Article

Theory and Practice of Coarse-Grained Molecular Dynamics of Biologically Important Systems

Adam Liwo [1,*], Cezary Czaplewski [1], Adam K. Sieradzan [1], Agnieszka G. Lipska [1], Sergey A. Samsonov [1] and Rajesh K. Murarka [2]

[1] Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland; cezary.czaplewski@ug.edu.pl (C.C.); adam.sieradzan@ug.edu.pl (A.K.S.); agnieszka.lipska@ug.edu.pl (A.G.L.); sergey.samsonov@ug.edu.pl (S.A.S.)

[2] Department of Chemistry, Indian Institute of Science Education and Research Bhopal, Bhopal Bypass Road, Bhopal 462066, MP, India; rkm@iiserb.ac.in

* Correspondence: adam.liwo@ug.edu.pl; Tel.: +48-58-523-5194

Abstract: Molecular dynamics with coarse-grained models is nowadays extensively used to simulate biomolecular systems at large time and size scales, compared to those accessible to all-atom molecular dynamics. In this review article, we describe the physical basis of coarse-grained molecular dynamics, the coarse-grained force fields, the equations of motion and the respective numerical integration algorithms, and selected practical applications of coarse-grained molecular dynamics. We demonstrate that the motion of coarse-grained sites is governed by the potential of mean force and the friction and stochastic forces, resulting from integrating out the secondary degrees of freedom. Consequently, Langevin dynamics is a natural means of describing the motion of a system at the coarse-grained level and the potential of mean force is the physical basis of the coarse-grained force fields. Moreover, the choice of coarse-grained variables and the fact that coarse-grained sites often do not have spherical symmetry implies a non-diagonal inertia tensor. We describe selected coarse-grained models used in molecular dynamics simulations, including the most popular MARTINI model developed by Marrink's group and the UNICORN model of biological macromolecules developed in our laboratory. We conclude by discussing examples of the application of coarse-grained molecular dynamics to study biologically important processes.

Keywords: molecular dynamics; coarse graining; biological macromolecules

1. Introduction

Molecular dynamics (MD) is an extremely useful methodology to investigate the behavior of biomolecules and biomolecular systems and soft matter in general at a level of detail that is still inaccessible during experiments [1–5]. Although, as with all simulation-based methods, it can lead to unwanted artefacts and mis-conclusions, a careful researcher who consistently checks the obtained results against the available experimental data can make important advances in the field with its use. The most accurate treatment is that at the quantum-mechanical level [6,7]; however, this approach is nowadays restricted to small systems. As is commonly understood, MD implies carrying out simulations at the all-atom (AA) level. Owing to tremendous advances in computer techniques and algorithms [3], including the construction of the ANTON supercomputer by the D.E. Shaw group [8,9], millisecond-scale simulations are now feasible for small proteins and it is even possible to run simulations of entire viruses or cell fragments at the microsecond scale [10]. This is, however, still not enough to cover the time scale of biological events. Therefore, 17 years after Alder and Wainwright [1] had published the results of their first MD simulation, the reduction of atomistic representation of the system studies was attempted by Levitt and Warshell [11]. This reduction is termed *coarse graining* (CG). With coarse graining, it is

possible to extend the time and size scale of simulations by orders of magnitude, thereby covering the biological scales [12–14].

Coarse graining is, however, much more difficult to carry out and has many more caveats compared to the all-atom approaches. First, the degrees of freedom that are not present in a CG model (the fine-grained degrees of freedom) are ignored. This seems to be fine, because a high-school student, who is posed the problem of determining the time after which a brick dropped from height *h* will hit the ground, does not need to consider the atomic details of the brick and needs only to know the "coarse-grained" Newton laws of dynamics. However, the separation of the time scales of the motions of the coarse-grained and the fine-grained degrees of freedom is much more diffuse than that in the physics textbook example mentioned. As a result, the relationship between the time scale of coarse-grained MD simulations and the real or all-atom simulation time scale is largely obtained by comparing the times of events that can easily be traced (e.g., helix formation) or calculated rate constants with those from the experiment or all-atom simulations [15–18].

