Chimia 55 (2001) 856–860

ISSN 0009-4293

Molecular Dynamics Simulation of Biomolecular Systems

Wilfred van Gunsteren*, Dirk Bakowies, Roland Bürgi, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Alice Glättli, Tomas Hansson, Chris Oostenbrink, Christine Peter, Jed Pitera, Lukas Schuler, Thereza Soares, and Haibo Yu

Abstract: The group for computer-aided chemistry at the ETH Zürich focuses its research on the development of methodology to simulate the behavior of biomolecular systems and the use of simulation techniques to analyze and understand biomolecular processes at the atomic level. Here, the current research directions are briefly reviewed and illustrated with a few examples.

Keywords: Biomolecular simulation · Membranes · Methodology · Molecular dynamics · Proteins

1. Introduction

Computer simulation of the dynamics of biomolecular systems by the molecular dynamics (MD) technique yields the possibility of describing and understanding the structure–dynamics–function relationships of molecular processes in terms of interactions at the atomic level. Once the reliability of the molecular models, force fields, and computational procedures has been established by comparison of simulated properties with known experimental ones, computer simulation can be used to interpret experimental data, to analyze and understand molecular processes and to predict molecular properties that are inaccessible to experimental probes. In this article we briefly review recent research directions of our research group for computer-aided chemistry at the ETH Zürich. Using examples from our own work we shall illustrate the possibilities, the limitations and the perspectives of computer simulation studies in the field of biochemistry and molecular biology.

*Correspondence: Prof. Dr. W.F. van Gunsteren
Department of Chemistry
Swiss Federal Institute of Technology
ETH Hönggerberg
CH-8093 Zürich
Tel.: + 41 1 632 5501
Fax: + 41 1 632 1039
E-Mail: wfvgn@igc.phys.chem.ethz.ch

2. Levels of Modeling

When undertaking a biomolecular modeling study of a particular system of interest, the level of modeling, *i.e.* the space resolution, time scale and degrees of freedom of interest, must be considered. Fig. 1 shows a number of possibilities. Once the particular degrees of freedom to be explicitly considered and simulated have been chosen, the appropriate method to treat the variation along the degrees of freedom or the appropriate equations of motion to simulate the dynamics along the degrees of freedom must be chosen [1]. Our research mainly concentrates on classical statistical mechanical MD simulation techniques at the atomic level [2], with a few forays into quantum-dynamical (QD) simulation techniques applied to proton transfer [3] and into a more coarse-grained level of modeling using whole amino acids as particles used to predict protein structure from data bases [4].

Having chosen the degrees of freedom to be modeled, one should consider which forces govern the motion along these degrees of freedom. The choice of interaction function or force field in a biomolecular modeling study is a very important one, since it will ultimately determine the quality of the results of the study. A variety of force fields is available [5].

3. Force-field Development

The potential energy functions of the various biomolecular force fields have slightly different forms. Their main differences reside in the force field parameters used. Since these can be obtained in a variety of ways, by fitting of a range of molecular properties of small molecules against different sets of quantum-mechanical and experimental data regarding these molecules, different parameter sets may yield widely different results. Over the past decades we have continuously improved and refined the parameters of the GROMOS (GROningen MOlecular Simulation) force field [6][7]. Since we are mostly interested in the properties of biomolecules in the condensed phase, the GROMOS force field is based on fitting thermodynamic properties of small compounds to experimental data. In particular, the solvent model should be compatible with the biomolecular force field used for the solute. This is illustrated in Fig. 2, which shows how well the latest GROMOS parameters for CH_n atoms reproduce the heat of vaporization and density of a series of alkanes and at the same time their free energy of solvation in water [7]. The further development of the GRO-

Fig. 1. Levels of modeling in computational biochemistry and molecular biology.

