

# Validation of Molecular Simulation: An Overview of Issues

Wilfred F. van Gunsteren,\* Xavier Daura, Niels Hansen, Alan E. Mark,  
Chris Oostenbrink, Sereina Riniker, and Lorna J. Smith

[\*] W. F. van Gunsteren, S. Riniker, Laboratory of Physical Chemistry, Swiss Federal Institute of Technology, ETH, 8093 Zurich (Switzerland), E-mail: wfvgn@ethz.ch

X. Daura, Institute of Biotechnology and Biomedicine, Universitat Autònoma de Barcelona (UAB), 08193, Barcelona (Spain) and Catalan Institution for Research and Advanced Studies (ICREA), 08010, Barcelona (Spain)

N. Hansen, Institute of Thermodynamics and Thermal Process Engineering, University of Stuttgart, Pfaffenwaldring 9, 70569, Stuttgart (Germany)

A. E. Mark, School of Chemistry and Molecular Biosciences, University of Queensland, St. Lucia, QLD 4072 (Australia)

C. Oostenbrink, Institute of Molecular Modeling and Simulation, University of Natural Resources and Life Sciences, Vienna (Austria)

L. J. Smith, Department of Chemistry, University of Oxford, Inorganic Chemistry, Laboratory, South Parks Road, Oxford OX1 3QR (UK)

**Date:** June 18, 2017

**For submission to:** Angewandte Chemie

**Keywords:** molecular dynamics simulation, Monte Carlo simulation, experimental data, errors, pseudo-validation, correlation, causality

**Abbreviations used:** MD, molecular dynamics; SD, stochastic dynamics, NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RMSD, root-mean-square difference; RMSF, root-mean-square fluctuation; SPC, simple point charge.



## ***Abstract***

Computer simulation of molecular systems enables structure—energy—function relationships of molecular processes to be described at the sub-atomic, atomic, supra-atomic or supra-molecular level and plays an increasingly important role in chemistry, biology and physics. To interpret results of such simulations appropriately, the quality of the calculated properties must be evaluated. This depends on (1) the degrees of freedom simulated, (2) the accuracy of the molecular model, interaction function or force field, (3) the equations of motion, integration scheme or other method used to sample degrees of freedom as well as the degree of sampling, (4) the boundary conditions, (5) the simulation software, and (6) how the software is used. When validating simulations against values of experimental observable quantities  $Q^{exp}$ , one must also consider (1) the accuracy of  $Q^{exp}$ , (2) the accuracy of the function  $Q(\mathbf{r}^N)$  used to calculate  $Q$  based on a molecular configuration  $\mathbf{r}^N$  of  $N$  particles, (3) the sensitivity of the function  $Q(\mathbf{r}^N)$  to the configuration  $\mathbf{r}^N$ , (4) the relative time scales of the simulation and experiment, (5) the degree to which the calculated and experimental properties are equivalent, and (6) the degree to which the system simulated matches the experimental conditions. Validation by comparison between experiment and simulation is seldom straightforward. Experimental data is limited in scope and generally corresponds to averages over both time and space. A critical analysis of the various factors that influence the apparent degree of (dis)agreement between simulations and experiment is presented and illustrated using examples from the literature. What can be done to enhance the validation of molecular simulation is also discussed.



## **1. Introduction**

The computer simulation of molecular systems is playing a growing role in academic and industrial research [1-6]. This is due to its ability to probe motional details at a microscopic level coupled with the rapid rise in computing power. Using the molecular dynamics (MD) simulation method, for example, it is possible to study the behaviour of liquids, electrolytes and solutions of polymers, such as proteins, DNA, RNA, and polysaccharides. Lipid membranes, liquid crystals, amorphous solids, and in fact any combination of the above, such as solutions at surfaces or in confined spaces can be studied. Processes such as melting, adsorption, segregation, formation of molecular complexes, protein folding and denaturation can be analysed. Protein stability, host-ligand binding, and membrane permeability can be investigated not only qualitatively, but ever more quantitatively. Such studies lead not only to improved understanding and insight, but to practical results such as optimised proteins or materials. In short, computer simulations of molecular systems contribute to (1) the understanding and (2) the interpretation of experiment, (3) (semi-quantitative) estimates of observables that may be measured experimentally, and (4) the capability to represent molecular behaviour, i.e. values of observable and non-observable quantities  $Q$ , in spatial regions and time windows that are difficult or impossible to access by experimental probes [3].

The widespread use of molecular simulations and computer software makes their validity in regard to particular applications of central importance [7], not only for researchers performing simulations but for those using results in the literature. No matter what techniques are employed simulating the physical world always involves a degree of approximation, uncertainty and compromise [6]. Before addressing how to validate molecular simulation methodology, software and applications, one may ask whether validation of an answer to a question is always possible. This is definitively not the case. Some questions are based on unprovable premises or involve the choice of an arbitrary quantity as a reference. In other cases a correlation between the values of two observable quantities measured under very different conditions may be due to hidden variables and not be directly related to the properties considered. It is easy to confuse correlation with causality. In general, molecular simulation is best used to address questions that are clear and concrete. In such cases the answers from simulations can be normally well validated. Often however, results of molecular simulation can only be partly or incompletely validated. It is these cases that are the primary focus of this overview.

Validation in the current context is the confirmation of conclusions drawn from



molecular simulations by independent observations. It is based on the use of empirical facts and logic. Since the number of independent empirical facts at the microscopic, atomic or molecular, level of resolution is limited, the validation of the results of simulations of high-dimensional models, i.e. containing many variables and parameters, is almost always incomplete. Here, we analyse how and to which extent molecular simulations can be validated. In Section 2 six issues regarding the quality of a simulation are considered [7].

1. Theory and model governing the degrees of freedom relevant to the question investigated.
2. The interaction function or Hamiltonian used to obtain the forces on the atoms, particles, or along the modelled degrees of freedom.
3. Convergence of ensemble or time averages of (observable) quantities, their statistics and the extent of sampling of configurational (or phase) space.
4. Spatial and thermodynamic boundary conditions applied in the simulation of the finite system.
5. Reliability of the simulation software used.
6. How the simulation software is used.

The first four issues (the degrees of freedom simulated and theory applied, e.g. classical or quantum mechanics, the Hamiltonian function or operator, the equations of motion solved and the boundaries equations chosen) define the model ( $M$ ) used to simulate the system ( $S$ ).

Section 3 addresses the validation of results of molecular simulation using experimental data, i.e. values  $Q^{exp}$  measured for observable quantities  $Q$ . If the average values of  $Q$  as derived from experiment,  $\langle Q \rangle_{exp}$ , and in the simulation,  $\langle Q \rangle_{sim}$ , agree to a sufficient degree, the simulation can be considered validated by experiment. For this, the following six issues are relevant [8].

1. The accuracy of  $Q^{exp}$  values.
2. The accuracy of the function  $Q(\mathbf{r}^N, \mathbf{p}^N)$  relating an observable quantity  $Q$  that depends on the Cartesian coordinates  $\mathbf{r}^N \equiv (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$  and conjugate momenta  $\mathbf{p}^N \equiv (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N)$  of the  $N$  atoms or particles in the system, to the statistical-mechanical phase-space variables  $\mathbf{r}^N$  and  $\mathbf{p}^N$  of the system. Here the dependence of  $Q$  on the momenta  $\mathbf{p}^N$  is not considered in detail.
3. The possible insensitivity of the function  $Q(\mathbf{r}^N)$  to a variation of configurations  $\mathbf{r}^N$  or to the configurations  $\mathbf{r}^N$  dominating the configurational ensemble.



4. Almost all experimental measurements involve averaging over molecules and/or over a given time window.  $\langle Q \rangle_{sim}$  and  $\langle Q \rangle_{exp}$  may be averaged (denoted by  $\langle \dots \rangle$ ) differently with respect to time or spatial extent.
5. The quantity  $Q$  derived from simulation may differ slightly from the quantity  $Q'$  measured experimentally leading to  $\langle Q \rangle_{sim} \neq \langle Q' \rangle_{exp}$ .
6. Slightly different systems  $S$  and  $S'$  may be compared leading to  $\langle Q(S) \rangle_{sim} \neq \langle Q(S') \rangle_{exp}$ .

These twelve issues are illustrated with examples from our own work pertaining to a small set of observables  $Q$ , measurement techniques  $MT$  and systems  $S$ . They are chosen to be instructive. They relate to classical statistical-mechanical models ( $M$ ) that involve classical MD simulation, mostly applied to small molecules and peptides. This minimises problems with conformational sampling inherent in larger systems or systems treated quantum-chemically. The issues illustrated relate directly to other choices of  $Q$ ,  $MT$ ,  $M$  and  $S$ . For example, in quantum-chemical calculations the choice of degrees of freedom (e.g. which electrons), theory applied (e.g. Born-Oppenheimer approximation, time-independent Schrödinger equation), ways to solve the Schrödinger equation (e.g. extent of basis set and configuration interaction), boundary conditions (e.g. vacuum or a mean-field dielectric outside the molecule(s) considered), the sampling of configurational space of the nuclei, which experimental data to compare with, etc. also determine the quality of the calculation and thus are relevant to validation. In Section 4 ways to improve the validation of molecular simulations are discussed. Section 5 contains some conclusions.

There is a growing awareness of the importance of validation in maintaining the integrity and utility of both the scientific literature and various research data banks [9,10]. This is reflected in recent literature covering a range of disciplines [11-15]. The present article aims to enhance the awareness of validation issues in molecular simulation.

## **2. The Quality of a Simulation**

Any model or theory used in a computer simulation involves four choices (see Figure 2 in Ref. [16]). All involve assumptions and approximations.

1. Which degrees of freedom are explicitly modelled, i.e. varied in the model and which theory, e.g. quantum statistical mechanics, classical statistical mechanics or approximations thereof, governs the motion or the probability distribution along the chosen degrees of freedom?
2. How are the interaction between parts of the system and the corresponding forces



along the degrees of freedom of the model represented, e.g. a Hamilton operator or an interaction function? The functional form of the potential energy function, the way it depends on parameters, which data, i.e. observable or non-observable quantities, were used to calibrate the model and how this calibration is performed, all influence the quality of the model.

3. Which method is used to generate configurations of the system along the degrees of freedom of the model? Methods can be based on equations of motion, such as the Schrödinger equation to propagate wave functions, Newton's equations of motion or variations thereof (e.g. the Langevin equation) to propagate particles, or continuum equations (e.g. the Navier-Stokes equation) to propagate concentrations or particle fluxes. Configurational sampling methods such as Metropolis Monte Carlo (MC) or replica-exchange can also be used to obtain configurations with the proper statistical-mechanical weights. The sampling methodology should produce, or closely approximate, the statistical-mechanical configurational ensemble corresponding to the chosen thermodynamic boundary conditions. Configurations generated arbitrarily or energy minimisation will not have the appropriate Boltzmann weights to allow a proper comparison to experiment.
4. Which boundary conditions are chosen? Only a finite number of particles or degrees of freedom can be modelled. This implies that the system has a boundary, the conditions of which must be specified. These boundary conditions, e.g. embedding the system in a vacuum or in a lattice of copies of the system (so-called periodic boundary conditions), will affect the results of a simulation. The thermodynamic boundary conditions, such as constant energy or temperature, constant volume or pressure, constant particle number or chemical potential, pH, ionic strength also must be specified.

In addition to the formulation of a model, the implementation of the model in computer software and the way this software is used will influence the results. Below we present specific examples illustrating the principles underlying each of these aspects. Note, while computers may get faster and new simulation and analysis techniques allow calculations to be performed with ever higher accuracy, these six aspects will remain relevant to all simulation studies.

### ***2.1. Quality of Theory and Model Governing the Degrees of Freedom***



Molecular systems are normally represented in terms of nuclear and electronic degrees of freedom, governed by quantum statistical mechanics, or atomic, supra-atomic (united atomic) or supra-molecular degrees of freedom, governed by classical statistical mechanics (see Table 1 in Ref. [17]). In principle, a finer-grained model, e.g. a sub-atomic one representing nuclei and electrons, should yield more accurate predictions than a model at a coarser-grained level, e.g. an atomic, supra-atomic or supra-molecular one. This is illustrated for models of liquid water at ambient temperature and pressure in Table 1.

Table 1. Liquid state properties of water at 1 atm and 298 K as obtained from experiment and from MD simulations of three models, COS/D2 [18], SPC [19], and CGW [20].