A related problem is the construction of the equations of motion. In almost all approaches, the extended sites are represented by point masses, just as atomic nuclei in all-atom simulations [12]. However, this representation of CG sites does not seem to be satisfactory, because these object are, generally, non-spherical. Replacing the full inertia matrix with a diagonal one, composed of point masses, is likely to influence the dynamics. Rigid-body quaternion dynamics [19,20] or treating the sites as stretchable rods [21] are plausible ways to take into account site anisotropy.

The most significant challenge, though, is the development of the coarse-grained force fields [12–14,22,23]. Usually, the functional expressions are copied from all-atom force fields, which results in insufficient capacity to model the structures of the systems under study (e.g., protein structures). The reason for this is that the fine-grained degrees of freedom, although not considered in CG models, can give rise to strong coupling between the coarse-grained degrees of freedom [24–26]. A variety of approaches have been designed to tackle these problems, including the multiscale force-matching approach [27] and scale-consistent expansion of the potential of mean force into Kubo cluster cumulants [24–26]. Nevertheless, the force fields with all-atom-like expressions for energy components can be used to model the dynamics and to estimate the properties of the systems studied, provided that they are applied with care.

The purpose of this article is to provide the reader with an overview of coarse-grained molecular dynamics and its applications in the study of biologically important systems. We do not describe here the approaches based on the Monte Carlo (MC) methodology, some of which, such as the CABS model of proteins [28] and the SimRNA model of nucleic acids [29], are very successful in modeling protein structures, simulating folding pathways, and protein–protein docking [30,31] or simulating RNA folding. We start from the theory, by sketching the derivation of the equations of motion and showing that the effective forces acting on coarse-grained degrees of freedom can be separated into the mean potential forces and the friction and stochastic forces that arise from the motion of the omitted degrees of freedom. Next, we discuss the equations of motion used in different CG MD schemes and the numerical algorithms of solving them. Subsequently, we describe the approaches to coarse-grained force field derivation and discuss selected coarse-grained force fields. Lastly, we discuss the applications of CG MD, illustrating them with examples from our research, and conclude by discussing possible directions of the developments of CG MD approaches.

2. Theory and Methodology

2.1. Origin of CG Dynamics: Separation of the Coarse-Grained and Fine-Grained Motions

Let us consider a system, such as a chunk of a polysaccharide chain, which is coarse-grained, by selecting some variables (e.g., coordinates of the glycosidic-oxygen atoms), the other ones being ignored (Figure 1).
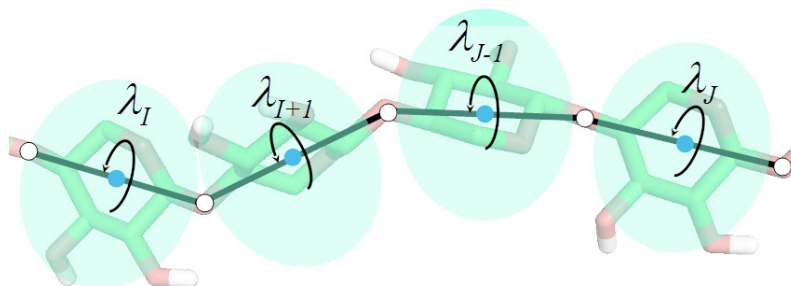
Figure 1. An illustration of coarse graining with the example of a chunk of poly-α-D-glucose chain. The spheroids (anchored in the glycosidic-oxygen atoms) represent united sugar rings and the angles λ for the rotation about the O···O virtual bond axes are the secondary degrees of freedom, which are not present in the CG model.

Let $\mathbf{R} = [\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M]^T$ and $\mathbf{P} = [\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_M]^T$ denote the coarse-grained degrees of freedom and the associated momenta, while $\mathbf{r}_i = [\mathbf{r}_{i1}, \mathbf{r}_{i2}, \ldots, \mathbf{r}_{in_i}]^T$ and $\mathbf{p}_i = [\mathbf{p}_{i1}, \mathbf{p}_{i2}, \ldots, \mathbf{p}_{in_i}]^T$ are the fine-grained degrees of freedom and momenta corresponding to the atoms of site i. At this point, we assume that the $\mathbf{R}$s and the $\mathbf{P}$s are the coordinates of the centers of mass of the sites and the corresponding momenta. By using the Mori–Zwanzig projection-operator formalism and Liouville equations [15,32,33], Kinjo and Hyodo [16] derived the equations of motion for the coarse-grained degrees of freedom (Equation (1)). Equation (1) was derived under the assumption that the coarse-grained degrees of freedom are the mass centers of the coarse-grained sites; however, its extension to generalized coordinates is straightforward.

