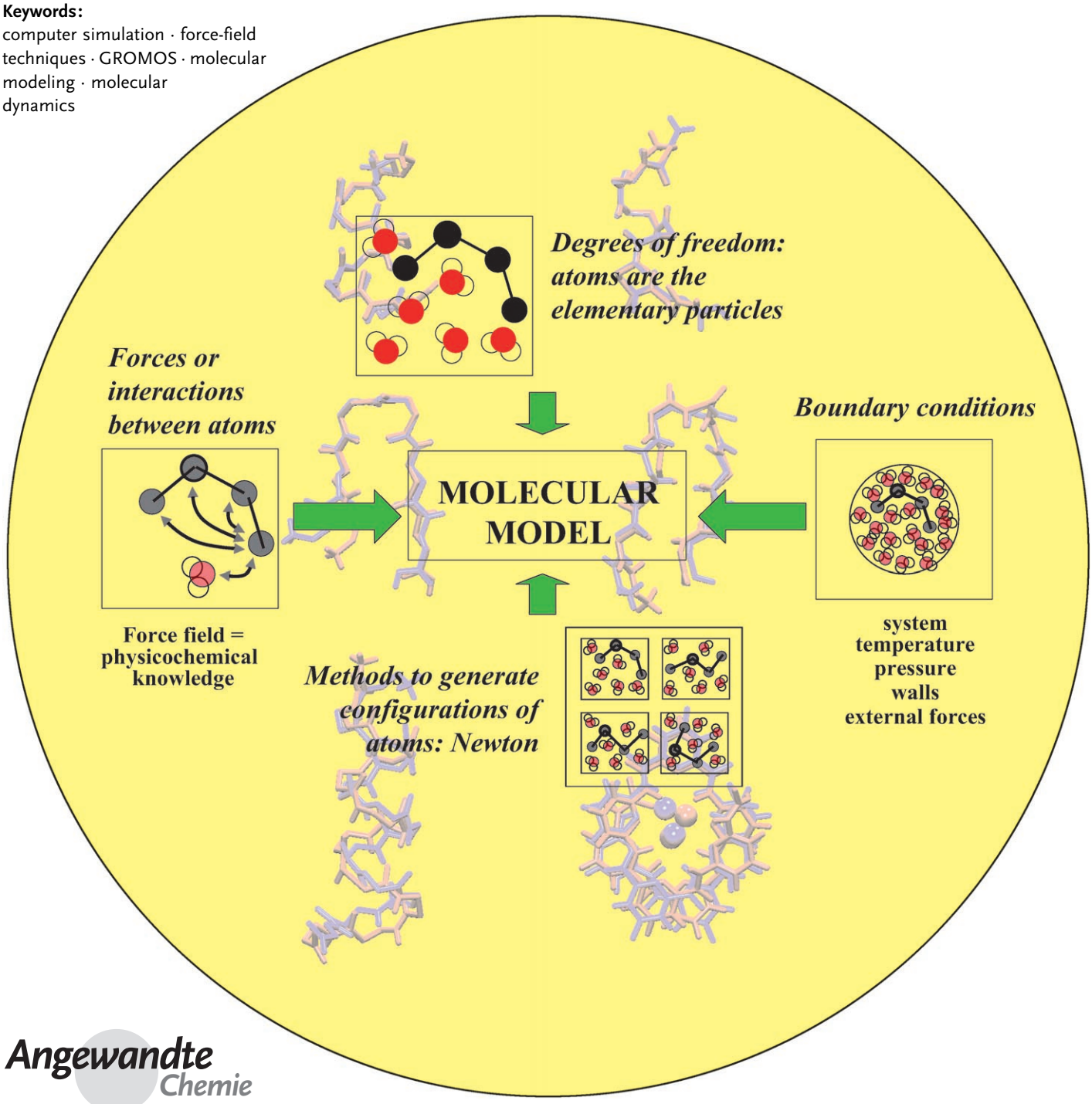
Molecular Dynamics

DOI: 10.1002/anie.200502655

Biomolecular Modeling: Goals, Problems, Perspectives

Wilfred F. van Gunsteren, Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P. Geerke, Alice Glättli, Philippe H. Hünenberger, Mika A. Kastenholz, Chris Oostenbrink, Merijn Schenk, Daniel Trzesniak, Nico F. A. van der Vegt, and Haibo B. Yu*

Keywords:
computer simulation · force-field techniques · GROMOS · molecular modeling · molecular dynamics



Computation based on molecular models is playing an increasingly important role in biology, biological chemistry, and biophysics. Since only a very limited number of properties of biomolecular systems is actually accessible to measurement by experimental means, computer simulation can complement experiment by providing not only averages, but also distributions and time series of any definable quantity, for example, conformational distributions or interactions between parts of systems. Present day biomolecular modeling is limited in its application by four main problems: 1) the force-field problem, 2) the search (sampling) problem, 3) the ensemble (sampling) problem, and 4) the experimental problem. These four problems are discussed and illustrated by practical examples. Perspectives are also outlined for pushing forward the limitations of biomolecular modeling.

From the Contents

1. Introduction

The role of computation in biology, biological chemistry, and biophysics has shown a steady increase over the past few decades. The continuing growth of computing power (in particular in the context of personal computers) has made it possible to analyze, compare, and characterize large and complex data sets that are obtained from experiments on biomolecular systems. This has in turn led to the formulation of models for biomolecular processes that are amenable to simulation or analysis on a computer. When undertaking a biomolecular modeling study of a particular system of interest, the level of modeling, that is, the spatial resolution, time scale, and degrees of freedom of interest, must be considered (Table 1).

Which level of modeling is chosen to describe a particular biomolecular process depends on the type of process. In this Review we focus on three of the four biomolecular processes illustrated in Figure 1: 1) polypeptide folding, 2) molecular complexation (e.g. protein–ligand, DNA–ligand, protein–DNA, etc.), 3) partitioning of molecules between different environments, such as lipid membranes, water, mixtures (e.g. water/urea, ionic solutions), and apolar solvents, and 4) the formation of lipid membranes or micelles out of mixtures of their components. These four processes play a fundamental role in the behavior of biomolecular systems and share the common feature that they are driven by weak, nonbonded interatomic interactions. Such interactions govern the thermodynamic properties of the condensed phase in which the four processes occur. Therefore, these processes are most promisingly modeled at the atomic or molecular level (third row in Table 1). Since the temperature range of interest basically lies between room and physiological temperatures, and energies involved in these processes are on the order of 1–10 $k_B T$ (which corresponds to tens of kJ mol^{-1}, k_B is the Boltzmann constant), the processes are largely determined by the laws of classical statistical mechanics. Although quantum mechanics governs the interactions between the electrons of the atoms and molecules as well as the motions of light particles such as protons, the nonbonded interactions can be very well described by a classical potential-energy function or force field as part of a classical Hamiltonian of the system of interest.[1]

Figure 2 shows the four choices to be made when modeling a biomolecular system: 1) which atomic or molecular degrees of freedom are explicitly considered in the model, 2) which interaction function or force field is used to describe the energy of the system as a function of the chosen degrees of freedom, 3) how these, generally many, degrees of freedom are to be sampled, and 4) how the spatial boundaries and external forces are modeled. As already mentioned, we mainly consider atomic and molecular degrees of freedom with the corresponding classical force fields and classical Newtonian dynamics to sample the degrees of freedom. System sizes that can be considered range up to 10^5 or

[*] Prof. Dr. W. F. van Gunsteren, Dr. D. Bakowies, R. Baron, Dr. I. Chandrasekhar, M. Christen, Prof. Dr. X. Daura, Dr. P. Gee, D. P. Geerke, Dr. A. Glättli, Dr. P. H. Hünenberger, M. A. Kastenholz, Dr. C. Oostenbrink, Dr. M. Schenk, D. Trzesniak, Dr. N. F. A. van der Vegt, Dr. H. B. Yu
Laboratory of Physical Chemistry
Swiss Federal Institute of Technology
ETH
8093 Zurich (Switzerland)
Fax: (+41) 44-632-1039
E-mail: wfvgn@igc.phys.chem.ethz.ch

Prof. Dr. X. Daura
ICREA, Institute of Biotechnology and Biomedicine
Universitat Autònoma de Barcelona
08193 Bellaterra (Barcelona) (Spain)

Dr. C. Oostenbrink
Pharmaceutical Sciences/Pharmacochemistry
Vrije Universiteit
De Boelelaan 1083 P262, 1081 HV Amsterdam (The Netherlands)

Dr. N. F. A. van der Vegt
Max-Planck-Institute for Polymer Research
Ackermannweg 10, 55128 Mainz (Germany)

Dr. H. B. Yu
Department of Chemistry
University of Wisconsin
1101 University Ave, Madison, WI 53706 (USA)

Table 1: Examples of levels of modeling in computational biochemistry and molecular biology.

Methods	Degrees of freedom	Properties, processes	Time scale
quantum dynamics	atoms, nuclei, electrons	excited states, relaxation, reaction dynamics	picoseconds
quantum mechanics (ab initio, density functional, semiempirical, valence bond methods)	atoms, nuclei, electrons	ground and excited states, reaction mechanisms	no time scale
classical statistical mechanics (MD, MC, force fields)	atoms, solvent	ensembles, averages, system properties, folding	nanoseconds
statistical methods (database analysis)	groups of atoms, amino acid residues, bases	structural homology and similarity	no time scale
continuum methods (hydrodynamics and electrostatics)	electrical continuum, velocity continuum etc.	rheological properties	supramolecular
kinetic equations	populations of species	population dynamics, signal transduction	macroscopic

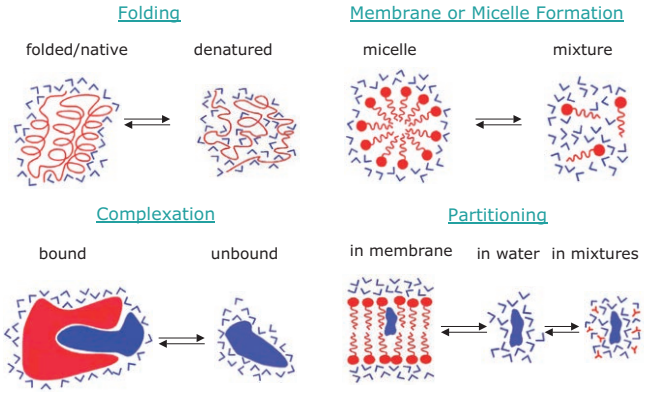


Figure 1. Four biomolecular processes that are governed by thermodynamic equilibria.

10^6 atoms or particles, which is still very small compared to Avogadro's number, that is, macroscopic sizes. For such small systems, the modeling of the boundary or surface will have a large effect on the calculated properties. Such surface effects can be minimized by using periodic boundary conditions, where the box that contains the molecular system is surrounded by an infinite number of copies of itself (Figure 3). This avoids surface effects at the expense of introducing periodicity artefacts.[2–5]

Present day biomolecular modeling is limited in its application by the four problems highlighted in Table 2: 1) the force-field problem, 2) the search (sampling) problem, 3) the ensemble (sampling) problem, and 4) the experimental problem. These four problems are the focus of the present Review and will be discussed and illustrated in Sections 2–5 by using examples from our own work. We stress that the aim of this Review is not to review the contributions of various research groups to the field.

The key reason why computer simulation is used in the study of biomolecular systems in spite of the above-mentioned

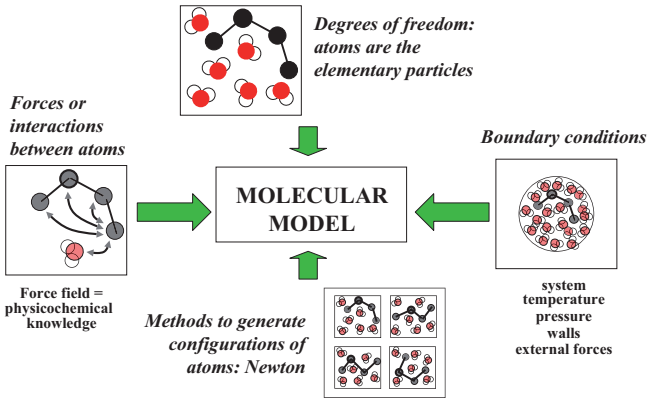


Figure 2. Four basic choices in the definition of a model for molecular simulation.

Wilfred F. van Gunsteren was born in 1947 in Wassenaar (The Netherlands). In 1968 he gained a BSc in physics at the Free University of Amsterdam; in 1976 he was awarded a "Meester" in Law, and in 1976 a PhD in nuclear physics. After postdoc research at the University of Groningen and at Harvard University he was, 1980–1987, senior lecturer and, until August 1990, Professor for Physical Chemistry at the University of Groningen. In 1990 he became professor of Computer Chemistry at the ETH Zürich. He is holder of a gold medal for research of the Royal Netherlands Chemical Scoiety. His main interests center on the physical fundamentals of the structure and function of biomolecules.

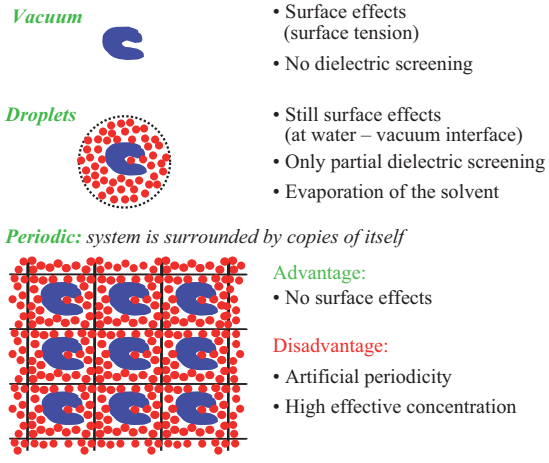


Figure 3. Three types of spatial boundary conditions used in molecular simulation.

Table 2: Four basic problems of biomolecular modeling.

1. force-field problem	A) very small (free) energy differences, many interactions B) entropic effects C) variety of atoms and molecules
2. search problem	A) convergence B) alleviating factors C) aggravating factors
3. ensemble problem	A) entropy B) averaging C) nonlinear averaging
4. experimental problem	A) averaging B) insufficient number of data C) insufficient accuracy of data

limitations to its accuracy resides in the fourth of the four reasons listed in Table 3: Only a very limited number of properties of a biomolecular system is actually accessible to experimental measurement, whereas in a computer simulation not only averages, but also distributions and time series of any definable quantity can be determined. Thus, computer simulation represents a complement to experiment by providing the detailed conformational and other distributions that determine the space and time averages obtained experimentally. As such, it is an indispensable tool to interpret experimental data. Moreover, it can be used to predict properties under environmental conditions that are difficult or expensive to realize. In the next four sections we illustrate the use, power, and limitations of biomolecular modeling in conjunction with experimental efforts with regard to the four processes of interest (Figure 1).

Table 3: Four reasons why computer simulation is used in science.

Simulation can replace or complement an experiment:	
1. experiment is impossible	collision of stars or galaxies weather forcast
2. experiment is dangerous	flight simulation explosion simulation
3. experiment is expensive	high pressure simulation wind channel simulation
4. experiment is blind	many properties cannot be observed on very short time scales and very small space scales

2. The Force-Field Problem

A biomolecular force field generally consists of potential-energy terms representing covalent interactions between atoms (such as bond-stretching, bond-angle bending, improper and proper dihedral-angle torsion) on the one hand and nonbonded interactions on the other hand between atoms in different molecules and between atoms in a molecule that are separated by more than two or three covalent bonds.[6,7]

Since nonbonded interactions govern the thermodynamic equilibria and processes of interest depicted in Figure 1, we focus on the formulation and parametrization of these potential-energy terms. Three problems dominate the topic of force-field development (Table 2, point 1, A–C).

A first problem is that the (free) energy differences driving the processes of Figure 1 are of the order of 1–10 $k_B T$ (which corresponds to tens of kJ mol^{-1}). These relatively small energies result from a summation over very many (10^6–10^8) atom pairs: A system of $N = 1000$ atoms has about $^1/_2 N(N-1) = 500\,000$ pairs of atoms contributing to the nonbonded interaction. To reach the requested accuracy for the total nonbonded energy, the accuracy of the individual terms in the summation (the atom-pair energies) must be higher. This difficulty becomes increasingly severe when trying to derive a force field of high accuracy for larger systems, that is, for larger values of N.

A second problem is to appropriately account for entropic effects. Since we are not interested in biomolecular systems at a temperature of −273.15 °C (0 K), we have to consider the contribution of entropy S to the free energy $F = U - TS$ of the system of interest. It is well known that entropy plays an essential role in all four of the processes shown in Figure 1. Changes in free energy that drive processes may result from changes in internal energy (U) or in entropy (S), which may work together or against each other depending on the relative strengths of the nonbonded interactions between the various components (atoms, molecules) of the system.[8,9] Figure 4 illustrates the phenomenon of energy–entropy compensation: two conformations x_1 and x_2 of a molecule may have $U(x_1) \ll U(x_2)$, while $F(x_1) > F(x_2)$ if at a given temperature $S(x_1) \ll S(x_2)$. The entropy is a measure of the extent of conformational space (x) accessible to the molecular system at a given temperature T.

Figure 4 also illustrates that searching for and finding the global energy minimum of a biomolecular system is meaningless when its entropy accounts for a sizeable fraction of its free energy. For example, $F = -24$ kJ mol^{-1}, $U = -41$ kJ mol^{-1}, and $TS = -17$ kJ mol^{-1} for liquid water at room temperature and pressure. The properties of water in

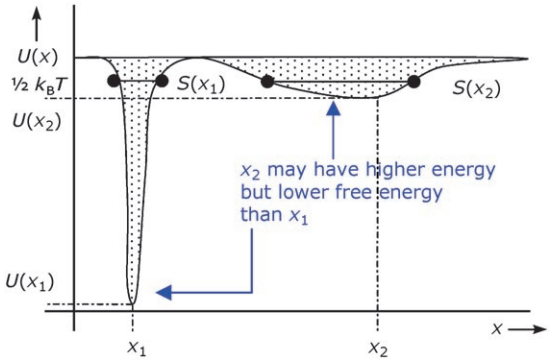


Figure 4. Energy–entropy compensation at finite temperatures.

the condensed phase can therefore only be described through a conformational distribution, which in turn can be generated by computer simulation. Similar considerations apply to biomolecular systems: an energy-minimized structure of a protein corresponds to a possible conformation at 0 K, and lacks information on the conformational distribution of the protein at physiological temperatures. This state of affairs has consequences for the development of force fields: if a force field is to be used in computer simulations above 0 K, its parameters should be derived or calibrated taking into account entropic effects for it to be consistent. In other words, calibration of force-field parameters involves computer simulations to generate configurational ensembles, which makes it a more costly task than when only single minimum-energy conformations or measured average structures are used.

A third problem in the development of a biomolecular force field is the enormous variety of chemical compounds for which adequate force-field parameters should be derived. If the force-field parameters are (to some extent) transferable between atoms or groups of atoms in different molecules, this problem may be (at least partially) alleviated. In general, putting the force-field terms on a physical (instead of a purely statistical) basis and keeping them simple and local will enhance the transferability of parameters from one compound to another. In addition, by keeping them computationally simple, the efficiency of biomolecular simulation can be enhanced, which facilitates the sampling of configurational space.

