Theor Chem Acc (2006) 116: 297–306
DOI 10.1007/s00214-005-0070-4

REGULAR ARTICLE

Xavier Daura

Molecular dynamics simulation of peptide folding

Received: 20 April 2005 / Accepted: 28 September 2005 / Published online: 21 December 2005

Abstract The simulation of peptide folding with atomic resolution has evolved remarkably during the last 7 years, i.e., from absolute skepticism on the capability of classical molecular dynamics (MD) methodology to reproduce complex biological phenomena such as the folding of even simple oligopeptides (6–15 residues) to the seemingly realistic representation of the thermodynamics and kinetics of folding of a rapidly increasing number of polypeptides (over 20 residues). Four factors permitted this rapid progress: the breakthrough of a second generation of force fields, a rapid and steady increase of (commodity) computer performance, a move from local computational resources to large distributed clusters and, last but not less important, a decision of particular groups to spend a large computational effort on projects that most other groups trusted unrealizable at the time. The present account goes over some aspects of peptide folding and its simulation with MD techniques while sweeping through the simulation landmarks of the last 7 years and conjecturing on the future.

1 Introduction

Peptide chains exist in an equilibrium between different conformations as a function of environment – number of molecules for each of the molecular species present in the system – and thermodynamic – temperature, pressure – conditions. If a peptide adopts a structurally ordered, densely populated conformation, it is commonly said to have a folded state, identified with this conformer, in equilibrium with the unfolded state, represented by diverse, sparsely populated conformers. Although, in principle, this equilibrium exists for chains of any size, its thermodynamics and kinetics are typically different for oligopeptides forming secondary structure elements, polypeptides (over 20 residues) with some degree of tertiary structure and proteins (over 50 residues) with a substantial solvent-excluded volume (hydrophobic core). This can be broadly explained with reference to the different dimensionalities of the free enthalpy hypersurfaces of these three classes of polymers of amino acids.

Historically, research on peptide and protein structure, whether by experimental or computational means, has focused on the folded state only. This is because an array of methods exists for the detailed investigation of the folded state of a peptide or protein, and because the folded state is commonly the functionally active one. There are, however, two scenarios in which a full characterization of the unfolded state becomes as essential as the determination of the folded conformation. The first is in the study of the physical, chemical, or biological properties of oligopeptides and short polypeptides. The folded conformation of such peptides is, in general, only marginally more stable than the lowest free enthalpy unfolded conformation. Moreover, the unfolded state of a peptide is often more populated than the folded state. As a result, any macroscopic observable of a peptide is weighted with both the folded and the unfolded states. Interpreting such observables in terms of only the folded conformation is, therefore, incorrect. The second scenario is related to the study of peptide and protein folding (to be distinguished from structure prediction). The description of an equilibrium requires knowledge about each of the states involved. Therefore, it is fundamental to describe not only the folded state but also the unfolded state accurately in order to draw conclusions on the nature and mechanisms of peptide and protein folding.

More recently, some efforts have been made to characterize the unfolded state of specific peptides and proteins [1–16]. Obtaining quantitative microscopic information from the unfolded state is difficult not only experimentally but also computationally. While experiments suffer from the low population and transient nature of individual unfolded conformations, theoretical studies are bound by the computational time required to sample the conformations of the unfolded state

X. Daura
Catalan Institution for Research and Advanced Studies (ICREA),
Institute of Biotechnology and Biomedicine (IBB),
Universitat Autónoma de Barcelona,
08193 Bellaterra, Spain
E-mail: Xavier.Daura@uab.es
Tel.: +34-93-5812805
Fax: +34-93-5812011

with appropriate weights. In most cases, experimental studies report on low-resolution average properties of the unfolded state [4,6,8,13–15] or focus on particular, not always well-defined substates like, for example, molten globules [17–19], metastable intermediates [20], or transition states [21–26], (see also reviews [27–31] and Adv Prot Chem, vol 62 on Unfolded Proteins). Computationally, sampling the unfolded state is bound to sampling the equilibrium ensemble of folding/unfolding pathways. Given the size of the problem, the challenge is to reduce the number of degrees of freedom of the model without perturbing its capacity to reproduce the behavior of the real system. The number of degrees of freedom can be kept to a computationally tractable number by either simplifying the physical model for the protein and/or its environment [32–43], or limiting the size of the molecular system to that of atomic-detail models of oligo- and polypeptides in solution [44–51]. Alternatively, for an atomic-detail model of a protein in solution, if the native three-dimensional structure is known one may construct a projection of the free enthalpy on particular coordinates using a predefined set of unfolded conformations and biased-sampling [52], or one may perform unfolding simulations under denaturing conditions and assume that the sampled pathways are relevant to folding under native conditions [53]. Arguments in favor and against each of these simplifying approaches are abundant in the mentioned literature.

The focus of this paper is on the molecular dynamics (MD) simulation of oligo- and polypeptide folding with atomic resolution.

2 1998–2005: Simulation landmarks

Although the simulation of peptide dynamics has already a long trajectory [54,55], before 1998 it was generally accepted that the folding of even small oligopeptides forming secondary structure elements – helices and hairpins – could not be tackled with MD simulation methods [56]. Reasons for this skepticism were the estimated timescales of folding, far beyond the computationally accessible timescales at the time, and the general feeling that the empirical force fields and other approximations intrinsic to MD simulation meant that this methodology would not be, in the foreseeable future, applicable to complex phenomena like folding.

Around 1997, however, the groups of van Gunsteren and Kollman made significant advances with somewhat different approaches. In the immediately preceding years, second generations of the most widely used force fields for biomolecular simulation had been developed [57–60]. In addition, it had become apparent that the experimentally estimated folding times depended strongly on the resolution of the technique used as well as on the model applied to interpret the data, and that actual folding times could well be shorter than estimated. Studying the structural properties of a synthetic β-peptide in methanol at different temperatures, Daura et al. [61] observed, as a matter of chance, the unfolding and refolding of the experimentally determined left-handed helix at high temperature in a short simulation. This observation encouraged the authors to attempt an extensive study of the conformational behavior of the peptide at a range of temperatures over a timescale of 50 ns, an order of magnitude larger than common simulations of the time [44]. The results of this study demonstrated, for the first time, that the MD technique could be used to simulate not only the folding of a peptide but also the equilibrium between the unfolded and folded states, with the root mean square difference (RMSD) between the backbones of the NMR model and the helix populated in the simulations being as low as 0.02 nm. Certainly, neither the peptide nor the solvent were biologically relevant, but there was nothing in the methods or the force field that could make the folding of an α-peptide in water less tractable, the physical principles being the same. This was shortly after a demonstration by Duan and Kollman [45], who had taken a complementary approach. They attempted the folding of a 36-residue polypeptide, villin headpiece subdomain, in water with a phenomenally long (1 μs) simulation. The lower bound to the folding time of this polypeptide had been estimated in 10 μs (current estimates are around 5 μs [62]). Nevertheless, the most populated conformer in the simulation contained many of the features of the NMR model structure and the authors suggested that it could correspond to a metastable folding intermediate. This simulation time record (still the longest continuous simulation of a polypeptide in explicit solvent to date) could be achieved thanks to the use of large supercomputing resources, an optimized parallel code developed by the same authors, and a reduced box size in combination with a simplified long-range interaction scheme. The papers by Daura et al. [44] and Duan and Kollman [45] showed that the simulation of the folding of small proteins with atomic resolution was not a chimera but, rather, a question of time [63]. The difficulty to access large supercomputing resources meant that, in most cases, the study of peptide folding by MD simulation was approached from extensive simulations aimed at reproducing the folding/unfolding equilibrium of small, quick folders (oligopeptides) [46,48,64–66].