$$\frac{d}{dt}\mathbf{P}_i(t) = -\nabla_{\mathbf{R}_i} W(\mathbf{R}) - \frac{1}{k_B T} \sum_j \int_0^t \left\langle \left[\delta\mathbf{F}_i^Q(\tau)\right]\left[\delta\mathbf{F}_j(\tau)^Q\right]^T \right\rangle \mathbf{v}_i(\tau)d\tau + \delta\mathbf{F}_i^Q(t) \quad (1)$$

where $W(\mathbf{R})$ is the potential of mean force, $\delta\mathbf{F}_i^Q$ is the fluctuating force (around the mean force value) acting on the ith coarse-grained site, k_B is the Boltzmann constant, T is the absolute temperature, $\mathbf{v}_i = \mathbf{P}_i/M_i$, where M_i is the mass corresponding to the ith coarse-grained site, is the velocity of the ith coarse-grained site, $\nabla_{\mathbf{R}_i} = \left(\frac{\partial}{\partial X_i}, \frac{\partial}{\partial Y_i}, \frac{\partial}{\partial Z_i}\right)$ is the gradient operator in the Cartesian coordinates of the ith site, and $\langle \cdots \rangle$ denotes the ensemble average. The first term on the right-hand side of Equation (1) is the mean force acting on the ith coarse-grained site, the second term corresponds to the friction forces and the last term refers to the fluctuating (stochastic) forces. The potential of mean force and the fluctuating forces are defined by Equations (2) and (3), respectively. The potential of mean force will be discussed in Section 2.3.

$$W(\mathbf{R}) = -k_B T \ln \frac{\int_{\mathbf{r}} \exp\left[-\frac{V(\mathbf{R};\mathbf{r})}{k_B T}\right]d\Omega_{\mathbf{r}}}{\int_{\mathbf{r}} d\Omega_{\mathbf{r}}} \quad (2)$$

where $V(\mathbf{R};\mathbf{r})$ is the potential energy function of the system and $d\Omega_{\mathbf{r}}$ is the volume element corresponding to the fine-grained degrees of freedom.

$$\delta\mathbf{F}_i^Q(t) = e^{t\hat{\mathbf{Q}}\hat{\mathbf{L}}}\delta\mathbf{F}_i^Q(0) \quad (3)$$

where $\hat{\mathbf{L}}$ is the Liouville operator and $\hat{\mathbf{Q}} = 1 - \hat{\mathbf{P}}$, where $\hat{\mathbf{P}}$ is the projection operator of the coordinates and momenta of the all-atom system to the coarse-grained coordinates and momenta [15,16,32,33].

From Equation (1), it follows that the net motion of the CG degrees of freedom is governed by the mean forces (the forces averaged over the atoms that constitute the interaction sites), while the fine-grained degrees of freedom contribute to the motion of the

CG degrees of freedom through the friction forces (which depend on the time correlation of the fluctuations of the forces acting on the sites) and the fluctuating forces. The friction forces depend on the whole history of the correlation between the fluctuating forces and the velocities of the coarse-grained centers [17,34]. A number of studies were carried out with liquids or simple polymers to determine the friction term exactly based on all-atom simulations. The reader is referred to the excellent review by Klippenstein et al. [17] for details.