Properties	Experiment	Sub-atomic level model (COS/D2, 1 H <sub>2</sub> O) 2014, 4 sites	Atomic level model (SPC, 1 H <sub>2</sub> O) 1981, 3 sites	Supra-molecular level model (CGW, 5 H <sub>2</sub> O) 2011, 2 sites
$\rho$ (kg m <sup>-3</sup> ) density	997	999	972	995
$\epsilon(0)$ dielectric permittivity	78.4	78.9	66.6	73.7
$\gamma$ (10 <sup>-3</sup> Nm <sup>-1</sup> ) surface tension	71.6	63.6	53.4	51.2
$C_p$ (J mol <sup>-1</sup> K <sup>-1</sup> ) heat capacity	75.3	88.9	68.9	83.2
$\alpha_p$ (10 <sup>-4</sup> K <sup>-1</sup> ) thermal expansion	4.2	4.9	9.0	26.0
$\kappa_T$ (10 <sup>-5</sup> atm <sup>-1</sup> ) compressibility	4.6	4.4	4.7	11.1
$D$ (10 <sup>-9</sup> m <sup>2</sup> s <sup>-1</sup> ) diffusion	2.3	2.2	4.2	6.9

Overall, the sub-atomic polarisable COS/D2 model [18] agrees better with experiment than the atomic SPC model [19]. This in turn performs better than the coarse-grained supra-molecular CGW model [20]. Accuracy comes at a cost. The atomic SPC water requires 100 times more computational effort than supra-molecular CGW water. The sub-atomic COS/D2 water is three times more expensive than simulation of atomic SPC water. However, a fine-grained model need not necessarily perform better than a coarse-grained one. Simulations of liquid water based on treating nuclear motion classically and the



electrons using quantum-mechanical density-functional theory [21-23] do not necessarily show better agreement with experiment for bulk water at ambient temperature and pressure than the COS/D2 and the SPC models based on an empirical force field and classical mechanics.

Some representations omit the degrees of freedom of the solvent (so-called implicit solvation models) and the intra-solute terms of the potential energy function are modified to account for mean solvent effects. These have been used to model peptides and proteins. However, the solvent may influence the internal dynamics leading to inappropriate results. This is illustrated by a comparison of the residence time of particular conformations of the Proline side chain in the cyclic peptide Antamanide (Table 2). When simulating this molecule without explicit solvent degrees of freedom using stochastic dynamics (SD) with a friction coefficient of  $19 \text{ ps}^{-1}$  to approximate the solvent chloroform, the residence time for Proline ring puckering (3 ps) was about 1/10 the time inferred from NMR (30 ps) [24]. Using the implicit solvation model this discrepancy could be redressed by increasing the friction coefficient to  $1000 \text{ ps}^{-1}$  or increasing the rotational barriers of the torsional-angle terms for the Proline rings by about  $2.5 \text{ kJmol}^{-1}$  ( $k_B T$ , where  $k_B$  is Boltzmann's constant and  $T$  the temperature). Neither solution can be justified physically. When Antamanide was simulated with chloroform represented explicitly, the puckering residence time was 24 ps (Table 2) [25]. The effect of solvent upon ring puckering was not well represented by the simple addition of stochastic and frictional forces on the solute atoms exposed to solvent. SD is a very simple approximation to implicit solvation and more sophisticated models exist. However, by definition no implicit solvent model can properly represent all the contributions made by individual solvent molecules. Such interactions, especially in the first solvation shells, play a fundamental role in phenomena such as the so-called hydrophobic effect. This is because entropic effects associated with the missing degrees of freedom cannot be represented in terms of the remaining degrees of freedom.



Table 2. Average residence times of Proline ring puckering in the cyclic peptide Antamanide as obtained from experiment, from implicit solvent stochastic dynamics (SD) simulations with different frictional coefficients and torsional-angle rotational barriers [24], and from molecular dynamics (MD) simulation in chloroform [25], using the GROMOS bio-molecular force field 37C4.

Simulation (stochastic or molecular) dynamics or experiment	GROMOS force field change	Friction coefficient ( $\text{ps}^{-1}$ )	Ring puckering residence time (ps)
NMR experiments			$\approx 30$
SD mean solvent	-	19	3
SD mean solvent	-	1000	25
SD mean solvent	Torsional barrier $k_B T$ up	19	25
MD explicit solvent	-	-	24

In some cases, the inclusion of additional degrees of freedom can be inappropriate. When performing classical MD simulations of molecules it can be tempting to model bond-stretching degrees of freedom as classical harmonic oscillators or using Morse functions. However, high-frequency bond vibrations are quantum-mechanical in nature. At ambient temperature and pressure they will primarily be in their quantum-mechanical ground state. In such cases, the most appropriate way to represent these degrees of freedom in a classical approximation is to use constraints and model the bonds as rigid [26].

Incorrect theory or modelling may yield irrelevant results. For example, the decomposition of the free-energy change corresponding to a change of the Hamiltonian of a system into two or more components that are defined spatially or with respect to individual terms (van der Waals, electrostatic) of the interaction function has no physical meaning. This is due to the definition of a free energy ( $F$ ) as the energy ( $U$ ) minus the entropy ( $S$ ) multiplied by the temperature ( $F=U-TS$ ). Since the entropy represents the extent of configuration space accessible to the system at a thermodynamic state point, it cannot meaningfully be split between or assigned to particular terms in a Hamiltonian, e.g. involving particular sets of particles or particular interaction terms. In contrast, the energy of a system may be decomposed, i.e. assigned to particular sets of particles or interaction



terms [27,28]. Other examples of faulty theory or modelling can be found in [7].

## **2.2. Quality of the Interaction Function or Hamiltonian**

The accuracy of a molecular model and interaction function or force field should be evaluated for as many different observable quantities or properties and for as many different types of molecules and environments as possible. Four categories of properties can be distinguished.

1. Thermodynamic properties such as
  - a. Heat of vaporisation
  - b. Excess free energy
  - c. Surface tension
  - d. Free energy of solvation, partitioning, mixing or binding
  - e. Heat capacity
  - f. Density
  - g. Isothermal compressibility
  - h. Thermal expansion
2. Dielectric properties such as
  - a. Dielectric permittivity
  - b. Molecular dipole moment, quadrupole moment
  - c. Debye relaxation time
3. Structural properties such as
  - a. Average atom positions or radial distribution functions ( $g(r)$ )
  - b. Molecular radius of gyration
  - c. Solvent accessible surface area
  - d.  $S^2$  NMR bond-vector order parameters
  - e. X-ray or neutron diffraction structure factors
  - f. Crystallographic refinement temperature factors
  - g. Degree of hydrogen bonding
4. Dynamical properties such as
  - a. Molecular diffusion
  - b. Various molecular rotational correlation times
  - c. Viscosity



A faithful representation of thermodynamic properties is important, as these constitute basic driving forces in a system. The dielectric permittivity of compounds in bio-molecular systems ranges from 2 to 80. This makes proper modelling of dielectric properties essential. Note, from a physical perspective structure follows energy. If the energetics are correct, the correct structure will follow. Calibrating models to have the correct energetics is crucial to maximise its applicability.

Unfortunately, the set of properties or quantities  $Q$  for which measured values  $Q^{exp}$  are available from experiment for validation is limited. This means that the number of different and independent values  $Q^{exp}$  is normally much smaller than the number of independent model parameters. This implies that the formulation and validation of a molecular model is an underdetermined problem. Just because a force field reproduces a wide range of experimental values  $Q^{exp}$  does not mean it is truly correct.

An example of such a case is the parametrisation of the GROMOS 53A6 parameter set [29]. While this parameter set was significantly better at reproducing experimental solvation free energies of the amino acid side chains in polar and a-polar solvents than early sets, used in protein simulations the prediction of secondary structure propensities was worse due to changes in intra-molecular interactions. Additional refinement was required [30].

Simulations involving different force fields may be used to detect possible force-field dependences in  $Q(r^N)$ . If a dependence is found, one or both of the force fields may be wrong. Observing no dependence does not prove either force field is correct.

### ***2.3. Sampling and Convergence and Statistics of Averaging***

To be able to use approaches from statistical mechanics to calculate experimental observables a molecular simulation must produce a configurational trajectory or a set of configurations that approximates a particular statistical-mechanical ensemble. The probability distribution of the configurations of the system should be Boltzmann-like for example. This can be achieved using MD or MC methods and tested (in part) by considering the distributions of particular quantities, e.g. the energy.

No simulation will ever sample the complete  $3N$ -dimensional configuration space of a complex molecular system of  $N$  particles. What one needs to know is whether those parts of configuration space accessible at the particular thermodynamic state point of temperature and pressure, are sufficiently well sampled to obtain converged ensemble averages. Since it is impossible to determine from the part of configuration space visited



whether there are other relevant parts of configuration space not yet visited, this question can never be answered definitively.

Even if a property or quantity  $Q(\mathbf{r}^N(t))$  (or  $Q(\mathbf{r}^N(\log(t)))$  which emphasises large values of  $t$ ) shows no change in time after a given period, this neither means that the simulation has reached equilibrium nor that the property  $Q$  is adequately sampled. For example, in protein simulations, it is customary to monitor the atom-positional root-mean-square deviation (RMSD) from the initial, X-ray or NMR derived, structure as function of time. A particular value of the RMSD for a molecule consisting of  $N$  atoms corresponds to a sphere in  $3N$  dimensions around the reference structure. The surface of this sphere rapidly becomes very large as the RMSD value increases. Thus a great variety of configurations can have identical RMSD values. Unless the RMSD is small, the protein may undergo major changes in conformation and energy without the RMSD changing significantly.

Another example of the effect of the mathematical properties of a function  $Q(\mathbf{r}^N)$  upon the ensemble average  $\langle Q \rangle$  and its statistics is the calculation of averages of exponential functions, e.g.  $\langle e^{+Q/k_B T} \rangle$  or  $\langle e^{-Q/k_B T} \rangle$ , e.g. with  $Q$  being an energy ( $E$ ) or an energy difference. The running average  $\langle e^{+E/k_B T} \rangle_t$ , i.e. the average over the trajectory configurations up to and including time  $t$ , will show positive jumps when an energy larger than any sampled previously is met. This is due to the fact that the exponential function greatly enhances the weight of configurations of larger energy. In contrast, configurations with a lower energy contribute little to the ensemble average. Since the likelihood of sampling a configuration with a higher energy than all those sampled previously decreases with time, the running average  $\langle e^{+E/k_B T} \rangle_t$  will show less and less such jumps. Optically, it may look like the ensemble average is converged. However, in reality if few high-energy configurations have been sampled, the apparent convergence simply reflects poor statistics. Since different quantities  $Q(\mathbf{r}^N)$  are different functions of the configuration of a system, their ensemble averages will have different convergence and statistical properties.

Trajectory averages will only be truly representative when the equilibration time (in MD) or number of configuration changes (in MC),  $\tau_{\text{equil}}$ , is larger than the relaxation time  $\tau_{\text{relax}}(Q)$  of the quantity  $Q$ ,

$$\tau_{\text{equil}} > \tau_{\text{relax}}(Q), \quad (1)$$

and when the sampling time,  $\tau_{\text{sample}}$ , is much longer than  $\tau_{\text{relax}}(Q)$ ,

$$\tau_{\text{sample}} \gg \tau_{\text{relax}}(Q). \quad (2)$$

If conditions (1) and (2) are not fulfilled, the running average  $\langle Q(\mathbf{r}^N) \rangle_t$  of the quantity  $Q(\mathbf{r}^N(t))$  will drift as function of time or show erratic behaviour due to the occurrence of rare



events [2]. The relaxation time  $\tau_{relax}(Q)$  may be long because:

1. The system jumps rapidly (but rarely) between alternative configurations. The flipping of Phenylalanine side-chain rings in a protein by  $180^\circ$  is fast (picoseconds) but occurs infrequently (microseconds). In such cases the trajectory average can be very sensitive to the number of events sampled.
2. The system configurations may change intrinsically slowly. In this case a trajectory average may display a continuous change as function of time.

The relaxation and dynamics of quantities calculated from trajectory configurations can be analysed by different means, see Ref. [7].

1. For a system at equilibrium one may monitor the time series of a quantity  $Q(\mathbf{r}^N(t))$ , its running average  $\langle Q(\mathbf{r}^N) \rangle_t$ , its fluctuations  $(\langle (Q(t) - \langle Q \rangle_t)^2 \rangle_t)^{1/2}$ , or calculate its autocorrelation function  $\langle Q(t')Q(t'+t) \rangle_t$ . The decay time of the autocorrelation function or the build-up rates of the trajectory averages give an indication of  $\tau_{relax}(Q)$ .
2. When starting from a non-equilibrium initial configuration, the rate different properties  $Q(\mathbf{r}^N(t))$  approach equilibrium can be used to estimate  $\tau_{relax}(Q)$ .
3. If different simulations of the same system started from different initial configurations and velocities under the same boundary conditions do not yield the same trajectory average  $\langle Q \rangle$ ,  $\tau_{relax}(Q)$  is larger than the simulation time.

An overview of the relaxation behaviour and convergence characteristics of trajectory averages for a variety of quantities  $Q$  that can be calculated for proteins in aqueous solution can be found in Ref. [31]. For a molecular property such as the diffusion coefficient or the rotational relaxation time, where it is possible to average over many identical molecules, converged values can be obtained from relatively short simulations. For properties for which the calculation depends on the whole system, such as the dielectric permittivity or the viscosity, such averaging is not possible and longer simulation periods are needed. Some properties, such as the entropy, are dependent on the full extent of phase space available to the system. These are not easily accessible through simulation as in principle an infinitely long simulation is required to scan the entire space [30,31].