In late 1998, a paper by Schaefer et al. [67] opened new expectations on a long-debated simplifying approach, i.e., the use of (improved) implicit-solvent representations in biomolecular simulation in general and in peptide folding in particular. This level of modeling permitted a significant timescale jump and was quickly adopted by a number of groups studying peptide folding [38,68–72]. (Interestingly, it did not have a comparable impact on MD simulations of folded proteins.) In particular, Ferrara et al. [37,73] used this fast approach to improve folding statistics with multiple long simulations. The presumed correspondence between implicit- and explicit-solvent thermodynamics and kinetics is, however, a matter of active discussion [74–83].

In parallel to these developments, two papers prepared the terrain for two future important lines of work. On the one hand, Sugita and Okamoto [84] developed a formulation for replica-exchange MD. This built on the replica [85] and multicanonical [86] Monte Carlo algorithms, the latter having been already adapted to MD [87] and later used in peptide

folding simulations [47]. In replica-exchange MD, system replicas at different temperatures (over a defined range) are run simultaneously and independently. Every so many steps, a pair of replicas at neighboring temperatures are exchanged with a Metropolis-based probability. The algorithm permits an efficient sampling of conformational space with appropriate weights. On the other hand, Voter [88] proposed a parallel-replica MD algorithm for efficient parallelization (minimal communication between nodes) of the simulation of rare events. A (large) number of replicas are also run in parallel, with different initial conditions (randomized momenta). When one of the replicas makes a transition to another state, all the replicas are reset to that state. The algorithm preserves the right kinetics. The door to (commodity) distributed computing was open.

In 2001, García and Sanbonmatsu [49] applied the replica-exchange MD algorithm to the study of peptide folding. Since then, this method and its derivatives [89–92] have become standards for the study of peptide-folding thermodynamics, both in explicit- and implicit-solvent environments [93–97]. Although the method allows a very efficient sampling of conformational space, effectively crossing over barriers, it does not lack inconveniences. Thus, the kinetics are scrambled by the exchange procedure and, in explicit solvent, efficient exchange requires in general the use of constant volume and a small spacing between temperatures, i.e., many replicas. Distinct formalisms, with derivations based to a varying extent on Markov chains, have been already proposed for the recovery of kinetic information on a system by reusing a Boltzmann-weighted distribution of states (e.g., from replica-exchange simulations) [98–100].

Building on previous efforts to apply large-scale distributed computing to biomolecular simulation [88,101,102], in 2002 Snow et al. [50] reported a kinetic study of the folding of a 23-residue polypeptide in implicit solvent by multiple, tens of thousands, relatively short MD simulations, using parallel-replica MD. As impressive as it is, this approach is neither without problems. The thermodynamics of the system are biased and the ability of the approach to evaluate the folding kinetics of proteins, as opposed to peptides, has been questioned [103] (although a kinetic model that could potentially circumvent this problem has been already proposed [104]). In addition, the complex infrastructure/logistics required by this approach means that only one group is actively using it at the moment [105].

3 A view on peptide folding

Looking at the existing literature, an outsider would easily conclude that the knowledge on how peptides fold is extense. A myriad of seemingly consistent conclusions have been written down, some compatible some not, most of them defendable but few irrefutable or generalizable. Indeed, there have been many interpretations of peptide folding on the basis of experimental and theoretical data, but there is still little true knowledge. Even when detailed computational models seem to reproduce accurately the experimental data, offering a plausible microscopic picture of the process, it is difficult to interpret the trajectories without introducing yet more models and assumptions in the analysis.

In the following paragraphs I will treat some aspects of peptide folding, mostly in connection with our own work. A significant part of our experience is based on observations from MD simulations of the reversible folding of non-natural peptides (especially β-peptides [106]) in non-aqueous solvents [7,107]. Whether results on these systems can be extrapolated to natural peptides in water is, of course, questionable. As a matter of fact, Kritzer et al. [108] have recently suggested that β-peptide folding might be governed by different biophysical forces than α-peptide folding. Finally, note that protein folding does not necessarily proceed as a simple extension of peptide folding, and I will only refer to the latter process.

3.1 The framework

I will start by defining unambiguously, though in some cases arbitrarily, some terminology. *Configuration* will be used to describe a set of distinct coordinates of the atoms of a molecular system. *Structure* will be used to describe a set of distinct internal coordinates of the atoms of a peptide. *Conformation*, on the other hand, will be used as an abstraction representing an (in principle infinite) ensemble of structures of a peptide with high structural similarity and identical macroscopic properties. From these definitions it follows that the *configurational space* (continuous) of a peptide is infinite, while the *conformational space* (discrete) is not. The term *accessible conformational space* will be based on thermodynamics rather than kinetics. Thus, it will refer to the space of likely conformations, conformations with non-zero probability, at equilibrium.