Methods	Degrees of freedom	Properties, processes	Time scale
Quantum dynamics	Atoms, electrons	Excited states, relaxation, reaction dynamics	Picoseconds
Quantum mechanics ⇒ ab initio ⇒ density functional ⇒ semi empirical ⇒ valence bond	Atoms, electrons	Ground states, reaction mechanisms	No time scale
Classical statistical mechanics ⇒ MD, MC ⇒ force fields	Atoms, solvent	Ensembles, averages, system properties, folding	Nanoseconds
Statistical methods ⇒ data base analysis	Groups of atoms, amino acid residues, bases	Structural homology, similarity	No time scale
Continuum methods ⇒ hydrodynamics	Continuum of a given type, electric, velocity, etc.	Rheological properties	Supramolecular
Kinetic equations	Populations of species	Population dynamics, signal transduction	Macroscopic

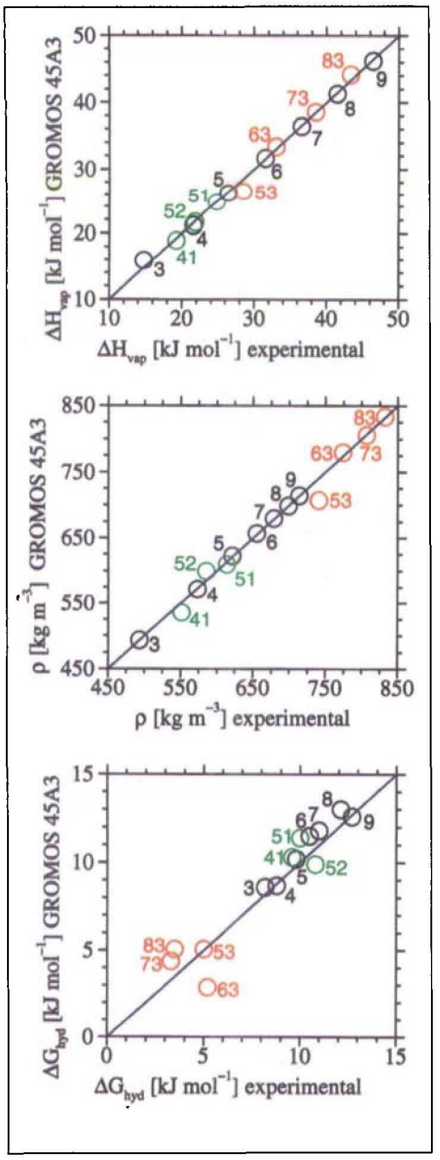

Fig. 2. Comparison of simulated (NPT) and experimental data (ΔH_{vap}: heat of vaporization; ρ: density; ΔG_{hyd}: free energy of hydration) for alkanes. The GROMOS force field version 45A3 was used [7]. Compounds: propane (3), butane (4), pentane (5), hexane (6), heptane (7), octane (8), nonane (9), isobutane (41), isopentane (51), neopentane (52), cyclopentane (53), cyclohexane (63), cycloheptane (73), cyclooctane (83). Black: linear alkanes; Green: branched alkanes; Red: cyclo-alkanes. Data taken from [7].

MOS force field [6][7] for biomolecules and typical solvents [8] and its application to proteins, DNA and liquid assemblies as test cases [9] takes a considerable effort of our group.

4. Development of Methodology

Most chemical quantities of interest, such as binding constants, solubilities, adsorption coefficients and chemical potentials, are directly related to the free energy. Since the development of computer simulation methods, attempts have been undertaken to efficiently calculate free energy differences using a variety of statistical-mechanical formulae and procedures. We have been investigating a number of techniques [10], of which the so-called one-step perturbation method using soft-core potentials and non-physical reference states turns out to be very efficient [11]. In Fig. 3 it is shown that a *single* MD simulation of a 0.6 nm radius neutral soft-core cavity in water yields

the free energy of hydration of a range of different (in size and shape), polar *and* non-polar molecules [11]. The technique is especially suitable for application in protein–ligand binding studies. Fig. 3 shows that experimental relative binding free energies for a number of estrogen receptor ligands can be reproduced using the GROMOS force field and the one-step perturbation technique (2.5 kJmol^{-1}) to within about RT [12].