2.1. Functional Form of the Force-Field Terms

Most biomolecular force fields are composed of terms that possess a rather simple functional form.[6] The GROMOS force field, for example, consists of the following terms [Eqs. (1)–(8)]:[7,10]

$$V^{\mathrm{bond}}(\boldsymbol{r};K_{\mathrm{b}},b_0) = \sum_{n=1}^{N_{\mathrm{b}}} \tfrac{1}{4}K_{b_n}[b_n^2-b_{0_n}^2]^2 \tag{1}$$

$$V^{\mathrm{angle}}(\boldsymbol{r};K_\theta,\theta_0) = \sum_{n=1}^{N_\theta} \tfrac{1}{2}K_{\theta_n}[\cos(\theta_n)-\cos(\theta_{0_n})]^2 \tag{2}$$

$$V^{\mathrm{har}}(\boldsymbol{r};K_\xi,\xi_0) = \sum_{n=1}^{N_\xi} \tfrac{1}{2}K_{\xi_n}[\xi_n-\xi_{0_n}]^2 \tag{3}$$

$$V^{\mathrm{trig}}(\boldsymbol{r};K_\varphi,\delta,m) = \sum_{n=1}^{N_\varphi} K_{\varphi_n}[1+\cos(\delta_n)\cos(m_n\varphi_n)] \tag{4}$$

$$V^{\mathrm{LJ}}(\boldsymbol{r};\mathrm{C}_{12},\mathrm{C}_6) = \sum_{\text{pairs } i,j}\left[\frac{\mathrm{C}_{12}(i,j)}{r_{ij}^{12}}-\frac{\mathrm{C}_6(i,j)}{r_{ij}^{6}}\right] \tag{5}$$

$$V^{C}(\boldsymbol{r};q) = \sum_{\text{pairs } i,j}\frac{q_i\,q_j}{4\pi\varepsilon_0\varepsilon_1}\frac{1}{r_{ij}} \tag{6}$$

$$V^{\mathrm{RF}}(\boldsymbol{r};q) = \sum_{\text{pairs } i,j}\frac{q_i\,q_j}{4\pi\varepsilon_0\varepsilon_1}\frac{(-\frac{1}{2}C_{rf}\,r_{ij}^2)}{R_{rf}^3} \tag{7}$$

$$V^{\mathrm{RF_c}}(\boldsymbol{r};q) = \sum_{\text{pairs } i,j}\frac{q_i\,q_j}{4\pi\varepsilon_0\varepsilon_1}\frac{(\frac{1}{2}C_{rf}-1)}{R_{rf}} \tag{8}$$

The first four equations describe the four types of covalent (bonded) interactions mentioned before, while the last four specify the nonbonded interactions: the van der Waals interaction cast in the form of a Lennard–Jones term, the electrostatic Coulomb interaction between (partial) atomic charges q_i, the distance-dependent and distance-independent (constant) interactions arising from the dipolar reaction field (RF) induced by the charge distribution inside the cut-off sphere through the continuous dielectric medium outside this cut-off sphere. Since this force field covers a variety of molecules (including polypeptides, polysaccharides, nucleic acids, lipids), it contains a large set of parameters:[7] 52 types of bonds [Eq. (1)], 54 types of bond angles [Eq. (2)], 3 types of improper (harmonic) dihedral angles [Eq. (3)], 41 types of proper torsional (trigonometric) dihedral angles [Eq. (4)], van der Waals interactions of 53 types of atoms [Eq. (5)], and many different sets of atomic charges for the typical polar or charged groups of atoms in the molecules mentioned above [Eqs. (6)–(8)].[7,10]

The functional forms are chosen such that they are easy to compute. The nonbonded interactions only contain pair terms, and the more complex three- and four-body covalent terms [Eqs. (3) and (4)] are much fewer in number than the nonbonded pair terms. The solvent part of this biomolecular force field only contains nonbonded terms, the intramolecular degrees of freedom of solvent molecules are kept frozen. The major computational effort resides in evaluating the nonbonded interactions.

2.2. Calibration of Force-Field Parameters

Having specified the functional form of the interaction terms, the formidable task of finding appropriate, consistent values for the hundreds of force-field parameters remains to be addressed. This task involves the choice of type of data, type of systems, thermodynamic phase, and properties to be used as the calibration set for specific force-field parameters. The choices made for the GROMOS force field are summarized in Table 4. Since biomolecular systems are generally in the condensed phase, data for the condensed phase (experimental and theoretical) are used whenever possible. Furthermore, to maximize the transferability of parameters between groups of atoms in different molecules, only data for small molecules are used. When using data from large molecules such as proteins (e.g. from the protein data bank) properties of groups of atoms may be dependent on their particular environment in the folded molecule. Furthermore, the protein data bank contains structures measured at widely different thermodynamic conditions (pH value, ionic strength, etc.). Finally, certain properties will be strongly related to specific force-field parameters and only weakly to others. This situation offers the opportunity to reduce the calibration effort by optimizing specific subsets of parameters separately against a limited set of properties.

Table 4: Choice of calibration sets of data, systems, properties, and thermodynamic phase for the derivation of the GROMOS biomolecular force-field parameter values.[7]

Type of data	Type of system	Phase	Type of properties	Force-field parameter
structural data (exptl)	small molecules	crystalline solid phase	molecular geometry: bond lengths, bond angles	b_0, θ_0, ξ_0
spectroscopic data (exptl)	small molecules	gas phase	molecular vibrations: force constants	K_b, K_θ, K_ξ
thermodynamic data (exptl)	small molecules, mixtures, solutions	condensed phase	heat of vaporization, density, partition coefficient, free energy of solvation	van der Waals: $C_{12}(i,j)$, $C_6(i,j)$, q_i(final)
dielectric data (exptl)	small molecules	condensed phase	dielectric permittivity, relaxation	charges q_i
transport data (exptl)	small molecules	condensed phase	diffusion and viscosity coefficients	$C_{12}(i,j)$, $C_6(i,j)$, q_i
electron densities (theor.)	small molecules	gas phase	quantum-chemical calculation of atom charges	charges q_i(initial)
energy profiles (theor.)	small molecules	gas phase	quantum-chemical calculation of torsional-angle rotational profiles	K_ϕ, δ, m

The following strategy was applied in developing the GROMOS force field (Table 4).[7] The geometric parameters for the covalent interaction terms were obtained from the crystal structures of small molecules, and the corresponding vibrational force constants came from infrared spectroscopic data on small molecules in the gas phase. The nonbonding interaction parameters C_{12}, C_6, and q were obtained by fitting heats of vaporization, densities of pure liquids, and free energies of solvation of small solutes in polar and apolar solvents as obtained from molecular dynamics (MD) simulations to data obtained by experiment. Dielectric permittivities and diffusion properties of liquids were used as secondary data in this parametrization. Electron densities obtained from quantum-mechanical calculations were only used to obtain an initial guess for partial atomic charges, because they may depend strongly on the environment (gas phase or condensed phase) as a result of polarization effects. Torsional-angle parameters were derived by fitting torsional-energy profiles to quantum-mechanical data, thus leaving the set of parameters for the nonbonding interactions untouched.

A biomolecular force field forms, in principle, a consistent set of parameters for both solute molecules (proteins, lipids, saccharides, nucleotides) and solvent molecules (water, alcohols, DMSO, chloroform, etc.). Changing a subset of parameters by taking them from other force fields or models may introduce inconsistencies and inaccuracy.

The simulation of peptide or protein folding (Figure 1) necessitates that the relative free energies of solvation of the 20 amino acid residues in polar solution (water) versus nonpolar solution (cyclohexane) as obtained from simulation compares well with the corresponding experimental data, since these differences are likely to largely determine the driving force for folding. Gibbs free energies of solvation obtained with the GROMOS 53A6 force field[7] are shown in Figure 5. The average absolute deviations from experiment for the 18 amino acid side chains (except Gly and Pro) are 1.0 kJ mol^{-1} in water and 2.0 kJ mol^{-1} in cyclohexane. Both values are smaller than the value of $k_B T$, which makes the 53A6 GROMOS force field suitable for studies on protein folding.

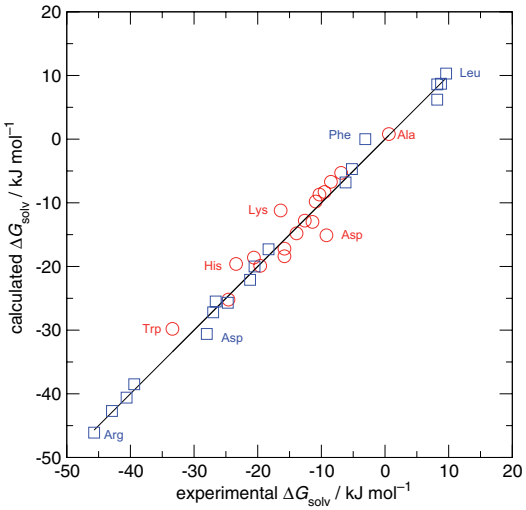


Figure 5. Comparison of the calculated (MD simulation using the GROMOS 53A6 force field) and experimental Gibbs free energies of solvation in cyclohexane (circles) and in water (squares) of 18 amino acid analogues (no Gly and Pro).[7]

2.3. Long-Range Forces

Electrostatic interactions play a major role in biomolecular systems. Compared to covalent and van der Waals interactions, their range is relatively long, because electrostatic interactions between molecules or parts of molecules at a distance r from each other decrease only slowly with increasing value of r:

1. The interaction energy between two charged molecules is proportional to r^{-1}, while the corresponding force, the (negative) spatial derivative of the energy, is proportional to r^{-2}.
2. The interaction energy between a neutral molecule with a dipole moment and a charged molecule is proportional to r^{-2}, while the corresponding force is proportional to r^{-3}.
3. The interaction energy between two neutral molecules with dipole moments is proportional to r^{-3}, while the corresponding force is proportional to r^{-4}.

Continuing this multipole expansion with quadrupole moments, octupole moments, etc. shows that even the interaction between two neutral molecules without dipole moments, but with quadrupole moments is longer ranged (proportional to r^{-5}) than the van der Waals dispersion interaction, which is proportional to r^{-6}. If we consider the electrostatic energy of a single charge, dipole, or quadrupole with all charges, dipoles, and quadrupoles surrounding it, we have to integrate the electrostatic interaction $V^{\mathrm{el}}(r)\,4\pi r^2$ from r to infinity, where $4\pi r^2\,\mathrm{d}r$ is the volume of the spherical shell between r and $r+dr$ surrounding the central charge, dipole, or quadrupole [Eq. (9)].

$$\int_0^\infty V^{\mathrm{el}}(r)\,4\pi r^2\,\mathrm{d}r \tag{9}$$

However, the integral (9) only converges under the conditions of Equation (10):

$$V^{\mathrm{el}}(r) \sim r^{-n},\ n > 3 \tag{10}$$

Thus, the total electrostatic energy of ionic systems depends on the spatial boundary conditions that restrict the range of the integral (9) in practical calculations. The gradual decrease of pair energies and forces with the interatomic separation r means that the results of simulations will depend on the way long-range interactions are treated in the force and energy calculations.

Two techniques are predominantly used currently to evaluate long-range (electrostatic) interactions in biomolecular systems (Figure 6). In the so-called lattice-sum methods the system is put into a particularly shaped box (cubic, rectangular, triclinic, truncated octahedral) and surrounded by an infinite number of identical copies of itself. In this way the boundary problem is moved to infinity, but, unfortunately, is not removed. Moreover, an artificial periodicity is enforced upon the system. Lattice-sum methods are the Ewald summation,[12] the particle–particle/particle–mesh (P^3M) method[13] and the particle–mesh–Ewald (PME) method.[14]

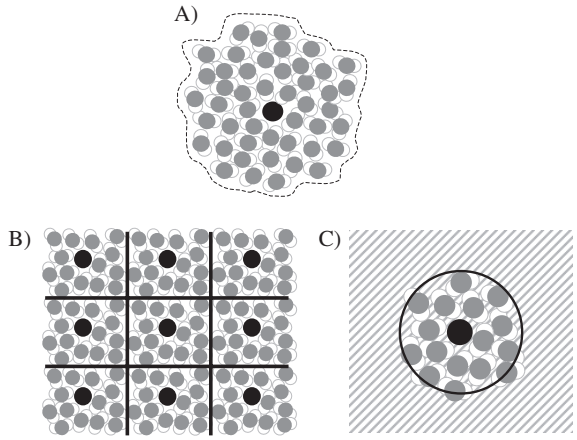


Figure 6. Two methods for calculating long-range electrostatic energies and forces in a molecular system: A) real system with explicit solvent; B) periodicity used in the Ewald, P^3M, and PME methods, and C) continuum approximations beyond a given cut-off distance.

An alternative method is to approximate the medium beyond a given cut-off distance R_{rf} from a specific atom or molecule by a dielectric continuum of uniform permittivity $\varepsilon_{\mathrm{rf}}$ and ionic strength I_{rf}.[15,16] Such a dielectric continuum produces a reaction field in response to the charge distribution inside the cut-off sphere with radius R_{rf}, which can easily be calculated for every atom or molecule [see Equations (7) and (8)]; the constant C_{rf} depends on R_{rf}, $\varepsilon_{\mathrm{rf}}$, and I_{rf}.[10,16]

Both methods are approximations of different type. The reaction-field approach is a mean-field approximation of the real charge distribution beyond a distance R_{rf}, and treats its dielectric response in a spherically symmetric way. It does not introduce artificial periodicity. The lattice-sum approach does not involve averaging, but treats interactions beyond the box size as periodic. Both approximations and their effects have been investigated for various systems.[3–5,17–19]

The effect of different treatments of the long-range electrostatic interactions on the free energy of hydration of a charged (ionic) solute in water is illustrated in Table 5.[20] The results show that the method used to handle long-range forces and its parameters (e.g. R_{rf}) are of great importance when parametrizing a force field. For example, the nonbonding parameters of the OPLS force field[28,29] have generally been obtained from calculations with cut-off radii $R_{\mathrm{c}} = 0.95$–1.5 nm, with the longer-range van der Waals interactions being included through correction formulae (see for example, Ref. [30]). The GROMOS force field was calibrated using $R_{\mathrm{c}} = 1.4$ nm and a reaction field force.

2.4. Testing Biomolecular Force Fields

Having developed a biomolecular force field through calibration of its parameters to reproduce a variety of properties of small molecules, this force field remains to be tested by application to biomolecular systems containing different, larger molecules in the condensed phase. Tests should include proteins, saccharides, or nucleotides in aqueous solution, and comparisons should be made for simulated

Table 5: Computation of methodology-independent ionic hydration free energies from molecular dynamics simulations with explicit solvent.[a]

Method	N_w	R_{rf}	ΔF_{sim}	ΔF_{ncb}	ΔF_{per}	ΔF_{sum}	ΔF_{srf}	ΔF^0_{hyd}
P^3M	4	–	−45.8	0.0	−293.2	−61.9	−3.54	−390.82
P^3M	8	–	−118.7	0.0	−259.7	−69.8	−1.97	−436.62
P^3M	16	–	−164.8	0.0	−221.5	−74.5	−1.04	−448.15
P^3M	32	–	−209.9	0.0	−183.9	−77.1	−0.54	−457.67
P^3M	64	–	−249.0	0.0	−150.0	−78.4	−0.27	−464.02
P^3M	128	–	−279.8	0.0	−121.1	−79.1	−0.14	−466.45
P^3M	256	–	−303.8	0.0	−97.1	−79.4	−0.07	−466.71
P^3M	512	–	−324.8	0.0	−77.6	−79.6	−0.03	−468.35
P^3M	1024	–	−340.5	0.0	−61.8	−79.7	−0.02	−468.35
P^3M	2048	–	−353.5	0.0	−49.2	−79.7	−0.01	−468.73
RF	512	0.8	−275.9	−128.1	−1.47	−76.8	0.0	−468.55
RF	512	1.0	−298.6	−102.7	−3.76	−77.4	0.0	−468.77
RF	512	1.2	−311.0	−85.7	−6.76	−77.6	0.0	−467.49
RF	1024	0.8	−278.4	−128.1	−0.38	−76.8	0.0	−469.97
RF	1024	1.0	−300.8	−102.7	−1.26	−77.4	0.0	−468.49
RF	1024	1.2	−315.6	−85.7	−2.72	−77.6	0.0	−468.03
RF	2048	0.8	−277.9	−128.1	−0.07	−76.8	0.0	−469.20
RF	2048	1.0	−301.7	−102.7	−0.32	−77.4	0.0	−468.48
RF	2048	1.2	−318.4	−85.7	−0.87	−77.6	0.0	−468.98

[a] Standard hydration free energy ΔF_{hyd} of the sodium cation calculated for different system sizes (number of water molecules N_w) by using the P^3M[13,21,22] or reaction-field[15,16] (with different cutoff radii R_{rf}) methods for the treatment of electrostatic interactions. The SPC water model[23] was used together with the Lennard–Jones ion–water interaction parameters of Straatsma and Berendsen.[24] The simulations were carried out at constant volume, in periodic cubic boxes of edge $L=[(N_w+1)\rho^{-1}]^{1/3}$ with $\rho=33.427\ nm^{-3}$. For the P^3M method a spherical hat charge-shaping function of width 0.4 nm (or 0.26 nm for $N_w \leq 32$) was used,[25] together with an assignment function of order three, a finite-difference operator of order two, three alias vectors for the calculation of the optimal influence function, and a grid spacing of 0.05 nm (or 0.0166 nm for $N_w \leq 32$).[21] For the RF method, the reaction-field radius was set to R_{rf} and the solvent permittivity to 66.6. A cut-off truncation was applied based on the oxygen atom as the molecular center. The raw charging free energies ΔF_{sim} (calculated from the simulations using the scheme proposed by Hummer et al.,[26] based on three ionic charge states of 0, 0.5, and 1 *e*) are corrected (based on a solvent permittivity of 66.6 and an approximate ionic radius of 0.2 nm) for the effect of non-Coulombic interactions ΔF_{ncb}, for artificial periodicity ΔF_{per}, for the use of an improper summation scheme for the electrostatic potential ΔF_{sum} (conversion from P-sum to M-sum convention[27]), for the effect of the interfacial potential at the ionic surface on the average potential within the computational box ΔF_{srf}, for the work of cavity formation (ΔF_{cav}; 5.67 kJ mol^{-1}), and for the compression work ΔF_{cmp} (7.95 kJ mol^{-1}; corresponding to the standard-state correction for a gas at a reference pressure of 1 bar), leading to final standard (intrinsic) values ΔF^0_{hyd}.

properties with available experimental values of measurable ones. Here, one may think of comparing simulated conformational distributions in crystals with averages derived from measured X-ray diffraction data. NOE (nuclear Overhauser effect) intensities, ^3J-coupling constant, and chemical shift values calculated from simulations of solutes in solution may be compared with the corresponding averages derived from NMR experiments. However, reproduction by simulation of a particular folded structure derived from experiment is neither a necessary[31] nor a sufficient condition for a force field to be correct. The force field should be able to reproduce the conformational distribution of the solute as a function of the thermodynamic conditions; for example, it should predict the correct melting temperature of a particular fold.

A particular challenge to biomolecular force fields is the prediction of the fold of a polypeptide in solution as a function of its amino acid residue composition and the type of solvent. How well this challenge is met by the GROMOS force field is illustrated in Figures 7 and 8. Using the GROMOS force-field parameter sets, 43A1 and 45A3 left-handed or right-handed helices of different types as well as β turns were found as dominant conformations in MD simulations of β- and α-peptides in methanol or water (Figure 7), a result that is in agreement with the dominant conformations derived from NOE data.[32–37] However, since these parameter sets were shown to underestimate the magnitude of the hydration (Gibbs) free energies for amino acid analogues,[7,38,39] a reparametrization was carried out which led to the 53A6 set.[7] Figure 8 illustrates the improvement obtained by including solvation free energies of polar compounds into the calibration set in the context of the prediction of the correct fold of a β-dodecapeptide containing many polar side chains. While the 45A3 set could not reproduce the experimentally observed helix, the 53A6 set did. Properties that are less sensitive to small (free) energy differences were, however, well reproduced by both parameter sets.[36]

Thus, a particular force field can only be (in)validated by investigating molecular or system properties that are sensitive to the particular simulation parameters and conditions.