Consider a peptide in conformational equilibrium. Suppose the peptide is part of a microscopic system in contact with a reservoir at constant temperature and pressure (exchange of heat and work). The conformational space accessible to the peptide (and its associated probability distribution) depends on the rest of the system or environment, **E**, and the thermodynamic conditions, **C**. I will call the most probable conformation of the peptide under **E**,**C** the *folded conformation* of the peptide. Any other conformers populated under **E**,**C** will be said to be *unfolded*. Note that this is irrespective of structural considerations, i.e., the most probable conformer may or may not be recognized as a secondary structure element, while there might be lower probability (unfolded) conformers recognizable as canonical secondary structures. In addition, the folded conformer is not necessarily unique, as it may change with **E**,**C**. The terms *state* and *free enthalpy* (or Gibbs free energy) refer to the system. Thus, the *folded state* under **C** is here defined as the ensemble of microstates of the system, states with specified atomic coordinates and momenta for all atoms, for which the peptide is in its folded conformation, while the *unfolded state* under **C**

is defined by all other microstates populated by the system. Likewise, I may define system states associated to each of the unfolded conformations of the peptide. For simplicity, I will use the term *conformational state* to refer to a system state associated to a particular conformation of the peptide. Using this nomenclature, the free enthalpy difference between any pair AB of states (e.g., free enthalpy of folding if A is the unfolded state and B the folded state) under **C** is given by:

$$\Delta G_{AB}^{(\mathbf{C})} = -\beta^{-1} \ln \left[P_B^{(\mathbf{C})} \Big/ P_A^{(\mathbf{C})} \right], \tag{1}$$

where $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T is the temperature, and $P_X^{(\mathbf{C})}$ is the probability of state X under **C**:

$$P_X^{(\mathbf{C})} = \sum_{\mathrm{i}} P_{X_i}^{(\mathbf{C})} = \frac{\sum_{\mathrm{i}} e^{-\beta H_{X_i}^{(\mathbf{C})}}}{\sum_{\mathrm{Y}} \sum_{\mathrm{j}} e^{-\beta H_{Y_j}^{(\mathbf{C})}}}, \tag{2}$$

with $H_{X_i}^{(\mathbf{C})}$ being the enthalpy of microstate X_i, belonging to state X, under **C**. Note that the sum over all conformational states Y (partition function) in Eq. (2) vanishes in Eq. (1). Note also that in principle the sums over microstates i, j in Eq. (2) are integrals over phase space. However, the impossibility to specify a microstate, a point in phase space, with infinite accuracy (theoretically) and precision (computationally) means that the phase space of the system is, for our purposes, discrete and we may approach these integrals as sums. The sum over microstates of the Boltzmann factor $e^{-\beta H_{X_i}^{(\mathbf{C})}}$ is very hard to estimate, even for a conformational state. Although only the low enthalpy microstates of X will significantly contribute to the sum, the Boltzmann factor is always positive and, therefore, we may severely underestimate the sum if any low free enthalpy microstates are missed. MD simulations are rather efficient at sampling the low enthalpy microstates, especially for rapidly relaxing environments like water. However, if the environment of the peptide contains other slowly relaxing molecular species, e.g., another peptide, in addition to the solvent, the estimation of the sum becomes impossible, as it involves the sampling of the conformational space accessible to the second peptide under the conditions that the enthalpy of the corresponding microstates is low and the first peptide is in the conformation that defines state X. Finally, the probabilities might be also estimated by population counting from an equilibrium distribution of conformational states. This is normally the approach used to analyze peptide-folding thermodynamics from simulation trajectories, the problem being again the difficulty to sample such equilibrium even for simple peptide/solvent systems. It has the basic advantage of not requiring any calculation.

For the sake of discussion, I will define yet another conformational state: the *native state*. This will be a system state associated to the experimentally observable (and biologically relevant) conformation of the peptide (native conformation), without consideration of its relative probability. By analogy to the folded/unfolded pair, I define the *denatured state* as the ensemble of populated microstates of the system for which the peptide is not in the native conformation. As with the folded conformer, the native conformer of a peptide may not be unique but will be dependent on **E**,**C**. For example, the native conformation of a peptide may be different in vacuum [109–112], in aqueous solution and when inserted in a membrane.

3.2 Bounds to a conformation

The concept of conformation is necessarily ill-defined. As an abstraction of a region of the configurational space of a peptide, it has no natural bounds. Thus, any set of criteria we use to define a conformation will be unavoidably arbitrary. However, some criteria conform better to the concept than others. In MD simulation studies, the assignment of sampled structures of a peptide to particular conformations is in general done through clustering. Clustering algorithms used in the context of peptide folding are diverse in nature and mathematical complexity. More important than the algorithm, however, are the descriptors used to determine whether two structures belong to the same conformation. The most common descriptors are global measures of (backbone) structure, e.g., RMSDs of atomic positions [113] or atomic distances [114], population densities in some type of projection of the configurational space sampled [115–117], strings of dihedral angles and associated secondary structure motifs [118], etc. In some cases, a set of local descriptors has been used, e.g., particular types of atomic interactions [69], occasionally in combination with specific energy terms [119]. An approach to conformation from phase space rather than configurational space would be desirable. Huisinga et al. [120], for example, devised a dynamics-based clustering which characterizes conformations in terms of their metastability with respect to fluctuations. Although an appealing idea, this type of analysis becomes impractical for large sets.

3.3 Thermodynamics versus kinetics

The most probable state of a system is, by definition (Eq. (1)), the one with the lowest free enthalpy. An apparently naive question then rises: is the native state under **C** the most probable state under **C** (i.e., are the states defined here as native and folded the same)? The answer is not straightforward. For example, it is common to read (although more in relation to proteins) that the native conformation of a peptide is the conformation with the lowest free enthalpy from those which are kinetically accessible. This sentence may be confusing. First, except for a peptide in vacuum, a peptide conformation does not have an associated free enthalpy. As already mentioned, free enthalpy is a system property. Thus, it is the peptide/**E** system that will (eventually) evolve toward its lowest free enthalpy state and not the peptide. In addition, to define the free enthalpy of the system as a function of, say, the peptide conformation, one must first fix the thermodynamic conditions (**C**). Second, we need to introduce the concept of kinetic inaccessibility. If a state is truly

(thermodynamically) inaccessible it will have zero probability and, hence, an infinitely high relative free enthalpy. We can argue, however, that the average life time of the system may be shorter than the minimum time it needs to reach its lowest free enthalpy conformational state and, therefore, a higher free enthalpy, kinetically more accessible, conformational state (native state) might be systematically more populated. In other words, the life time of the system may be much shorter than its relaxation time, i.e. the time required to reach an equilibrium distribution of (conformational) states. Thus, the question reduces to whether natural peptide/**E** systems are in equilibrium. From our (maybe still poor) understanding of the characteristics of free enthalpy landscapes of peptides in dilute solution (the particular case we are treating here) we can assume that there are no unsurmountable kinetic barriers in these landscapes and, therefore, conditions close to equilibrium may exist for naturally occurring life times. There is also a practical reason to avoid considering the possibility of non-equilibrium peptide/**E** systems: most available formalisms to study peptide-folding thermodynamics refer to equilibrium distributions of states. It does, in fact, make little sense to talk about probabilities and inaccessible states if the real system can never sample the conformational space with appropriate (equilibrium) weights. Hence, I will assume that what I have defined as folded and native corresponds to the same conformational state.