In practical implementations, the friction and the stochastic force terms are taken from the Langevin equation [35], which is equivalent to the assumption that the fluctuating forces and the coarse-grained velocities are correlated only over an infinitesimally small period of time (δ-correlated); in other words, it is assumed that the fine-grained degrees of freedom move much faster than the coarse-grained ones. This results in replacing the last two terms on the right-hand side of Equation (1) with the net friction and stochastic force terms, respectively (Equation (4)) [15,17,21,34,36,37].

$$\mathbf{G}\ddot{\mathbf{q}} = -\nabla_{\mathbf{q}}U(\mathbf{q}) - \mathbf{\Gamma}\dot{\mathbf{q}} + \mathbf{f}^{rand} \qquad (4)$$

where $\mathbf{G}$ is the inertia matrix, $\mathbf{q}$, $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ are the generalized coarse-grained coordinates, velocities and accelerations, respectively, U is the effective coarse-grained energy function (which originates in the potential of mean force; see Section 2.3) [24,25,27], $\mathbf{\Gamma}$ is the friction matrix, and $\mathbf{f}^{rand}$ are the random forces. From Equation (1), it can be inferred that the friction and stochastic force term are interrelated; in fact, they are related through the friction matrix, as given by Equation (5).

$$\mathbf{f}^{rand} = \left(\frac{k_BT}{\delta t}\mathbf{\Gamma}\right)^{\frac{1}{2}}\mathbf{N}(0,1) \qquad (5)$$

where $\mathbf{N}(0,1)$ is the multidimensional normal distribution with zero mean and unit variance–covariance matrix and δt is the time step.

The friction and stochastic forces can be divided into those due to the solvent (the degrees of freedom of which are wholly or in part averaged over in the CG approaches) and those due to the averaged over internal degrees of freedom of the solute, which give rise to the so-called internal friction. In implicit solvent CG models, all solvent degrees of freedom are averaged over. Therefore, the internal friction has a significantly smaller effect on the dynamics than the solvent friction [38]. Consequently, the stochastic and the friction forces are assumed to be due to the solvent. For explicit solvent CG models, the solvent contribution to friction is smaller but still significant, because the rotational degrees of freedom of the solvent molecules are averaged over and, therefore, the stochastic and friction forces should also be included in MD with these models. The solvent viscosity is usually scaled down in MD simulations with implicit solvent CG models and, consequently, the friction and stochastic forces in such simulations are only a fraction of those resulting from Equation (1), this giving rise to the tremendous speed-up of the CG dynamics with respect to the AA dynamics [37].

In the overdamped limit, when the friction is so high that the left-hand side of Equation (4) can be ignored, it becomes a system of first-order differential equations that describes the Brownian dynamics.

2.2. Implementation of Coarse-Grained MD

Equation (4) can be applied to any choice of coordinate system. In most of the applications, the coordinates of the centers of masses (CM) of the coarse-grained sites are selected and the sites are treated as point masses [27,39]. In this case, the matrix of inertia is diagonal (with site masses as elements). The friction matrix also is typically diagonal, the coefficients being the net friction coefficients of the interaction sites (Equation (6)) [3,4]. For implicit solvent CG models, it is assumed that only the solvent contributes to the friction.

In some approaches, hydrodynamic interactions with the solvent are considered [40–42], the friction matrix having usually the Rotne–Prager form [43].

$$M_I \ddot{\mathbf{R}}_I = -\nabla_{\mathbf{R}_I} U(\mathbf{R}) - \gamma_I \dot{\mathbf{R}}_I + \sqrt{\frac{k_B T \gamma_I}{\delta t}} \mathbf{N}(0, 1) \tag{6}$$

$$\mathbf{R}_I = \frac{\sum_{i \in \{I\}} m_i \mathbf{r}_i}{M_I} \tag{7}$$

$$M_I = \sum_{i \in \{I\}} m_i \tag{8}$$

where $\mathbf{R}_I$ is the position of the center of the mass of site I (with mass M_I), γ_I is the friction coefficient of site I, $\mathbf{N}(0, 1)$ is the 3D vector of normally distributed numbers with zero mean and unit variance, and δt is the time step. The friction coefficients are usually expressed by using the Stokes' law (Equation (9)) [44].

$$\gamma_i = 6\pi \rho_I \eta \tag{9}$$

where η is the solvent's viscosity and ρ_I is the Stokes' radius of a site, which is obtained by adding the effective dimension of the water molecule (1.4 Å) to the site's effective van der Waals radius; the radius is often scaled by the fraction of the solvent-accessible surface of a site. The solvent viscosity is usually scaled down to speed up simulations (typically by a factor of 100 [37]).