Note, a quantity  $Q(\mathbf{r}^N)$  that consists of  $N_c$  different components  $Q_i(\mathbf{r}^N)$  added together,

$$Q(\mathbf{r}^N) = \sum_{i=1}^{N_c} Q_i(\mathbf{r}^N), \quad (3)$$

may appear converged, even if its components  $Q_i$  are not in equilibrium. An example is the



relaxation of the total potential energy of a system ( $Q$ ) compared to that of its components ( $Q_i$ ), e.g. electrostatic, Lennard-Jones, solute internal, solute-solvent, etc. energies. Figure 1 of [31] shows that the Lennard-Jones energy of a system consisting of a protein in water appears converged after about 500 ps. However, the intra-protein Lennard-Jones energy is still decreasing and the protein-water Lennard-Jones energy is increasing. In free-energy calculations it is easier to converge a free energy difference  $\Delta F = \Delta U - T\Delta S$  than the corresponding differences in energy,  $\Delta U$ , and entropy,  $\Delta S$ , due to energy/entropy compensation [32-35].

## ***2.4. Spatial and Thermodynamic Boundary Conditions***

In molecular simulation finite systems are considered. This means that how it is coupled to the outside world must be defined. This coupling can be spatial in nature, e.g. the way the spatial boundaries of the system and the forces across the boundaries are defined, or thermodynamic in nature, e.g. the way the coupling to a heat or pressure bath is defined.

The definition of the spatial boundary conditions may influence the results of a simulation, in particular for relatively small systems in which long-range forces are present and for properties possessing a long correlation length. The earliest protein simulations employed vacuum boundary conditions, i.e. the solvent degrees of freedom were omitted (implicit solvation). This leads to sizeable distortive effects: (1) Surface tension reduces the surface area, making the protein more spherical; (2) The dielectric permittivity of vacuum (1) is much smaller than that of water at ambient temperature and pressure (about 80) meaning that dielectric screening by the solvent is missing in vacuum. An alternative is to apply periodic boundary conditions, i.e. surrounding the computational box with an infinite number of identical copies of itself. This artificial periodicity also has distortive effects [36,37]: It induces an attractive force between like charges and a repulsive force between opposite charges. As a consequence, it reduces the magnitude of the free energy of ionic solvation. The size of the artefact is related to: (1) the size of  $\epsilon^{-1}$ , where  $\epsilon$  is the dielectric permittivity of the region of interest (the solute), (2) the ratio of the size of the solute to that of the periodic box, and (3) the overall charge of the solute [36,37]. The enforced periodicity may induce cooperative effects in a system, e.g. the artificial formation of a helical turn [38].

Thermodynamic boundary conditions such as the coupling to a temperature bath or to a pressure bath can be achieved [3], using (1) constraints, (2) stochastic techniques, (3) a first-order weak coupling technique, or (4) the (second-order) extended Lagrangian technique. As discussed in Ref. [3], each method has its advantages and disadvantages. Possible artefacts must be considered. Using stochastic techniques the individual particle



velocities are randomly modified at each time step such that a Maxwell velocity distribution is obtained. However, the continuity of the particle velocity trajectories is lost. In addition, artefacts caused by repeating sequences of pseudo-random numbers must be avoided [39]. Using weak coupling the average temperature of all particles (or large subsets of particles, e.g. solutes or solvent) is controlled, which may lead to an inhomogeneous temperature distribution within the set of particles when some particles are more prone to algorithmic (e.g. non-bonded cut-off) noise than others. When using the extended Lagrangian technique in the form of a Nose-Hoover thermostat without a chain of thermostats, the second-order coupling of the system to the heat bath induces oscillations in the potential energy of the system the period of which depends on the strength of the coupling to the heat bath, see Figure 3 of [40].

## 2.5. Reliability of the Simulation Software Used

The quality of simulation software depends primarily on the care with which it was constructed and tested by the software developers and their understanding of the molecular models and algorithms implemented. Testing of simulation software can be done on various levels within different communities of researchers [7].

1. Elementary algorithmic tests for MD simulation software based on classical mechanics.
  - a. It can be numerically tested whether the force on each particle is equal to the negative of the numerical gradient of the potential energy for each term of the interaction function or force field.
  - b. The total (kinetic plus potential) energy of the system without external forces or coupling to temperature or pressure baths should remain constant. That is, the root-mean-square fluctuation (RMSF) of the total energy,  $E_{tot} = E_{kin} + E_{pot}$ , of the system,

$$\Delta E_{tot} = (\langle (E_{tot} - \langle E_{tot} \rangle)^2 \rangle)^{1/2}, \quad (4)$$

should be small compared to the RMSF of the kinetic energy,  $\Delta E_{kin}$ , or of the potential energy,  $\Delta E_{pot}$  [37],

$$\Delta E_{tot} \ll \Delta E_{kin}. \quad (5)$$

The relative fluctuation of the total energy,

$$\Delta E_{tot} / \langle E_{tot} \rangle, \quad (6)$$

is not a useful criterion for energy conservation as  $E_{tot}$  is only determined up to



a constant. This implies that the quantity (6) will be arbitrarily large or small depending on the choice of the origin of the energy scale.

- c. The total momentum of the system should remain constant in the absence of external forces and coupling to temperature and pressure baths.
  - d. The total angular momentum of the system should remain constant in the absence of external forces, coupling to temperature and pressure baths and periodic boundary conditions.
2. Properties of the appropriate statistical-mechanical ensemble should be reproduced by the simulated trajectories.
  3. For simple systems and interaction functions analytical solutions of the equations of motion or analytical expressions for ensemble averages, distributions or equations of state for liquids may be available. Simulations should reproduce the results from such analytical theory.
  4. The simulation software should reproduce standard results for well-studied systems (e.g. liquid argon or water, or the protein hen egg white lysozyme) from the literature.
  5. Before releasing software to a broad range of users, researchers developing simulation software and methodology should use the software they develop to undertake research projects within their own groups in order to facilitate the detection of errors.
  6. By making the software including source code available to the scientific community, the extent of de facto testing is greatly enhanced.

Users of simulation software packages can test the quality of the software themselves by executing the types of tests mentioned above. Another possibility is to compare results obtained using different software packages and to analyse possible differences in results with respect to their origins in the implemented codes (for an example see [42]). This option is not easily executed in practice, since it requires either the source code or a detailed and correct description of it in a manual to be available or the same algorithms, force fields, boundary conditions and implementation details to be present in both simulation software packages. It is also important that the value or meaning of an input parameter set by the user does not change (without warning), because this greatly hampers the interpretation of simulation results.



## 2.6. Reliability of the Use of the Simulation Software

Any software can be miss-used and nonsensical results produced. Incompatible values for input parameters can be chosen or parameter values used that violate the range of applicability of the molecular model, force field or algorithm. A multifunctional simulation code for bio-molecular systems will contain, apart from the many force-field parameter values, many input parameters to be set by the user. For example, the simulation code of the GROMOS bio-molecular simulation package has over 100 input variables which may adopt two or more values. It is not possible for software developers to test all possible combinations of input values or prevent users choosing inappropriate combinations.

Most errors can, however, be avoided by a careful consideration of the physical laws involved and the physical and chemical characteristics of the molecular system of interest. For example, if the pressure of the system,  $P(t)$ , is calculated using the virial formula,

$$P(t) = 2/3(E_{kin}(t) - W(t))/V(t), \quad (7)$$

the pressure will depend on the kinetic energy  $E_{kin}(t)$  (which determines the temperature), the virial  $W(t)$  (representing the forces between atoms or particles in the system), and the volume  $V(t)$  of the (periodic) computational box. When coupling the system to a temperature bath and a pressure bath, the coupling to the former should be tighter than to the latter in order to avoid resonance of  $P(t)$  and  $E_{kin}(t)$  induced by Eq. (7). The simulation parameter settings should be chosen such that the relaxation time for the temperature,  $\tau_T$ , is smaller than the relaxation time of the pressure,  $\tau_P$ ,

$$\tau_T \ll \tau_P. \quad (8)$$

MD integration algorithms are based on a truncated Taylor expansion. Thus the time step  $\Delta t$  should be chosen much smaller than the shortest oscillation period in the system. This depends on the masses of the particles, the curvature of the potential energy surface (force field) and whether geometric constraints, e.g. for bond lengths or angles, are applied [41].

An example of a poor choice of input parameters [43] is analysed in Ref. [40]. When using temperature replica-exchange, the period between exchanges of configurational replicas,  $\tau_{RE}$ , should be larger than the relaxation time  $\tau_T$  for the temperature,

$$\tau_{RE} \gg \tau_T. \quad (9)$$

Otherwise the potential energy distribution upon which the replica-exchange criterion depends, becomes too narrow, which induces artefacts [40].



### 3. Validation of Simulation Using Experimental Data

The interpretation of experimentally measured data is not easy. In 1938 Enrico Fermi was awarded the Nobel Prize for Physics for his 1934 “discovery” of new elements [44]. His interpretation of the nuclear reactions induced by neutrons colliding with Uranium, i.e. a mixture of 99.3%  $^{238}_{92}\text{U}$  and 0.7%  $^{235}_{92}\text{U}$ , was based on the expected decay of  $^{238}_{92}\text{U}$ . In reality the decay of the “impurity”  $^{235}_{92}\text{U}$  dominated the experiments and Fermi’s “discovery” appeared erroneous in 1939.

The interpretation of experimentally measured data for observable quantities  $Q(\mathbf{r}^N)$  in terms of the distribution  $P(\mathbf{r}^N)$  of molecular configurations or conformations is also not straightforward and requires a specification of a configuration and a conformation. Here the statistical-mechanical definition of a configuration  $\mathbf{r}^N$  consisting of the Cartesian coordinates of the system is used. This applies e.g. to the solute and solvent molecules combined. We refer to a group of closely related configurations of a solute molecule that have a particular set of geometric features as a conformation. For example, all configurations of a solute that possess slightly varying bond lengths, bond angles and torsional angles within a limited range, e.g.  $15^\circ$  from each other, may be considered a single conformation. This can be done when the differences between the configurations that belong to this conformation have no effect on the phenomenon that is considered. There are a number of aspects that make a structural interpretation of experimentally measured data not straightforward [8].

First, experimentally measured values  $Q^{exp}$  are subject to uncertainty or error. Some examples are presented in Section 3.1.

Second, as discussed in Ref. [45], quantities  $Q(\mathbf{r}^N)$  that can be calculated from configurations  $\mathbf{r}^N$  can be either observable quantities  $Q^{obs}(\mathbf{r}^N)$  that are directly measurable experimentally, or non-observable quantities  $Q^{der}(\mathbf{r}^N)$  that are derived from  $\langle Q^{obs} \rangle_{exp}$  values by applying a given procedure,  $f$  [8],

$$Q^{der} = f(\langle Q^{obs} \rangle_{exp}). \quad (10)$$

Peak location and intensity from X-ray diffraction or NMR spectroscopic experiments are examples of observable quantities  $Q^{obs}$ . Molecular structure, torsional angles, hydrogen bonds, NMR order parameters or atom-atom radial distribution functions are examples of derived quantities  $Q^{der}$ . The latter incorporate the assumptions and approximations associated with the procedure  $f$  used to convert  $\langle Q^{obs} \rangle_{exp}$  values into  $Q^{der}$  values. Depending on the validity of the assumptions and the reliability of the approximations involved, the  $Q^{der}$  values obtained may in reality carry little experimental information. As illustrated in



[46], if an inappropriate procedure  $f$  is used, the structural interpretation of the  $\langle Q^{obs} \rangle_{exp}$  values can be flawed. For some non-observable, derived quantities  $Q^{der}$  the assumptions and approximations involved in the procedure  $f$  relating  $Q^{der}$  to  $\langle Q^{obs} \rangle_{exp}$  are of minor importance. In such cases, the non-observable quantities may be useful for validation. For example, in the absence of spin diffusion, measured nuclear Overhauser (NOE) peak intensities reflect atom-atom distance bounds. Likewise,  $S^2$  bond-vector order parameters can be derived from NMR spectra based on well-defined assumptions. When using such derived quantities for validation, one must make allowance for the uncertainty introduced into the  $Q^{der}$  values due to the procedure  $f$ . This is illustrated in Section 3.2.

Third, experimentally measured properties correspond in general to averages  $\langle Q \rangle$  of a quantity  $Q$  over both space and time, while  $\langle Q \rangle$  is an average over the distribution  $P(Q(\mathbf{r}^N))$  of  $Q$ -values depending on configurations  $\mathbf{r}^N$ . Due to the averaging inherent in the measurement it is usually not possible to determine the  $Q$ -distribution  $P(Q(\mathbf{r}^N))$  or the configurational distribution  $P(\mathbf{r}^N)$  from  $\langle Q \rangle_{exp}$ , see Section 3.3. Furthermore, representing the conformational distribution  $P(\mathbf{r}^N)$  by one conformation, e.g. derived from X-ray crystallography, may lead to inconsistencies when calculating alternative experimental quantities that have a different dependence on the configurational distribution  $P(\mathbf{r}^N)$ , see Section 3.4. Examples of this can be found in Refs. [47-49]. NOE intensities may result from different molecular conformations [47]. Seemingly contradictory data from X-ray and NMR studies of *cis-trans* interconversion induced by the protein Cyclophilin could be explained from the conformational distribution obtained by MD simulation [48]. In some cases small-angle X-ray scattering profiles calculated from simulated ensemble averages agree better with experiment than profiles calculated from single X-ray crystal structures [49].