2.5. Perspectives in Force-Field Development

There is still room for improvement in current biomolecular force fields. First, the van der Waals parameters and partial charge distributions of charged moieties should be based on free energies of solvation, as has been done for those of apolar and polar neutral moieties.[7] Examples of such groups are the side chains of Arg, Lys, Asp, and Glu amino acid residues or phosphate groups occurring in DNA, RNA, and lipids. However, this is easier said than accurately done

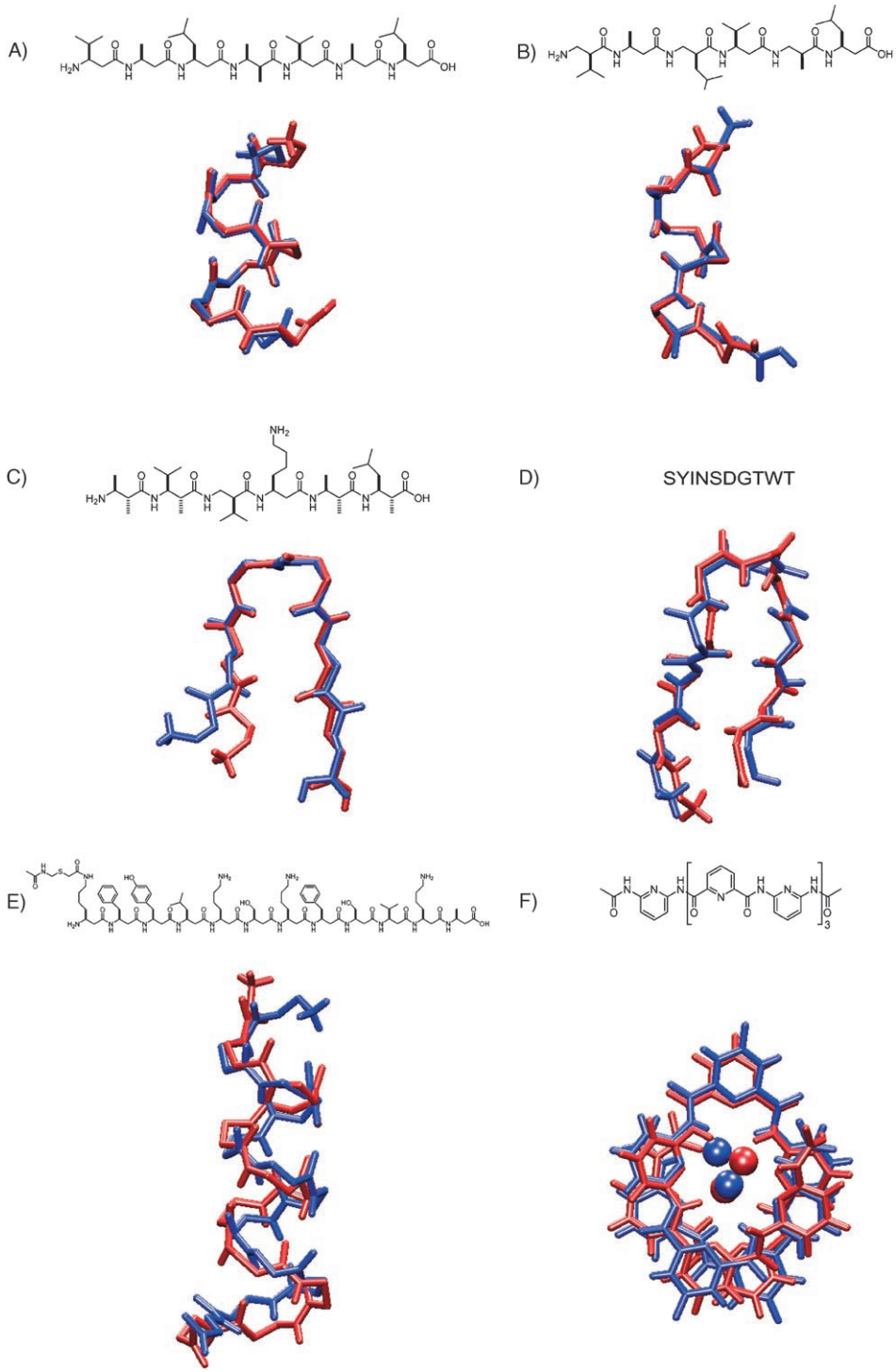


Figure 7. Folding of different polypeptides and peptoides into different folds in different solvents by MD simulation. The folded structure (red), modeled from the available NMR or X-ray experimental data, is superimposed on a folded structure (blue) representing the most populated conformation from the MD simulations of the folding/unfolding equilibrium.[32–37] The solvents are methanol (A–C, E), water (D), and water or chloroform (F). The versions of the GROMOS force field used are: 43A1 (A–D), 45A3 (F), and 53A6 (E).

because of the large solvation free energies of single ions (of the order of several hundred $k_B T$) and the technical difficulties in obtaining such data from experiment[40] and from simulations.[27]

Second, the properties of solvent mixtures, as these are used in experimental protein denaturation studies, should be evaluated as a function of their composition. The thermodynamic properties, such as energy and density of mixing, are of particular importance when the free energy of solvation and folding of solutes is to be calculated.[8,9,41–43] The properties of a mixture of two solvents need not be a linear function of its composition, as illustrated for water/dimethyl sulfoxide (DMSO) mixtures in Figure 9. Only the dielectric permittivity ε shows a linear behavior, while the other properties considered show different types of nonlinearity.[42]

Currently used biomolecular force fields treat the electronic polarization of the molecules in an average manner, which leads, however, to limited accuracy when systems under varying dielectric conditions are considered.[44–48] The limitation could be removed by the introduction of explicit polarizability into biomolecular force fields,[49,50] which requires, however, a more or less complete reparametrization of the force field, followed by extensive testing for realistic systems. Clearly, this is a formidable task. Recently, the first polarizable biomolecular force fields have been proposed.[51–53] Their performance in solvation free energy calculations or in reproducing folding equilibria has not yet been reported, and should be investigated.

To be able to efficiently simulate large biomolecular systems and slow processes, such as membrane or micelle formation, it would be helpful to formulate so-called coarse-grained molecular models, in which a number of covalently bound atoms are treated as a single particle or bead.[54–59] Such models can be made orders of magnitude faster in simulations than atomic models, at the expense of losing atomic detail. Models of this type have been successfully applied to membrane and micelle formation.[60] A comparison of the properties obtained from coarse-grained models with those from atomic models is required to evaluate the effect of the approximations and simplifications made.

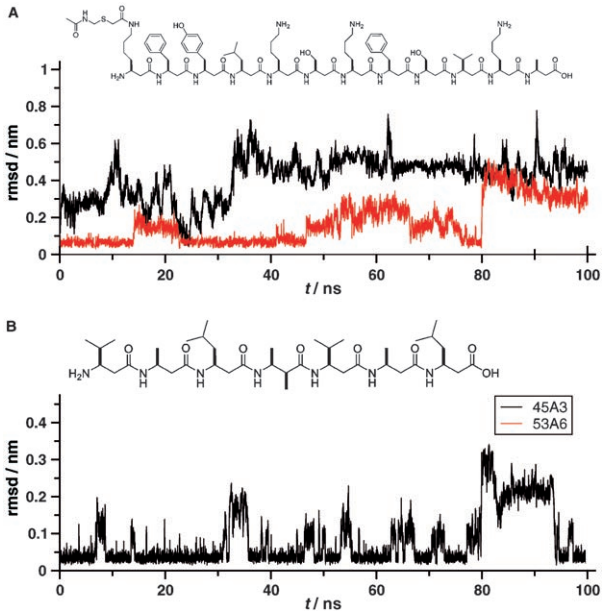


Figure 8. Root-mean-square deviation (rmsd) of the positions of backbone atoms in MD trajectory structures from the helical model structures derived from NMR data for two β-peptides in methanol. A) The peptide containing polar side chains only shows the experimental fold with the newer force-field parameter set 53A6.[36] B) The other peptide is equally well folded by using the old (45A3) and the new (53A6) force fields, and only data for the former are shown.

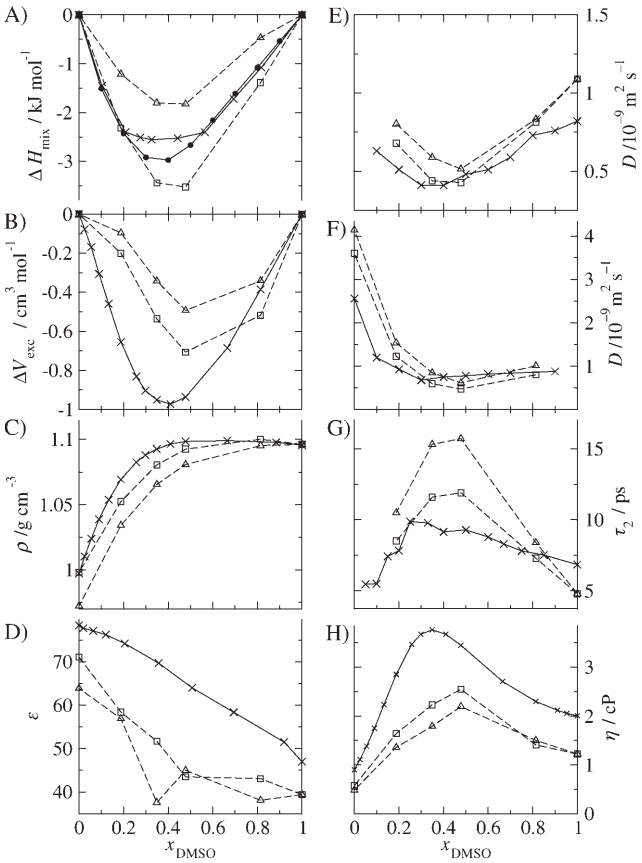


Figure 9. Properties of water/DMSO mixtures at 298 K and 1 atm as a function of the mole fraction of DMSO (x_{DMSO}) from MD simulations.[42] ΔH_{mix}: mixing enthalpy, ΔV_{exc}: excess volume, ρ: density, ε: relative dielectric permittivity, D: diffusion coefficient (panel E: DMSO, panel F: water), τ_2: rotational correlation time (DMSO), η: shear viscosity. Values from simulation: △ SPC water model, □ SPC/L water model; experimental values: × and ●.

3. The Search Problem

A biomolecular system is generally characterized by a very large number of degrees of freedom (10^4–10^6 or more). The motions along these degrees of freedom show a variety of characteristics, from highly harmonic to anharmonic, chaotic, and diffusive. Moreover, correlations are present that cover a wide range of time and spatial scales, from femtoseconds and tenths of nanometers to milliseconds and micrometers. The energy hypersurface of such a system, which is defined by the potential-energy function [e.g. Eqs. (1)–(8)], is therefore a very rugged surface, with energy basins and mountains of a wide range of depths or heights and spatial extent. This makes the search for the global energy minimum of such a high-dimensional function—or rather the search for those regions of the surface that contribute most to the free energy of the system—a daunting if not impossible task.

As mentioned before, the state of a biomolecular system cannot be described by a single global minimum energy configuration or structure, but only by a statistical-mechanical ensemble of configurations, in which the weight of a configuration x is given by the Boltzmann factor [Eq. (11); k_B is the Boltzmann constant and T the (absolute) temperature].

$$P(x) \sim \exp\left(-V(x)/k_{\mathrm{B}}\,T\right) \tag{11}$$

The exponential weighting in Equation (11) implies that high-energy regions of the energy hypersurface will not contribute configurations that are relevant to the state of the system, unless they are very numerous (entropy). The equilibrium properties of the system are dominated by those parts of configuration space, for which $V(x)$ is low. Therefore, one of the fundamental challenges to biomolecular modeling is to develop methodology to efficiently search the vast biomolecular energy surface for regions of low energy.

Below, we only mention and classify the major techniques that are currently used to search and sample configuration space.[61–63] There are also search and sampling techniques in which not only the molecular coordinates x serve as variables, but also their Boltzmann probabilities $P(x)$. A discussion and examples of these so-called probability search techniques can be found in Refs. [61, 64, 65]

3.1. Methods to Search and Sample Configuration Space

A variety of search methods is available, each with its own particular strengths and weaknesses, which depend on the form of the function $V(x)$ and the number and types of degrees of freedom in the system. Two basic types of search methods can be distinguished: systematic and heuristic.

Systematic or exhaustive search methods scan the complete or a significant fraction of the configuration space of the

biomolecular system. Particular subspaces can be excluded from the search without reduction in the quality of the solution found by applying rigorous arguments that mean that these subspaces cannot contain the desired solution.[66] Such arguments are based on a priori knowledge, often of physical or chemical nature, about the structure of the space or energy hypersurface to be searched. Systematic search techniques can only be applied to small molecules with only a few degrees of freedom,[67–70] because of the exponential growth of the required computing effort as the number of degrees of freedom included in the search increases.

Heuristic search methods, although visiting only a tiny fraction of the configuration space, aim at generating a representative (in the Boltzmann weighted sense) set of system configurations. These methods may generally be divided into three types (see also Table 7:

1. Nonstep methods, in which a series of system configurations is generated, which are independent of each other. An example is the so-called distance-geometry metric-matrix method,[72,73] which generates, at least in principle, an uncorrelated series of random configurations for a search problem that can be cast into a distance-based form.
2. Step methods that build a complete molecular or system configuration from configurations of fragments of the molecule or system in a step-wise manner. Examples are the build-up procedure of Gibson and Scheraga,[74,75] combinatorial build-up methods that make use of dynamic programming techniques,[76] and Monte Carlo (MC) chain-growing methods,[77,78] such as the so-called configurational bias Monte Carlo (CBMC) technique.[79]
3. Step methods, such as energy minimization (EM), Metropolis Monte Carlo (MC), molecular dynamics, and stochastic dynamics (SD),[80] that generate a new configuration of the complete system from the previous configuration. These methods can be further classified according to the way in which the step direction and step size are chosen (see Table 6). Energy minimization can be based on only energy values and random steps (simplex methods), or on energy and energy gradient values (steepest-descent and conjugate-gradient methods), or on second-order derivatives of the energy (Hessean matrix methods). In MC methods the step direction is taken at random, and the step size is limited by the Boltzmann acceptance criterion: when the energy of the system changes by $\Delta V < 0$, the step in configuration space is accepted, while for $\Delta V > 0$, the step is accepted with probability $\exp(-\Delta V/k_B T)$. In MD simulation the step is determined by the force (the negative of the local gradient $\partial V/\partial x$) and the inertia of the degrees of freedom, which serves as a short-time memory of the path followed so far. In SD simulations a random component is added to the force, the size of which is determined by the temperature of the system and the atomic masses and friction coefficients. In the potential-energy contour tracing (PECT) algorithm[81] and in the potential-energy annealing conformational search (PEACS) algorithm[71] the energy values are monitored and kept constant (PECT) or annealed (PEACS) to locate saddle points and pass over these. There exists a large variety of search procedures based on stepping through configuration space using a combination of the five mentioned basic elements (energy, gradient, Hessean, memory, and randomness) combined in one way or another.[61]

Table 6: Heuristic methods to search configuration space for configurations x with low energy $V(x)$.[a]

Reason for the change	Method				
	EM	MC	MD	SD	PEACS
energy	yes	yes	no	no	yes
energy gradient	yes	no	yes	yes	yes
second derevative of the energy	yes	no	no	no	no
memory	no	no	yes	yes	yes
randomness	yes	yes	no	yes	no

[a] EM: energy minimization, MC: Monte Carlo, MD: molecular dynamics, SD: stochastic dynamics, PEACS: potential energy annealing conformational search.[71]

The efficacy of search methods for biomolecular systems is severely restricted by the nature of the energy hypersurface $V(x)$ that is to be explored to find low-energy regions. The occurrence of a multitude of high-energy barriers between local minima means that the radius of convergence of the step methods is generally very small. Therefore, a variety of techniques has been developed to enhance the search and sampling power of searching methods. Three general types of search and sampling enhancement techniques are distinguished in Table 7.[61]

Table 7: Techniques to enhance the searching and sampling power of simulation methods.[a]

1. Deformation or smoothening of the potential-energy surface
a) omission of high-resolution structure factor data in structure refinement based on X-ray diffraction data
b) gradual introduction of longer range distance bounds - in structure refinement based on NOE data[82]
c) softening of the hard core of atoms in the nonbonding interaction ("soft-core" atoms)[83]
d) reduction of the ruggedness of the energy surface through a diffusion-equation type of scaling[83,84]
e) avoiding the repeated sampling of an energy well through local potential-energy elevation or conformational flooding[85,86]
f) softening of geometric restraints derived from experimental data (NMR, X-ray) through time averaging[87,88]
g) circumvention of energy barriers through an extension of the dimensionality of the Cartesian space (4D-MD)[89]
h) freezing of high-frequency degrees of freedom through the use of constraints[90]
i) coarse-graining the model by reduction of the number of interaction sites[54–59]

2. scaling the system parameters
a) temperature annealing[91]
b) mass scaling[92]
c) mean-field approaches[93]

3. multicopy searching and sampling
a) genetic algorithms[94]
b) replica-exchange and multicanonical algorithms[62]
c) cooperative search: SWARM [95]

[a] For details see Refs. [60–62] and references therein.

3.1.1. Deformation or Smoothening of the Potential-Energy Hypersurface to Reduce Barriers

a) Generally, a smoothening of the potential-energy function $V(x)$ allows for a faster search for its minima. This technique has been applied to different problems, such as structure determination based on X-ray diffraction or NMR spectroscopic data, conformational search, and protein-structure prediction. In method Ia of Table 7, the electron density of a biomolecular crystal is smoothed by the omission of high-resolution diffraction intensities when back calculating the electron density from these through Fourier transforms. This smoothening enhances the radius of convergence of the structure refinement.

b) When building a protein structure from atom–atom distance data obtained from NMR spectroscopy, the convergence of the configurational search process is enhanced by gradually introducing distance restraints that connect atoms at longer distances along the polypeptide chain in the potential-energy function. This is called a variable-target function method.[82]

c) The hard core of atoms, that is, the strong repulsive interaction between overlapping atoms, is responsible for many barriers on the energy hypersurface of a molecular system. These barriers can be removed by making the repulsive short-range interactions between atoms "soft".[96–99] Atoms with soft cores smooth the energy surface and lead to strongly enhanced sampling.[83]

d) In the deformation methods based on the diffusion equation,[83,84] the deformation of the energy surface during a simulation is made proportional to the local curvature (second derivative) of the surface, which leads to a preferential smoothening of the sharpest peaks and valleys in the surface and a very efficient search.

e) Incorporation of information on the energy hypersurface obtained during the search into the potential-energy function is another possibility to enhance sampling. Once a local energy minimum is found, it is removed from the energy surface by a suitable local deformation of the potential-energy function. This idea is the basis of the deflation method,[100] the local-elevation search method,[85] recently also called meta-dynamics[101], and the method of conformational flooding.[86]

f) Another way to introduce a memory into the search is the use of a potential-energy term which uses a running average of a coordinate over the trajectory or ensemble generated so far rather than its instantaneous value.[102] Application of this type of time-dependent or ensemble-dependent restraints in protein-structure determination based on NMR spectroscopic or X-ray data leads to a much enhanced sampling of the molecular configuration space.[87,88]

g) Barriers in the energy hypersurface can be circumvented by an extension of the dimensionality of the configuration space beyond the three Cartesian ones. The technique of energy embedding locates a low-energy conformation in a high-dimensional Cartesian space and gradually projects this conformation to three-dimensional Cartesian space while perturbing its energy and configuration as little as possible.[103] Variations on the original procedure have been proposed.[104–107] Dynamic search methods can also be used in conjunction with an extension of the dimensionality. Energy barriers in three-dimensional space can be circumvented by performing MD simulations in four-dimensional Cartesian space,[89] and free-energy changes can be calculated.[108]

h) A long-used standard technique to smooth the energy surface is to freeze the highest-frequency degrees of freedom of a system through the application of constraints.[90,109–113] Bond-length constraints are applied as standard in biomolecular simulation and allow for a four times longer time step.[114] High-frequency motion can also be eliminated by using soft constraints:[115] the (bond-) constraint lengths change adiabatically as a result of the forces.