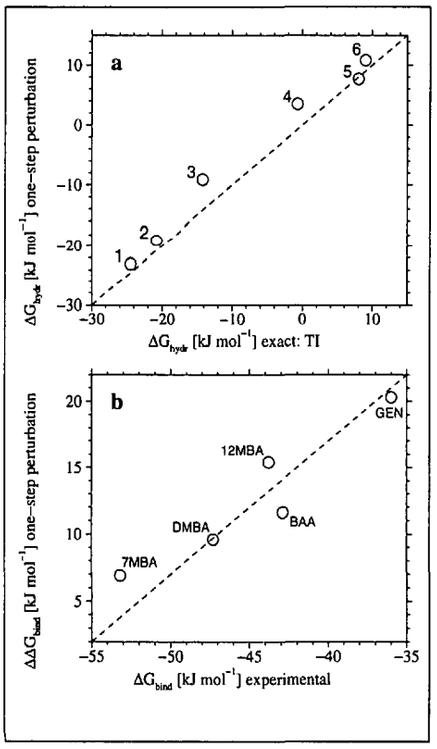

Fig. 3. a) Comparison of absolute free energies of hydration for a number of molecules using a *single* MD simulation of an unphysical reference state to exact results (TI = thermodynamic integration): H_2O (1), methanol (2), ethanol (3), chloroform (4), methane (5), propane (6). Data taken from [11]. b) Comparison of experimental and one-step perturbation results for the relative free energy of binding of ligands to the estrogen receptor ligand binding domain. Ligands: benz[a]anthracene-3,9-diol (BAA), 7,12-dimethyl-benz[a]anthracene-3,9-diol (DMBA), genistein (GEN), 7-methyl-benz[a]anthracene-3,9-diol (7MBA), 12-methyl-benz[a]anthracene-3,9-diol (12MBA). Data taken from [12].

Classical simulation has reached the stage where structural and thermodynamical properties of biochemical systems can now be reproduced and predicted with a reasonable degree of accuracy. However, where processes involve the breaking and making of chemical bonds or other forms of redistribution of the electron density, classical simulation is inadequate. Then, some sort of quantum-mechanical treatment of a limited number of degrees of freedom in the molecular system is required. We have investigated and tested different approaches in hybrid quantum-classical simulations of proteins such as HIV protease, metallothionein and *p*-hydroxybenzoate hydroxylase [13], and used them to study photoisomerization [14].

Another example of methodology development is the algorithm to perform MD simulation at constant pH presented in [15]. It generates a trajectory at a Boltzmann distributed ensemble of protonation states by a combination of MD and MC (Monte Carlo) simulation.

5. Interpretation of Experimental Data

The value of an observable for a molecular system that results from a measurement is generally an average over a Boltzmann ensemble of molecular conformations. This poses a problem with regard to the structural interpretation of experimental data. We have been investigating the effect of motional averaging with regard to NMR data [16], with regard to crystallographic X-ray data [17] and, recently, with regard to CD data [18]. The problem is illustrated in Fig. 4, where the distribution of particular atom–atom distances r_{ij} in a MD simulation of a β-heptapeptide is shown. According to NMR measurements the r_{ij}^{-6} average $\langle r_{ij}^{-6}\rangle^{-1/6}$ should be smaller than a given observed bound r^{obs}. Due to the r^{-6} averaging this experimental bound is easily fulfilled if only 25% of the trajectory conformations satisfy the bound. The molecule can be most of the time unfolded and yet the trajectory average $\langle r_{ij}^{-6}\rangle^{-1/6}$ satisfies the experimentally observed value r^{obs}, which, therefore, only conveys very limited information about the ensemble of conformers. In such cases MD simulation may help to investigate and analyze the different conformational ensembles that are compatible with experimental data [19].