3.4 Accessible conformational space

It has been proposed that fast folding, i.e., folding in biologically relevant timescales, may be achieved thanks to a bias of the unfolded state toward the folded state [36,121]. This has been often expressed as a preference of the system for native contacts. Supposedly, this bias could have been introduced by some evolutionary mechanism. But, what could this bias be? The answer must be necessarily in the Hamiltonian of the system. For example, imagine a peptide/water microscopic system with the peculiarity that all atoms interact only through a very basic short-range repulsion term that avoids atomic overlap (in addition to a unique bonding term for bound atoms). The entire conformational space of the peptide would be basically accessible, and all the conformational states of the system would have similar probabilities (distinguished only entropically). In a landscape like this, finding a particular conformation and ensuring its stability during a biologically meaningful time would be a hard task. The Hamiltonian governing a real peptide/water system is obviously much more complex than this and it differentially weights conformational states, effectively reducing the conformational space accessible to the peptide. How much is the space reduced and in which direction (if a specific one) is a question of active discussion. There are indications that in the unfolded state the average structural properties of a peptide may be close to those in the folded state [10,114,122]. This is not, however, incompatible with the conformations of the peptide in the unfolded state being at the same time structurally diverse (to be distinguished from random), with many of the populated conformers lacking native contacts [113,123–125].

Based on simulation studies of the equilibria of non-natural peptides (mostly) in non-aqueous solvents at different temperatures, we proposed that the number of conformational states accessible to a peptide is far, orders of magnitude, smaller than the number of theoretical conformational states – states resulting from the consideration of a fixed, e.g., 3, number of conformers per backbone angle – and that, opposite to the theoretical space, the accessible space does not grow exponentially with the length of the peptide [7]. In the picture that emerged from these simulations, the unfolded state is heavily dominated by a relatively small number of conformational states, and folding occurs, in average, through a small number of intermediates [107]. We suggested that there might be a correlation between the number of accessible (populated) conformational states and the average folding time [7,126] and, therefore, a conformational landscape like the one proposed would favor fast folding kinetics. We also suggested that the average life time of the folded state might grow faster with the number of residues than the average folding time, thus making folding a more efficient process for longer chains [107]. Theoretical and experimental kinetic data suggest that, for proteins, the scaling of the folding time with the number of residues (N) is roughly proportional to $e^{N^{1/2}}$ (the proportionality constant being in microseconds) [127,128]. On the other hand, the minimum folding time for a single-domain protein would be approximately $N/100\,\mu$s [129]. The dependence we observe for our reduced set of peptides is approximately linear in nature, in line with the latter fast folders (albeit with a different slope) [107]. Some aspects of our analysis of the accessible conformational space and its implications on folding have been, however, criticized. In particular, it has been argued that the average number of (unfolded) conformational states visited during folding may be small and the accessible conformational space still grow exponentially with the length of the peptide [130,131].

Currently, we are analyzing the accessible conformational spaces of two β-peptides in methanol, at different temperatures, in terms of network theory [132]. The populated conformational states are determined by clustering peptide structures sampled at regular time intervals during the simulation [113]. Each conformational state represents, thereafter, a node in the graph (network). Links between nodes correspond to bidirectional transitions between the conformational states (clusters) sampled during the simulation. Preliminary results from this analysis (R.V. Solé and X. Daura, in progress) indicate that the accessible conformational spaces of these peptides have a small-world topological organization (high clustering, small average path length) with an exponential degree distribution [most clusters can perform transitions to only one or two other clusters, whereas a few of them (hubs) have multiple connections]. Similar results have been recently reported by Rao and Caflisch [133] on a 20-residue antiparallel β-sheet peptide. This level of organization, which agrees with our previous observation of a small

average number of folding intermediates, explains to a large degree the fast folding of these so-called foldamers.

3.5 Specific folding energetics

For one of these peptides, the energetics have been analyzed using classical thermodynamics formulae that relate free enthalpy and entropy differences over a temperature range to enthalpies and heat capacities at constant pressure (X. Daura and W.F. van Gunsteren, in preparation). There are two aspects of the results which I wish to discuss here. The first one is that, for the temperature range studied (298–360 K, with 10 K intervals) the higher the temperature the more favorable the (system) enthalpy is to folding. Since the free enthalpy of folding increases with temperature, the entropy must counteract the enthalpy trend. The second refers to the contribution of different potential energy terms to folding: internal (peptide) bonding interactions, internal non-bonding interactions and solvent–solvent interactions favor folding at all temperatures, while peptide–solvent non-bonding interactions strongly disfavor folding. While this is not unexpected, it called our attention to the fact that the contribution of solvent–solvent interactions to folding is similar in magnitude to the contribution of the peptide's internal non-bonding interactions. It goes without saying that these results may not be extendable to peptides in water.

4 Outlook

There are three choices implicit in any MD simulation of peptide folding.

4.1 Type of degrees of freedom

The atomic representation of peptide and environment will probably take the lead in peptide folding studies in the coming years. Quantum dynamics will be out of question for a long while, and it is not even clear that quantum effects (non-accountable classically) may play an important role in peptide-folding dynamics. Simple models seem, with current methods and computational performances, an unnecessary reduction in regard to peptide folding. Implicit-solvent models, together with an atomic representation of the peptide, may be used, instead, to explore the long time limit.

4.2 Force field

The force field has been, traditionally, a cause of distrust in classical MD. Starting with the second-generation force fields already mentioned, important progress has been made [134–137]. Long-range interactions (e.g., Ewald-type sums or reaction field corrections for electrostatics and long-range cutoffs or corrections for van der Waals) have been, in general, already incorporated in the parametrization procedures. Nevertheless, there is still work for force-field developers [138–142]. Water models with three interaction sites have been proven difficult to improve [143]. TIP5P [144, 145] is currently the best water model available, but is somewhat inconsistent with the peptide description and is computationally expensive. Polarizability has been long claimed to be the key addition to classical force fields. Although polarizable biomolecular force fields already exist [146–149], they require an additional computational effort and have not yet shown an unquestionable advantage over current non-polarizable ones (even for water).

Half-way between the sampling algorithm and the force field, pH stands as one of the missing parameters in standard MD. The pH is usually taken into account in an average way by fixing the protonation state of the titratable groups according to standard pK_A values. This is, however, a rather crude approximation, especially because pH is known to strongly affect folding. There exist a number of constant-pH algorithms, both for explicit-water [150–152] and implicit-water [153–155] MD simulations, but their properties have not been thoroughly investigated yet. In addition, they have an extra computational cost which has probably contributed to restrict their range of application so far.