The coarse-grained sites have, however, moments of inertia and, therefore, considering only the motion of their centers of masses is often insufficient (a good example is the dynamics of a liquid composed of anisotropic particles). The rotational dynamics of the coarse-grained particles is usually handled in rigid-body mode through the quaternion formalism [19,20,45]. We have developed an alternative formalism, in which the interaction sites are not assumed rigid but to have axial symmetry after averaging out the secondary degrees of freedom [21,36], which is the case of proteins and other biopolymers. The interaction sites are assumed to behave as uniformly stretchable rods, this resulting in a full, albeit constant, inertia matrix, which can be inverted once and its inverse used to obtain accelerations from forces (Equation (4)). We initially applied this formalism to the UNRES model of polypeptide chains developed in our laboratory [46]; subsequently, we extended it to the Unified Coarse-Grained (UNICORN) model of biological macromolecules, which also encompasses nucleic acids and polysaccharides [47]. The sites are assumed to be linked with virtual bonds, a virtual bond being the axis of a site. The resulting equations of motion can be expressed by Equation (10) [21,37].

$$\left(\mathbf{A}^T \mathbf{M} \mathbf{A} + \mathbf{I} \right) \ddot{\mathbf{q}} = -\nabla_{\mathbf{q}} U(\mathbf{q}) + \mathbf{A}^T \mathbf{\Gamma} \ddot{\mathbf{q}} + \mathbf{A}^T \mathbf{f}^{rand} \tag{10}$$

where $\mathbf{A}$ is the matrix of the (linear) mapping of the generalized coordinates $\mathbf{q}$ to the Cartesian coordinates of the site centers, $\mathbf{M}$ is the diagonal matrix of the masses of the sites, $\mathbf{I}$ is the diagonal matrix of the moments of inertia of the sites about the axes perpendicular to their virtual bond axes, $\mathbf{\Gamma}$ is the matrix of friction coefficients of the sites (which is usually assumed to be composed of diagonal elements only, but full friction tensors of the sites or the full Rotne–Prager tensor of hydrodynamic interactions can be introduced [42]) and $\mathbf{f}^{rand}$ is a vector composed of the vectors of random forces acting on the sites. Initially [21,36], we used the virtual bond vectors as generalized coordinates, this resulting in a full inertia matrix; recently [26,48], we replaced the vectors with the anchor points of the sites (the α-carbon and side-chain-center coordinates in UNRES), thereby reducing the inertia matrix.

Many algorithms have been designed to solve the equations of motion numerically, most of them being based on the Verlet, velocity Verlet and leap-frog schemes [3]. These algorithms are symplectic, i.e., they conserve the total energy, if the right-hand side of the equation of motion contains only the potential force term. For these algorithms, the error in trajectory is proportional to the square of the time step, which seems to be a

disadvantage when compared with the higher-order algorithm (e.g., the Runge–Kutta family algorithms or the Gear algorithm [3]) designed to solve the systems of differential equations. However, the energy conservation property, which is not a feature of the more sophisticated algorithms mentioned above, enables us to obtain reliable trajectories even with large time steps, especially when the multiple-time-step or time-split algorithms are applied, in which the Liouville operator is split into the part corresponding to fast- and slow-varying forces [3]. With the time-split feature, it is possible to extend the time step even to 10 fs [21].

With the friction and stochastic forces, not the total energy but the kinetic temperature (defined by Equation (11)) is to be conserved; these contributions to forces provide, therefore, a thermostat.

$$T_{kin} = \frac{E_k}{N_f k_B} \tag{11}$$

where E_k is the kinetic energy of the system and N_f is the number of the degrees of freedom. The Langevin thermostat is very stable and provides correct temperature distribution [49]. Therefore, its use has been recommended even in all-atom MD with explicit solvent [50]. In this case, the non-conservative forces can be considered as those due to the environment of the system that is subjected to MD simulations.