A fourth source of uncertainty when interpreting experimental data is their scarcity. The number of quantities  $N_Q$  that can be measured for a molecular system is generally very much smaller than the number of degrees of freedom  $N_{dof}$  of the system. This makes the problem of determining the configurational or conformational distribution  $P(\mathbf{r}^N)$  from the set of  $\langle Q \rangle_{exp}$  values highly underdetermined [46,50,51].

In some experiments only a single quantity is measured, e.g. wide-angle X-ray scattering or atomic force microscopy. The ability to use such data in validation studies is limited as it is not possible to relate the measured values  $\langle Q^{obs} \rangle_{exp}$  to specific degrees of freedom within the model and because of the high probability of compensation of errors. A quantity that is relatively insensitive to the configurational distribution  $P(\mathbf{r}^N)$  is the small-angle X-ray scattering intensity, from which the radius of gyration of a molecule can be derived. As shown in Ref. [50] very different configurational distributions can give rise to a



given experimental observation.

When relating the average of a quantity  $Q$  over a given configurational distribution  $P(\mathbf{r}^N)$ , whether originating from simulation ( $\langle Q \rangle_{sim}$ ) or experiment ( $\langle Q \rangle_{exp}$ ), to the configurational distribution itself, three general cases can be distinguished, see Ref. [45].

1.  $\langle Q \rangle$  does not reflect the shape of  $P(\mathbf{r}^N)$ , as  $\langle Q \rangle$  is insensitive to configuration  $\mathbf{r}^N$ , see Section 3.2 for an example on  $^3J$ -couplings derived from NMR data [52].
2.  $\langle Q \rangle$  does not reflect the shape of  $P(\mathbf{r}^N)$ , as  $\langle Q \rangle$  is determined by rarely sampled configurations with low (insignificant) Boltzmann weights, see Section 3.3 for an example on CD spectra [53].
3.  $\langle Q \rangle$  reflects the configurations dominating  $P(\mathbf{r}^N)$ .

Only in case 3 does  $\langle Q \rangle_{sim}$  carry information relevant to the interpretation of  $\langle Q \rangle_{exp}$  at the molecular level and thus may be used for validation. Nevertheless, agreement between  $\langle Q \rangle_{sim}$  and  $\langle Q \rangle_{exp}$  may be obtained if

1.  $\langle Q \rangle$  is insensitive to  $P(\mathbf{r}^N)$ , that is,  $\langle Q \rangle_{sim}$  matches  $\langle Q \rangle_{exp}$  irrespective of the  $P(\mathbf{r}^N)$  being simulated [16].
2. There are compensating errors in the simulation procedure, model or experimental set-up [54].
3. The experimental data of interest,  $\langle Q \rangle_{exp}$ , have been used to bias the simulation or to parametrise the force field used.
4.  $\langle Q \rangle$  is sensitive to the configurational distribution  $P(\mathbf{r}^N)$ .

Again, only in case 4 can any agreement between  $\langle Q \rangle_{sim}$  and  $\langle Q \rangle_{exp}$  be used to validate the simulation and/or to interpret experiment. Failure to observe a correlation between simulated and experimental results can result if

1. the simulation is insufficiently accurate, i.e. (i) relevant degrees of freedom were omitted, (ii) the force field was insufficiently accurate, (iii) approximations made when generating the configurational ensemble were too crude, (iv) inappropriate spatial or thermodynamic boundary conditions or (v) faulty software was used, (vi) inappropriate use of software;
2. the measured  $\langle Q \rangle_{exp}$  values were inaccurate;
3. the function  $Q(\mathbf{r}^N)$  relating observable to configuration was inappropriate;
4.  $\langle Q \rangle_{sim}$  and  $\langle Q \rangle_{exp}$  were averaged differently with respect to time or spatial extent;



5. related but slightly different quantities  $Q$  and  $Q'$  were compared;
6. different molecular systems  $S$  and  $S'$  were compared.

Clearly, the interpretation of values of experimental observables in terms of molecular configurations is not straightforward. In particular, the degree of agreement between  $\langle Q \rangle_{sim}$  and  $\langle Q \rangle_{exp}$  is easily over-interpreted. This may lead researchers to make inappropriate claims regarding validity of a simulation (if there is an apparent correlation between simulation and experiment) or, equally problematic, to disregard simulation results unjustly (if no correlation is evident).

### 3.1. Uncertainty or Inaccuracy of Values of $Q^{exp}$

All experimentally measured or derived values  $Q^{exp}$  of a quantity  $Q$  have a degree of uncertainty  $\Delta Q^{exp}$ .  $\Delta Q^{exp}$  is often estimated based on a description of the procedures used in the experiment. An alternative way to infer  $\Delta Q^{exp}$  is to compare results from different researchers. A compilation of experimental values for the area per lipid in bilayers consisting of different phospholipids is given in Ref. [55] and shown in Figure 1.

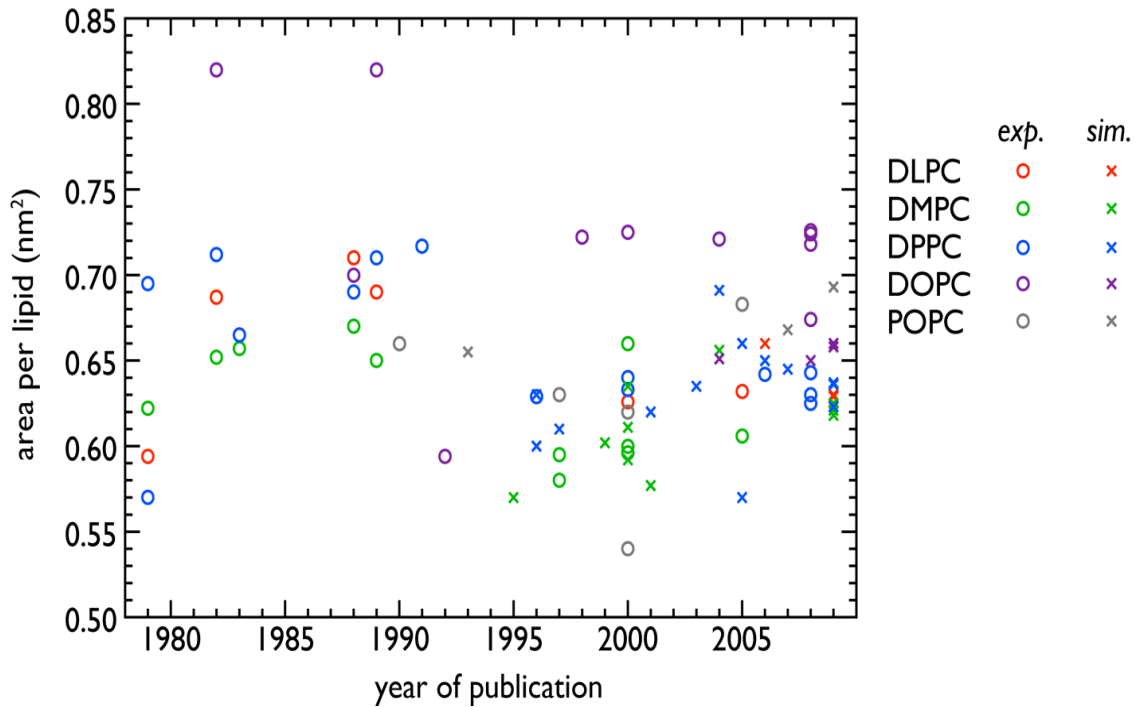


Figure 1. Area per lipid (in nm²) as obtained from experiment (circles) and from MD simulations (crosses) in the fluid phase of DLPC (red), DMPC (green), DPPC (blue), DOPC (purple), POPC (grey) bilayers [55].

The experimental estimates of the area per lipid of DPPC in the fluid phase vary between



0.57 and 0.73 nm<sup>2</sup> (about 25%) with an uncertainty of 0.15 nm<sup>2</sup>.

Another example is the difference in energy between the *trans* and the *gauche* conformer of *n*-butane in the liquid phase, the gas phase and calculated using *ab-initio* quantum-chemical methods [56]. This energy difference is used to calibrate force fields and is important as the value chosen will affect the structure and mobility of lipids. As seen in Figure 2, the variation between the measurements (1.2 kJmol<sup>-1</sup> or  $\frac{1}{2}k_B T$ ) is larger than the uncertainty reported in any of the studies. Figure 2 contains data published until the year 2000. Since then other studies, in particular quantum-chemical calculations aimed at determining the length at which unbranched alkane chains no-longer predominantly adopt an extended conformation, have appeared, see e.g. [57-59]. Due to the assumptions and approximations required to derive the *trans/gauche* energy difference for a single torsional angle from experimental or calculated data, the uncertainty is still about 2 kJmol<sup>-1</sup>.

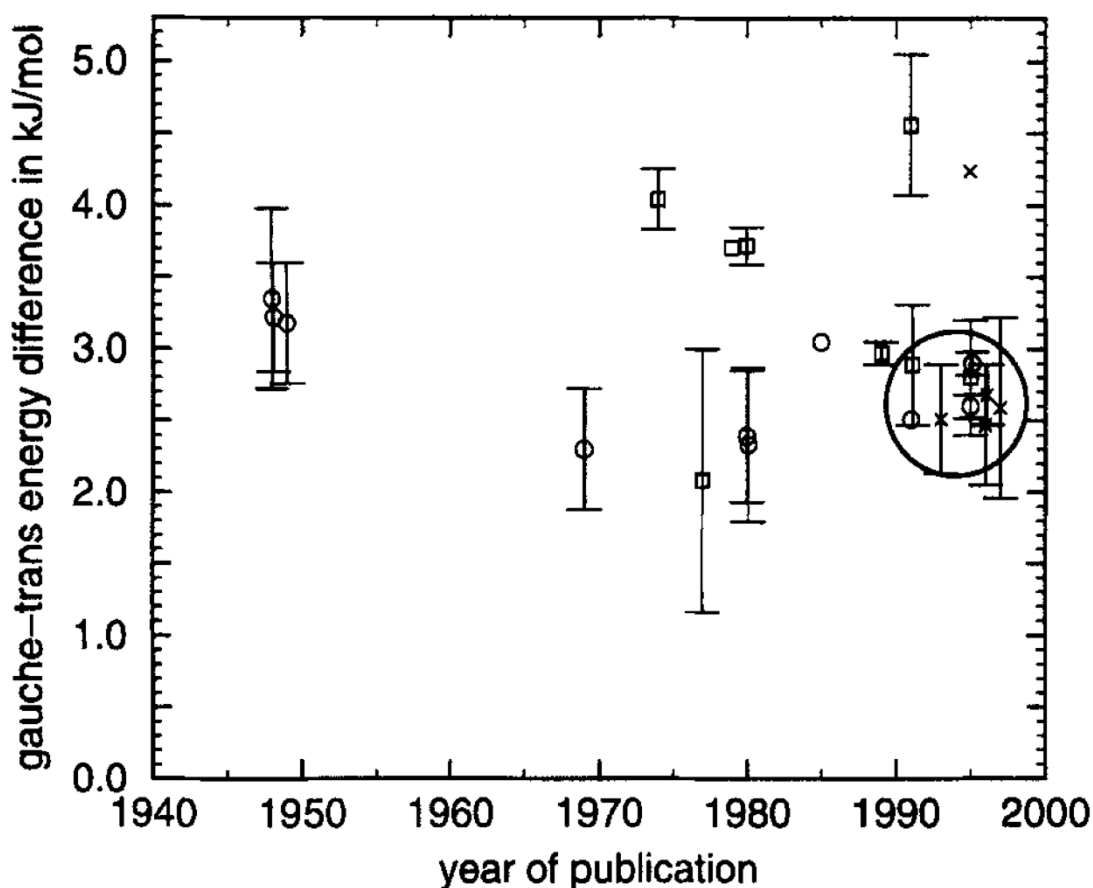


Figure 2. *n*-butane *gauche-trans* energy difference derived from experiments in the liquid phase (circles), gas phase (squares) and from *ab initio* quantum-chemical calculations (crosses). Error bars are shown whenever the data was available. The average of the encircled data was chosen as target value for the GROMOS 45A3 aliphatic torsional-angle parameter optimization. From Ref. [56] and references therein.



More examples can be found in Ref. [16]. In one, the variation in the free energy of a ligand binding to a protein is larger than the deviation between simulated and measured values. In another, atom-atom distance bounds inferred from NMR NOE intensities measured in 1993 and then again in 2001 converge towards the corresponding distances observed in a simulation of the protein hen egg white lysozyme [60].