4.3 Sampling of conformational space

Two recently implemented techniques, replica-exchange MD and parallel-replica MD, permit already the study of folding thermodynamics and kinetics, respectively, for reasonably sized polypeptides. Standard MD is still unique in that it delivers both types of information for free. Its timescale limitations are, however, frustratingly obvious. Nevertheless, ensemble-type simulations – many, very long (for current standards) simulations – of peptides in explicit solvent may be within reach in only a few years. The GRID initiatives may decisively contribute to it (http://gridcafe.web.cern.ch/gridcafe/).

Acknowledgements The Spanish MEC/FEDER is acknowledged for financial support; grant ref. BIO2003-02848. Thanks to R.V. Solé and W.F. van Gunsteren for letting me discuss unpublished results. Thanks to W.F. van Gunsteren and his group for their continuous support.

References

1. Shortle D, Simons KT, Baker D (1998) Clustering of low-energy conformations near the native structures of small proteins. Proc Natl Acad Sci US Am 95:11158–11162
2. Pappu RV, Srinivasan R, Rose GD (2000) The Flory isolated-pair hypothesis is not valid for polypeptide chains: implications for protein folding. Proc Natl Acad Sci US Am 97:12565–12570
3. Wong KB, Clarke J, Bond CJ, Neira JL, Freund SMV, Fersht AR, Daggett V (2000) Towards a complete description of the structural and dynamic properties of the denatured state of barnase and the role of residual structure in folding. J Mol Biol 296:1257–1282
4. Bai YW, Chung J, Dyson HJ, Wright PE (2001) Structural and dynamic characterization of an unfolded state of poplar apoplastocyanin formed under nondenaturing conditions. Protein Sci 10:1056–1066

5. Plaxco KW, Gross M (2001) Unfolded, yes, but random? Never! Nat Struct Biol 8:659–660
6. Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. Science 293:487–489
7. van Gunsteren WF, Burgi P, Peter C, Daura X (2001) The key to solving the protein-folding problem lies in an accurate description of the denatured state. Angew Chem Int Ed 40:351–355
8. Choy WY, Mulder FAA, Crowhurst KA, Muhandiram DR, Millett IS, Doniach S, Forman-Kay JD, Kay LE (2002) Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J Mol Biol 316:101–112
9. Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wirmer J, Duchardt E, Ueda T, Imoto T, Smith LJ, Dobson CM, Schwalbe H (2002) Long-range interactions within a nonnative protein. Science 295:1719–1722
10. Zagrovic B, Snow CD, Khaliq S, Shirts MR, Pande VS (2002) Native-like mean structure in the unfolded ensemble of small proteins. J Mol Biol 323:153–164
11. Lei HX, Smith PE (2003) The role of the unfolded state in hairpin stability. Biophys J 85:3513–3520
12. Fitzkee NC, Fleming PJ, Gong HP, Panasik N, Street TO, Rose GD (2005) Are proteins made from a limited parts list? Trends Biochem Sci 30:73–80
13. Dyer RB, Maness SJ, Franzen S, Fesinmeyer RM, Olsen KA, Andersen NH (2005) Hairpin folding dynamics: the cold-denatured state is predisposed for rapid refolding. Biochemistry 44:10406–10415
14. Platt GW, McParland VJ, Kalverda AP, Homans SW, Radford SE (2005) Dynamics in the unfolded state of beta(2)-microglobulin studied by NMR. J Mol Biol 346:279–294
15. Pletneva EV, Gray HB, Winkler JR (2005) Many faces of the unfolded state: conformational heterogeneity in denatured yeast cytochrome c. J Mol Biol 345:855–867
16. Vendruscolo M, Dobson CM (2005) Towards complete descriptions of the free-energy landscapes of proteins. Philos Trans R Soc Lond A Math Phys Eng Sci 363:433–450
17. Kobayashi T, Ikeguchi M, Sugai S (2000) Molten globule structure of equine beta-lactoglobulin probed by hydrogen exchange. J Mol Biol 299:757–770
18. Chakraborty S, Ittah V, Bai P, Luo L, Haas E, Peng ZY (2001) Structure and dynamics of the alpha-lactalbumin molten globule: fluorescence studies using proteins containing a single tryptophan residue. Biochemistry 40:7228–7238
19. Kim YJ, Kim YA, Park N, Son HS, Kim KS, Hahn JH (2005) Structural characterization of the molten globule state of apomyoglobin by limited proteolysis and HPLC-mass spectrometry. Biochemistry 44:7490–7496
20. Troullier A, Reinstadler D, Dupont Y, Naumann D, Forge V (2000) Transient non-native secondary structures during the refolding of alpha-lactalbumin detected by infrared spectroscopy. Nat Struct Biol 7:78–86
21. Matouschek A, Kellis JT, Serrano L, Fersht AR (1989) Mapping the transition-state and pathway of protein folding by protein engineering. Nature 340:122–126
22. Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. 1. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 224:771–782
23. Sosnick TR, Dothager RS, Krantz BA (2004) Differences in the folding transition state of ubiquitin indicated by phi and psi analyses. Proc Natl Acad Sci US Am 101:17377–17382
24. Horng JC, Cho JH, Raleigh DP (2005) Analysis of the pH-dependent folding and stability of histidine point mutants allows characterization of the denatured state and transition state for protein folding. J Mol Biol 345:163–173
25. Raleigh DP, Plaxco KW (2005) The protein folding transition state: what are phi-values really telling us? Prot Pept Lett 12:117–122
26. Tollinger M, Kay LE, Forman-Kay JD (2005) Measuring pKa values in protein folding transition state ensembles by NMR spectroscopy. J Am Chem Soc 127:8904–8905
27. Evans PA, Radford SE (1994) Probing the structure of folding intermediates. Curr Opin Struct Biol 4:100–106
28. Plaxco KW, Dobson CM (1996) Time-resolved biophysical methods in the study of protein folding. Curr Opin Struct Biol 6:630–636
29. Callender RH, Dyer RB, Gilmanshin R, Woodruff WH (1998) Fast events in protein folding: the time evolution of primary processes. Annu Rev Phys Chem 49:173–202
30. Dobson CM, Hore PJ (1998) Kinetic studies of protein folding using NMR spectroscopy. Nat Struct Biol 5:504–507
31. Dyson HJ, Wright PE (1998) Equilibrium NMR studies of unfolded and partially folded proteins. Nat Struct Biol 5:499–503
32. Onuchic JN, LutheySchulten Z, Wolynes PG (1997) Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 48:545–600
33. Chan HS, Dill KA (1998) Protein folding in the landscape perspective: Chevron plots and non-Arrhenius kinetics. Proteins-Struct Funct Genet 30:2–33
34. Thirumalai D, Klimov DK (1999) Deciphering the timescales and mechanisms of protein folding using minimal off-lattice models. Curr Opin Struct Biol 9:197–207
35. Derreumaux P (2000) Ab initio polypeptide structure prediction. Theor Chem Acc 104:1–6
36. Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 25:331–339
37. Ferrara P, Caflisch A (2000) Folding simulations of a three-stranded antiparallel beta-sheet peptide. Proc Natl Acad Sci US Am 97:10780–10785
38. Wang HW, Sung SS (2000). Molecular dynamics simulations of three-strand beta-sheet folding. J Am Chem Soc 122:1999–2009
39. Mirny L, Shakhnovich E (2001) Protein folding theory: from lattice to all-atom models. Annu Rev Biophys Biomol Struct 30:361–396
40. De Mori GMS, Colombo G, Micheletti C (2005) Study of the villin headpiece folding dynamics by combining coarse-grained Monte Carlo evolution and all-atom molecular dynamics. Proteins-Struct Funct Bioinform 58:459–471
41. Ding F, Buldyrev SV, Dokholyan NV (2005) Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model. Biophys J 88:147–155
42. Irback A, Mohanty S (2005) Folding thermodynamics of peptides. Biophys J 88:1560–1569
43. Liwo A, Khalili M, Scheraga HA (2005) Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. Proc Natl Acad Sci US Am 102:2362–2367
44. Daura X, Jaun B, Seebach D, van Gunsteren WF, Mark AE (1998) Reversible peptide folding in solution by molecular dynamics simulation. J Mol Biol 280:925–932
45. Duan Y, Kollman PA (1998) Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 282:740–744
46. Takano M, Yamato T, Higo J, Suyama A, Nagayama K (1999) Molecular dynamics of a 15-residue poly(L-alanine) in water: helix formation and energetics. J Am Chem Soc 121:605–612
47. Higo J, Galzitskaya OV, Ono S, Nakamura H (2001) Energy landscape of a beta-hairpin peptide in explicit water studied by multicanonical molecular dynamics. Chem Phys Lett 337:169–175
48. Hummer G, Garcia AE, Garde S (2001) Helix nucleation kinetics from molecular simulations in explicit solvent. Proteins-Struct Funct Genet 42:77–84
49. Garcia AE, Sanbonmatsu KY (2001) Exploring the energy landscape of a beta hairpin in explicit solvent. Proteins-Struct Funct Genet 42:345–354
50. Snow CD, Nguyen N, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420:102–106
51. Colombo G, De Mori GMS, Roccatano D (2003) Interplay between hydrophobic cluster and loop propensity in beta-hairpin formation: a mechanistic study. Protein Sci 12:538–550