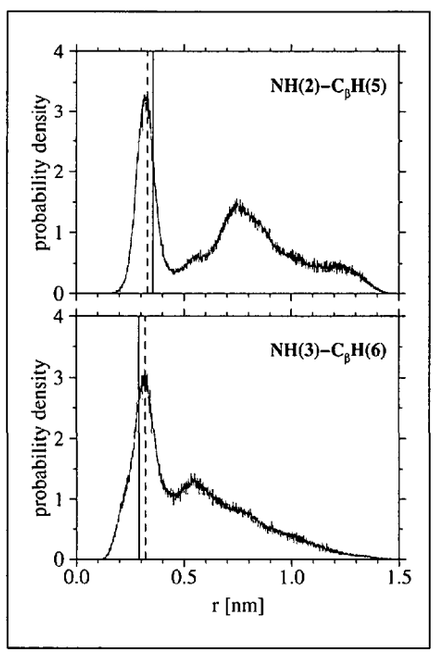

Fig. 4. Distribution of atom–atom distances from a 200 ns MD simulation of a β-heptapeptide in methanol at 340 K. Atom pairs NH(2)-HCβ(5) and NH(3)-HCβ(6). Residue number in parentheses. The dashed vertical line indicates the distance upper bound r^{obs} inferred from NMR measurements. The solid vertical line indicates the r_{ij}^{-6} average $\langle r_{ij}^{-6}\rangle^{-1/6}$ as obtained from the MD trajectory.

6. Understanding Molecular Processes

Molecular dynamics simulation provides a microscopic picture of unlimited resolution in time, space and energy. System parameters can be changed at will to study particular cause-effect relationships, leading to an enhanced understanding of biomolecular systems. A variety of systems, properties and processes have been studied in our group.

Fig. 5 illustrates the folding/unfolding equilibrium of a β-heptapeptide in methanol as observed in a 200 nsec MD simulation [20]. The folding pathways could be analyzed in atomic detail. Similar MD simulations of a variety of peptides show that the denatured state comprises much fewer conformations than expected from a simple counting of conformational degrees of freedom [21]. Fig. 6a shows that the number of conformations found along a trajectory grows as function of time, but in a sub-linear manner for the β-hep-