### 3.2. Accuracy of the Function $Q(\mathbf{r}^N)$ Relating Observable $Q$ to Structure $\mathbf{r}^N$

A value  $Q(\mathbf{r}^N)$  for a particular configuration  $\mathbf{r}^N$  can be calculated using theory that connects an observable quantity  $Q$  to electronic and nuclear coordinates and properties of molecules. Such a function  $Q(\mathbf{r}^N)$ , even if based on quantum-chemical theory or a set of physically reasonable assumptions, will still involve approximations and rely on values for a given set of parameters. For example, the intensity  $I_{hkl}$  of a diffracted X-ray beam as function of the crystal lattice or Miller indices  $h$ ,  $k$ , and  $l$  is related to the square of the structure factor amplitude  $F_{hkl}(\mathbf{r}^N)$  obtained from a spatial Fourier transform of the electron density of the system [61]. NOE intensity  $I_{ij}(\mathbf{r}^N)$  for a pair of nuclear spins belonging to atoms  $i$  and  $j$  can be obtained from a relaxation matrix calculation [62].  $^3J_{ij}$ -couplings between atoms  $i$  and  $j$  connected by three covalent bonds constitute another example. These can be estimated from the torsional angle  $\theta_{ij}$  using the Karplus relation [63]

$$^3J_{ij}(\theta_{ij}) = a \cos^2(\theta_{ij}) + b \cos(\theta_{ij}) + c \quad . \quad (11)$$

The parameters  $a$ ,  $b$ , and  $c$  depend on the atoms involved in the  $^3J$ -coupling and their covalently bound neighbours. These are normally calibrated using experimental  $^3J$ -couplings thought to be associated with particular  $\theta$ -values in a set of reference structures. Figure 3 [64] shows  $^3J$ -coupling curves for various parameter sets available in the literature [65-69]. These show differences up to 3 Hz for particular angles  $\theta$ . The uncertainty with which  $^3J$ -couplings can be measured experimentally is in general lower than 0.5 Hz. However, using the Karplus relation introduces an uncertainty of 1-2 Hz which must be considered when using such data to validate simulation results.



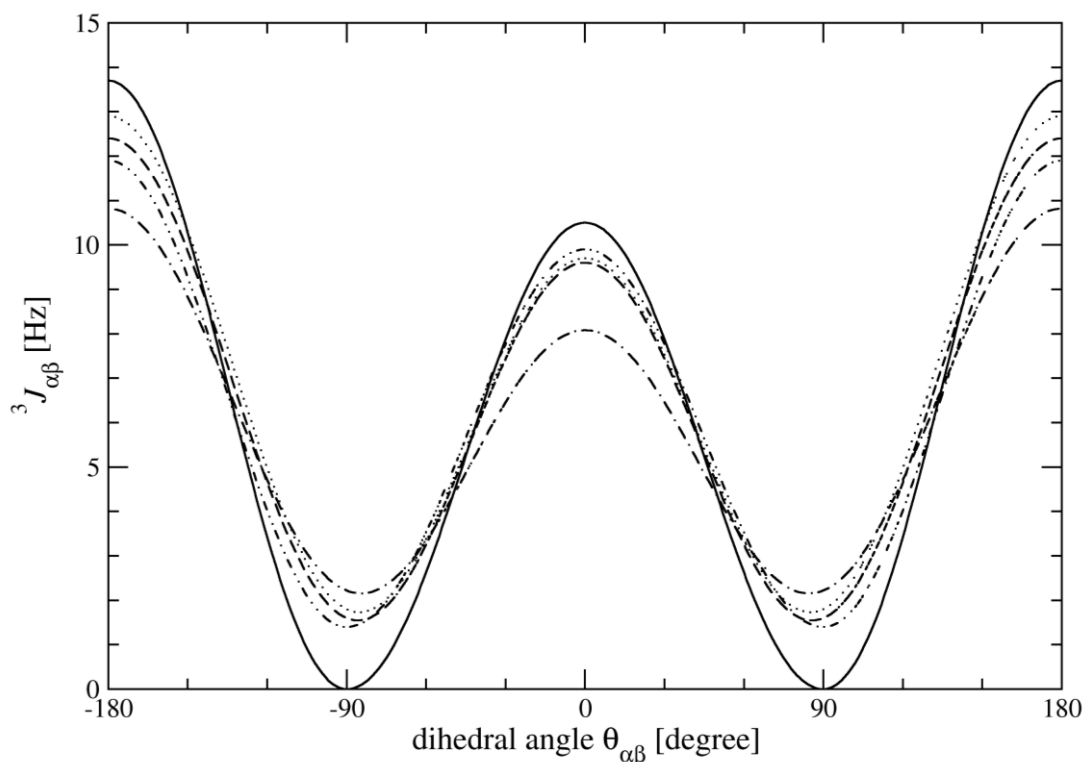


Figure 3. Karplus curves of Eq. (11) for  ${}^3J_{\alpha\beta}$  as function of the angle  $\theta_\beta$  using different values of the parameters  $a$ ,  $b$ , and  $c$  from the literature. Solid line: Ref. [65]. Dot-dot-dashed line: Ref. [66]. Dashed line: Ref. [67]. Dotted line: Ref. [68]. Dot-dashed line: Ref. [69].

### 3.3. Insensitivity of the Function $Q(\mathbf{r}^N)$ to a Variation of the Dominant Configurations $\mathbf{r}^N$

In some cases the quantity  $Q(\mathbf{r}^N)$  is insensitive to differences in the underlying conformational distribution. For example, the conformational distributions of a  $\beta$ -hepta-peptide generated at temperatures 298K, 340K, 350K and 360K were used to estimate average  ${}^3J$ -couplings. Despite the conformational distributions at the four temperatures being very different containing 97%, 50%, 39% and 25%  ${}^3J_{14}$ -helical structures, respectively [52], the agreement with experiment was of equal quality [16].

A quantity  $Q(\mathbf{r}^N)$  can also be dominated by a small fraction of the configurational ensemble. For example, the circular dichroism (CD) spectra of two  $\beta$ -hexa-peptides in methanol solution were found to be very similar, even though they were designed to have very different conformational ensembles [53]. Using NMR one peptide was shown to form a  ${}^3J_{14}$ -helix. The other was doubly methylated at the  $\alpha$ -carbon atom. This was expected to prevent helix formation. To resolve this, 100 ns MD simulations of each molecule were performed and average CD spectra calculated from the trajectories. The calculated spectra were similar to those measured experimentally. For the peptide methylated at the  $\alpha$ -carbon the CD spectrum was dominated by a conformation present for only 13% of the time. This



result depends on the accuracy of the method used to calculate the CD spectrum (function  $Q(\mathbf{r}^N)$ ) given a molecular structure, and illustrates the difficulty of interpreting experimental CD spectra in terms of a conformational model, especially for small peptides. If the value of a quantity  $Q$  is dominated by a minority of configurations in the ensemble, the utility of this quantity when validating an ensemble generated in a simulation is limited.

### 3.4. Different averaging in $\langle Q \rangle_{sim}$ and $\langle Q \rangle_{exp}$

A value  $Q_{exp}$  of a quantity  $Q$  obtained using a particular measurement technique  $MT$  will be an average over a specific time. Different experimental quantities can be sensitive to different time windows. For example, NMR implicitly involves time averaging due to the finite duration of the radio-frequency pulses and mixing times as well as the nature of the underlying phenomenon, e.g. magnetisation transfer. The time window of sensitivity for NOE measurements depends on the rotational tumbling time and thus on the size of the molecule of interest and the solvent environment. In order to directly compare  $\langle Q \rangle_{sim}$  with  $\langle Q \rangle_{exp}$  the time windows over which  $Q$  is averaged should be comparable. This is not the case when results from many short MD simulations are compared to experimental protein folding rates [70,71]. Such comparisons are problematic not only due to the different time scales in the simulation and experiment, but also because different definitions of when the protein is folded are used and assumptions must be made regarding the probability of folding at long time scales [51].

### 3.5. Differently defined but related quantities $Q$ and $Q'$

Even if two quantities  $Q$  and  $Q'$  are both related to a particular phenomenon or observation, if they are defined in different ways, their averages  $\langle Q \rangle_{exp}$  and  $\langle Q' \rangle_{exp}$  or  $\langle Q \rangle_{sim}$  and  $\langle Q' \rangle_{exp}$  may differ. Often they should not be the same.

For example, the free enthalpy of folding or denaturation of a protein can be derived from experiment by changing the thermodynamic conditions in different ways. In thermal denaturation experiments, measurements made in the narrow temperature range where unfolding occurs are then extrapolated to ambient temperature, often using a modified Gibbs-Helmholtz equation [72],

$$Q = \Delta G_u(T) = \Delta H_m(1 - T/T_m) + \Delta C_p[(T - T_m) - T \ln(T/T_m)], \quad (12)$$

where  $\Delta G_u(T) = \Delta H_u(T) - T\Delta S_u(T)$ ,  $\Delta H_u(T)$ ,  $\Delta S_u(T)$  and  $\Delta C_p(T)$  are the free enthalpy ( $G$ ), enthalpy ( $H$ ), entropy ( $S$ ) and heat capacity ( $C_p$ ) differences between the denatured state and the native state at temperature  $T$ ,  $\Delta H_m$  is the value of  $\Delta H_u$  at  $T_m$  and  $T_m$  is the midpoint



for thermal denaturation. Eq. (12) assumes that the change in heat capacity between folded and unfolded state,  $\Delta C_p$ , to be independent of temperature. Alternatively, if a chemical denaturant is used, extrapolation to zero denaturant concentration is required, e.g. by assuming a linear dependence of the observed free enthalpy on the concentration  $[c]$ ,

$$Q' = \Delta G_u([c]) = \Delta G_u([c] = 0) + m[c] \quad (13)$$

The quantities  $T_m$  and  $\Delta G_u([c] = 0)$  both of which reflect fold stability were measured for 15 single-residue mutants in which each individual backbone NH was replaced by oxygen in the WW domain of the Pin1 protein [73]. Figure 4 shows that the change in melting temperature upon mutation,  $-\Delta T_{mw}^m$ , is not well correlated with the change in the free enthalpy of folding on mutation inferred from varying the Gd-Cl concentration,  $\Delta\Delta G_{mw}^c$ . This is unsurprising as the nature of the unfolded state induced by raising the temperature may differ from that induced by the co-solvent. One should also not expect the free enthalpy of folding  $Q''$  as derived from a simulation at a given thermodynamic state point (constant temperature and zero or constant co-solvent concentration) by calculating the ratio of folded versus unfolded conformations, counted according to a particular geometric criterion, to be equal to  $Q$  or  $Q'$  as defined in Eqs. (12) and (13), respectively [74].

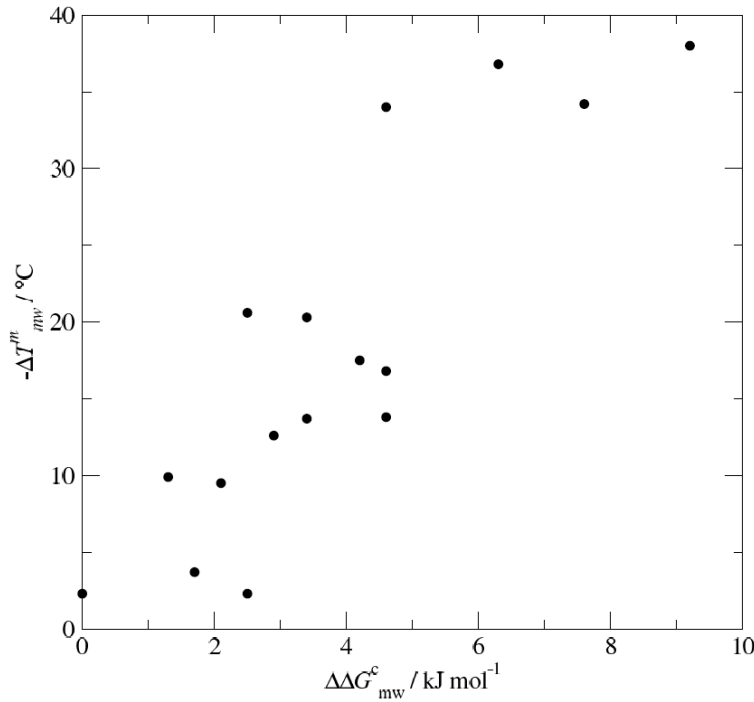


Figure 4. Differences  $\Delta T_{mw}^m$  between midpoints  $T_m$  of thermal denaturation curves, versus differences  $\Delta\Delta G_{mw}^c$  of folding free enthalpy differences as obtained by chaotrope



denaturation experiments [73] for 15 mutants ( $m$ ) and the wild type ( $w$ ) Pin1 WW domain [74].

Another example of related but differently defined quantities  $Q$  and  $Q'$  are the crystallographic B-factor  $B_i$  derived from X-ray diffraction intensities by fitting the radius of the electron density at position  $i$  to match the corresponding electron density profile, and the mean-square atom-positional fluctuation  $MSF_i$  of atom  $i$  obtained from a simulation.  $B_i$  constitutes an average for a specific position  $i$  over different atoms passing through this position, while  $MSF_i$  is an average for a specific atom  $i$  over its trajectory of positions. This means that the relation

$$B_i = (8\pi^2/3) MSF_i \quad (14)$$

will only hold for small amplitude atomic motion.