52. Shea JE, Brooks CL (2001) From folding theories to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. Annu Rev Phys Chem 52:499–535
53. Daggett V (2002) Molecular dynamics simulations of the protein unfolding/folding reaction. Acc Chem Res 35:422–429
54. Daggett V, Kollman PA, Kuntz ID (1991) A molecular-dynamics simulation of polyalanine – an analysis of equilibrium motions and helix coil transitions. Biopolymers 31:1115–1134
55. Tobias DJ, Mertz JE, Brooks CL (1991) Nanosecond time scale folding dynamics of a pentapeptide in water. Biochemistry 30: 6054–6058
56. Shakhnovich EI (1997) Theoretical studies of protein-folding thermodynamics and kinetics. Curr Opin Struct Biol 7:29–40
57. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) A 2nd generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules. J Am Chem Soc 117:5179–5197
58. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118:11225–11236
59. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102:3586–3616
60. Daura X, Mark AE, van Gunsteren WF (1998) Parametrization of aliphatic CHn united atoms of GROMOS96 force field. J Comput Chem 19:535–547
61. Daura X, vanGunsteren WF, Rigo D, Jaun B, Seebach D (1997) Studying the stability of a helical beta-heptapeptide by molecular dynamics simulations. Chem Eur J 3:1410–1417
62. Kubelka J, Eaton WA, Hofrichter J (2003) Experimental tests of villin subdomain folding simulations. J Mol Biol 329:625–630
63. Berendsen HJC (1998) Protein folding – a glimpse of the holy grail? Science 282:642–643
64. Chipot C, Pohorille A (1998) Folding and translocation of the undecamer of poly-L-leucine across the water-hexane interface: a molecular dynamics study. J Am Chem Soc 120:11912–11924
65. Bonvin A, van Gunsteren WF (2000) beta-Hairpin stability and folding: molecular dynamics studies of the first beta-hairpin of tendamistat. J Mol Biol 296:255–268
66. Colombo G, Roccatano D, Mark AE (2002) Folding and stability of the three-stranded beta-sheet peptide betanova: insights from molecular dynamics simulations. Proteins-Struct Funct Genet 46:380–392
67. Schaefer M, Bartels C, Karplus M (1998) Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. J Mol Biol 284:835–848
68. Lazaridis T, Karplus M (1999) Effective energy function for proteins in solution. Proteins-Struct Funct Genet 35:133–152
69. Dinner AR, Lazaridis T, Karplus M (1999) Understanding beta-hairpin formation. Proc Natl Acad Sci US Am 96:9068–9073
70. Ferrara P, Apostolakis J, Caflisch A (2002) Evaluation of a fast implicit solvent model for molecular dynamics simulations. Proteins-Struct Funct Genet 46:24–33
71. Simmerling C, Strockbine B, Roitberg AE (2002) All-atom structure prediction and folding simulations of a stable protein. J Am Chem Soc 124:11258–11259
72. Feig M, Brooks CL (2004) Recent advances in the development and application of implicit solvent models in biomolecule simulations. Curr Opin Struct Biol 14:217–224
73. Ferrara P, Apostolakis J, Caflisch A (2000) Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations. J Phys Chem B 104:5000–5010
74. Bursulaya BD, Brooks CL (2000) Comparative study of the folding free energy landscape of a three-stranded beta-sheet protein with explicit and implicit solvent models. J Phys Chem B 104:12378–12383
75. Schafer H, Daura X, Mark AE, van Gunsteren WF (2001) Entropy calculations on a reversibly folding peptide: changes in solute free energy cannot explain folding behavior. Proteins-Struct Funct Genet 43:45– 56
76. Shen MY, Freed KF (2002) Long time dynamics of met-enkephalin: comparison of explicit and implicit solvent models. Biophys J 82:1791–1808
77. Zhou RH, Berne BJ (2002) Can a continuum solvent model reproduce the free energy landscape of a beta-hairpin folding in water? Proc Natl Acad Sci US Am 99:12777–12782
78. Nymeyer H, Garcia AE (2003) Simulation of the folding equilibrium of alpha-helical peptides: a comparison of the generalized born approximation with explicit solvent. Proc Natl Acad Sci US Am 100:13934–13939
79. Zhou RH (2003) Free energy landscape of protein folding in water: explicit vs. implicit solvent. Proteins-Struct Funct Genet 53:148–161
80. Rhee YM, Sorin EJ, Jayachandran G, Lindahl E, Pande VS (2004) Simulations of the role of water in the protein-folding mechanism. Proc Natl Acad Sci US Am 101:6456–6461
81. Stultz CM (2004) An assessment of potential of mean force calculations with implicit solvent models. J Phys Chem B 108:16525–16532
82. Wagoner J, Baker NA (2004) Solvation forces on biomolecular structures: a comparison of explicit solvent and Poisson–Boltzmann models. J Comput Chem 25:1623–1629
83. Snow CD, Sorin EJ, Rhee YM, Pande VS (2005) How well can simulation predict protein folding kinetics and thermodynamics? Annu Rev Biophys Biomol Struct 34:43–69
84. Sugita Y, Okamoto Y (1999) Replica-exchange molecular dynamics method for protein folding. Chem Phys Lett 314:141–151
85. Swendsen RH, Wang JS (1986) Replica Monte-Carlo simulation of spin-glasses. Phys Rev Lett 57:2607–2609
86. Berg BA, Neuhaus T (1991) Multicanonical algorithms for 1st order phase-transitions. Phys Lett B 267:249–253
87. Nakajima N, Nakamura H, Kidera A (1997) Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. J Phys Chem B 101:817–824
88. Voter AF (1998) Parallel replica method for dynamics of infrequent events. Phys Rev B 57:R13985–R13988
89. Sugita Y, Okamoto Y (2000) Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. Chem Phys Lett 329:261–270
90. Rhee YM, Pande VS (2003) Multiplexed-replica exchange molecular dynamics method for protein folding simulation. Biophys J 84:775–786
91. Paschek D, Garcia AE (2004) Reversible temperature and pressure denaturation of a protein fragment: a replica exchange molecular dynamics simulation study. Phys Rev Lett 93(23):238105
92. Affentranger R, Tavernelli I, Di Iorio EE (2005) A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. (submitted)
93. Zhou RH, Berne BJ, Germain R (2001) The free energy landscape for beta hairpin folding in explicit water. Proc Natl Acad Sci US Am 98:14931–14936
94. Garcia AE, Onuchic JN (2003) Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. Proc Natl Acad Sci US Am 100:13898–13903
95. Pitera JW, Swope W (2003) Understanding folding and design: Replica-exchange simulations of "Trp-cage" fly miniproteins. Proc Natl Acad Sci US Am 100:7587–7592
96. Rao F, Caflisch A (2003) Replica exchange molecular dynamics simulations of reversible folding. J Chem Phys 119:4035–4042