3.1.2. Scaling of System Parameters To Enhance Sampling

a) The technique of simulated temperature annealing[91] involves simulation or search at a high temperature T, followed by gradually cooling the system. By raising the temperature, the system may more easily surmount energy barriers, so a larger part of configurational space can be searched. The technique of simulated temperature annealing has been widely used in combination with MC, MD, and SD simulations. An example of potential-energy annealing can be found in Ref. [71].

b) Scaling of atomic masses can be used to enhance sampling. In the classical partition function and in case no constraints are applied, the integration over the atomic momenta can be carried out analytically, separately from the integration over the coordinates. Thus, the atomic masses do not appear in the configurational integral, which means that the equilibrium (excess) properties of the system are independent of the atomic masses. This freedom can be exploited in different ways to enhance the sampling. By increasing the mass of specific parts of a molecule, its relative inertia is enhanced, which eases the surmounting of energy barriers,[92] and may allow for larger time steps.

c) Enhanced sampling by a mean-field approximation is obtained by separating the biomolecular system into two parts (A and B), each of which moves in the average field of the other. The initial configuration of the system consists of N_A identical copies of part A and N_B identical copies of part B. The positions of corresponding atoms in the identical copies may be chosen to be identical. The force on atoms of each copy of part A exerted by the atoms in all copies of part B is scaled by a factor $N_B{}^{-1}$ to obtain the mean force exerted by part B on the individual atoms of part A. Likewise, the force on atoms of each copy of part B exerted by the atoms in all copies of part A is scaled by a factor $N_A{}^{-1}$. The forces between different copies of part A are zero, and so are the forces between different copies of part B. The MD simulation involves the integration of Newton's Equation of motion, $\boldsymbol{f} = m\boldsymbol{a}$, for all copies of parts A and B simultaneously. Thus, one obtains N_A individual trajectories of part A in the mean

field of part B and vice versa. This situation comes at the loss of correct dynamics: Newton's third law, $\boldsymbol{f}_{AB} = -\boldsymbol{f}_{BA}$ is violated. The technique only enhances efficiency when the system is partitioned into parts of very different sizes (for example, $A \ll B$) and the bigger part is represented by one copy: $N_B = 1$. Enhanced searching and sampling procedures based on a mean-field approximation have been proposed in different forms,[93,116–127] and have been applied to the diffusion of CO molecules in the field of a protein molecule,[93,118] to the conformational equilibrium of protein side chains,[117] to the determination of protein-loop conformations,[123] and to the search for the minimum-energy conformations of polypeptides[126,127] and nucleic acid segments[128] as well as to the search for binding sites in enzymes.[129–131]

3.1.3. Multicopy Simulation with a Given Relationship between the Copies

In the mean-field approach, multiple copies of a part of the system were simulated. This idea has also been used in other ways to enhance searching and sampling (Table 7).

a) In genetic algorithms[132] a pool of copies of the biomolecular system in different configurations is considered, and new configurations are created and existing ones deleted by mutating and combining (parts of) configurations according to a given set of rules.
b) In the so-called replica-exchange algorithm multiple copies of the system are simulated by MC, MD, or SD, each at a distinct temperature. From time to time copies from simulations close in temperature are exchanged through an exchange probability based on the Boltzmann factor [Eq. (11)]. This leads, within the limit of infinite sampling, to Boltzmann-distributed (canonical) ensembles for each temperature.[133] So-called multicanonical algorithms are a generalization of this procedure.[62] These types of algorithms have been used to simulate proteins in vacuo.[133] The inclusion of solvent degrees of freedom may impair the efficiency of the algorithm.[134] Dynamic information is lost in the exchanges and for short sampling times the entropy content is likely to be biased at the lower and upper ends of the temperature range considered.
c) The so-called SWARM type of MD[95] is based on the idea of combining a collection (or swarm) of copies of the system each with its own trajectory into a cooperative multicopy system that searches configurational space. To build such a cooperative multicopy system, each copy is, in addition to physical forces arising from $V(x)$, subject to (artificial) forces that drive the trajectory of each copy towards an average of the trajectories of the swarm of copies. This effect is analogous to the intelligent and efficient behavior of a whole swarm of insects which can be achieved even in the absence of any particular intelligence or forethought of the individuals. SWARM-MD is less attracted by local minima and is more likely to follow an overall energy gradient toward the global energy minimum.

This overview of methods to search and sample configurational space is rather limited. More extensive reviews can be found in Refs. [61–63] Since biomolecular configurational space is too large to be exhaustively sampled, one generally has to use heuristic search methods in biomolecular modeling studies. The overview (Tables 6 and 7) of types of methods and tricks that can be used and combined to obtain a powerful search method may assist in the choosing of a combination of methods and tricks that will be particularly suited to the specific problem or energy hypersurface of interest.

3.2. Convergence of Simulated Properties

The time scales involved in the dynamics of different properties of biomolecular systems range from femtoseconds to seconds or even longer. Limited computing power means that current MD simulations of biomolecular systems cover nanoseconds to tens or hundreds of nanoseconds, depending on the system size. This poses the question as to whether such time periods are long enough to yield reliable trajectory averages for the different molecular or system properties. Trajectory averages will generally only be representative when the equilibration period of a simulation τ_{equil} is longer than the relaxation time $\tau_{relax}(Q)$ of the property Q [Eq. (12)] and when the sampling period τ_{sample} is much longer than $\tau_{relax}(Q)$ [Eq. (13)].

$$\tau_{equil} > \tau_{relax}(Q) \tag{12}$$

$$\tau_{sample} \gg \tau_{relax}(Q) \tag{13}$$

If conditions (12) or (13) are not fulfilled, the trajectory average $\langle Q(t)\rangle_t$ of the property Q will display a drift with time or erratic behavior.[135–137]

The time scale associated with the change or relaxation of a particular physical quantity calculated for a particular biomolecular system will depend on 1) the type of molecular system, 2) the thermodynamic state point, and 3) the particular quantity or property. This relationship is illustrated in Figure 10 for the dynamics of β-heptapeptides in methanol solution. At 298 K the dominant conformer is the 3_{14}-L helix, which completely unfolds and refolds only a few times within 100 ns (panel A). At 340 K many more (un)folding events are observed (panel B). Reducing the solvent viscosity to 1/3 (Panel C) or to 1/10 (Panel D) enhances the rate of the (un)folding process considerably, and leads to improved convergence of the statistics for the (un)folding equilibrium. Panel E illustrates that the presence of longer and charged side chains in the polypeptide solute slows down the (un)folding process, thereby reducing the sampling (compare with panel B).

Not only can different molecules or thermodynamic state points show different relaxation times, but different properties also can do so. The potential energy of the solute and the square of its total dipole moment (related to its dielectric permittivity) relax faster than, for example, the average atom-positional root-mean-square fluctuations for all atoms.[140,141] System properties, such as the free energy of folding,

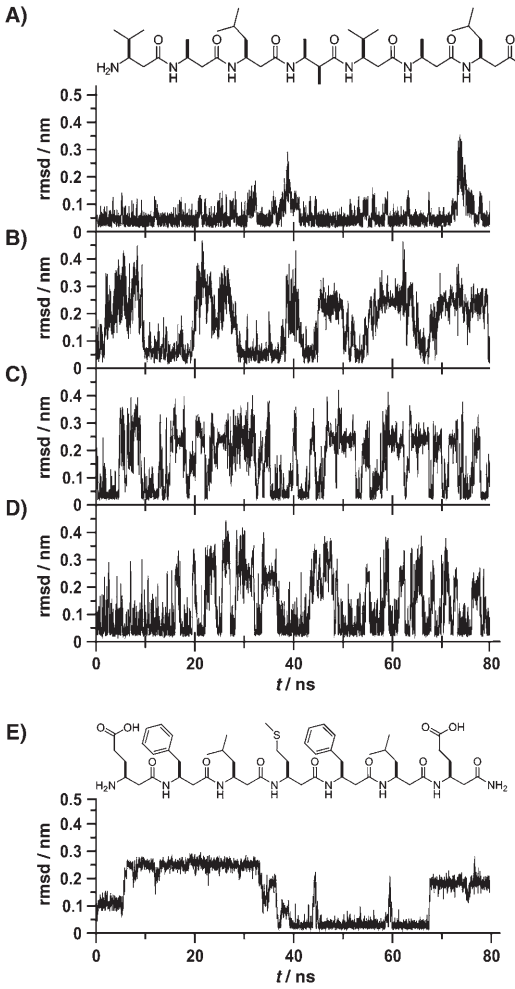


Figure 10. Root-mean-square deviation of the positions of backbone atoms in MD trajectory structures from the helical model structures derived from NMR data for two β-heptapeptides of identical chain lengths in methanol at 1 atm. The peptide with apolar side chains is simulated at 298 K (A) and at 340 K (B–D).[32,138] The viscosity of the methanol solvent is reduced by a factor of 3 (C) and by a factor of 10 (D) through mass scaling. Raising the temperature or reducing the solvent viscosity increases the rate of (un)folding. The peptide with a few polar side chains is simulated at 340 K in normal methanol (E). The polar side chains reduce the rate of (un)folding.[139]

converge even slower—in general more slowly than molecular properties.[1,142]

The relaxation and dynamics of the various properties of a biomolecular system can be analyzed in different ways:[135–137]

1. For equilibrium simulations one may monitor the time series of a property $Q(t)$, its average value $\langle Q(t)\rangle_t$, or fluctuations $\langle(Q(t)-\langle Q(t)\rangle_t)^2\rangle_t^{1/2}$, or calculate its autocorrelation function $\langle Q(t')Q(t'+t)\rangle_{t'}$. The decay time of the autocorrelation function or the build-up rate of the trajectory averages gives an indication of the magnitude of $\tau_{relax}(Q)$.[137]
2. When starting a simulation from a non-equilibrium initial state, the rate of relaxation of $Q(t)$ toward equilibrium, measured over many non-equilibrium trajectories, will give an indication of $\tau_{relax}(Q)$.
3. If different MD simulations starting from different initial states do not converge to the same trajectory average for property Q, it can be concluded that $\tau_{relax}(Q)$ is longer than the simulation period.[143]

Further examples of different relaxation times of properties in various systems can be found in Refs. [135–137, 140–143].

3.3. Alleviation of the Search and Sampling Problems

Although the search and sampling problem with regard to biomolecular systems looks formidable at first sight, the characteristics of the energy hypersurface may alleviate these problems. Simulations of (un)folding equilibria of polypeptides in solution using a thermodynamically calibrated force field and an explicit representation of the solvent molecules have shown that the unfolded or denatured state of these polypeptides contains much fewer conformations significantly populated at equlilibrium than there are possible polypeptide conformations. In the case of peptides possessing 20 rotatable torsional angles in their backbone, the use of physical, realistic force fields representing the particular (nonbonding) interactions between the various residues results in a reduction from 10^9 possible conformers to 10^3 relevant conformers.[144–146] This is illustrated in Figure 11, where the number of conformations visited during a MD simulation of a polypeptide and of another polymer of equal length are shown. This number grows sublinearly for the β-heptapeptide in methanol and levels off at about 200 conformations (Figure 11 A), whereas the number of visited (relevant) conformations for a poly(hydroxybutanoate) molecule of similar length in chloroform grows linearly with time (Figure 11 B), as would be expected considering the number of possible conformations for either molecule is about 10^9. The difference is due to the presence of hydrogen-bond donor and acceptor atoms in the β-heptapeptide, which restrict (through favorable hydrogen bonding) the conformational space accessible to the molecule

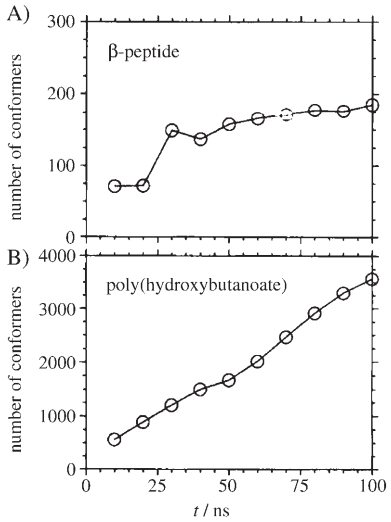


Figure 11. Number of conformations as a function of time from MD stimulation: A) A β-heptapeptide in methanol at 340 K;[138] b) $(Val\text{-}Ala\text{-}Leu)_2$-3-hydroxybutanoate in chloroform at 298 K.[147] For the definition of a conformation (cluster of structures) we refer to Ref. [138].

at the given temperature. In the absence of hydrogen-bond donors in poly(hydroxybutanoate), this restriction is not present.[147]

When a realistic force field is used, most of the configuration space will have a very high energy, which indicates that the configuration space to be searched or sampled to predict the most stable fold of a polypeptide or protein does not grow exponentially with the system size or chain length.[144]

3.4. Aggravation of the Search and Sampling Problems

The free energy F of a system of N particles in a volume V at temperature T is a $6N$-dimensional integral over all particle coordinates $\boldsymbol{r}$ and momenta $\boldsymbol{p}$ of the Boltzmann factor of the system Hamiltonian (kinetic plus potential energy) [Eq. (14); h is Planck's constant)].

$$F_{NVT} = -k_B T \ln\left((N!\,h^{3N})^{-1} \int\int \exp(-H(\boldsymbol{p},\boldsymbol{r})/k_B T)\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{r}\right) \quad (14)$$

The factor $N!$ must be omitted when dealing with distinguishable particles. Since the integrand—the Boltzmann factor—is everywhere positive, the omission of configurations in the integral leads to systematic (not canceling) errors. Equation (14) shows that not only will the global minimum energy structure or configuration determine the free energy or be representative for the configurational ensemble, but other configurations of higher energy and greater abundance will also do so. In other words, both energy U and entropy S contribute to the free energy [Eq. (15)].

$$F = U - TS \quad (15)$$

Solvent degrees of freedom may contribute significantly to the free energy of folding.[148] From a 200-ns MD simulation of the (un)folding equilibrium of a β-heptapeptide in methanol, the differences between the enthalpy H_{solute} and entropy S_{solute} of the peptide in the folded conformation and in the unfolded conformations could be calculated as $\Delta H^{folding}_{solute} = -64\ kJ\,mol^{-1}$ and $T\Delta S^{folding}_{solute} = -157\ kJ\,mol^{-1}$, which yields $\Delta G^{folding}_{solute} = +93\ kJ\,mol^{-1}$. However, the Gibbs free energy of folding as calculated for the whole system (peptide plus solvent) from the ratio between the folded and unfolded conformations appeared to be $\Delta G^{folding} = -8\ kJ\,mol^{-1}$. Thus, changes in solute free energy alone cannot explain the observed folding behavior. This underlines the important role of the solvent in peptide folding and that entropy calculations including solvent degrees of freedom are needed. Unfortunately, extensive sampling of solvent degrees of freedom aggravates the sampling problem.

Figure 12 illustrates another example of the observation that an ensemble of relevant conformations needs to be sampled to obtain accurate estimates of the free energy. It shows a superposition of the configurations of a protein–ligand complex that contribute most to the free energy of binding of the biphenyl ligand to the ligand-binding domain of the estrogen receptor. The configurations show sizeable variation, and choosing only one of them to represent the ensemble would lead to an inaccurate estimate of the binding free energy.[149] In contrast to the assumptions of many standard ligand–protein docking algorithms, this example illustrates that inclusion of protein degrees of freedom in the sampling is probably necessary to obtain accurate results. This, unfortunately, aggravates the search and sampling problem of docking algorithms.

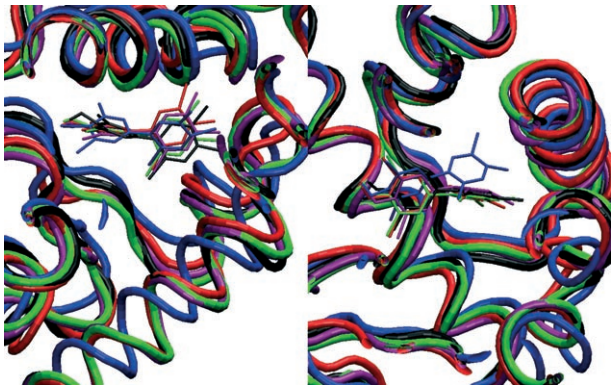


Figure 12. Calculation of the free energy of binding of 16 hydroxylated polychlorinated biphenyls to the ligand-binding domain of the estrogen receptor using MD simulation and the one-step perturbation technique.[149] Five protein–ligand structures that contribute most to the free energy of binding of a particular ligand are shown.

A last example of the aggravation of the sampling problem is the dependence of the magnitude of the hydrophobic effect on the size of a hydrophobic cluster and the composition of the solvent. Understanding the hydrophobic effect at the molecular level will help to understand the driving forces for protein folding in which this effect is thought to play an essential role. Figure 13 shows the free energy of cluster growth for clusters with 2 to 46 methane

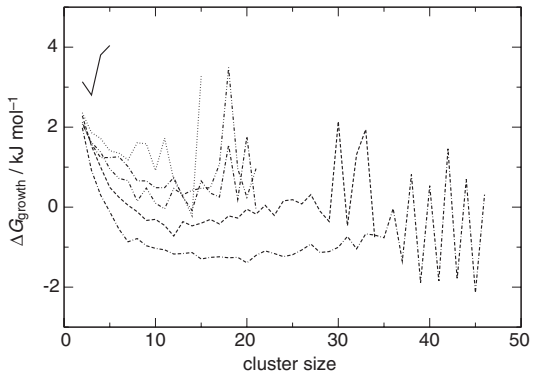


Figure 13. Gibbs free energy of methane cluster growth as a function of the cluster size for MD simulations of different methane/urea/water mixtures.[150] N_m is the number of methane molecules, N_u the number of urea molecules, and N_w the number of water molecules. Solid line: $N_m = 10$, $N_u = 0$, $N_w = 990$; dotted line: $N_m = 40$, $N_u = 0$, $N_w = 960$; dot-dashed line: $N_m = 50$, $N_u = 150$, $N_w = 800$; double dot-dashed line: $N_m = 50$, $N_u = 250$, $N_w = 700$; dashed line: $N_m = 50$, $N_u = 0$, $N_w = 950$; dot-double dashed line: $N_m = 64$, $N_u = 250$, $N_w = 686$; Large fluctuations at the end of the curves stem from poor statistics for large cluster sizes.

molecules in urea/water mixtures at different urea concentrations.[150] Methane aggregation and cluster formation only becomes favorable at cluster sizes of at least five methane molecules. A second observation is that high urea concentrations result in slightly enhanced clustering of methane molecules, rather than in a reduction of the hydrophobic interactions. This result hints against a mechanism of protein denaturation by urea through a weakening of hydrophobic interactions.[151] The sampling problem is aggravated by the observed dependence of the free energy on the cluster size and urea concentration, because both these degrees of freedom need to be varied to obtain meaningful results.

3.5. Perspectives Regarding the Search and Sampling Problem

There is certainly still room to enhance the search and sampling efficiency of biomolecular simulation techniques; however, the past ten years have already shown encouraging progress that we expect to be of benefit also for the study of larger, more interesting and relevant biomolecular systems. In particular the technique of smoothening the potential-energy surface can enhance sampling through the use of soft-core atoms, local-elevation, and diffusion-equation types of deformation of the energy surface, and on another level through the formulation of coarse-grained models. The efficiency of local-elevation types of sampling[85] of the conformational space of a dipeptide in aqueous solution is illustrated in Figure 14. A much larger part of the Ramachandran map is covered in a much shorter simulation time than when using a standard MD simulation.

The so-called single-step perturbation methodology allows ligand-binding free energies or solvation free energies

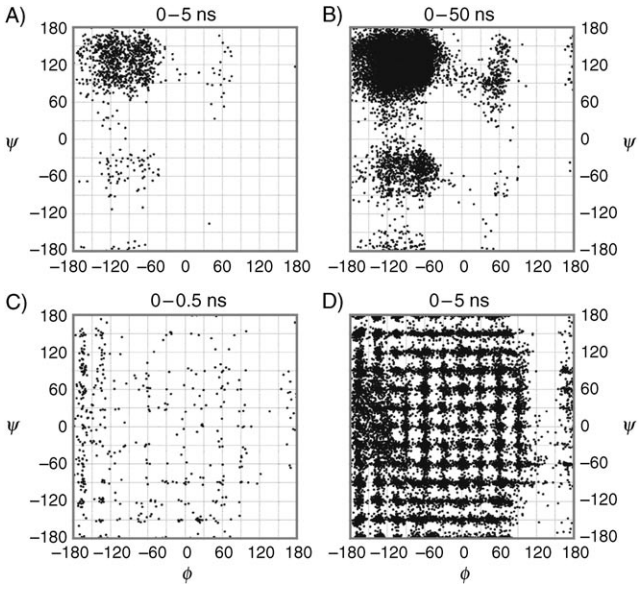


Figure 14. Ramachandran maps obtained from MD simulations of an alanine dipeptide in (SPC/E) water. φ/ψ distributions obtained by standard MD for 5 ns (A) and 50 ns (B), and by local-elevation (LE) MD for 0.5 ns (C) and 5 ns (D),[85] showing the much more rapid sampling of the LE-MD algorithm. Upon revisiting a conformation the LE potential energy is raised by 2.0 kJ mol^{-1}.

to be obtained for a great many compounds. This method uses only a few simulations involving nonphysical reference states with soft-core atoms chosen to widen the configurational ensemble and offers orders of magnitude gains in efficiency compared to standard (thermodynamic integration or perturbation) free energy calculations.[149,152–156] This is illustrated in Figure 15, where binding (Gibbs) free energies of 16 hydroxylated polychlorinated biphenyls to the estrogen receptor as

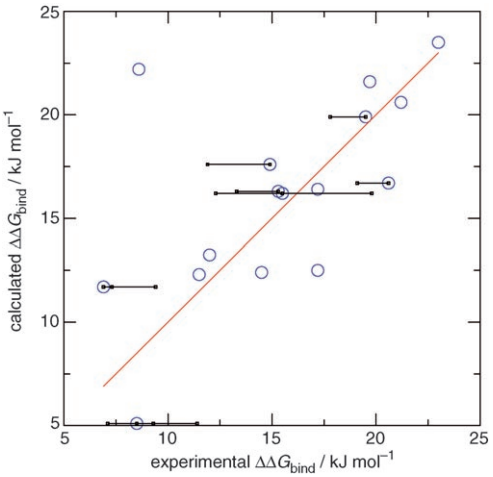


Figure 15. Experimental versus calculated relative Gibbs free energies of binding for 16 hydroxylated polychlorinated biphenyls to the estrogen receptor binding domain in solution.[149] The horizontal lines connect different experimental values for one compound. The calculated values were obtained from only two simulations and the one-step perturbation method. The GROMOS 43 A1 force field was used.

calculated using the single-step perturbation technique from only two MD simulations are compared to the corresponding experimental values.[149] The average deviation of the simulated values from the experimental ones is, at 2.5 kJ mol^{-1}, smaller than the variation of 4.2 kJ mol^{-1} in the experimental values themselves. Thus, the force field used and the sampling technique based on soft-core atoms are able to reproduce the experimental values within the accuracy of their measurement.