### **3.6. Similar but Slightly Different Molecular Systems $S$ and $S'$**

Even if the composition of two systems  $S$  and  $S'$  is only slightly different, e.g. by a few atoms, the values of  $\langle Q(S) \rangle$  and  $\langle Q(S') \rangle$  may still differ significantly. Figure 5 shows the backbone atom-positional RMSD from the experimentally observed  $3_{14}$ -helical fold of a  $\beta$ -hepta-peptide with only hydrophobic side chains in methanol as function of simulation time. Results for three different ionisation states are shown. The most probable ionisation state according to the estimated proton concentration in the methanol solution is the doubly protonated state [75]. At low pH both termini are protonated. At intermediate pH the N-terminus is deprotonated and the C-terminus is still protonated. At high pH both termini are deprotonated. The  $3_{14}$ -helical fold is the only dominant conformation at low pH corresponding to the experimental conditions. At the other pH conditions the  $3_{14}$ -helical fold is barely sampled. Similar differences may be observed for a protein in a crystal versus in solution.



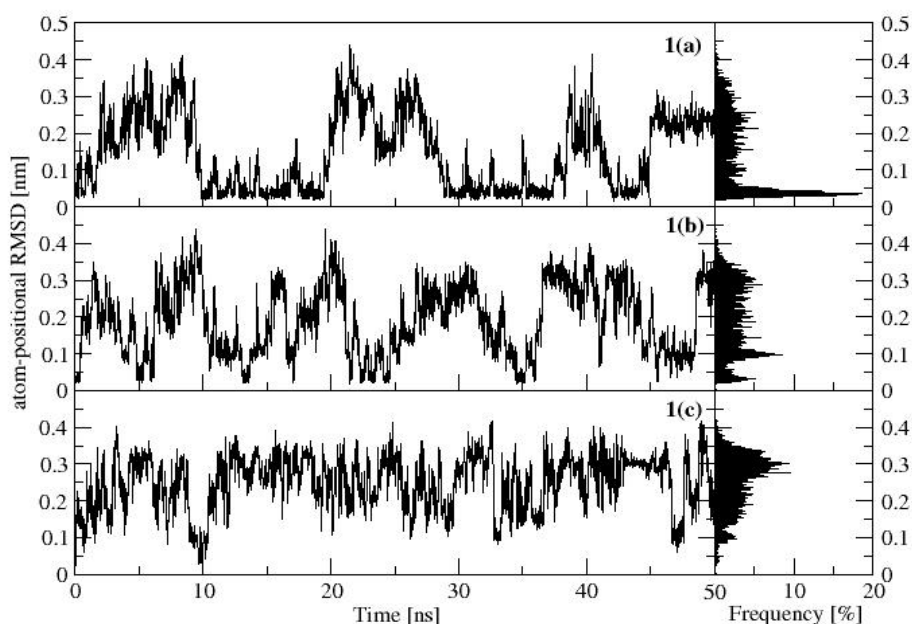


Figure 5. Time course and distribution of the atom-positional root-mean-square deviation (RMSD) from the  $3_{14}$ -helical conformation of a  $\beta$ -hepta-peptide in methanol simulated in different ionisation states [75]. Top: low pH, corresponding to experiment. Middle: high pH. Bottom: intermediate pH.

## 4. Discussion

In Sections 2 and 3 a range of issues that must be considered when validating a molecular simulation were categorised and illustrated. There are, however, other sources of artefacts [76] and ways validation can be affected.

### 4.1. Simulation Artefacts

When studying protein-protein or protein-DNA interfaces and interactions, it is tempting, in view of the many degrees of freedom to be considered, to employ an implicit solvent model. However, water molecules at surfaces of bio-molecules do not show bulk water properties. They may for example play a structural role in molecular complexes [77]. Using an implicit solvation model in such cases may introduce artefacts.

Due to the abundance of structures in the Protein Data Bank (PDB) [78], it can also be tempting to calibrate force-field parameters by fitting the structure of peptide fragments or even whole proteins to structures deposited in the PDB. Not only are the structures of



proteins and other molecules derived, not measured data, but the collection of structures in the PDB is not a valid set of structures from a statistical-mechanical perspective. The structures in the PDB lack proper Boltzmann weights and their inclusion in the PDB is somewhat arbitrary. The distribution of conformations in the PDB is not representative of all possible conformations. This may lead to artefacts when using force fields based on PDB data.

The cut-off radius applied when evaluating non-bonded interactions is a force-field parameter which, if varied, may hamper comparison between simulations. More generally, if the simulation settings (description of long-range interactions, reaction-field versus lattice sums, twin-range, atomic or group-based cut-off, tin-foil versus non-conducting boundary conditions, (multi-step) integration algorithm, time step, etc.) are very different from those used in the force-field parametrisation, artefacts may arise. The TIP4P-EW water model, a revision of the earlier TIP4P model for use with lattice-sum electrostatics, serves as an example [79].

As noted previously, treating covalent bonds as classical harmonic oscillators instead of constraints may introduce non-physical effects in a simulation, because of their quantum nature [26]. A more severe approximation is to restrict the range of atom-atom distances between non-covalently bound atoms by using harmonic springs, as done in so-called elastic-network models. Long-distance interactions between atoms are not harmonic and using such models the protein motions are unrealistic.

One may be tempted to choose a small computational box in order to reduce the computational effort. However, the smaller the box, the more the results of a simulation will be determined by the boundary conditions chosen [76].

#### **4.2. Compensation of Errors**

Compensation of errors may occur in the simulation model, procedure or experimental set-up [54] obfuscating flaws. For example, a crystal of a cytidine derivative was simulated at constant pressure (1 atm) at low ( $T = 113$  K) and at room ( $T = 289$  K) temperature. The lengths of the three axes  $a$ ,  $b$ , and  $c$  of the crystal unit cell and its volume are compared to the corresponding experimental values in Table 3. The almost perfect agreement with experiment for the volume results from a compensation of errors. At  $T = 113$  K the  $a$ - and  $b$ -axes are too short and the  $c$ -axis too long (-1.3%, -0.5%, +1.7%), while at  $T = 289$  K the  $a$ -axis is too long and the  $b$ - and  $c$ -axes are too short (+0.7%, -0.8%, -0.9%). The relative errors compensate each other when calculating the volume by multiplying  $a$ ,  $b$ , and  $c$ . Such compensation of errors is most likely to occur for complex models that map a



large number of variables and parameters on to one or a few quantities.

Table 3. Axes lengths and volume of 4 unit cells of a crystal of a cytidine derivative at 1 atm and two temperatures as obtained from experiment and from MD simulation [54, 80].

	T = 113 K		T = 289 K	
4 unit cells	experiment	MD	experiment	MD
$2a$ (nm)	1.788	1.765	1.778	1.791
$2b$ (nm)	2.065	2.055	2.074	2.057
$c$ (nm)	2.758	2.806	2.864	2.837
Volume ( $4abc$ ) ( $10^3$ nm <sup>3</sup> )	10.18	10.18	10.56	10.47

Another example of compensation of errors is the use of the MARTINI Lennard-Jones (LJ) coarse-grained (CG) water model employed in CG MD simulations of lipid membranes, bilayers and vesicles. The parameters  $\epsilon$  (5.0 kJmol<sup>-1</sup>) and  $\sigma$  (0.47 nm) of the LJ potential energy function in the MARTINI water model [81] are such that the system at room temperature (298 K) and pressure (1 atm) would be in the solid state according to the known LJ phase diagram [82] (see also Figure 5 of [83]). As is analysed in Fig. 1 of [83], increasing the time step  $\Delta t$  in an MD simulation at constant volume and constant total energy leads to noise due to the finite difference integration algorithm reflected in the non-conservation of the total energy. At  $\Delta t = 20$  fs the fluctuation in the total energy (which should be zero) is larger than the fluctuation in the potential or kinetic energy. The figure also shows that for larger  $\Delta t$  the average velocities are lowered and the integration error in the leap-frog scheme is rapidly increasing. However, when coupling the system to temperature and pressure baths, as is standardly done in MD simulations, the finite time-step integration error is compensated by a heat flow from the temperature bath and a heat flow towards the pressure bath, see Fig. 3 of [83]. The analysis of [83] shows that the time step should be  $\Delta t \leq 10$  fs in order to avoid compensating heat flows in and out of the system.

#### **4.3. Pseudo-Validation Based on the Use of the Same Data for Calibration and Testing**

In some studies, [86-89], the experimental data  $\langle Q \rangle_{exp}$  that were used to bias the simulation when generating a conformational distribution  $P(\mathbf{r}^N)$ , were subsequently also used to compare  $\langle Q(P(\mathbf{r}^N)) \rangle_{sim}$  with  $\langle Q \rangle_{exp}$ . Such comparisons do not validate the simulation nor demonstrate that the conformational distribution  $P(\mathbf{r}^N)$  reflects experiment. In fact, restraining  $Q(\mathbf{r}^N)$  in the simulation to match  $\langle Q \rangle_{exp}$  can introduce artefacts in  $P(\mathbf{r}^N)$ , even



though the correct average  $\langle Q \rangle_{sim} = \langle Q \rangle_{exp}$  is maintained, enhancing the likelihood of inappropriate conclusions being drawn. An example is the underestimation of the mobility of a side chain of the protein Tendamistat due to the use of a restraining function that does not account for averaging of configurations [90].

Protein structures in the PDB are derived from experimental X-ray diffraction or NMR spectroscopic data using a particular model or force field for the bond lengths and bond angles of a molecule. These models are reflected in the structures obtained. Using the same force field in a simulation may enhance the agreement with the PDB structures. However, this cannot be considered validation as the calibration and test data are the same.

#### ***4.4. Pseudo-Validation Based on Comparison to Non-Observed Data***

Ref. [50] illustrates the difference between comparing simulation results to observed data (X-ray intensities) as opposed to derived data (radius of gyration). There are ample examples in the literature of the inappropriate use of non-observed or derived quantities  $Q^{der}$  in validation studies. In Ref. [91] fluorescence measurements on the Trp-cage protein are used to infer changes in enthalpy. These are then used to infer changes in the number of hydrogen bonds as function of temperature which are ultimately related to results from simulations. In another example the validity of simulations performed with  $^3J$ -coupling restraints were assessed by comparing to proposed hydrogen-bond energies instead of values of observable quantities [92]. In such cases the basic comparison is between alternative modelling procedures, not between simulation and experiment. By contrast, where the comparison is to observed experimental quantities such as infrared spectra [93,94] or NMR spectra [95,96], a direct validation of the simulation models by experiment is possible.

#### ***4.5. Validation of Simulations Based on Simulated Data***

Often attempts are made to validate a simulation of a model at a lower level of resolution by comparison to a simulation of a model at a higher level of resolution, which may be physically more correct. For example, one might validate a MD simulation at the atomic level by comparing to a quantum-chemistry model, or compare the results of a supra-atomic (CG) simulation to that at an atomic (FG) level. In such cases it is important that the high-resolution model itself is properly validated. In Sections 2.1 and 2.2 reasons why a given high-resolution model may be inaccurate or an inappropriate reference in some cases were discussed [6]. The example of liquid water (Section 2.1) illustrates that a quantum-chemical model calculation may not always be an appropriate basis against which to validate a classical simulation. For other examples see [97,98].



Calibrating low-resolution models solely on reference to data from a high-resolution model holds multiple risks. In particular, errors in the high-resolution model will be propagated to the lower resolution model, compounding its limitations.

#### ***4.6. Validation of Simulations Using Post-Simulation Corrections to Averages***

For a system with electrostatic interactions, it is generally not practicable to make the system large enough for the finite-system-size associated errors in thermodynamic quantities to be smaller than the thermal energy and allow the use of non-periodic boundary conditions. As a consequence, systems of microscopic size are often simulated using periodic boundary conditions. Periodic boundary conditions also give rise to artefacts in thermodynamic quantities. Analytical expressions for various terms can be used to correct the average value of various quantities derived from such simulations. However, as the underlying configurational ensemble was still produced using uncorrected forces, the ability to use such post-simulation corrected values of properties for validation is limited. This is, unless corrections to the physical forces during a molecular dynamics simulation are applied on-the-fly [99].

#### ***4.7. Statistical Precision, Correlation and Causality***

When comparing simulation results with experimental or other data, one may perform a statistical analysis to determine the correlation between sets of data or variables, see for example Ref. [100]. A range of statistical parameters can be used to assess the performance of a model or simulation, e.g. correlation coefficients or mean absolute error. The impact of errors in the underlying data must also always be considered [101]. In Figure 2, the error bars on the individual measurements are much smaller than the variation between different experiments. It has been argued that a minimum potency range of three log units with a minimum of approximately 50 data points are needed to properly validate predicted ligand binding data. If fewer data points or a narrower potency range is used, the probability of there being chance correlations is high [101].

Understanding the propagation of error or uncertainty when data are processed using particular functions or procedures is important. Error propagation should also be considered when there is uncertainty in the values of parameters of a model that is a complex function of parameters and variables.