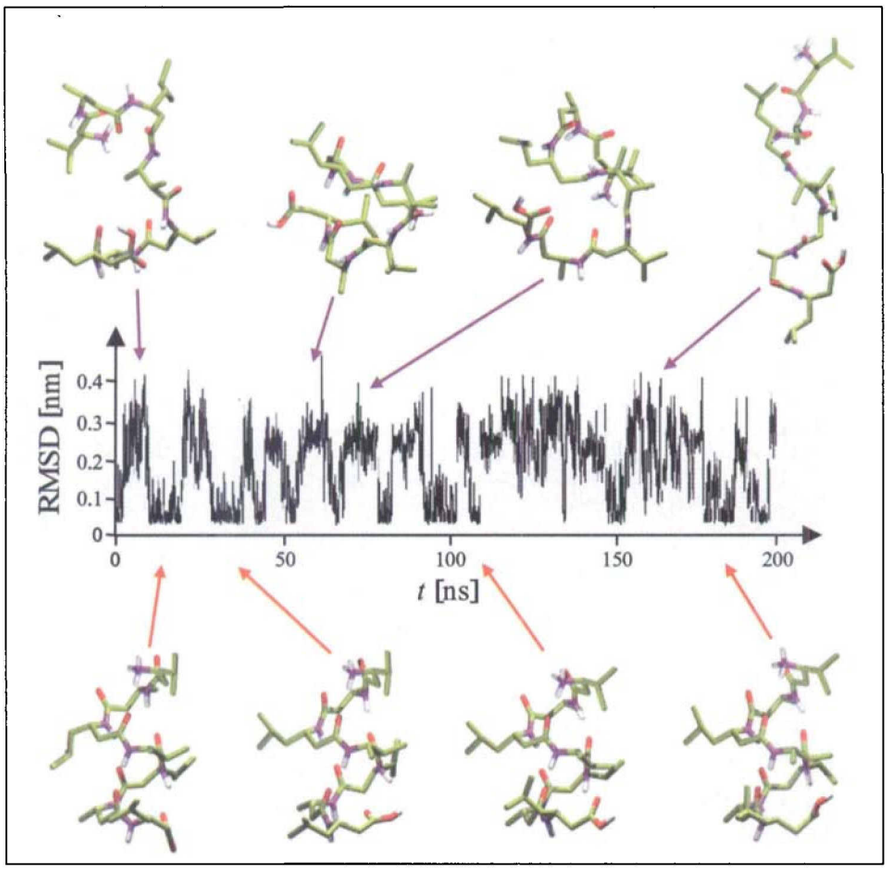

Fig. 5. The upper series of structures show snapshots of the various unfolded conformations at different time points along the 200 nsec simulation of a β-heptapeptide in methanol. The lower series shows the folded conformations. The helical conformations that are repeatedly adopted are also evident from the change in the root-mean-square deviation (RMSD) of the backbone atom positions (residues 2 to 6) during the simulation from the helical model structure derived from NMR experiments. Data taken from [21].

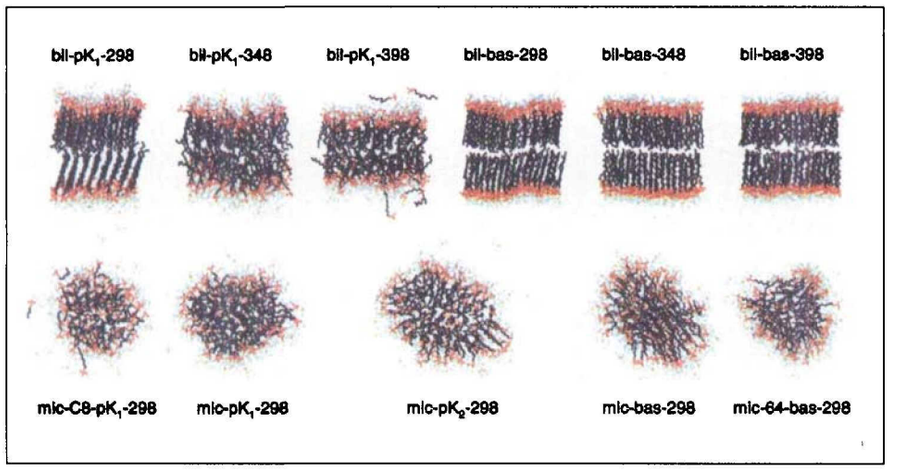

Fig. 7. *n*-Dodecylphosphate structures obtained at the end of simulations of systems containing lipids, water and counterions over 3 to 5 nsecs. Data taken from [30]. Lipid hydrocarbon tails appear in black, head groups in intermediate gray shades. Ions and solvent molecules close to them and to the lipids are shown in light colors. The display is laid out in five sections. Upper left: bilayer simulations at $pK_1 \approx 2$ and at 298 K, 348 K or 398 K. Upper right: bilayer simulations at basic pH = 11.2 and at 298 K, 348 K or 398 K. Lower left: unstable micelles (90 lipids) at pK_1 and at 298 K with different carbon chain lengths (8 and 12 units). Lower middle: micelle (90 lipids) and $pK_2 \approx 7$ and 298 K. Lower right: micelles (90 and 64 lipids) at basic pH and at 298 K. Thus, *e.g.* mic-C8-pK_1-298 means, the micellar system of 90 lipids with chain lengths of 8 carbons per lipid (default is always C_{12} = dodecylphosphate) was built as the structure that reflects pH = pK_1 and has been simulated at T = 298 K. Any bilayer system of 128 lipids is abbreviated by (bil) correspondingly, and for micelles the labels of systems different from 90 lipids contain the aggregation number.

tapeptide in methanol [20]. For a polyhydroxybutanoate molecule of similar length in chloroform, however, the number of conformations grows linear with time (Fig. 6b). The difference is probably due to the presence of hydrogen bond donor and acceptor atoms in the β-heptapeptide, which restrict the conformational space accessible to the molecule at the given temperature.