A reduction of the solvent viscosity (see Figure 10) may also lead to considerably enhanced sampling without affecting the nondynamic equilibrium properties of the system.

	number of amino acids in the protein	folding time (exptl/sim) in seconds	number of *possible* structures	number of *relevant* (observed) structures
peptide	10	10^{-8}	$3^{20} \approx 10^{9}$	10^{3}
protein	100	10^{-2}	$3^{200} \approx 10^{90}$	10^{9}

Assuming that the number of relevant unfolded structures is proportional to the folding time, only 10^9 protein structures need to be simulated instead of 10^{90} structures.

⇒ Folding mechanism is simpler than generally expected: searching through only 10^9 structures

⇒ Protein folding on a computer is possible before 2010

Figure 16. A surprising result after the simulation of many polypeptides: The number of unfolded conformations visited in MD simulations of (un)folding equilibria of a host of polypeptides and peptoids is much smaller than theoretically possible.[144–146]

The observation that the unfolded state of polypeptides contains much fewer relevant conformations than possible conformations opens up the possibility to simulate the reversible (un)folding of small proteins within not too many years (Figure 16).

4. The Ensemble Problem

Biomolecular modeling is hindered by the fact that the behavior of biomolecular systems is governed by statistical mechanics. If mechanics were applicable only, one could characterize such systems in terms of (global) minimum-energy structures. Statistical mechanics leads to the concept of the entropy of a system, that is, the negative derivative of the free energy F with respect to temperature [Eq. (16)]:

$$S_{\mathrm{NVT}} = -\left(\frac{\partial F}{\partial T}\right)_{\mathrm{NV}} = (U-F)/T \tag{16}$$

The entropy together with the energy of a system [Eq. (17)], that is, the average of the Hamiltonian of the system over the momenta and coordinates of all degrees of freedom, determines the free energy [Eq. (14)] of the system.

$$U_{\mathrm{NVT}} = \left(\frac{\partial(F/T)}{\partial(1/T)}\right)_{\mathrm{NV}} = \langle H(\boldsymbol{p},\boldsymbol{r})\rangle_{\boldsymbol{p},\boldsymbol{r}} \tag{17}$$

The Boltzmann average $\langle\cdots\rangle_{\mathbf{p},\mathbf{r}}$ of a quantity $Q(\boldsymbol{p},\boldsymbol{r})$, which depends on the (atomic) coordinates and momenta, is defined by Equation (18):

$$\langle Q(\boldsymbol{p},\boldsymbol{r})\rangle_{\boldsymbol{p},\boldsymbol{r}} = \frac{\int\int Q(\boldsymbol{p},\boldsymbol{r})\exp[-H(\boldsymbol{p},\boldsymbol{r})/k_{\mathrm{B}}T]\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{r}}{\int\int \exp[-H(\boldsymbol{p},\boldsymbol{r})/k_{\mathrm{B}}T]\mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{r}} \tag{18}$$

The state of a system is generally characterized not by one configuration or structure, but by a Boltzmann ensemble of configurations or structures. This complicates biomolecular modeling, because it is easier to think of and handle single structures than to consider configurational ensembles. However, a number of (experimental) observations can only be understood by an analysis in terms of alternative structures present in an ensemble and in terms of entropy.

4.1. Free Energy, Energy, and Entropy of Solvation

Since the free energy F can be written as $F = U - TS$, different combinations of energy (U) and entropy (S) values may result in the same free energy. This can be illustrated through a calculation of the free energy of solvation ΔF_{S} of a solute in a binary solvent from MD simulations.[8] Figure 17 shows the Gibbs free energy ΔG_{S} (the equivalent of the quantity ΔF_{S} under constant pressure) of solvating a methane molecule in a mixture of water and a co-solvent (NaCl, DMSO and acetone) as a function of the co-solvent concentration or mole fraction. In NaCl, the value of ΔG_{S} increases as the co-solvent concentration increases, whereas in DMSO and acetone the value of ΔG_{S} decreases as the mole fraction of the co-solvent increases. The Gibbs free energy ΔG_{S} can be broken down into an energetic contribution, the change in the solute–solvent energy upon solvation (ΔU_{uv}), and an entropic contribution—the solute–solvent entropy change upon solvation ($-T\Delta S_{\mathrm{uv}}$) [Eq. (19)]:[8,157–159]

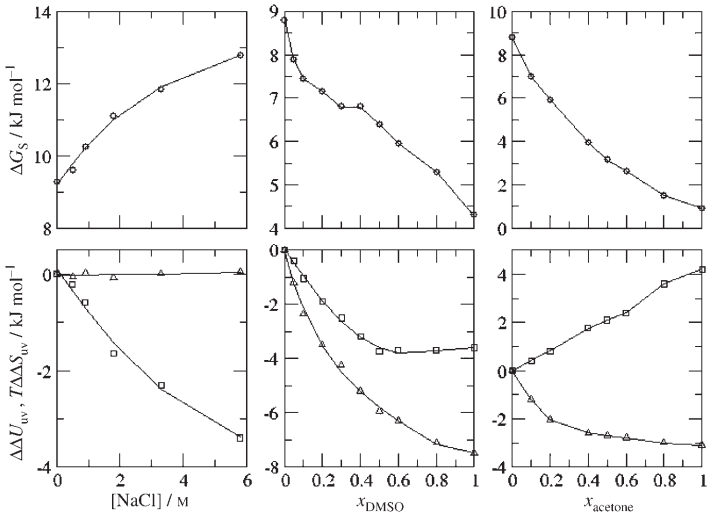


Figure 17. Gibbs free energy ΔG_s for the solvation of methane (upper panels) and solute–solvent energy $\Delta\Delta U_{uv}$ (triangles) and entropy $T\Delta\Delta S_{uv}$ (squares) relative to neat water (lower panels) as a function of the salt concentration NaCl (left panels), the mole fraction of DMSO (middle panels), and the mole fraction of acetone (right panels).[8] The calculations were performed by using the particle-insertion technique.

$$\Delta G_{\mathrm{S}} = \Delta U_{\mathrm{uv}} - T\Delta S_{\mathrm{uv}} \tag{19}$$

The lower panels in Figure 17 show these energetic and entropic contributions relative to neat water (indicated by $\Delta\Delta$) also as a function of the co-solvent concentration or mole fraction. The solvation of methane in aqueous solution is disfavored with increasing NaCl concentration as a result of an increasingly unfavorable solute–solvent entropy of solvation. On the other hand, solvation of methane in aqueous solution is favored both with increasing DMSO or acetone mole fraction, but the relative roles of solute–solvent energy and entropy are quite different, even though DMSO and acetone are structurally rather similar molecules. Solvation in DMSO is dominated by a favorable energy with increasing DMSO concentration, whereas solvation in acetone is dominated by a favorable entropy with increasing acetone concentration. Thus, comparable curves for ΔG_{S} are due to quite different solvation mechanisms. This example illustrates that entropy should be properly taken into account in biomolecular modeling studies.

The varying roles of energy and entropy in hydrophobic solvation is illustrated in Table 8 for a set of hydrophobic molecules in different aqueous solutions. Increasing the size of the hydrophobic solute in NaCl causes an increase in the $\Delta\Delta G_{\mathrm{S}}$ value as a consequence of an increase in $-T\Delta\Delta S_{\mathrm{uv}}$. In a

Table 8: Thermodynamic data [kJ mol^{-1}] for solute transfer from pure water to co-solvent/water mixtures at 298 K and 1 atm as calculated from MD simulations.[9] Mole fractions are given in percent. Two different models (I and II) were used for acetone.

Solute	NaCl (11%)			Urea (15%)			DMSO (10%)[a]			Acetone(I) (10%)			Acetone(I) (50%)			Acetone(II) (10%)		
	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$
helium	2.0	−0.4	−2.4	1.3	−0.1	−1.4	0.0	−0.2	−0.2	−0.4	−0.2	0.2	−1.7	−0.6	1.1	−0.1	−0.2	−0.1
neon	2.8	0.2	−2.6	1.1	−0.7	−1.8	−0.3	−0.7	−0.4	−0.7	−0.4	0.3	−2.4	−1.0	1.4	−0.4	−0.6	−0.2
argon	3.7	−0.1	−3.8	0.8	−2.3	−3.1	−0.9	−1.8	−0.9	−1.3	−1.0	0.3	−4.4	−2.1	2.3	−1.1	−1.7	−0.6
krypton	4.0	−0.2	−4.2	0.4	−3.2	−3.6	−1.2	−2.3	−1.1	−1.6	−1.2	0.4	−5.2	−2.6	2.6	−1.5	−2.3	−0.8
xenon	4.4	−1.2	−5.6	−0.5	−4.7	−4.2	−1.5	−2.7	−1.2	−2.2	−1.7	0.5	−6.8	−3.6	3.2	−2.1	−3.4	−1.3
methane	4.2	−0.2	−4.4	0.5	−3.1	−3.6	−1.1	−2.1	−1.0	−1.7	−1.2	0.5	−5.4	−2.6	2.8	−1.4	−2.2	−0.8
ethane	5.6	−0.8	−6.4	−0.9	−6.0	−5.1	−2.8	−4.4	−1.6	−2.9	−2.4	0.5	−8.4	−4.6	3.8	−2.6	−4.3	−1.7
propane	5.4	−0.3	−5.7	−2.2	−7.3	−5.1	−4.8	−5.8	−1.0	−5.8	−3.9	1.9	−13.0	−6.5	6.5	−4.8	−6.2	−1.4
n-butane	7.4	−1.2	−8.6	−2.7	−10.6	−7.9	−5.8	−7.9	−2.1	−6.4	−5.6	0.8	−15.7	−9.3	6.4	−4.3	−6.0	−1.7
iso-butane	5.7	−1.2	−6.9	−2.3	−10.2	−7.9	−6.5	−6.7	−0.2	−6.7	−4.9	1.8	−15.2	−7.7	7.5	−5.9	−7.7	−1.8
neo-pentane	8.2	−0.6	−8.8	−2.2	−10.9	−8.7	−6.0	−8.3	−2.3	−6.7	−4.3	2.4	−15.8	−7.1	8.7	−5.3	−9.7	−4.4

[a] To expedite statistical sampling, MD simulation runs were performed using an atomic mass of 15.9994 amu for the sulfur and water hydrogen atoms of DMSO. An MD time step of 4 fs was used.

urea/water mixture, $\Delta\Delta G_S$ is relatively independent of the solute size, because of energy–entropy compensation. In DMSO, $\Delta\Delta G_S$ becomes more favorable with increasing solute size, because of a strongly favorable energetic contribution, which is slightly counteracted by the entropy term $-T\Delta\Delta S_{uv}$. The picture again changes on changing from DMSO to acetone in that the entropy term now coacts with the solvation energy term, an effect that appears to increase with the acetone concentration. These results show that similar free energy changes, and thus the processes driven by them, can be due to very different mechanisms. A proper modeling of entropic effects is required to capture such molecular processes.

4.2. Temperature Dependence of Folding Equilibria

The folding-unfolding equilibrium of a polypeptide is not only determined by energy changes upon (un)folding, but also by entropy changes. This will make the corresponding conformational ensemble and folding pathways temperature-dependent. Figure 18 shows the most populated conformations of a β-heptapeptide in methanol at three different temperatures as obtained from MD simulations.[138] The quasivertical arrows between the rows indicate corresponding conformations at the different temperatures, while the horizontal arrows indicate the most dominant (un)folding pathways (from and) to the most stable (helical) conformation C1. Both the conformational ensemble and the (un)folding pathways are temperature-dependent, as expected, and this dependence can be investigated by MD simulation.

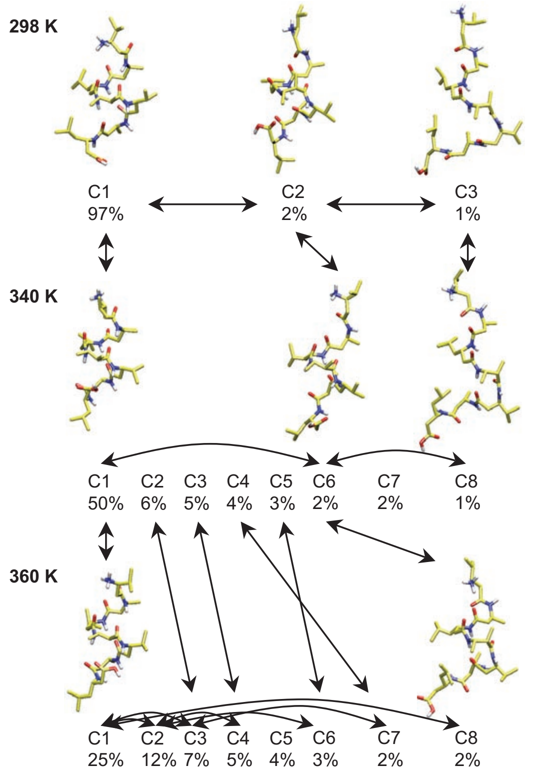


Figure 18. Most populated conformations of a β-heptapeptide in methanol from MD simulations at three different temperatures.[138] The quasi-vertical arrows indicate corresponding conformations at the different temperatures, while the horizontal arrows represent the dominant (un)folding pathways leading to the most stable (helical) conformation C1.

4.3. Different Conformations may Contribute to the Ensemble Averages

If different conformations are present in the ensemble of polypeptide conformations in solution, the measured average value of an observable, such as an NOE intensity or a ^3J-coupling constant, may not correspond to any realistic (that is, energetically accessible) conformation of the solute. This problem may be aggravated by a nonlinear dependence of the observable upon solute conformation. In such a case, structure determination based on NMR data will only lead to consistency with the experimentally measured observable values if conformational ensembles instead of single structures are considered. An example of such a case is the β-nonapeptide shown in Figure 19. Single-structure refinement based on 65 NOE intensities and 4 ^3J-coupling constants measured for the (unprotected) solute in methanol resulted in a 12/10-helical model structure. The hydrogen bonds charac-

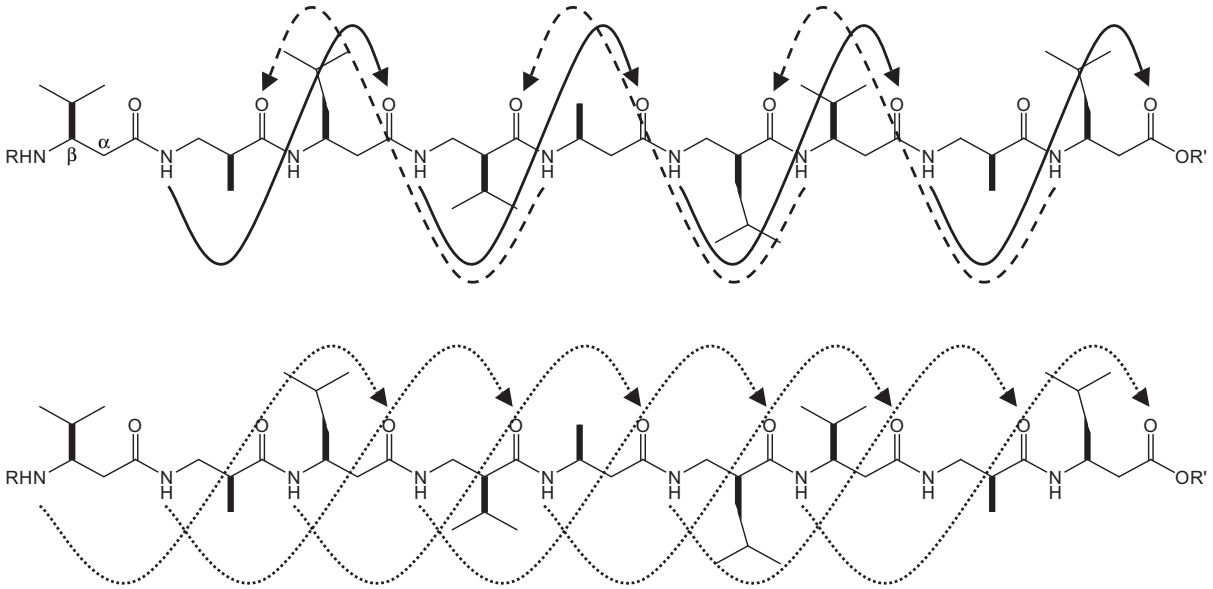


Figure 19. Structural formula of a β-nonapeptide with arrows indicating the hydrogen-bonding patterns characteristic of 12/10- and 3_{14}-helices. The 12/10-helix (top) is characterized by 10- (solid arrows) and 12-membered (dashed arrows) hydrogen-bonded rings, whereas the 3_{14}-helix (bottom) is characterized by 14-membered hydrogen-bonded rings (dotted arrows).

teristic for this type of helix are indicated in Figure 19 (dashed and solid arrows). However, three weak NOE intensities (2HN-5Hβ, 6HN-3Hβ, 7HN-4Hβ_{Re}) were omitted in the single-structure refinement, because of their incompatibility with the 12/10 helix.[160] A 100-ns MD simulation of the β-nonapeptide in methanol (without applying any NOE or ^3J-couping constant restraints) did essentially reproduce all measured NOE intensities and ^3J values.[161] Analysis of the MD trajectory showed that the three mentioned NOE intensities are due to a small percentage of an alternative 3_{14}-helical conformation (Figure 19, dotted arrows) in the conformational ensemble. This example shows that unrestrained MD simulation using a consistent, thermodynamically calibrated force field for both solute and solvent and including solvent degrees of freedom explicitly can contribute significantly to a correct interpretation of experimental data in terms of conformational distributions.

4.4. Perspectives in Calculating Entropies

While the determination of free-energy differences by MD simulation has become a standard procedure, for which a variety of techniques have been developed, total entropies and entropy differences are still seldom computed. However, entropy is the key property for understanding phenomena such as hydrophobic interactions, solvation, ligand binding, etc. Unfortunately, determining absolute entropies and entropy differences from MD simulation is not an easy task: it requires in principle the complete sampling of phase space. Generally, one can distinguish two types of methods to obtain reasonable estimates for entropies from MD simulations.

One type of methods focuses on conformational entropies, which consider not all, but only internal (conformational) nondiffusive degrees of freedom,[162–165] for example, in a protein.[166]

The other type of methods extends techniques that are successfully used to estimate free-energy differences from the calculation of entropy differences. To obtain free-energy differences between two states of a system or between two systems, the evaluation of the complete partition functions is not really needed. It is sufficient to extensively sample the relevant parts of phase or configuration space where the two states or systems differ. In contrast, the corresponding techniques to obtain entropy differences suffer from the fact that they require an accurate estimate of an ensemble average that includes the complete Hamiltonian operator H of the system in the two states, not only the part of the Hamiltonian operator that differs between the two states or systems ($\partial H/\partial\lambda$; see Figure 20).[142] The complete Hamiltonian operator is a sum over very many terms, of which only a few differ between the two states of the system. It therefore takes a very long time to obtain a precise ensemble. The most accurate methods are methods 3 and 4 of Figure 20. Method 4 yields only accurate solute–solvent entropies, not solvation entropies, which involve all solvent terms. This methodology was used to obtain the data shown in Table 8 and Figure 17.

The first type of methods is only suitable for estimating solute conformational entropies; it is of little help for diffusive systems such as solutions. The approaches used in the second type of methods (Figure 20) have recently been reviewed and evaluated.[142] Despite the progress made in developing methods, the possibility to accurately compute entropies is not good. None of the techniques considered seems suitable for the calculation of the entropy of ligand–protein binding or the entropy of polypeptide folding.