In addition, an observed correlation may be caused by hidden, i.e. not analysed and detected, variables. An example of this is the use of harmonic mode analysis or essential dynamics to detect functionally important motional modes in protein simulations. A repeating pattern can indicate functional motions. However, cosine-like essential modes



can also arise from random noise [102].

Finally, when a correlation between data sets is observed, a possible mechanism should be identified that would explain the correlation through causality. Otherwise, the interpretation of the correlation may be flawed or the correlation fortuitous.

#### ***4.8. Validation Based on Popularity?***

One might think that popularity of a molecular model, force field, simulation algorithm, boundary condition or simulation software, i.e. its number of users or downloads, may reflect its quality and thus may serve as criterion for validation. One only has to compare different newspapers or fast-food vs. other restaurants to grasp that there is no a-priori correlation between popularity and quality.

Software may be popular because it is free, fast or easy to use. None of these characteristics are indicative of the quality of the software, and thus cannot be used as criterion for validation. Likewise, the number of citations, while a measure of short-term popularity, is not necessarily correlated with the quality of a publication [103].

#### ***4.9 What Can Be Done to Improve the Quality of Validation?***

##### ***4.9.1 The Simulator***

From the analysis presented in Section 2 it should be possible to deduce which choices of degrees of freedom, theory, molecular model and interaction function, sampling methodology, boundary conditions, software and its input parameters may induce artefacts and thus should be avoided. Once a trajectory or configurational ensemble has been generated, it should be analysed with an eye on statistical relevance and to detecting artefacts, under-sampling, and fortuitous results [7,104]. For MD simulations, the convergence of time series of various single-molecule or system properties should be considered. For MC simulations, the proper sampling of configurational space must be demonstrated. When applying Hamiltonian or temperature replica-exchange simulation, it should be checked whether all replicas actually exchange and sufficiently often to ensure Boltzmann sampling. In simulations aimed at the calculation of the free energy of binding of a series of ligands to a protein, thermodynamic cycle closure should be demonstrated [105,106]. Dynamical aspects may be checked by the calculating NMR relaxation parameters from the MD trajectory and comparing these to measured values [99]. Strategies to validate simulations of biological membranes can be found in [107].

Simulations should also be repeated with different initial configurations or velocities in order to determine if there is any dependence of the trajectories or ensembles



upon the initial conditions. For example, simulations performed at slightly different temperatures should in general yield similar results, unless the system is simulated close to a phase transition temperature.

Finally, the true uncertainty in the values of observable quantities obtained should be estimated. This requires going beyond a mere statistical estimate such as block averaging.

#### **4.9.2 *The Experimentalist***

The lack of experimental data for a defined set of compounds and properties, including uncertainty estimates, not only hampers validation but also model and force-field development. The systematic gathering of experimental data for particular molecular systems of general interest is key to further improving molecular simulation. Such data should be obtained using many different experimental techniques and by different research groups. It should also include uncertainty estimates. Unfortunately, while the importance of computational models for the future of scientific research is widely acknowledged, the systematic production of experimental data for model calibration and validation is not encouraged by research objectives focused on immediate impact.

#### **4.9.3 *Reporting of Simulated and Experimental Results***

In order to reach a wide community of scientists, simulation and experimental results must be published in the literature. To allow the validation of a simulation or experiment on the basis of a publication, the quality of the latter should be evaluated [103,108,109].

1. Clarity of text, tables and figures.
2. Reproducibility of the results from the data specified.
3. Clear reporting of the origin of reference data. Widely different values for the same property are sometimes reported and propagated in the literature [110].
4. Obvious, cheap tests that could confirm or repudiate a model, theory, simulation or measurement, for example to detect hidden variables, should be presented.

Software suppliers and vendors should allow for proper validation of the software by the following actions, see Table 2 of Ref. [7].

1. Specify the models and algorithms implemented.
2. Completely specify the force fields provided, including functional form, parameters, assignments, constants, and ensure that different force-field versions are identified.



3. Provide source code and version identification or, alternatively, provide standard results for checking.

Software users should allow their simulation results to be validated (Table 2 of Ref. [7]).

1. Refer to the theory and model used in the simulation.
2. Provide the version of the force field used and outline any modifications.
3. Present the time or ensemble evolution of key properties so that the degree of sampling and convergence can be judged.
4. Give the version number of the software that was used specifying any modifications.
5. Specify the values of key input parameters so that the risk of artefacts can be judged.

Publishers, editors and reviewers of papers should enforce these guidelines for simulation software suppliers and users. Administrators of molecular data bases such as the PDB should be supported in efforts to ensure their integrity. Adhering to such guidelines would help ensure that simulation methods are applied in an appropriate manner and the results are validated. Only then will it be possible to know that the “novel”, unexpected or strange results are not due to a flawed model, an inadequate force field, insufficient sampling, inadequate boundary conditions, software bugs or input errors, but instead indicative of a new phenomenon.

#### **4.9.4 University Management, Funding Agencies, Scientific Societies, Government**

Proper validation inevitably rests upon repetition of simulations or experiments by different researchers. It was in this way that the erroneous interpretation by Fermi of his nuclear reaction experiments was detected [44]. The need to validate research by repetition by others spans all fields (see e.g. Refs. [111-114]). While universities and funding agencies should encourage work to be “novel” and “innovative”, they must demand that the approaches on which such work is based is appropriately validated. Validation standards may be formulated by scientific societies or organisations. Work that definitively validates a novel result is a crucial part of the foundation on which rigorous future research is built. Claims of major advances readily attract attention. Invalidating such claims does not. This means that popularity measures such as citation indices, Hirsch-indices and the like should not be used as the sole or even the primary consideration when determining which research is to be funded. Widely used is not the same as well validated and such metrics do not necessarily reflect the innate quality of an item of research [103].



## **5. Conclusions**

A variety of issues concerning the validation of molecular simulation results have been analysed and categorised. They are presented with an eye to facilitating the validation of simulation results. The primary, two most important issues that determine the predictive power of a molecular simulation are the interaction function or force field used and the degree of sampling of the relevant configurational degrees of freedom of the system of interest. These will determine the accuracy of the various properties calculated from the trajectory or configurational ensemble. However, choosing inappropriate thermodynamic or spatial boundary conditions or simulation input parameters may undermine the accuracy of a simulation based on a properly calibrated force field and extensive sampling.

Validation of simulation results by comparison to experimental data is not straightforward. The amount of experimental data available for a particular system is limited, meaning the problem is often underdetermined. Experimental data, as well as the functions that connect a configuration to an observable quantity, are also subject to uncertainty and error. Finally, some quantities, the values of which are commonly considered to be experimental data, are in fact not directly observable. For example, the structure of a molecule in solution or a radial distribution function of a liquid cannot be observed directly. Such non-observable quantities should not be used as the sole means of validating a simulation. This is because they may represent the quality of the function or procedure, by which the value of the non-observable quantity was derived from values of observable quantities, rather than the quality of the simulation.

Since experimental values are generally averages over space and time, a comparison with values calculated from a simulation should be based on comparable averaging. This implies that averaged values of a quantity should not be compared to instantaneous values of that quantity.

Validation of molecular simulation studies would be enhanced if publishers, editors, reviewers and authors of manuscripts ensured that all information needed to reproduce the data was reported in the manuscript. Equally, funding agencies should encourage the repetition of research essential to maintain the integrity of the scientific literature.

Finally, we note that popularity is not a suitable validation criterion. While a trend towards fact-free politics is evident in many countries, the academic community must avoid any move towards fact-free or validation-free science.

## **Acknowledgements**



LJS would like to acknowledge the use of the University of Oxford Advanced Research Computing (ARC) facility in carrying out some of this work. <http://dx.doi.org/10.5281/zenodo.22558>. NH thanks the German Research Foundation (DFG) for financial support within the Cluster of Excellence in Simulation Technology (EXC 310/2) at the University of Stuttgart.

## References

- [1] R.W. Hockney, J.W. Eastwood, *Computer Simulation Using Particles*, McGraw-Hill, New York, 1981.
- [2] M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.
- [3] W.F. van Gunsteren, H.J.C. Berendsen, *Angew. Chem. Int. Ed. Engl.* **1990**, 29, 992 - 1023; *Angew. Chem.* **1990**, 102, 1020 - 1055.
- [4] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, Academic Press, London, 1996.
- [5] M. Karplus, J.A. McCammon, *Nat. Struct. Biol.* **2002**, 9, 646 – 652.
- [6] H.J.C. Berendsen, *Simulating the Physical World: Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, Cambridge University Press, Cambridge, 2007.
- [7] W.F. van Gunsteren, A.E. Mark, *J. Chem. Phys.* **1998**, 108, 6109 - 6116.
- [8] W.F. van Gunsteren, J.R. Allison, X. Daura, J. Dolenc, N. Hansen, A.E. Mark, C. Oostenbrink, V.H. Rusu, L.J. Smith, *Angew. Chem. Int. Ed.* **2016**, 55, 15990 – 16010; *Angew. Chem.* **2016**, 128, 16222 – 16244.
- [9] G. Miller, *Science* **2006**, 314, 1856 - 1857.
- [10] A. Extance, *Chemistry World* **2015**, 3, 16 March.
- [11] A. Wlodawer, W. Minor, Z. Dauter, M. Jaskolski, *FEBS J.* **2008**, 275, 1 – 21.
- [12] R. Carlson, T. Hudlicky, *Helv. Chim. Acta* **2012**, 95, 2052 – 2062.
- [13] R.P. Joosten, K. Joosten, G.N. Murshudov, A. Perrakis, *Acta Crystallogr. Sect. D* **2012**, 68, 484 – 496.
- [14] A. Wlodawer, W. Minor, Z. Dauter, M. Jaskolski, *FEBS J.* **2013**, 280, 5705 – 5736.
- [15] P. D. Adams, K. Aertgeerts, C. Bauer, J. A. Bell, H. M. Berman, T. N. Bhat, J. M. Blaney, E. Bolton, G. Bricogne, D. Brown, S. K. Burley, D. A. Case, K. L. Clark, T. Darden, P. Emsley, V. A.



- Feher, Z. Feng, C. R. Groom, S. F. Harris, J. Hendle, T. Holder, A. Joachimiak, G. J. Kleywegt, T. Krojer, J. Marcotrigiano, A. E. Mark, J. L. Markley, M. Miller, W. Minor, G. T. Montelione, G. Murshudov, A. Nakagawa, H. Nakamura, A. Nicholls, M. Nicklaus, R. T. Nolte, A. K. Padyana, C. E. Peishoff, S. Pieniazek, R. J. Read, C. Shao, S. Sheriff, O. Smart, S. Soisson, J. Spurlino, T. Stouch, R. Svobodova, W. Tempel, T. C. Terwilliger, D. Tronrud, S. Velankar, S. C. Ward, G. L. Warren, J. D. Westbrook, P. Williams, H. Yang, J. Young, *Structure* **2016**, *24*, 502 – 508.
- [16] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenzholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt, H. B. Yu, *Angew. Chem. Int. Ed.* **2006**, *45*, 4064 – 4092; *Angew. Chem.* **2006**, *118*, 4168 – 4198.
- [17] K. Meier, A. Choutko, J. Dolenc, A.P. Eichenberger, S. Riniker, W.F. van Gunsteren, *Angew. Chem. Int. Ed.* **2013**, *52*, 2820 - 2834; *Angew. Chem.* **2013**, *125*, 2 - 19.
- [18] S.J. Bachmann, W.F. van Gunsteren, *J. Chem. Phys.* **2014**, *141*, 22D515.
- [19] 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.
- [20] S. Riniker, W.F. van Gunsteren, *J. Chem. Phys.* **2011**, *134*, 084110.
- [21] B. Guillot, *J. Mol. Liq.* **2002**, *101*, 219 – 260.
- [22] Z. Ma, Y. Zhang, M.E. Tuckerman, *J. Chem. Phys.* **2012**, *137*, 044506.
- [23] K. Forster-Tonigold, A. Gross, *J. Chem. Phys.* **2014**, *141*, 064501.
- [24] R.M. Brunne, W.F. van Gunsteren, R. Brüschweiler, R.R. Ernst, *J. Am. Chem. Soc.* **1993**, *115*, 4764 – 4768.
- [25] J.W. Peng, C.A. Schiffer, P. Xu, W.F. van Gunsteren, R.R. Ernst, *J. Biomol. NMR* **1996**, *8*, 453 – 476.
- [26] I.G. Tironi, R.M. Brunne, W.F. van Gunsteren, *Chem. Phys. Letters* **1996**, *250*, 19 – 24.
- [27] A.E. Mark, W.F. van Gunsteren, *J. Mol. Biol.* **1994**, *240*, 167 – 176.
- [28] P.E. Smith, W.F. van Gunsteren, *J. Phys. Chem.* **1994**, *98*, 13735 – 13740.
- [29] C. Oostenbrink, A. Villa, A.E. Mark, W.F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656 – 1676.
- [30] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W.F. van Gunsteren, *Eur. Biophys. J.* **2011**, *40*, 843 – 856.
- [31] W.F. van Gunsteren, P.H. Hünenberger, A.E. Mark, P.E. Smith, I.G. Tironi, *Comput. Phys.*



*Commun.* **1995**, *91*, 305 – 319.