Inclusion of the friction and stochastic forces in numerical integration of the equations of motion is not straightforward. In the simplest instance, they can be included explicitly and this is the method of choice when the friction matrix is not diagonal [37]. However, the stochastic forces vary rapidly, which can result in problems with temperature conservation when the time step is large. The most elaborate algorithm has been developed by Ciccoti and coworkers [51,52], in which the friction and stochastic forces are pre-integrated over the time step, this resulting in exponential terms in the Verlet-family integrators. However, this algorithm becomes prohibitively expensive with the full friction matrix, because calculating the matrix exponential requires its diagonalization. Thermostating can also be achieved without the stochastic and friction term through the use of the Berendsen [53,54], Nosé–Hoover [55] or Nosé–Poincaré [56] thermostats; the latter two also enable us to rigorously control temperature conservation [52].

Apart from full-blown equations of motion, also the discontinuous MD, which is reminiscent of the early work of Alder and Wainwright [1], has been implemented with some CG models. An example is the knowledge-based PRIME 20 CG model of polypeptide chains [57,58].

2.3. Effective Potential Energy Functions

As demonstrated in Section 2.1, the mean forces acting on the coarse-grained degrees of freedom (the negative of the gradient of the potential of mean force) are the conservative forces that govern the motion of the system in coarse-grained representation. A natural consequence of this observation is that the prototype of the effective energy function of a coarse-grained system is the potential of mean force, in which the secondary degrees of freedom are integrated out (Equation (2)). This definition is, indeed, applied to derive the physics-based force fields [24,27,39] and can even be extended to derive the statistical force fields, the components of which are obtained by Boltzmann inversion of the distribution functions extracted from structural databases [59–61]. However, the exact PMF is specific for a given system and prohibitively expensive to determine. Therefore, approximations have to be developed, in which the effective energy function is split into transferable components, as in all-atom force fields. This approach was originated by the seminal work by Levitt and Warshel [11], who proposed the first physics-based coarse-grained model of proteins and derived the CG terms by Boltzmann averaging out the all-atom energy of model systems. The two major approaches are the bottom-up and the top-down approach [12,26,39]. In the first one, the energy terms are obtained from atomistically detailed interactions, as in the work by Levitt and Warshel [11], while in the top-down approach, a force field is constructed so as to reproduce the measurable properties. A

number of methods combine the bottom-up approach in the derivation of coarse-grained energy terms and the top-down approach to put them together into a working force field. In this section, we summarize some of the coarse-grained force fields and methods of their derivation, referring the reader to dedicated review articles [23,61–64] and books [12–14] for details.

The most straightforward method of the construction of a coarse-grained force field is to import the respective expressions from the all-atom force fields. The effective energy function is then a sum of local terms, corresponding to virtual bond stretching (U_s), virtual bond angle bending (U_θ), torsional terms corresponding to the rotation about the virtual bonds (U_{tor}) and the pairwise interaction terms that consist of the electrostatic (usually Coulombic) and nonbonded terms, the latter usually expressed by the Lennard–Jones potential, which can be modified to introduce a softer repulsion term or dependence on site orientation (the most commonly used form of the latter is the Gay–Berne potential [65]). The solvent is either treated explicitly (usually by coarse graining several water molecules into one center) or is implicit in the potentials; in the second case, an additional term V_{solv} can be added to account for, e.g., exposure of a site to the solvent. The whole energy expression is given by Equation (12). Expressions of this type are referred to as the neoclassical expressions. Such expressions are implemented in the most popular MARTINI force field [66–70].

$$U = \sum_i \frac{1}{2} k_i^d (d_i - d_i^\circ)^2 + \sum_i \frac{1}{2} k_i^\theta (\theta_i - \theta_i^\circ)^2 + \sum_i \sum_n a_i^{(n)} [1 + \cos(n\gamma_i)] + b_i^{(n)} [1 + \sin(n\gamma_i)] +$$

$$332 \sum_{i<j} \frac{q_i q_j}{D r_{ij}} + 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + V_{solv} \qquad (12)$$

where d_i, d_i° and k_i^d are the length, equilibrium length and force constant of the ith virtual bond, θ_i, θ_i° and k_i^θ are the actual and equilibrium value and the force constant of the ith virtual bond angle, γ_i is the ith virtual bond dihedral angle, $a_i^{(n)}$ and $b_i^{(n)}$ being the coefficients in the expressions for the torsional potentials, q_i is the partial charge on the ith site, D is the relative dielectric permittivity, σ_{ij} and ϵ_{ij} are the constants of the Lennard–Jones potential for the interaction of site i with site j, and r_{ij} is the distance between these two sites. The coefficient of 332 in the expression for the Coulombic energy is introduced to express the energy in kcal/mol if the distance is expressed in ångstroms and the charges are expressed in electron charge units.