Four Ways to Compute Entropy Differences

Coupling parameter λ approach

Hamiltonian is made function of λ: $H_a(\vec{p},\vec{r}) = H(\vec{p},\vec{r},\lambda_a) \rightarrow$ state a

$H_b(\vec{p},\vec{r}) = H(\vec{p},\vec{r},\lambda_b) \rightarrow$ state b

Free energy depends on λ:

$$F_{NVT}(\lambda) = -k_b T \ln\left(\left[h^{3N} N!\right]^{-1} \iint \exp\left(-H(\vec{p},\vec{r},\lambda)/k_b T\right) d\vec{p} d\vec{r}\right)$$

1. Entropy difference from energy difference and thermodynamic integration (TI) of the free energy

Free Energy Difference and End States Energy Difference

$$\Delta F_{ba}^{TI} = F(\lambda_b) - F(\lambda_a) = \int_{\lambda a}^{\lambda b} \frac{dF}{d\lambda} d\lambda = \int_{\lambda a}^{\lambda b} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda$$ accurate

$$\Delta U_{ba} = \langle H \rangle_{\lambda_b} - \langle H \rangle_{\lambda_a}$$

$$= U(\lambda_b) - U(\lambda_a) = \Delta U_{ba}^{end}$$ not accurate

$$\Delta S_{ba}^{TI} = \frac{\Delta U_{ba}^{end} - \Delta F_{ba}^{TI}}{T}$$ not so accurate because of inaccuracy of energy

2. Entropy difference directly from TI

using $S = -\left(\frac{\partial F}{\partial T}\right)_{N,V}$ $\Rightarrow$ $\Delta S_{ba}^{TI} = \frac{1}{k_b T^2} \int_{\lambda a}^{\lambda b} \left[\left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \langle H \rangle_\lambda - \left\langle \frac{\partial H}{\partial \lambda} H \right\rangle_\lambda \right] d\lambda$

and $\left(\frac{\partial S}{\partial \lambda}\right)_{N,V,T}$ correlation between $\frac{\partial H}{\partial \lambda}$ and H not so accurate

($\frac{\partial H}{\partial \lambda}$: only λ-dependent terms; H: all terms)

3. Entropy difference from finite temperature difference

using $S = -\left(\frac{\partial F}{\partial T}\right)_{N,T}$ $\Rightarrow$ $\Delta S_{ba}^{\Delta T} = -\frac{\Delta F_{ba}^{TI}(T+\Delta T) - \Delta F_{ba}^{TI}(T-\Delta T)}{2\Delta T}$

difference between almost equal accurate values

4. Solvation entropy difference from solute-solvent entropy difference (using TI) and **end states solvent-solvent energy difference**

$$\Delta S_{ba}^{TI,all} = \frac{1}{k_B T^2} \int_{\lambda_a}^{\lambda_b} \left[\left\langle \frac{\partial H_{uv}}{\partial \lambda} \right\rangle_\lambda \langle H_{uv} \rangle_\lambda - \left\langle \frac{\partial H_{uv}}{\partial \lambda} H_{uv} \right\rangle \right] d\lambda + \frac{1}{T} \left[\langle H_{vv} \rangle_{\lambda_b} - \langle H_{vv} \rangle_{\lambda_a} \right]$$

$$= \Delta S_{ba}^{TI,uv} + \frac{\Delta U_{ba}^{end,vv}}{T}$$

accurate (only solute-solvent terms); not so accurate (all solvent terms); solvent: v, solute: u

Figure 20. Four ways to compute entropy differences by using the coupling parameter technique and MD simulations.[142]

5. The Experimental Problem

Experimental data play an essential role in biomolecular modeling. First, they form the basis on which classical force fields are built (see Table 4). Without the experimental data mentioned there, classical force-field development would be virtually impossible. Quantum-chemical theoretical data alone do not suffice to build a force field. Second, the simulation methodology and the force field used can be validated and tested by comparison of simulated or calculated values for various molecular or system properties with experimentally measured ones.

Three problems arise with respect to roles of experimental data in biomolecular modeling: 1) Almost every experiment involves an averaging over time and the space or molecules, and, therefore, does not yield direct information on all configurations constituting a simulation trajectory. 2) Experimental data for biomolecular systems are scarce relative to the number of degrees of freedom involved. This makes the problem of deriving the conformational ensemble from experimental data for a biomolecular system under-determined. Many different ensembles may reproduce the same set of experimental data. 3) The experimental data may be of insufficient accuracy to be used to (in)validate simulation predictions. These three types of experimental problems with regard to biomolecular modeling will be illustrated in the following sections with examples.

5.1. The Averaging Limitation

A measurement of a quantity $Q(\boldsymbol{r})$ that can be expressed as a function of a molecular or system configuration $\boldsymbol{r}$ does not yield a value $Q^{obs} = Q(\boldsymbol{r})$ that depends on a single configuration $\boldsymbol{r}$, but yields an average over many molecules in the real (macroscopic) system and over the duration of the measurement [Eq. (20); the symbol $\langle \cdots \rangle$ denotes averaging].

$$Q^{obs} = \langle Q(\boldsymbol{r}) \rangle_{\text{molecules,time}} \tag{20}$$

In other words, only an average over the distribution $P(Q)$ of the quantity Q is measured (see Figure 21). The distribution itself is not measured. The detailed information on the distribution is lost by averaging, and very different distributions may yield the same average. The sensitivity of an average $\langle Q \rangle$ value of a particular quantity Q to the shape of the distribution $P(Q)$ or the shape of the conformational distribution $P(\boldsymbol{r})$ can be very different for different quantities Q.

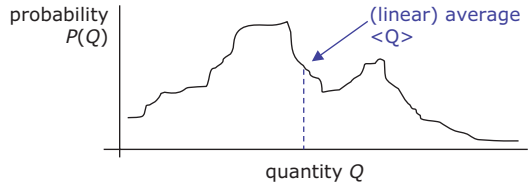


Figure 21. The averaging problem: the conformational distribution over which an average is measured cannot be derived from this average.

As an example, we consider the conformational distributions of an octapeptide $(Aib)_6$-Leu-Aib in DMSO solution and in the crystal (Figure 22). Since this peptide contains mainly achiral helix promoting Aib residues, the molecule is expected to be able to adopt both an *R*- and an *L*-helical conformation in solution. Indeed NOE intensities and ^3J values compatible with either *R*- or *L*-helices are observed in NMR experiments in solution.[167] In the crystal only the *R*-helical form is found.[168] However, the calculated ^3J values for the crystal structure are with 4.2 Hz significantly lower than the values of 6.9 Hz measured in solution. This observation

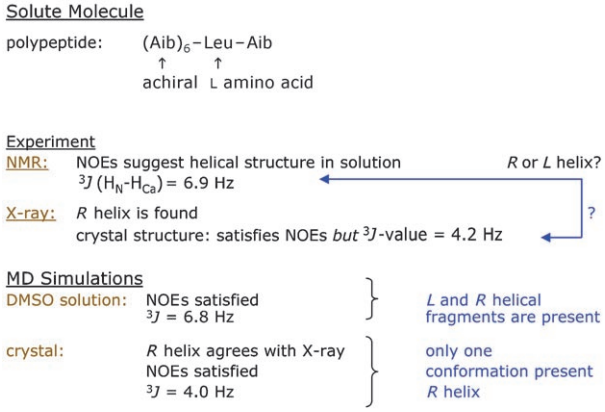


Figure 22. ^3J Coupling constants are dependent upon the conformational arrangement of an octapeptide, while in contrast NOE intensities are not sensitive.[169,170]

hints at different conformational distributions in solution and in the crystal. To investigate these, MD simulations of the octapeptide in DMSO solution and in the crystalline form were carried out.[169,170] These reproduced accurately the measured NOE intensities, ^3J values, and X-ray structure, and showed the differences in the conformational ensembles (Figure 23). In solution transient *R*- and *L*-helical fragments

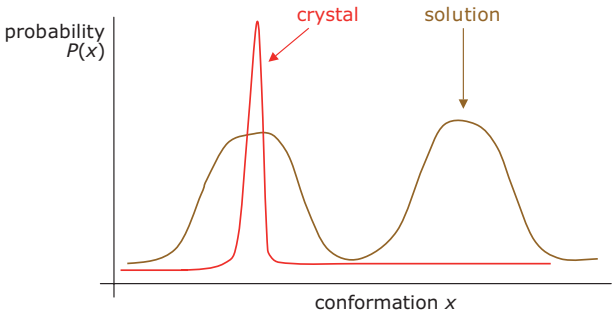


Figure 23. The conformational distribution in solution and in the crystal can be different. In the text, an example of an octapeptide is discussed.[169,170]

are present which led to a broad conformational ensemble with $\langle^3J\rangle = 6.8$ Hz, which coincides with the experimental value. In the crystal, a rather narrow *R*-helical conformational ensemble is found with $\langle^3J\rangle = 4.0$ Hz close to the ^3J value found from the X-ray structure. The simulations illustrate that the NOE intensities are not very sensitive to the shape of the conformational ensemble, as long as helical fragments are present; however, they cannot distinguish between the solution and crystal ensembles. In contrast, the ^3J-coupling constants reflect in this case the rather large differences between the two conformational ensembles.

In other cases, ^3J-coupling constants may be extremely insensitive to the underlying conformational distribution. This is illustrated in Figure 24, where simulated ^3J values for a β-heptapeptide in methanol are compared to experimental

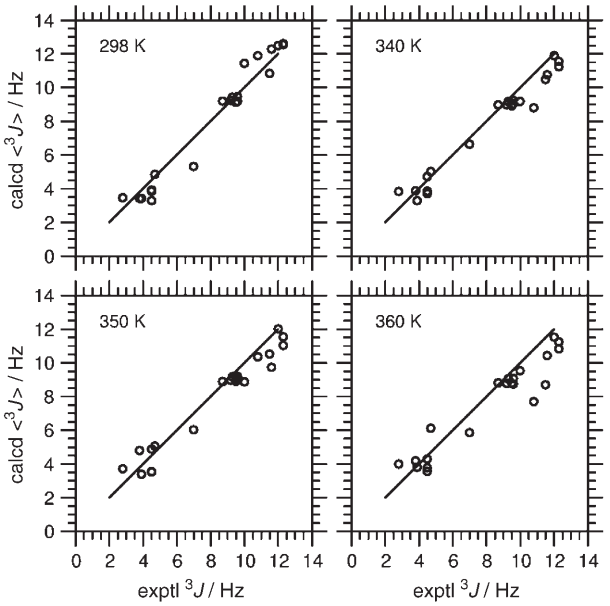


Figure 24. Comparison of the 21 experimental average ^3J-coupling constants measured at 298 K with the corresponding averaged ^3J-coupling constants calculated for the trajectory structures of 50-ns MD simulations of a β-heptapeptide in methanol at four different temperatures. The four conformational distributions are rather different: they contain 97%, 50%, 39%, and 25% 3_{14}-helical structures, respectively.[171]

values.[171] Four conformational distributions (generated at 298 K, 340 K, 350 K, and 360 K) that are rather different have been used: they contain 97%, 50%, 39% and 25% 3_{14}-helical structures, respectively.[171] However, the agreement with experiment is of equal quality for each of the four different distributions.

The loss of information through averaging in the measurement is not restricted to NMR experiments. Similar observations were made for circular dichroism (CD) spectra.[172] Figure 25 shows the measured CD spectra for two β-hexapeptides in methanol solution. The CD spectra are very similar, although the conformational ensembles for the two solutes were expected to be rather different since the double methylation at the α-carbon atom of peptide A inhibits the formation of a 3_{14}-helix, whereas peptide B, which only differs from peptide A in this respect, can and does adopt such a helix. This was confirmed by NOE intensities measured in NMR experiments. To resolve this puzzle, 100-ns MD simulations of each molecule were carried out and average CD spectra were calculated from the MD trajectories.[172] The spectra are of similar shape as the experimental ones, only the amplitudes are smaller. The conformations dominating the corresponding conformational ensembles are shown in Figure 26 together with the corresponding CD spectra. For peptide A, the CD spectrum is dominated by the second most populated (13%) conformation, not by the most populated one. The only helical conformation of peptide B (18% population) exhibits a spectrum quite different from the other conformations and from experiment.

This raises the question as to whether a particular spectrum can be assigned to a particular structure. In

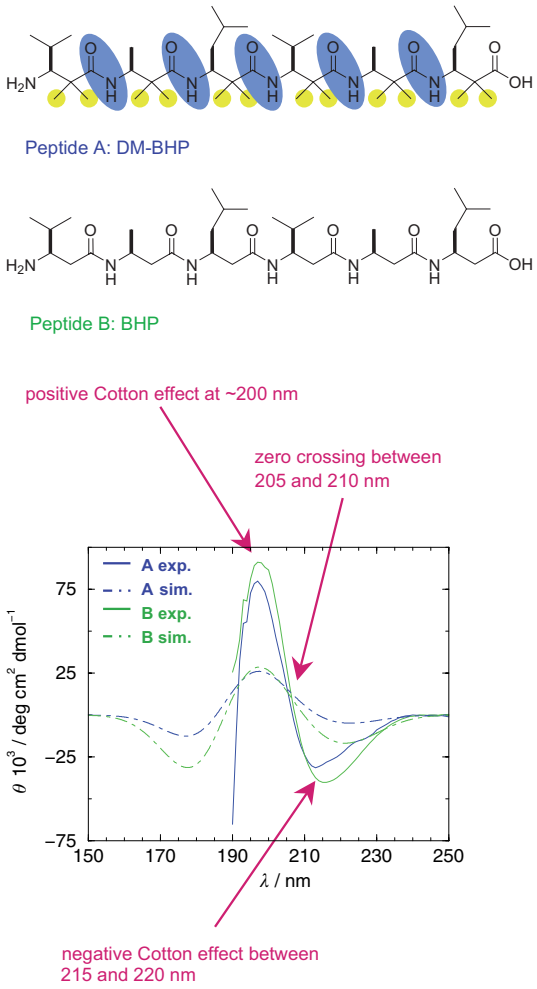


Figure 25. Experimental CD spectra and CD spectra calculated from 100-ns MD simulations of two β-hexapeptides in methanol at 298 K. Peptide B adopts a 3_{14}-helix, as confirmed by NMR experiments. Peptide A is doubly methylated at the α position and does not form a 3_{14}-helix, although a CD spectrum "typical" for a 3_{14}-helix is obtained.[172]

Figure 27 the six CD spectra with the largest amplitudes are displayed for each peptide together with the six MD trajectory structures that lead to these spectra. It is clear that the same spectrum can be generated by very different structures. Figure 28 shows a conformational clustering analysis of the combined MD trajectories of both peptides. It turns out that there is virtually no overlap between the two conformational ensembles of the peptides, although their CD spectra are very similar both in experiment and in simulation. This observation posts a cautionary note when interpreting CD spectra in terms of molecular conformations. The averaging problem means that no reliable conclusions about conformational preferences can be drawn from the measured CD spectra.

5.2. Insufficient Number of Experimental Data

If the number of experimental data on a biomolecular system is lower than the number of degrees of freedom, or if the different experimental data are correlated, they may not uniquely determine the solute conformation that dominates the conformational ensemble. An example of such a situation is illustrated in Figure 29 where NOE distances and ^3J-coupling constant values obtained from MD simulations[173] as well as from a set of 20 model structures derived during NMR structure refinement[174] are compared to the measured values for a β-hexapeptide in methanol. The NOE distances from the 100-ns MD simulations agree with the experimental values (only two small violations) and so do the ^3J values. The set of 20 NMR model structures also satisfies the experimental data, which simply reflects the fact that they were derived using the same data.[174] Figure 30 shows that there is relatively little overlap between the simulated conformational ensembles and the set of NMR model structures, yet the two sets of structures reproduce the experimental data. These data appear to be insufficient in number to uniquely determine

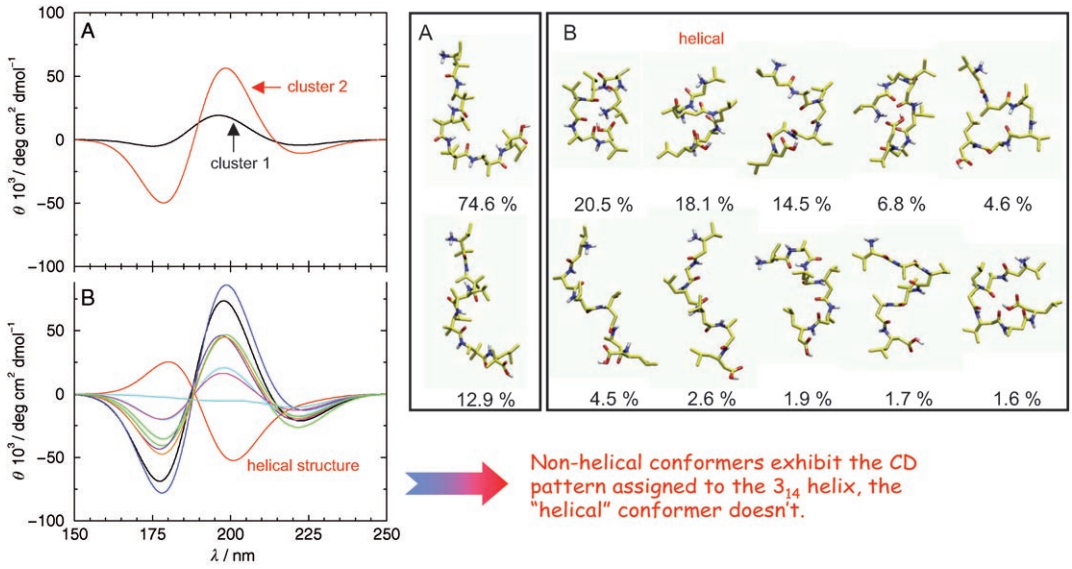


Figure 26. Dominant conformations in the MD trajectories of two β-hexapeptides A (methylated) and B (not methylated; see Figure 24) together with the corresponding CD spectra.[172] (Similarity criterion: RMSD of backbone atoms ≤0.09 nm; 10000 structures, 10 ps apart.)

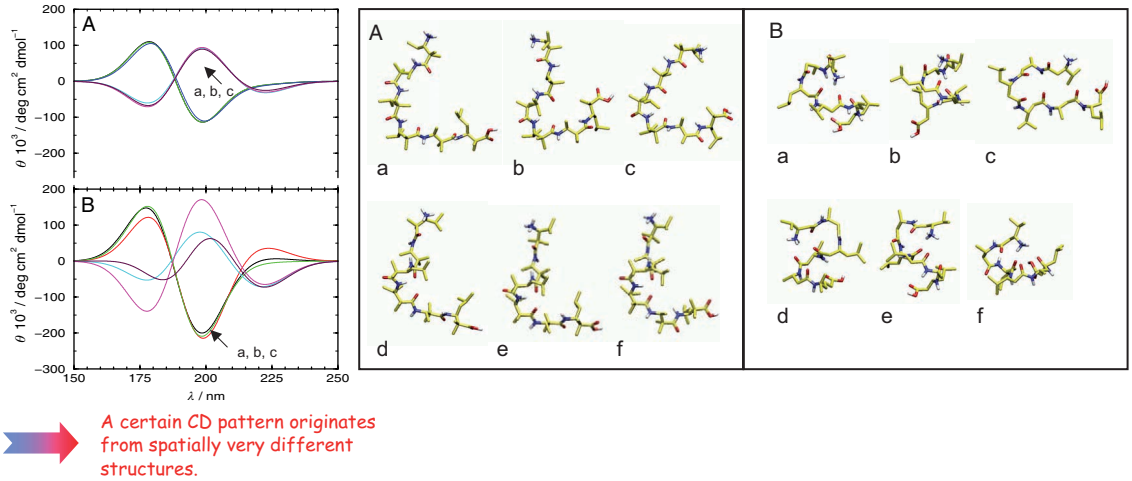


Figure 27. CD spectra for MD trajectory structures which show the largest CD signals, together with the corresponding structures of the two β-hexapeptides A and B.[172]

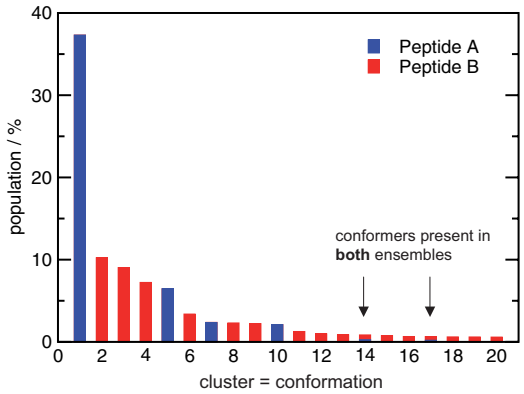


Figure 28. Conformational cluster analysis of the combined MD trajectories of both β-hexapeptides A (methylated, DM-BHP) and B (not methylated, BHP) confirm that, despite their similar CD spectra, there is virtually no overlap between the conformational ensembles of the two molecules.[172]

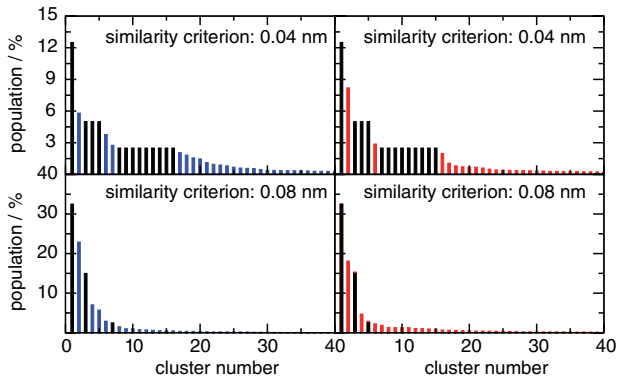
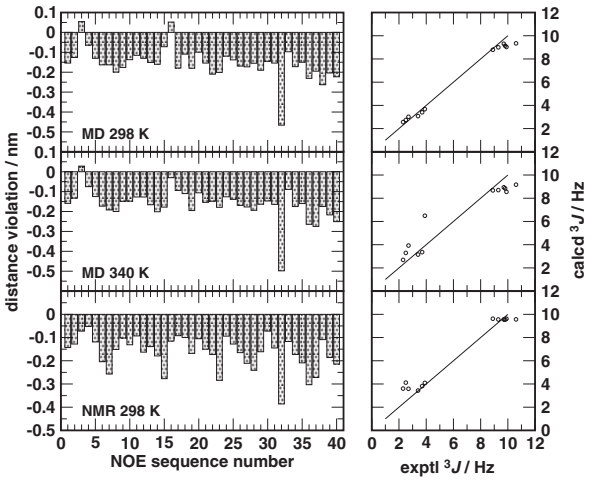


Figure 30. Conformational cluster analysis on combining the set of NMR model structures (black) for a β-octapeptide with an MD trajectory at 298 K (blue) or at 340 K (red), using two different similarity criteria for the atom-positional root-mean-square difference between the backbone atoms of the peptide structures.[173]



- 298 K:
  2 violations (~0.05 nm)

  average deviation from exptl J values: 0.44 Hz
- 340 K:
  1 violation (~ 0.03 nm)

  average deviation from exptl J values: 0.91 Hz
- NMR bundle
  no violation

  average deviation from exptl J values: 0.57 Hz

Figure 29. NOE distance-bound violations and ^3J-coupling constants for a β-octapeptide in methanol. The parameters were calculated from two 100-ns MD trajectories (not using the experimental data)[173] and from a set of 20 NMR model structures obtained by structure refinement based on the experimental NMR data.[174]

the dominant conformer: the MD trajectories suggest a 2.5_{12}-P helix to be dominant, whereas the set of NMR model structures suggest a 2_8-P-helix.[173]

5.3. Accuracy of Experimental Data

The accuracy of experimental data is finite and often insufficient to (in)validate simulation results. An example of this situation is found in Figure 15, where the disagreement between computed and measured binding free energies for a set of ligands to the estrogen receptor turns out to be smaller than the variation between different experimental values for the same ligands.