97. Ohkubo YZ, Brooks CL (2003) Exploring Flory's isolated-pair hypothesis: statistical mechanics of helix-coil transitions in polyalanine and the C-peptide from RNase A. Proc Natl Acad Sci US Am 100:13916–13921
98. Swope WC, Pitera JW, Suits F (2004) Describing protein folding kinetics by molecular dynamics simulations. 1. Theory. J Phys Chem B 108:6571–6581
99. Swope WC, Pitera JW, Suits F, Pitman M, Eleftheriou M, Fitch BG, Germain RS, Rayshubski A, Ward TJC, Zhestkov Y, Zhou R (2004) Describing protein folding kinetics by molecular dynamics simulations. 2. Example applications to alanine dipeptide and beta-hairpin peptide. J Phys Chem B 108:6582–6594
100. Andrec M, Felts AK, Gallicchio E, Levy RM (2005) Protein folding pathways from replica exchange simulations and a kinetic network model. Proc Natl Acad Sci US Am 102:6801–6806
101. Shirts M, Pande VS (2000) Computing – screen savers of the World unite! Science 290:1903–1904
102. Shirts MR, Pande VS (2001) Mathematical analysis of coupled parallel simulations. Phys Rev Lett 86:4983–4987
103. Fersht AR (2002) On the simulation of protein folding by short time scale molecular dynamics and distributed computing. Proc Natl Acad Sci US Am 99:14122–14125
104. Singhal N, Snow CD, Pande VS (2004) Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. J Chem Phys 121:415–425
105. Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, Larson SM, Rhee YM, Shirts MR, Snow CD, Sorin EJ, Zagrovic B (2003) Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. Biopolymers 68:91–109
106. Seebach D, Beck AK, Bierbaum DJ (2004) The world of beta- and gamma-peptides comprised of homologated proteinogenic amino acids and other components. Chem Biodivers 1:1111–1239
107. Daura X, Glattli A, Gee P, Peter C, Van Gunsteren WF (2002) Unfolded state of peptides. Adv Prot Chem 62:341–360
108. Kritzer JA, Tirado-Rives J, Hart SA, Lear JD, Jorgensen WL, Schepartz A (2005) Relationship between side chain structure and 14-helix stability of beta(3)-peptides in water. J Am Chem Soc 127:167–178
109. Wolynes PG (1995) Biomolecular folding in vacuo!!!? Proc Natl Acad Sci US Am 92:2426–2427
110. Daura X, Mark AE, van Gunsteren WF (1999) Peptide folding simulations: no solvent required? Comput Phys Commun 123:97–102
111. Velazquez I, Reimann CT, Tapia O (1999) Proteins in vacuo: relaxation of unfolded lysozyme leads to folding into native and non-native structures. A molecular dynamics study. J Am Chem Soc 121:11468–11477
112. Levy Y, Jortner J, Becker OM (2001) Solvent effects on the energy landscapes and folding kinetics of polyalanine. Proc Natl Acad Sci US Am 98:2188–2193
113. Daura X, van Gunsteren WF, Mark AE (1999) Folding-unfolding thermodynamics of a beta-heptapeptide from equilibrium simulations. Proteins-Struct Funct Genet 34:269–280
114. Zagrovic B, Pande VS (2004) How does averaging affect protein structure comparison on the ensemble level? Biophys J 87:2240–2246
115. de Groot BL, Daura X, Mark AE, Grubmuller H (2001) Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds. J Mol Biol 309:299–313
116. Hamprecht FA, Peter C, Daura X, Thiel W, van Gunsteren WF (2001) A strategy for analysis of (molecular) equilibrium simulations: Configuration space density estimation, clustering, and visualization. J Chem Phys 114:2079–2089
117. Ikeda K, Galzitskaya OV, Nakamura H, Higo J (2003) beta-hairpins, alpha-helices, and the intermediates among the secondary structures in the energy landscape of a peptide from a distal beta-Hairpin of SH3 domain. J Comput Chem 24:310–318
118. Corcho FJ, Canto J, Perez JJ (2004) Comparative analysis of the conformational profile of substance P using simulated annealing and molecular dynamics. J Comput Chem 25:1937–1952
119. Bursulaya BD, Brooks CL (1999) Folding free energy surface of a three-stranded beta-sheet protein. J Am Chem Soc 121:9947–9951
120. Huisinga W, Best C, Roitzsch R, Schutte C, Cordes F (1999) From simulation data to conformational ensembles: structure and dynamics-based methods. J Comput Chem 20:1760–1774
121. Onuchic JN, Wolynes PG (2004) Theory of protein folding. Curr Opin Struct Biol 14:70–75
122. Daura X, Antes I, van Gunsteren WF, Thiel W, Mark AE (1999) The effect of motional averaging on the calculation of NMR-derived structural properties. Proteins-Struct Funct Genet 36:542–555
123. Cardenas AE, Elber R (2003) Kinetics of cytochrome C folding: atomically detailed simulations. Proteins-Struct Funct Genet 51:245–257
124. Kohn JE, Millett IS, Jacob J, Zagrovic B, Dillon TM, Cingel N, Dothager RS, Seifert S, Thiyagarajan P, Sosnick TR, Hasan MZ, Pande VS, Ruczinski I, Doniach S, Plaxco KW (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. Proc Natl Acad Sci US Am 101:12491–12496
125. Wei GH, Mousseau N, Derreumaux P (2004) Complex folding pathways in a simple beta-hairpin. Proteins-Struct Funct Bioinform 56:464–474
126. van Gunsteren WF, Burgi R, Peter C, Daura X (2001) Comment on the communication "The key to solving the protein-folding problem lies in an accurate description of the denatured state" by van gunsteren et al. – reply. Angew Chem Int Ed 40:4616–4618
127. Li MS, Klimov DK, Thirumalai D (2004) Thermal denaturation and folding rates of single domain proteins: size matters. Polymer 45:573–579
128. Naganathan AN, Munoz V (2005) Scaling of folding times with protein size. J Am Chem Soc 127:480–481
129. Kubelka J, Hofrichter J, Eaton WA (2004) The protein folding 'speed limit'. Curr Opin Struct Biol 14:76–88
130. Dinner AR, Karplus M (2001) Comment on the communication "The key to solving the protein-folding problem lies in an accurate description of the denatured state" by van gunsteren et al. Angew Chem Int Ed 40:4615–4616
131. Cavalli A, Haberthur U, Paci E, Caflisch A (2003) Fast protein folding on downhill energy landscape. Protein Sci 12:1801–1803
132. Strogatz SH (2001) Exploring complex networks. Nature 410:268–276
133. Rao F, Caflisch A (2004) The protein folding network. J Mol Biol 342:299–306
134. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. J Phys Chem B 105:6474–6487
135. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong GM, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang JM, Kollman P (2003) A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J Comput Chem 24:1999–2012
136. Mackerell AD (2004) Empirical force fields for biological macromolecules: overview and issues. J Comput Chem 25:1584–1604
137. Oostenbrink C, Villa A, Mark AE, Van Gunsteren WF (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. J Comput Chem 25:1656–1676
138. Hu H, Elstner M, Hermans J (2003) Comparison of a QM/MM force field and molecular mechanics force fields in simulations of alanine and glycine "dipeptides" (Ace-Ala-Nme and Ace-Gly-Nme) in water in relation to the problem of modeling the unfolded peptide backbone in solution. Proteins-Struct Funct Genet 50:451–463
139. Mu YG, Kosov DS, Stock G (2003) Conformational dynamics of trialanine in water. 2. Comparison of AMBER, CHARMM, GROMOS, and OPLS force fields to NMR and infrared experiments. J Phys Chem B 107:5064–5073