The role of entropy in the folding process has been investigated in [22]. It turns out that explicit simulation of solvent molecules is essential to obtain correct folding/unfolding equilibria [23]. The role of solvent viscosity on protein dynamics [24] and exchange pathways for water molecules bound inside a protein [25] were analyzed. Dielectric properties of proteins can be investigated using MD simulations [26]. The relative stability between protein mutants and between different DNA double helices can be reproduced by MD simulation [27].

Mechanistic questions regarding the photocycle of the photoactive yellow protein were answered through MD simulation [28]. Conformational search techniques can be used to analyze the flexibility of the active sites of enzymes with an eye to inhibitor design [29].

Finally, the stability of lipid bilayers and of micelles can be studied as function of pH. This is illustrated in Fig. 7 where snapshots from MD trajectories at different temperatures and pH values are shown [30].

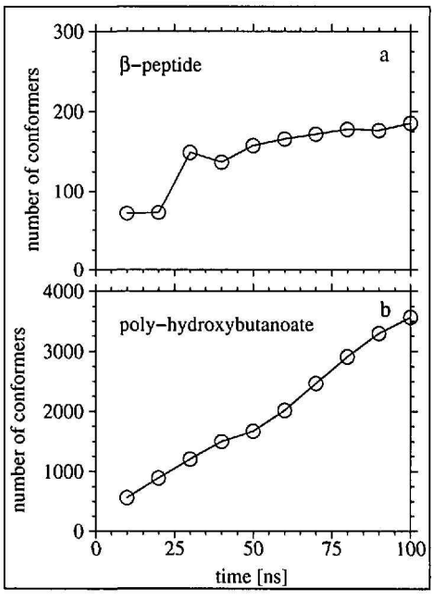

Fig. 6. Number of conformations as function of time: a) for a β-heptapeptide in methanol at 340 K [20]; b) for $(Val\text{-}Ala\text{-}Leu)_2$-3-hydroxybutanoate in chloroform at 298 K. For the definition of a conformation (cluster) we refer to [20].

7. Conclusion

Due to the continuing increase of computing power, MD techniques can be used to simulate ever larger systems over ever longer times. This means that their range of application is ever growing. However, only if the force field used is of high quality, can meaningful trajectories, ensembles and properties be obtained. As long as the accuracy of the biomolecular force fields continues to increase, ever more precise results will be obtained.

Acknowledgement

We thank the ETH Zürich for its continuing, generous support.

Received: July 17, 2001

[1] W.F. van Gunsteren, L.J. Smith, 'Molecular dynamics and stochastic dynamics simulations', to appear in NATO ASI Series (**2001**), (INSERM NATO School, 27 May–6 June 1996, Loutraki, Greece).

[2] W.R.P. Scott, P.H. Hünenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Krüger, W.F. van Gunsteren, *J. Phys. Chem.* **1999**, *A 103*, 3596.

[3] S.R. Billeter, W.F. van Gunsteren, *Comp. Phys. Comm.* **1997**, *107*, 61; S.R. Billeter, W.F. van Gunsteren, *J. Phys. Chem.* **1998**, *A 102*, 4669; S.R. Billeter, W.F. van Gunsteren, *J. Phys. Chem.* **2000**, *A 104*, 3276.

[4] P. Ulrich, W.R.P. Scott, W.F. van Gunsteren, A.E. Torda, *Proteins* **1997**, *27*, 367; F.A. Hamprecht, W.R.P. Scott, W.F. van Gunsteren, *Proteins* **1997**, *28*, 522.

[5] 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/Escom Science Publishers, Dordrecht, The Netherlands. **1997**, p. 3.