It is not easy to determine the accuracy of NOE distance bounds derived from NMR experiments on proteins in aqueous solution.

In other words, how large must a violation of such NOE distance bounds be in an MD simulation to constitute a significant disagreement between simulation and experiment? Figure 31 shows the distribution of NOE distance bounds violations for 3.5-ns MD trajectories of the protein

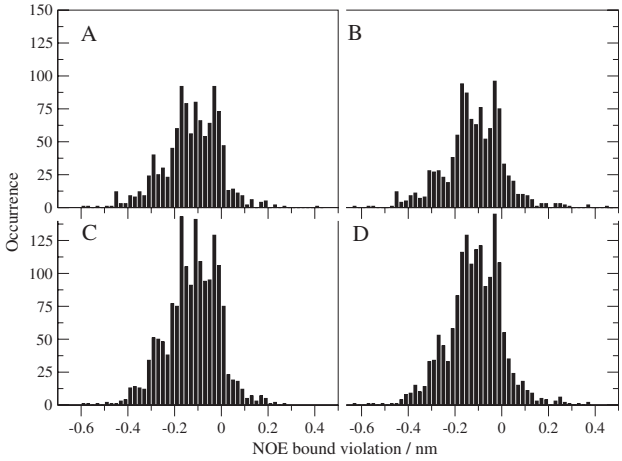


Figure 31. Distribution of r^{-3}-averaged ^{1}H-^{1}H NOE distances in two 3.5-ns MD simulations of hen egg white lysozyme in aqueous solution[175] relative to two sets of NOE distance bounds derived from NMR data. Positive values represent violations of the experimental NOE upper distance bounds. A, B) 1079 NOE distances calculated from the MD trajectories relative to the NOE distance bounds of Ref. [176] C, D) 1630 NOE distances calculated from the same MD trajectories relative to the more recent NOE distance bounds of Ref. [177] Left panels: MD simulation based on the GROMOS 43A1 force field. Right panels: MD simulation based on the GROMOS 45A3 force field.

hen egg white lysozyme in aqueous solution.[175] Two MD simulations based on different force-field parameter sets (Panels A and C, versus B and D) are analyzed in terms of two sets of experimentally determined NOE distance bounds (Panels A and B versus C and D). The experiment in 1993 produced 1079 distance bounds[176] while the more recent experiment in 2001 produced considerably more distance bounds, namely 1630.[177] A comparison of both experimental data sets with the single MD trajectories shows that the more recent, more abundant experimental data agree slightly better with the simulations (lower mean violation, fewer large violations) than the older data set. This result shows that experimental data may approach theoretical ones over time, which may serve as a cautionary note when drawing conclusions about (insufficient) quality of simulation results from observed discrepancies between simulated and measured data.

5.4. Perspectives in Comparing Simulated and Measured Data

Simulation studies are normally verified by a comparison of simulated and experimentally measured properties of the system considered.[178] The results of such a comparison between simulation and experiment can be classified as follows.[137,179,180]

Case 1: Agreement between simulation and experiment. This may arise from the following reasons:

a) The simulation adequately reflects the experimental system.
b) The property examined is insensitive to the details of the simulated trajectory. Variation of the simulation parameters would not change the agreement.
c) A compensation of errors has occurred. This situation can easily emerge if only a few (global or system) properties for a system with very many degrees of freedom are calculated and compared.

Case 2: No agreement between simulation and experiment is obtained as a consequence of one or both of the following reasons:

a) The simulation does not reflect the experimental system. The theory or model is incorrect, the simulated property is not converged, the software is at fault, or the software is incorrectly used.
b) The experimental data are incorrect or incorrectly interpreted, or both.

When comparing simulated with experimental results, the same properties should be compared. This is not always possible and sometimes only related properties are compared. For example, atom-positional mean-square fluctuations and crystallographic *B* factors both measure atomic mobility and disorder but are differently defined: the former measures the spatial distribution of a particular atom, whereas the latter measures the extent of the occupation of a given position in space by any atom that happens to be in that position.[181] As another example, protein folding rates as measured from simulated temperature renaturation may differ from those measured in an (un)folding induced by a co-solvent admixture.

With time, simulations will produce more accurate values for the various molecular and system properties, because of improved force fields and more extended equilibration and sampling. It is hoped that the improvement in experimental accuracy through improved measuring techniques may keep up with the improvements in modeling.

With regard to experimental data to be used in the development of a force field, a precise measurement of thermodynamic data such as heat of vaporization, heat of mixing, etc. for a wide variety of compounds at physiological conditions would be most beneficial for the further development of biomolecular modeling.

6. Perspectives in Biomolecular Modeling

The essential driving force behind the growth and development of the field of biomolecular modeling was, is, and will be the steady and rapid increase of computing power. Figure 32 shows there has been an average increase in computing power by a factor of 10 about every 5 years over the past few decades. This trend will probably continue in the near future, based on the on-going application of parallel computing. The possibility of parallel computing can be

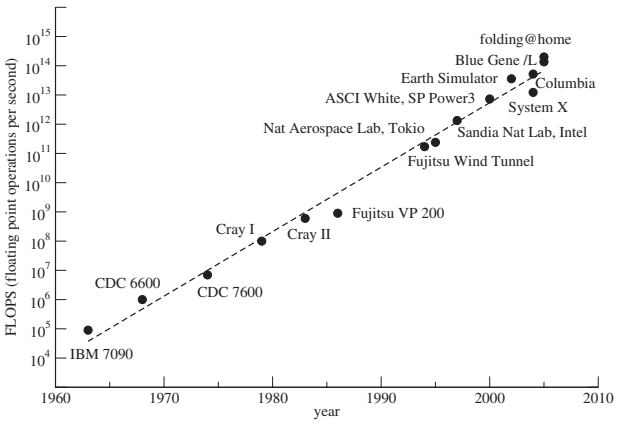


Figure 32. Development of computing power of the most powerful computers.

exploited in biomolecular simulation, since the most time-consuming part is the force or interaction calculation, which can be carried out in parallel for all atoms in the system. In particular, the advent of new hardware designed to solve the protein-folding problem through classical dynamical simulation opens up the possibility of more accurate simulations and new applications.[182]

A second driving force behind biomolecular modeling is the advancement of modeling techniques. For example, efficient algorithms to compute long-range electrostatic forces have become available.[13,21,27] Methods have been developed to extend and enhance sampling,[61,63,85,155,183] and biomolecular force fields have been refined.[7,184]

These driving forces have led to simulations of ever larger systems or over ever longer time periods (see Table 9). Practical applications of simulation address a variety of systems and processes: molecular complexation, ligand binding, polypeptide folding, transport across membranes, membrane formation, crystallization. Extrapolation of the efficiency increase in simulation by a factor 10 every 5 years leads to the predictions listed in Table 9. However, these are rather senseless predictions. First, computing power is unlikely to continue to grow for ever at the rate observed up until now. Second, when simulating ever-larger systems in atomic detail, more and more pair interactions need to be added to obtain the system energy. To obtain the same overall accuracy for a large system as for a small one, the accuracy of the pair interactions must be much higher for the large system. However, this accuracy is limited by the approximations on which a force-field description of the system rests. Third, one may question the value of a detailed atomic description of a macroscopic system. In other words, it still remains mandatory to formulate simple and approximate models (Figure 33) that contain just the necessary degrees of freedom to adequately represent the phenomenon of interest.

Table 9: History and extrapolated future of computer simulations of molecular dynamics. The future is deduced from extrapolation based on an observed increase of computing speed of a factor 10 every 5 years over the past decades (see Figure 31).

Year	Molecular system (type, size)	Length of the simulation [s]
1957	first molecular dynamics simulation (hard discs)	
1964	atomic liquid (argon)	10^{-11}
1971	molecular liquid (water)	5×10^{-12}
1977	protein in a vacuum	2×10^{-11}
1983	protein in water	2×10^{-11}
1989	protein–DNA complex in water	10^{-10}
1997	polypeptide folding in solvent	10^{-7}
2001	micelle formation	10^{-7}
200x	folding of a small protein	10^{-3}
And the future ...		
2001	biomolecules in water (ca. 10^4 atoms)	10^{-8}
2029	biomolecules in water (folding sooner?)	10^{-3}
2034	*E. coli* bacteria (ca. 10^{11} atoms)	10^{-9}
2056	mammalian cell (ca. 10^{15} atoms)	10^{-9}
2080	biomolecules in water (as fast as nature)	10^{6}
2172	human body (ca. 10^{27} atoms)	1

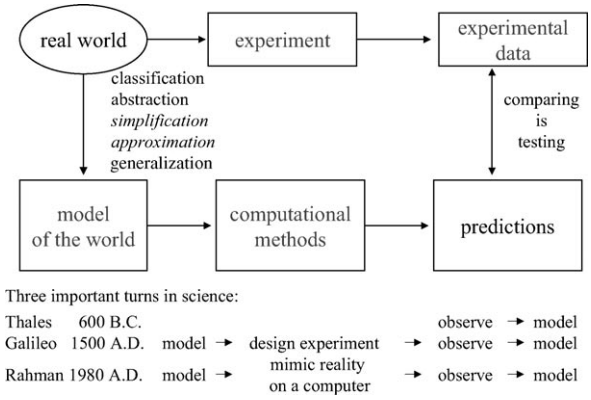


Figure 33. Computational physics, chemistry, and biology involve the formulation and testing of (mathematical) models of the real world.

The question remains, along which lines should current biomolecular models be extended, improved, or simplified? First, an appropriate description of enzyme reactions requires the inclusion of electronic degrees of freedom, one level up from the theoretical level of classical MD methods in Table 1. Hybrid quantum-classical (QM/MM) modeling will be further developed to this end.[49,185–187] To simulate proton-transfer reactions, it may be necessary to employ quantum-dynamical methodology,[188–191] which requires even more computing power than QM/MM calculations.[192] Second, at the classical level of modeling, improvements will come from the introduction of polarizability in biomolecular force fields,[44–48] from the incorporation of co-solvent effects through explicit simulation,[193] and from techniques to extend the sampling power of simulations.[61–63] Third, simplification of models by averaging over atomic degrees of freedom (coarse-graining)[54–59] will allow the simulation of slower processes, such as membrane formation.[60,194]

The reason for using simulation and modeling was indicated in Table 3: to provide a microscopic picture of unrivaled resolution in time, space, and energy that comple-

ments the limited set of properties obtainable from experiment. Second, system parameters can be changed at will during the modeling study to study particular cause–effect relationships, thus leading to enhanced understanding of biomolecular systems.

The challenge when modeling a biomolecular system lies in a proper balance between the choices to be made regarding degrees of freedom, force field, sampling, and boundary conditions (see Figure 2). These choices will depend on three factors (Figure 34).

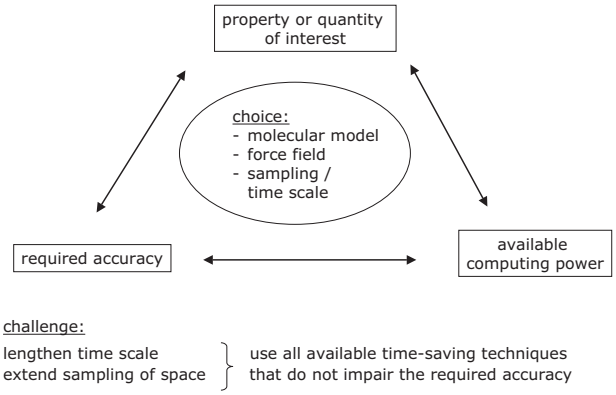


Figure 34. The choice of molecular model (degrees of freedom), force field, and extent of sampling depends on the property of interest, the required accuracy of the result, and the available computing power to generate the Boltzmann ensemble.

1. The properties of the biomolecular system one is interested in should be listed and the size of the configuration space (or time scale) to be searched and sampled should be estimated.
2. The required accuracy of the properties should be specified.
3. The available computing power should be estimated.

If the model selected is too simple, the phenomena of interest may be lost or the accuracy may be insufficient. If the model is too elaborate, sampling of the required extent of configuration space may be impossible. It is the art of biomolecular modeling to sail safely between these *Scylla* and *Charybdis*.[195]

The National Centre of Competence in Research (NCCR) in Structural Biology of the Swiss National Science Foundation is gratefully acknowledged for financial support. We thank Daniela Kalbermatter for her help in editing the manuscript.

Received: July 28, 2005

[1] M. P. Allen, D. J. Tildesley, *Computer simulation of liquids*, Oxford University Press, New York, **1987**.
[2] B. A. Luty, W. F. van Gunsteren, *J. Phys. Chem.* **1996**, *100*, 2581 – 2587.
[3] P. H. Hünenberger, J. A. McCammon, *J. Chem. Phys.* **1999**, *110*, 1856 – 1872.
[4] P. H. Hünenberger, J. A. McCammon, *Biophys. Chem.* **1999**, *78*, 69 – 88.
[5] W. Weber, P. H. Hünenberger, J. A. McCammon, *J. Phys. Chem. B* **2000**, *104*, 3668 – 3675.
[6] P. H. Hünenberger, W. F. van Gunsteren in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, *Vol. 3* (Eds.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson), Kluwer, Dordrecht, The Netherlands, **1997**, pp. 3 – 82.
[7] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656 – 1676.
[8] N. F. A. van der Vegt, W. F. van Gunsteren, *J. Phys. Chem. B* **2004**, *108*, 1056 – 1064.
[9] N. F. A. van der Vegt, D. Trzesniak, B. Kasumaj, W. F. van Gunsteren, *ChemPhysChem* **2004**, *5*, 144 – 147.
[10] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag, Zürich, **1996**.
[11] J. D. Jackson, *Classical Electrodynamics*, Wiley, New York, **1962**.
[12] P. Ewald, *Ann. Phys.* **1921**, *64*, 253 – 287.
[13] R. W. Hockney, J. W. Eastwood, *Computer simulation using particles*, 2nd ed., Institute of Physics Publishing, Bristol, **1988**.
[14] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577 – 8593.
[15] J. A. Barker, R. O. Watts, *Mol. Phys.* **1973**, *26*, 789 – 792.
[16] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 5451 – 5459.
[17] S. Boresch, O. Steinhauser, *J. Chem. Phys.* **2001**, *115*, 10793 – 10807.
[18] M. Bergdorf, C. Peter, P. H. Hünenberger, *J. Chem. Phys.* **2003**, *119*, 9129 – 9144.
[19] C. Peter, W. F. van Gunsteren, P. H. Hünenberger, *J. Chem. Phys.* **2003**, *119*, 12205 – 12223.
[20] M. A. Kastenholz, P. H. Hünenberger, *J. Phys. Chem. B* **2004**, *108*, 774 – 778.
[21] P. H. Hünenberger in *Simulation and theory of electrostatic interactions in solution: Computational chemistry, biophysics and aqueous solution* (Eds.: L. R. Pratt, G. Hummer) AIP, New York, **1999**, pp. 17 – 83.
[22] B. A. Luty, I. G. Tironi, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *103*, 3014 – 3021.
[23] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans in *Intermolecular Forces* (Ed.: B. Pullman), Reidel, Dordrecht, **1981**, pp. 331 – 342.
[24] T. P. Straatsma, H. J. C. Berendsen, *J. Chem. Phys.* **1988**, *89*, 5876 – 5886.
[25] P. H. Hünenberger, *J. Chem. Phys.* **2000**, *113*, 10464 – 10476.
[26] G. Hummer, L. R. Pratt, A. E. Garcia, *J. Phys. Chem.* **1996**, *100*, 1206 – 1215.
[27] M. A. Kastenholz, P. H. Hünenberger, *J. Chem. Phys.*, in press.
[28] W. L. Jorgensen, J. D. Madura, C. J. Swenson, *J. Am. Chem. Soc.* **1984**, *106*, 6638 – 6646.
[29] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225 – 11236.
[30] L. D. Schuler, W. F. van Gunsteren, *Mol. Simul.* **2000**, *25*, 301 – 319.
[31] A. Glättli, W. F. van Gunsteren, *Angew. Chem.* **2004**, *116*, 6472 – 6476; *Angew. Chem. Int. Ed.* **2004**, *43*, 6312 – 6316.
[32] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *J. Mol. Biol.* **1998**, *280*, 925 – 932.
[33] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew. Chem.* **1999**, *111*, 249 – 253; *Angew. Chem. Int. Ed.* **1999**, *38*, 236 – 240.
[34] X. Daura, K. Gademann, H. Schäfer, B. Jaun, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2001**, *123*, 2393 – 2404.