140. Okur A, Strockbine B, Hornak V, Simmerling C (2003) Using PC clusters to evaluate the transferability of molecular mechanics force fields for proteins. J Comput Chem 24:21–31
141. Gnanakaran S, Garcia AE (2005) Helix-coil transition of alanine peptides in water: force field dependence on the folded and unfolded structures. Proteins-Struct Funct Bioinform 59:773–782
142. Sorin EJ, Pande VS (2005) Empirical force-field assessment: the interplay between backbone torsions and noncovalent term scaling. J Comput Chem 26:682–690
143. Glattli A, Daura X, van Gunsteren WF (2002) Derivation of an improved simple point charge model for liquid water: SPC/A and SPC/L. J Chem Phys 116:9811–9828
144. Mahoney MW, Jorgensen WL (2000) A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. J Chem Phys 112:8910–8922
145. Rick SW (2004) A reoptimization of the five-site water potential (TIP5P) for use with Ewald sums. J Chem Phys 120:6085–6093
146. Cieplak P, Caldwell J, Kollman P (2001) Molecular mechanical models for organic and biological systems going beyond the atom centered two body additive approximation: aqueous solution free energies of methanol and N-methyl acetamide, nucleic acid base, and amide hydrogen bonding and chloroform/water partition coefficients of the nucleic acid bases. J Comput Chem 22:1048–1057
147. Kaminski GA, Stern HA, Berne BJ, Friesner RA (2004) Development of an accurate and robust polarizable molecular mechanics force field from ab initio quantum chemistry. J Phys Chem A 108:621–627
148. Patel S, Brooks CL (2004) CHARMM fluctuating charge force field for proteins: I parameterization and application to bulk organic liquid simulations. J Comput Chem 25:1–15
149. Patel S, Mackerell AD, Brooks CL (2004) CHARMM fluctuating charge force field for proteins: II – Protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. J Comput Chem 25:1504–1514
150. Borjesson U, Hunenberger PH (2001) Explicit-solvent molecular dynamics simulation at constant pH: methodology and application to small amines. J Chem Phys 114:9706–9719
151. Baptista AM, Teixeira VH, Soares CM (2002) Constant-pH molecular dynamics using stochastic titration. J Chem Phys 117:4184–4200
152. Burgi R, Kollman PA, van Gunsteren WF (2002) Simulating proteins at constant pH: an approach combining molecular dynamics and Monte Carlo simulation. Proteins-Struct Funct Genet 47:469–480
153. Dlugosz M, Antosiewicz JM (2004) Constant-pH molecular dynamics simulations: a test case of succinic acid. Chem Phys 302:161–170
154. Lee MS, Salsbury FR, Brooks CL (2004) Constant-pH molecular dynamics using continuous titration coordinates. Proteins-Struct Funct Bioinform 56:738–752
155. Mongan J, Case DA, McCammon JA (2004) Constant pH molecular dynamics in generalized born implicit solvent. J Comput Chem 25:2038–2048