[6] 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, in 'Biomolecular simulation: The GROMOS96 manual and user guide', Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland. **1996**, p. 1; X. Daura, A.E. Mark, W.F. van Gunsteren, *J. Comput. Chem.* **1998**, *19*, 535; W.F. van Gunsteren, X. Daura, A.E. Mark, in 'Encyclopedia of computational chemistry', Vol. 2, Eds. P. von Ragué Schleyer, N.L. Allinger, T. Clark, J. Gasteiger, P.A. Kollman, H.F. Schaefer III, P.R. Schreiner, Wiley & Sons, Chichester, U.K. **1998**, p. 1211; L.D. Schuler, W.F. van Gunsteren, *Mol. Simulation* **2000**, *25*, 301.

[7] L.D. Schuler, X. Daura, W.F. van Gunsteren, *J. Comput. Chem.* **2001**, in press.

[8] I.G. Tironi, W.F. van Gunsteren, *Molecular Physics* **1994**, *83*, 381; H. Liu, F. Müller-Plathe, W.F. van Gunsteren, *J. Am. Chem. Soc.* **1995**, *117*, 4363; I.G. Tironi, P. Fontana, W.F. van Gunsteren, *Mol. Simulation* **1996**, *18*, 1; R. Walser, A.E. Mark, W.F. van Gunsteren, M. Lauterbach, G. Wipff, *J. Chem. Phys.* **2000**, *112*, 10450.

[9] A.M.J.J. Bonvin, M. Sunnerhagen, G. Otting, W.F. van Gunsteren, *J. Mol. Biol.* **1998**, *282*, 859; L.J. Smith, C.M. Dobson, W.F. van Gunsteren, *Proteins* **1999**, *36*, 77; U. Stocker, W.F. van Gunsteren, *Proteins* **2000**, *40*, 145; I. Chandrasekhar, W.F. van Gunsteren, *Current Science* **2001**, submitted; I. Chandrasekhar, W.F. van Gunsteren, *Europ. Biophys. J.* **2001**, submitted.

[10] P.E. Smith, W.F. van Gunsteren, *J. Chem. Phys.* **1994**, *100*, 577; T.C. Beutler, A.E. Mark, R.C. van Schaik, P.R. Gerber, W.F. van Gunsteren, *Chem. Phys. Letters* **1994**, *222*, 529; T.C. Beutler, D.R. Béguelin, W.F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 3787; H. Liu, A.E. Mark, W.F. van Gunsteren, *J. Phys. Chem.* **1996**, *100*, 9485; X. Daura, P.H. Hünenberger, A.E. Mark, E. Querol, F.X. Avilés, W.F. van Gunsteren, *J. Am. Chem. Soc.* **1996**, *118*, 6285; W.F. van Gunsteren, X. Daura, A.E. Mark, 'Computation of free energy', to appear in NATO ASI Series (**2001**), (INSERM NATO School, 27 May–6 June 1996, Loutraki, Greece); R. Bürgi, F. Läng, W.F. van Gunsteren, *Mol. Simulation* **2001**, in press; J.W. Pitera, W.F. van Gunsteren, *Mol. Simulation* **2001**, in press.

[11] H. Schäfer, W.F. van Gunsteren, A.E. Mark, *J. Comput. Chem.* **1999**, *20*, 1604; J.W. Pitera, W.F. van Gunsteren, *J. Phys. Chem. A* **2001**, submitted.

[12] C. Oostenbrink, J.W. Pitera, M.M.H. van Lipzig, J.H.N. Meerman, W.F. van Gunsteren, *J. Med. Chem.* **2000**, *43*, 4594.

[13] W.F. van Gunsteren, H. Liu, F. Müller-Plathe, *J. Mol. Structure (Theochem)* **1998**, *432*, 9; C.D. Berweger, W. Thiel, W.F. van Gunsteren, *Proteins* **2000**, *41*, 299; S.R. Billeter, C.F.W. Hanser, T.Z. Mordasini, M. Scholten, W. Thiel, W.F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2001**. *3*, 688.

[14] C.D. Berweger, W.F. van Gunsteren, F. Müller-Plathe, *Angew. Chem. Int. Ed.* **1999**, *38*, 2609.