[35] C. M. Santiveri, M. A. Jiménez, M. Rico, W. F. van Gunsteren, X. Daura, *J. Pept. Sci.* **2004**, *10*, 546 – 565.
[36] C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, W. F. van Gunsteren, *Eur. Biophys. J.* **2005**, *34*, 273 – 284.
[37] I. Huc, V. Maurizot, H. Gornitzka, J.-M. Leger, *Chem. Commun.* **2002**, 578 – 579.
[38] A. Villa, A. E. Mark, *J. Comput. Chem.* **2002**, *23*, 548 – 553.
[39] M. R. Shirts, J. W. Pitera, W. C. Swope, V. S. Pande, *J. Chem. Phys.* **2003**, *119*, 5740 – 5761.
[40] Y. Marcus, *Ionic solvation*, Wiley, Chichester, **1985**.
[41] L. J. Smith, H. J. C. Berendsen, W. F. van Gunsteren, *J. Phys. Chem. A* **2004**, *108*, 1065 – 1071.
[42] D. P. Geerke, C. Oostenbrink, N. F. A. van der Vegt, W. F. van Gunsteren, *J. Phys. Chem. B* **2004**, *108*, 1436 – 1445.
[43] D. Trzesniak, N. F. A. van der Vegt, W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2004**, *6*, 697 – 702; Erratum: D. Trzesniak, N. F. A. van der Vegt, W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2004**, *6* (3rd August, 2004).
[44] T. A. Halgren, W. Damm, *Curr. Opin. Struct. Biol.* **2001**, *11*, 236 – 242.
[45] S. W. Rick, S. J. Stuart, *Rev. Comput. Chem.* **2002**, 89 – 146.
[46] B. Guillot, *J. Mol. Liq.* **2002**, *101*, 219 – 260.
[47] J. W. Ponder, D. A. Case, *Adv. Protein Chem.* **2003**, *56*, 27 – 85.
[48] H. B. Yu, W. F. van Gunsteren, *Comput. Phys. Commun.* **2005**, *172*, 69 – 85.
[49] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227 – 249.
[50] D. Van Belle, I. Couplet, M. Prevost, S. J. Wodak, *J. Mol. Biol.* **1987**, *198*, 721 – 735.
[51] J. L. Banks, G. A. Kaminski, R. H. Zhou, D. T. Mainz, B. J. Berne, R. A. Friesner, *J. Chem. Phys.* **1999**, *110*, 741 – 754.
[52] S. Patel, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1 – 15.
[53] S. Patel, A. D. MacKerell, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1504 – 1514.
[54] A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackofsky, H. A. Scheraga, *Protein Sci.* **1993**, *2*, 1715 – 1731.
[55] J. C. Shelley, M. Y. Shelley, *Curr. Opin. Colloid Interface Sci.* **2000**, *5*, 101 – 110.
[56] J. Barschnagel, K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Müller-Plathe, M. Murat, W. Paul, S. Santos, U. W. Suter, V. Tries, *Adv. Polym. Sci.* **2000**, *152*, 41 – 156.
[57] M. Müller, K. Katsov, M. Schick, *J. Polym. Sci. Part B* **2003**, *41*, 1441 – 1450.
[58] A. Liwo, M. Khalili, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2362 – 2367.
[59] S. J. Marrink, A. H. de Vries, A. E. Mark, *J. Phys. Chem. B* **2004**, *108*, 750 – 760.
[60] S. J. Marrink, H. J. Risselada, A. E. Mark, *Chem. Phys. Lip.* **2005**, *135*, 223 – 244.
[61] W. F. van Gunsteren, T. Huber, A. E. Torda in *Proc. Eur. Conf. Comput. Chem. (ECCC 1)*, American Institute of Physics, Conference Proceedings **1995**, *330*, 253 – 268.
[62] Y. Okamoto, *J. Mol. Graphics Modell.* **2004**, *22*, 425 – 439.
[63] K. Tai, *Biophys. Chem.* **2004**, *107*, 213 – 220.
[64] P. Koehl, M. Delaure, *J. Mol. Biol.* **1994**, *239*, 249 – 275.
[65] T. Huber, A. E. Torda, W. F. van Gunsteren, *Biopolymers* **1996**, *39*, 103 – 114.
[66] J. Desmet, M. DeMaeyer, B. Hazes, I. Lasters, *Nature* **1992**, *356*, 539 – 542.
[67] M. Lipton, W. C. Still, *J. Comput. Chem.* **1988**, *9*, 343 – 355.
[68] "Protein Structure and Engineering": D. D. Bensen, G. R. Marshall, *NATO ASI Ser. Ser. A* **1989**, *183*, 97 – 109.
[69] M. Saunders, K. N. Houk, Y.-D. Wu, W. C. Still, M. Lipton, G. Chang, W. C. Guida, *J. Am. Chem. Soc.* **1990**, *112*, 1419 – 1427.
[70] D. G. Covell, R. L. Jernigan, *Biochemistry* **1990**, *29*, 3287 – 3294.
[71] R. C. van Schaik, W. F. van Gunsteren, H. J. C. Berendsen, *J. Comput.-Aided Mol. Des.* **1992**, *6*, 97 – 112.
[72] G. M. Crippen, T. F. Havel, *Distance geometry and molecular conformation*, Wiley, New York, **1988**.
[73] T. F. Havel, *Biopolymers* **1990**, *29*, 1565 – 1585.
[74] K. D. Gibson, H. A. Scheraga, *J. Comput. Chem.* **1987**, *8*, 826 – 834.
[75] H. A. Scheraga in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, *Vol. 2* (Eds.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson), Escom Science Publishers, Leiden, **1993**, pp. 231 – 248.
[76] S. Vajda, C. DeLisi, *Biopolymers* **1990**, *29*, 1755 – 1772.
[77] J. Harris, S. A. Rice, *J. Chem. Phys.* **1988**, *88*, 1298 – 1306.
[78] B. Velikson, T. Garel, J.-C. Niel, H. Orland, J. C. Smith, *J. Comput. Chem.* **1992**, *13*, 1216 – 1233.
[79] D. Frenkel, G. C. A. M. Mooij, B. Smit, *J. Phys. Condens. Matter* **1992**, *4*, 3053 – 3076.
[80] W. F. van Gunsteren in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, *Vol. 2* (Eds.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson), Escom Science Publishers, Leiden, **1993**, pp. 3 – 36.
[81] R. M. J. Cotterill, J. K. Madsen in *Characterising Complex Systems* (Ed.: H. Bohr), World Scientific, Singapore, **1990**, p. 177 – 191.
[82] W. Braun, N. Go, *J. Mol. Biol.* **1985**, *186*, 611 – 626.
[83] T. Huber, A. E. Torda, W. F. van Gunsteren, *J. Phys. Chem. A* **1997**, *101*, 5926 – 5930.
[84] A. Piela, J. Kostrowicki, H. A. Scheraga, *J. Phys. Chem.* **1989**, *93*, 3339 – 3346.
[85] T. Huber, A. E. Torda, W. F. van Gunsteren, *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695 – 708.
[86] H. Grubmüller, *Phys. Rev. E* **1995**, *52*, 2893 – 2906.
[87] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, *J. Mol. Biol.* **1990**, *214*, 223 – 235.
[88] W. F. van Gunsteren, R. M. Brunne, P. Gros, R. C. van Schaik, C. A. Schiffer, A. E. Torda in *Methods in Enzymology: Nuclear Magnetic Resonance*, *Vol. 239* (Eds.: T. L. James, N. J. Oppenheimer), Academic Press, New York, **1994**, pp. 619 – 654.
[89] R. C. van Schaik, H. J. C. Berendsen, A. E. Torda, W. F. van Gunsteren, *J. Mol. Biol.* **1993**, *234*, 751 – 762.
[90] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Phys.* **1977**, *34*, 1311 – 1327.
[91] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *Science* **1983**, *220*, 671 – 680.
[92] B. Mao, A. R. Friedmann, *Biophys. J.* **1990**, *58*, 803 – 805.
[93] R. Elber, M. Karplus, *J. Am. Chem. Soc.* **1990**, *112*, 9161 – 9175.
[94] R. Unger, J. Moult, *J. Mol. Biol.* **1993**, *231*, 75 – 81.
[95] T. Huber, W. F. van Gunsteren, *J. Phys. Chem. A* **1998**, *102*, 5937 – 5943.
[96] M. Levitt, *J. Mol. Biol.* **1983**, *170*, 723 – 764.
[97] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, W. F. van Gunsteren, *Chem. Phys. Lett.* **1994**, *222*, 529 – 539.
[98] M. Zacharias, T. P. Straatsma, J. A. McCammon, *J. Chem. Phys.* **1994**, *100*, 9025 – 9031.
[99] V. Hornak, C. Simmerling, *J. Mol. Graphics Modell.* **2004**, *22*, 405 – 413.
[100] G. M. Crippen, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **1969**, *64*, 42 – 49.
[101] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562 – 12566.
[102] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, *Chem. Phys. Lett.* **1989**, *157*, 289 – 294.
[103] G. M. Crippen, *J. Comput. Chem.* **1982**, *3*, 471 – 476.
[104] E. O. Purisima, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 2782 – 2786.
[105] G. M. Crippen, *J. Phys Chem.* **1987**, *91*, 6341 – 6343.
[106] G. M. Crippen, T. F. Havel, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 222 – 227.

[107] P. L. Weber, R. Morrison, D. L. Hare, *J. Mol. Biol.* **1988**, *204*, 483 – 487.
[108] T. C. Beutler, W. F. van Gunsteren, *J. Chem. Phys.* **1994**, *101*, 1417 – 1422.
[109] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327 – 341.
[110] H. C. Andersen, *J. Comput. Phys.* **1983**, *52*, 24 – 34.
[111] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952 – 962.
[112] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463 – 1472.
[113] V. Kräutler, W. F. van Gunsteren, P. H. Hünenberger, *J. Comput. Chem.* **2001**, *22*, 501 – 508.
[114] W. F. van Gunsteren, M. Karplus, *Macromolecules* **1982**, *15*, 1528 – 1544.
[115] M. Christen, W. F. van Gunsteren, *J. Chem. Phys.* **2005**, *122*, 144106.
[116] J. E. Straub, M. Karplus, *J. Chem. Phys.* **1991**, *94*, 6737 – 6739.
[117] A. Roitberg, R. Elber, *J. Chem. Phys.* **1991**, *95*, 9277 – 9287.
[118] G. Verkhivker, R. Elber, Q. H. Gibson, *J. Am. Chem. Soc.* **1992**, *114*, 7866 – 7878.
[119] G. Verkhivker, R. Elber, W. Nowak, *J. Chem. Phys.* **1992**, *97*, 7838 – 7841.
[120] A. Ulitsky, R. Elber, *J. Chem. Phys.* **1993**, *98*, 3380 – 3388.
[121] P. Amara, D. Hsu, J. E. Straub, *J. Phys. Chem.* **1993**, *97*, 6715 – 6721.
[122] J. P. Ma, D. Hsu, J. E. Straub, *J. Chem. Phys.* **1993**, *99*, 4024 – 4035.
[123] Q. Zheng, R. Rosenfeld, S. Vajda, C. DeLisi, *Protein Sci.* **1993**, *2*, 1242 – 1248.
[124] Q. Zheng, R. Rosenfeld, D. J. Kyle, *J. Chem. Phys.* **1993**, *99*, 8892 – 8896.
[125] K. A. Olszewski, L. Piela, H. A. Scheraga, *J. Phys. Chem.* **1992**, *96*, 4672 – 4676.
[126] K. A. Olszewski, L. Piela, H. A. Scheraga, *J. Phys. Chem.* **1993**, *97*, 260 – 266.
[127] K. A. Olszewski, L. Piela, H. A. Scheraga, *J. Phys. Chem.* **1993**, *97*, 267 – 270.
[128] C. Simmerling, J. L. Miller, P. A. Kollman, *J. Am. Chem. Soc.* **1998**, *120*, 7149 – 7155.
[129] H. Y. Liu, Z. H. Duan, Q. M. Luo, Y. Y. Shi, *Proteins Struct. Funct. Genet.* **1999**, *36*, 462 – 470.
[130] J. Zhu, H. B. Yu, H. Fan, H. Y. Liu, Y. Y. Shi, *J. Comput.-Aided Mol. Des.* **2001**, *15*, 447 – 463.
[131] J. Zhu, H. Fan, H. Y. Liu, Y. Y. Shi, *J. Comput.-Aided Mol. Des.* **2001**, *15*, 979 – 996.
[132] D. E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison-Wesley, Reading, **1989**.
[133] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141 – 151.
[134] R. Zhou, B. J. Berne, R. Germain, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14931 – 14936.
[135] W. F. van Gunsteren, P. H. Hünenberger, A. E. Mark, P. E. Smith, I. G. Tironi, *Comput. Phys. Commun.* **1995**, *91*, 305 – 319.
[136] W. F. van Gunsteren, A. E. Mark, *J. Chem. Phys.* **1998**, *108*, 6109 – 6116.
[137] "Dynamics, Structure and Function of Biological Macromolecules": W. F. van Gunsteren, D. Bakowies, W. Damm, T. Hansson, U. Stocker, X. Daura, *NATO ASI Ser. Ser. A* **2001**, *315*, 1 – 26.
[138] X. Daura, W. F. van Gunsteren, A. E. Mark, *Proteins Struct. Funct. Genet.* **1999**, *34*, 269 – 280.
[139] A. Glättli, X. Daura, P. Bindschädler, B. Jaun, Y. R. Mahajan, R. I. Mathad, M. Rueping, D. Seebach, W. F. van Gunsteren, *Chem. Eur. J.* **2005**, *11*, 7276 – 7293.
[140] J. W. Pitera, M. Falta, W. F. van Gunsteren, *Biophys. J.* **2001**, *80*, 2546 – 2555.
[141] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger, *J. Chem. Phys.* **2001**, *115*, 1125 – 1136.
[142] C. Peter, C. Oostenbrink, A. van Dorp, W. F. van Gunsteren, *J. Chem. Phys.* **2004**, *120*, 2652 – 2661.
[143] L. J. Smith, X. Daura, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2002**, *48*, 487 – 496.
[144] W. F. van Gunsteren, R. Bürgi, C. Peter, X. Daura, *Angew. Chem.* **2001**, *113*, 363 – 367; *Angew. Chem. Int. Ed.* **2001**, *40*, 351 – 355.
[145] W. F. van Gunsteren, D. Bakowies, R. Bürgi, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, A. Glättli, T. Hansson, C. Oostenbrink, C. Peter, J. Pitera, L. Schuler, T. Soares, H. B. Yu, *Chimia* **2001**, *55*, 856 – 860.
[146] X. Daura, A. Glättli, P. Gee, C. Peter, W. F. van Gunsteren, *Adv. Protein Chem.* **2002**, *62*, 341 – 360.
[147] P. J. Gee, F. A. Hamprecht, L. D. Schuler, W. F. van Gunsteren, E. Duchardt, H. Schwalbe, M. Albert, D. Seebach, *Helv. Chim. Acta* **2002**, *85*, 618 – 632.
[148] H. Schäfer, X. Daura, A. E. Mark, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2001**, *43*, 45 – 56.
[149] C. Oostenbrink, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2004**, *54*, 237 – 246.
[150] C. Oostenbrink, W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2005**, *7*, 53 – 58.
[151] L. J. Smith, R. M. Jones, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2005**, *58*, 439 – 449.
[152] H. Liu, A. E. Mark, W. F. van Gunsteren, *J. Phys. Chem.* **1996**, *100*, 9485 – 9494.
[153] J. W. Pitera, W. F. van Gunsteren, *J. Phys. Chem. B* **2001**, *105*, 11264 – 11274.
[154] C. Oostenbrink, W. F. van Gunsteren, *J. Comput. Chem.* **2003**, *24*, 1730 – 1739.
[155] C. Oostenbrink, W. F. van Gunsteren, *Chem. Eur. J.* **2005**, *11*, 4340 – 4348.
[156] C. Oostenbrink, W. F. van Gunsteren, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6750 – 6754.
[157] H.-A. Yu, M. Karplus, *J. Chem. Phys.* **1988**, *89*, 2366 – 2379.
[158] B. Guillot, Y. Guissani, *J. Chem. Phys.* **1993**, *99*, 8075 – 8094.
[159] E. Gallicchio, M. M. Kubo, R. M. Levy, *J. Phys. Chem. B* **2000**, *104*, 6271 – 6285.
[160] M. Rueping, J. V. Schreiber, G. Lelais, B. Jaun, D. Seebach, *Helv. Chim. Acta* **2002**, *85*, 2577 – 2593.
[161] D. Trzesniak, A. Glättli, B. Jaun, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2005**, 127, 14320 – 14329.
[162] M. Karplus, J. N. Kushick, *Macromolecules* **1981**, *14*, 325 – 332.
[163] J. Schlitter, *Chem. Phys. Lett.* **1993**, *215*, 617 – 621.
[164] H. Schäfer, A. E. Mark, W. F. van Gunsteren, *J. Chem. Phys.* **2000**, *113*, 7809 – 7817.
[165] J. Carlsson, J. Aqvist, *J. Phys. Chem. B* **2005**, *109*, 6448 – 6456.
[166] H. Schäfer, L. J. Smith, A. E. Mark, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2002**, *46*, 215 – 224.
[167] M. Bellanda, E. Peggion, R. Bürgi, W. F. van Gunsteren, S. Mammi, *J. Pept. Res.* **2001**, *57*, 97 – 106.
[168] A. Bavoso, E. Benedetti, B. DiBlasio, V. Pavone, C. Pedone, C. Toniolo, G. M. Bonora, F. Formaggio, M. Crisma, *J. Biomol. Struct. Dyn.* **1988**, *6*, 803 – 817.
[169] R. Bürgi, X. Daura, A. Mark, M. Bellanda, S. Mammi, E. Peggion, W. F. van Gunsteren, *J. Pept. Res.* **2001**, *57*, 107 – 118.
[170] H. B. Yu, M. Ramseier, R. Bürgi, W. F. van Gunsteren, *ChemPhysChem* **2004**, *5*, 633 – 641.
[171] X. Daura, I. Antes, W. F. van Gunsteren, W. Thiel, A. E. Mark, *Proteins Struct. Funct. Genet.* **1999**, *36*, 542 – 555.
[172] A. Glättli, X. Daura, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2002**, *124*, 12972 – 12978.
[173] A. Glättli, W. F. van Gunsteren, *Angew. Chem.* **2004**, *116*, 6472 – 6476; *Angew. Chem. Int. Ed.* **2004**, *43*, 6312 – 6316.

[174] K. Gademann, A. Häne, M. Rueping, B. Jaun, D. Seebach, *Angew. Chem.* **2003**, *115*, 1573 – 1575; *Angew. Chem. Int. Ed. Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 1534 – 1537.
[175] T. A. Soares, X. Daura, C. Oostenbrink, L. J. Smith, W. F. van Gunsteren, *J. Biomol. NMR* **2004**, *30*, 407 – 422.
[176] L. J. Smith, M. J. Sutcliffe, C. Redfield, C. M. Dobson, *J. Mol. Biol.* **1993**, *229*, 930 – 944.
[177] H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, L. J. Smith, *Protein Sci.* **2001**, *10*, 677 – 688.
[178] W. F. van Gunsteren, A. E. Mark, *J. Chem. Phys.* **1998**, *108*, 6109 – 6116.
[179] "Modelling of Molecular Structures, Properties": W. F. van Gunsteren in *Studies in Physical Theoretical Chemistry*, *Vol. 71* (Ed.: J.-L. Rivail), Elsevier, Amsterdam, **1990**, pp. 463 – 478.
[180] W. F. van Gunsteren, A. E. Mark, *Eur. J. Biochem.* **1992**, *204*, 947 – 961.
[181] P. H. Hünenberger, A. E. Mark, W. F. van Gunsteren, *J. Mol. Biol.* **1995**, *252*, 492 – 503.
[182] B. G. Fitch, R. S. Germain, M. Mendell, J. Pitera, M. Pitman, A. Rayshubskiy, Y. Sham, F. Suits, W. Swope, T. J. C. Ward, Y. Zhestkov, R. Zhou, *J. Parall. Distr. Comp.* **2003**, *63*, 759 – 773.
[183] J. Norberg, L. Nilsson, *Q. Rev. Biophys.* **2003**, *36*, 257 – 306.
[184] A. D. MacKerell, *J. Comput. Chem.* **2004**, *25*, 1584 – 1604.
[185] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700 – 733.
[186] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580 – 10594.
[187] A. Warshel, *Computer Modelling of Chemical Reactions in Enzymes and Solution*, Wiley, New York, **1991**.
[188] H. J. C. Berendsen, J. Mavri, *J. Phys. Chem.* **1993**, *97*, 13464 – 13468.
[189] S. R. Billeter, W. F. van Gunsteren, *Comput. Phys. Commun.* **1997**, *107*, 61 – 91.
[190] H. J. C. Berendsen, J. Mavri in *Theoretical Treatments of Hydrogen Bonding* (Ed.: D. Hadzi), Wiley, New York, **1997**, pp. 119 – 141.
[191] S. R. Billeter, S. P. Webb, T. Iordanov, P. K. Agarwal, S. Hammes-Schiffer, *J. Chem. Phys.* **2001**, *114*, 6925 – 6936.
[192] S. J. Benkovic, S. Hammes-Schiffer, *Science* **2003**, *301*, 1196 – 1202.
[193] T. Hansson, C. Oostenbrink, W. F. van Gunsteren, *Curr. Opin. Struct. Biol.* **2002**, *12*, 190 – 196.
[194] J. C. Shillcock, R. Lipowsky, *J. Chem. Phys.* **2002**, *117*, 5048 – 5061.
[195] *Scylla* and *Charybdis* are two monsters from Greek mythology that guarded the narrow waters of the Straits of Messina destroying ships as they attempted to navigate through.