[32] W.P. Reinhardt, M.A. Miller, L.M. Amon, *Acc. Chem. Res.* **2001**, *34*, 607 – 614.

[33] C. Peter, C. Oostenbrink, A. van Dorp, W.F. van Gunsteren, *J. Chem. Phys.* **2004**, *120*, 2652 – 2661.

[34] B. Lai, C. Oostenbrink, *Theor. Chem. Acc.* **2012**, *131*, 1272.

[35] W. Huang, W.F. van Gunsteren, *J. Phys. Chem. B* **2015**, *119*, 753 – 763.

[36] B.A. Luty, W.F. van Gunsteren, *J. Phys. Chem.* **1996**, *100*, 2581 – 2587.

[37] P.H. Hünenberger, J.A. McCammon, *J. Chem. Phys.* **1999**, *110*, 1856 – 1872.

[38] D. Mohanty, R. Elber, D. Thirumalai, D. Beglov, B. Roux, *J. Mol. Biol.* **1997**, *272*, 423 – 442.

[39] D.S. Cerutti, R. Duke, P.L. Freddolino, H. Fan, T.P. Lybrand, *J. Chem. Theory Comput.* **2008**, *4*, 1669 – 1680.

[40] Z. Lin, W.F. van Gunsteren, *J. Chem. Phys.* **2015**, *143*, 034110.

[41] W.F. van Gunsteren, H.J.C. Berendsen, *Mol. Phys.* **1977**, *34*, 1311 – 1327.

[42] R. Baron, D. Trzesniak, A.H. de Vries, A. Elsener, S.J. Marrink, W.F. van Gunsteren, *ChemPhysChem*, **2007**, *8*, 452 – 461.

[43] E. Rosta, N. V. Buchete, G. Hummer, *J. Chem. Theory Comput.* **2009**, *5*, 1393 – 1399.

[44] M. Quack, *Angew. Chem. Int. Ed.* **2013**, *52*, 9362 – 9370; *Angew. Chem.* **2013**, *125*, 9530 – 9538.

[45] W.F. van Gunsteren, J. Dolenc, A.E. Mark, *Curr. Opin. Struct. Biol.* **2008**, *18*, 149 – 153.

[46] A. Glättli, W.F. van Gunsteren, *Angew. Chem. Int. Ed. Engl.* **2004**, *43*, 6312 – 6316; *Angew. Chem.* **2004**, *116*, 6472 – 6476.

[47] D. Trzesniak, A. Glättli, B. Jaun, W.F. van Gunsteren, *J. Am. Chem. Soc.* **2005**, *127*, 14320 – 14329.

[48] D. Trzesniak, W.F. van Gunsteren, *Protein Sci.* **2006**, *11*, 2544 – 2551.

[49] A.Y. Lau, B. Roux, *Structure* **2007**, *15*, 1203 – 1214.

[50] B. Zagrovic, G. Jayachandran, I.S. Millett, S. Doniach, V.S. Pande, *J. Mol. Biol.* **2005**, *353*, 232 – 241.

[51] B. Zagrovic, W.F. van Gunsteren, *Proteins Struct. Funct. Bioinf.* **2006**, *63*, 210 – 218.



- [52] X. Daura, I. Antes, W.F. van Gunsteren, W. Thiel, A.E. Mark, *Proteins: Struct. Funct. Genet.* **1999**, 36, 542 – 555.
- [53] A. Glättli, X. Daura, D. Seebach, W.F. van Gunsteren, *J. Am. Chem. Soc.* **2002**, 124, 12972 – 12978.
- [54] W. F. van Gunsteren in *Studies in Physical and Theoretical Chemistry, Vol. 71 of Modelling of Molecular Structures and Properties* (Ed.: J.-L. Rivail), Elsevier, Amsterdam, **1990**, pp. 463 – 478.
- [55] D. Poger, A.E. Mark, *J. Chem. Theory Comput.* **2010**, 6, 325 – 336.
- [56] L.D. Schuler, W.F. van Gunsteren, *Mol. Simul.* **2000**, 25, 301 – 319.
- [57] N.O.B. Lüttschwager, T.N. Wassermann, R.A. Mata, M.A. Suhm, *Angew. Chem. Int. Ed.* **2013**, 52, 463 - 466; *Angew. Chem.* **2013**, 125, 482 - 485.
- [58] J.N. Byrd, R.J. Bartlett, J.A. Montgomery Jr, *J. Phys. Chem. A* **2014**, 118, 1706 – 1712.
- [59] D.G. Liakos, F. Neese, *J. Chem. Theory Comput.* **2015**, 11, 2137 – 2143.
- [60] T. A. Soares, X. Daura, C. Oostenbrink, L. J. Smith, W. F. van Gunsteren, *J. Biomol. NMR* **2004**, 30, 407 – 422.
- [61] J. Drenth, *Principles of Protein X-ray Crystallography*, Springer, New York, **1994**.
- [62] R. Boelens, T. M. G. Koning, R. Kaptein, *J. Mol. Struct.* **1988**, 173, 299 – 311.
- [63] M. Karplus, *J. Chem. Phys.* **1959**, 30, 11 – 15.
- [64] D. Steiner, J. R. Allison, W. F. van Gunsteren, *J. Biomol. NMR* **2012**, 53, 223 – 246.
- [65] R.J. Abraham, K.A. McLauchlan, *Mol. Phys.* **1962**, 5, 513 – 523.
- [66] C.M. Deber, D.A. Torchia, E.R. Blout, *J. Am. Chem. Soc.* **1971**, 93, 4893 – 4897.
- [67] K.D. Kopple, G.R. Wiley, R. Tauke, *Biopolymers* **1973**, 12, 627 – 636.
- [68] A. de Marco, M. Llinás, K. Wüthrich, *Biopolymers* **1978**, 17, 617 – 636.
- [69] C. Pérez, F. Löhr, H. Rüterjans, J.M. Schmidt, *J. Am. Chem. Soc.* **2001**, 123, 7081 – 7093.
- [70] C.D. Snow, N. Nguyen, V.S. Pande, M. Gruebele, *Nature* **2002**, 420, 102 – 106.
- [71] C.D. Snow, E.J. Sorin, Y.M. Rhee, V.S. Pande, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, 34, 43 – 69.
- [72] A.D. Robertson, K.P. Murphy, *Chem. Rev.* **1997**, 97, 1251 – 1268.
- [73] S. Deeckongkit, P.E. Dawson, J.W. Kelly, *J. Amer. Chem. Soc.* **2004**, 126, 16762 - 16771.



- [74] A.P. Eichenberger, W.F. van Gunsteren, S. Riniker, L. von Ziegler, N. Hansen, *Biochim. Biophys. Acta, General Subjects* **2015**, 1850, 983 – 995.
- [75] P.J. Gee, W.F. van Gunsteren, *Proteins Struct. Funct. Bioinf.* **2006**, 63, 136 – 143.
- [76] D. Frenkel, *Eur. Phys. J. Plus* **2013**, 128, 10.
- [77] Z. Otwinowski, R.W. Schevitz, R.-G. Zhang, C.L. Lawson, A. Joachimiak, R.Q. Marmorstein, B.F. Luisi, P.B. Sigler, *Nature* **1988**, 335, 321 – 329.
- [78] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, *Nucleic Acids Res.* **2000**, 28, 235 – 242.
- [79] H.W. Horn, W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, T. Head-Gordon, *J. Chem. Phys.* **2004**, 120, 9665 – 9677.
- [80] W.F. van Gunsteren, H.J.C. Berendsen in *Molecular Dynamics and Protein Structure* (Ed.: J. Hermans), Polycrystal Book Service, P.O. Box 27, Western Springs, **1985**, pp. 5 - 14.
- [18] S.J. Marrink, A.H. de Vries, A.E. Mark, *J. Phys. Chem. B* **2004**, 108, 750 – 760.
- [82] J.-P. Hansen, J.-J. Weis, *Phys. Rev.* **1969**, 188, 314 - 318.
- [83] M. Winger, D. Trzesniak, R. Baron, W.F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2009**, 11, 1934 – 1941.
- [84] S.J. Marrink, X. Periole, D.P. Tieleman, A.H. de Vries, *Phys. Chem. Chem. Phys.* **2010**, 12, 2254 – 2256.
- [85] W.F. van Gunsteren, M. Winger, *Phys. Chem. Chem. Phys.* **2010**, 12, 2257 – 2258.
- [86] E. Paci, K. Lindorff-Larsen, C.M. Dobson, M. Karplus, M. Vendruscolo, *J. Mol. Biol.* **2005**, 352, 495 - 500.
- [87] K. Lindorff-Larsen, R.B. Best, M.A. DePristo, C.M. Dobson, M. Vendruscolo, *Nature* **2005**, 433, 128 - 132.
- [88] M. Vendruscolo, *Curr. Opin. Struct. Biol.* **2007**, 17, 15 - 20.
- [89] K.A. Scott, L.G. Randles, S.J. Moran, V. Daggett, J. Clarke, *J. Mol. Biol.* **2006**, 359, 159 - 173.
- [90] A.E. Torda, R.M. Scheek, W.F. van Gunsteren, *J. Mol. Biol.* **1990**, 214, 223 – 235.
- [91] A.T. Iavarone, A. Patriksson, D. van der Spoel, J.H. Parks, *J. Am. Chem. Soc.* **2007**, 129, 6726 - 6735.
- [92] J. Gsponer, H. Hopearuoho, A. Cavalli, C.M. Dobson, M. Vendruscolo, *J. Am. Chem. Soc.*



**2006**, 128, 15127 - 15135.

[93] Y.G. Mu, D.S. Kosov, G. Stock, *J. Phys. Chem. B* **2003**, 107, 5064 - 5073.

[94] S. Gnanakaran, R.M. Hochstrasser, A.E. Garcia, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 9229 - 9234.

[95] C. Peter, M. Rüping, H.J. Wörner, B. Jaun, D. Seebach, W.F. van Gunsteren, *Chem. Eur. J.* **2003**, 9, 5838 – 5849.

[96] S.A. Showalter, R. Brüschweiler, *J. Chem. Theory Comput.* **2007**, 3, 961 - 975.

[97] R.E. Plata, D.A. Singleton, *J. Am. Chem. Soc.* **2015**, 137, 3811 - 3826.

[98] M.G. Medvedev, I.S. Bushmarinov, J. Sun, J.P. Perdew, K.A. Lyssenko, *Science* **2017**, 355, 49 – 52.

[99] M.M. Reif, C. Oostenbrink, *Theor. Chem. Acc.* **2015**, 134, 2.

[100] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyan, S. Robinson, M.K. Dahlgren, J. Greenwood, D.L. Romero, C. Masse, J.L. Knight, T. Steinbrecher, T. Beuming, W. Damm, E. Harder, W. Sherman, M. Brewer, R. Wester, M. Murcko, L. Frye, R. Farid, T. Lin, D.L. Mobley, W.L. Jorgensen, B.J. Berne, R.A. Friesner, R. Abel, *J. Am. Chem. Soc.* **2015**, 137, 2695 - 2703.

[101] S. P. Brown, S.W. Muchmore, P.J. Hajduk, *Drug Disc. Today* **2009**, 14, 420 – 427.

[102] B. Hess, *Phys. Rev. E* **2000**, 62, 8438 – 8448.

[103] W.F. van Gunsteren, *Infazine Special Issue* **2016**, 1, 27-28, DOI: 10.3929/ethz-a-010745085

[104] J. Wong-ekkabut, M. Karttunen, *Biochim. Biophys. Acta* **2016**, 1858, 2529 – 2538.

[105] W.F. van Gunsteren, in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications* (Eds.: W.F. van Gunsteren, P.K. Weiner), Escom Science Publishers, Leiden, The Netherlands, **1989**, pp. 27 – 59.

[106] C. Chipot, A. Pohorille, Eds., *Free Energy Calculations. Theory and Applications in Chemistry and Biology*, Springer Science & Business Media, **2007**, pp. 1 – 518.

[107] D. Poger, B. Caron, A.E. Mark, *Biochim. Biophys. Acta* **2016**, 1858, 1556 – 1565.

[108] W.F. van Gunsteren, *Angew. Chem. Int. Ed.* **2013**, 52, 118 – 122; *Angew. Chem.* **2013**, 125, 128 – 132.

[109] W.F. van Gunsteren, *F1000Research* **2015**, 4, 1244.



- [110] C. Margreitter, M.M. Reif, C. Oostenbrink. *J. Comput. Chem.* **2017**, DOI: 10.1002/jcc.24733
- [111] C.G. Begley, L.M. Ellis, *Nature* **2012**, 483, 531 – 533.
- [112] T.M. Errington, E. Iorns, W. Gunn, F.E. Tan, J. Lomax, B.A. Nosek, *eLife* **2014**, 3, e04333.
- [113] Open Science Collaboration, *Science* **2015**, 349, aac4716.
- [114] M. Baker, D. Penny, *Nature* **2016**, 533, 452 – 454.