[15] R. Bürgi, P.A. Kollman, W.F. van Gunsteren, *Proteins* **2001**, submitted.

[16] W.R.P. Scott, A.E. Mark, W.F. van Gunsteren, *J. Biomol. NMR* **1998**, *12*, 501; X. Daura, K. Gademann, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, *Angew. Chem. Int. Ed.* **1999**, *38*, 236; X. Daura, I. Antes, W.F. van Gunsteren, W. Thiel, A.E. Mark, *Proteins* **1999**, *36*, 542; R. Bürgi, X. Daura, A. Mark, M. Bellanda, S. Mammi, E. Peggion, W.F. van Gunsteren, *J. Peptide Res.* **2001**, *57*, 107; R. Bürgi, J. Pitera, W.F. van Gunsteren, *J. Biomol. NMR* **2001**, *19*, 305; C. Peter, X. Daura, W.F. van Gunsteren, *J. Biomol. NMR* **2001**, submitted.

[17] C.A. Schiffer, W.F. van Gunsteren, *Proteins* **1999**, *36*, 501.

[18] P. Krüger, W.F. van Gunsteren, X. Daura, *Eur. Biophys. J.* **2001**, to be submitted.

[19] D. Seebach, J.V. Schreiber, S. Abele, X. Daura, W.F. van Gunsteren, *Helv. Chim. Acta* **2000**, *83*, 34; M. Bellanda, E. Peggion, R. Bürgi, W.F. van Gunsteren, S. Mammi, *J. Peptide Res.* **2001**, *57*, 97; X. Daura, K. Gademann, H. Schäfer, B. Jaun, D. Seebach, W.F. van Gunsteren, *J. Am. Chem. Soc.* **2001**, *123*, 2393; X. Daura, B. Jaun, D. Seebach, W.F. van Gunsteren, A.E. Mark, *J. Mol. Biol.* **1998**, *280*, 925.

[20] X. Daura, W.F. van Gunsteren, A.E. Mark, *Proteins* **1999**, *34*, 269.

[21] W.F. van Gunsteren, R. Bürgi, C. Peter, X. Daura, *Angew. Chem. Int. Ed.* **2001**, *40*, 351; W.F. van Gunsteren, R. Bürgi, C. Peter, X. Daura, *Angew. Chem. Int. Ed.* **2001**, submitted.

[22] H. Schäfer, X. Daura, A.E. Mark, W.F. van Gunsteren, *Proteins* **2001**, *43*, 45.

[23] X. Daura, A.E. Mark, W.F. van Gunsteren, *Comp. Phys. Comm.* **1999**. *123*, 97; C. Peter, X. Daura, W.F. van Gunsteren, *J. Am. Chem. Soc.* **2000**, *122*, 7461.

[24] R. Walser, W.F. van Gunsteren, *Proteins* **2001**, *42*, 414.

[25] D. Bakowies, W.F. van Gunsteren, *J. Mol. Biol.* **2001**, in press; D. Bakowies, W.F. van Gunsteren, *Proteins* **2001**, submitted.

[26] J.W. Pitera, M. Falta, W.F. van Gunsteren, *Biophys. J.* **2001**, *80*, 2546.

[27] S. Voordijk, T. Hansson, D. Hilvert, W.F. van Gunsteren, *J. Mol. Biol.* **2000**, *300*, 963; W. Czechtizky, X. Daura, A. Vasella, W.F. van Gunsteren, *Helv. Chim. Acta* **2001**, in press.

[28] I. Antes, W. Thiel, W.F. van Gunsteren, *Biochemistry* **2001**, submitted.

[29] X. Daura, E. Haaksma, W.F. van Gunsteren, *J. Comp.-Aided Mol. Design* **2000**, *14*, 507.

[30] L.D. Schuler, P. Walde, P.L. Luisi, W.F. van Gunsteren, *Europ. Biophys. J.* **2001**, in press.