The neoclassical expressions usually do not produce the force fields that can model the structures of the systems under study without external information. The reason is that the PMF cannot be broken down into pairwise terms, even if the parent all-atom energy function is pairwise [24,25,60]. Multibody terms need to be introduced for structure modeling. Moreover, the site–site interaction potentials of most coarse-grained force fields are usually too "sticky", this resulting in too compact modeled structures [71–75]. The "stickiness" arises from interaction imbalance, which can be caused by the absence of specific cross Lennard–Jones parameters for pairs of particles of different size or overly weak bond force constants [76]. This problem was resolved recently in the MARTINI 3 force field [77].

A method for the derivation of explicit multibody terms, which is based on the expansion of the PMF into Kubo cluster cumulant functions [78], has been developed in our laboratory [24,25]. Following this method, the PMF is expressed as a sum of elementary PMFs characteristic of single sites, pairs of sites and higher clusters of sites. The factors can be identified with single-body, pairwise and multibody terms in the effective energy function. Each factor can be expanded into the Kubo generalized cumulants [78] and the lowest-order cumulants can be used as templates of the analytical expressions for the respective energy terms. Notably, the cumulant-based derivation of the effective energy terms enables us to determine which effective energy terms should be scaled depending on temperature [79]; that this should be the case is easy to realize given the origin of

the CG energy functions in the PMF (Equation (2)). Recently [25,26], we developed a generalized mathematical formalism for the derivation of the formulas for the cumulants for the coarse-grained representations of polymer chains, by expanding the all-atom energy in the trigonometric functions of the collective rotation angles of the groups about the virtual bond axes (Figure 1). The obtained formulas implicitly include the atomic details of the system; in particular, they keep the correct dependence of the interaction energy of the non-radial sites on orientation and the correct dependence of the site–site interaction energy on local geometry, through multibody terms. We have termed this approach the scale consistent approach to force field derivation. The procedure has been summarized and illustrated with examples in our recent review article [26]. This feature solves part of the problem of force field "stickiness" pointed out in [75], because the force field contains explicit terms that couple the backbone-local and backbone-electrostatic conformational states, which prevent overly compact local chain fragments [80]. Another way of including the multibody terms is to introduce expressions derived based on structural regularities in crystal structures [60,81] or on a heuristic basis [82,83].

As mentioned, the parametrization of the effective energy function follows the bottom-up, the top-down approach or a combination of both. In the bottom-up approach, the individual components are parametrized based on direct numerical integration of all-atom energy surfaces [11,84–86] or from MD simulations of model systems [87–89]. In the top-down approach, all parameters are optimized simultaneously. One method is to reproduce the physicochemical properties such as the partition coefficients between the aqueous and lipid phase, the structural properties (e.g., the virtual bond and virtual bond dihedral angle distribution from the Protein Data Bank (PDB)) or the observables obtained from the all-atom MD simulations. This approach has been used to develop the MARTINI force field [66,67]. A systematic feedback between the CG and the corresponding AA system with iterative improvement of the parameters is achieved through the iterative Boltzmann inversion [90,91] and inverse Monte Carlo (IMC) method [92,93]. A more sophisticated approach is the relative entropy minimization [94–97], in which the Kullback–Leibler divergence [98] between the CG and AA ensembles is minimized. We developed a variant of this method, which we termed maximum-likelihood optimization [99], replacing the AA-simulated ensembles with the experimental ensembles, obtained from Nuclear Magnetic Resonance (NMR) studies of the training proteins at various temperatures, which were selected to bracket the folding–unfolding transition [100]. Including the experimental ensemble required "Gaussian smearing" of each of the experimental conformations so that it could be matched to the conformation of the simulated ensemble. This method was implemented in optimizing the UNRES force field for proteins [80,101] and the NARES-2P force field for nucleic acids [102].

Because one of the purposes of running MD simulations is obtaining information about the dynamic behavior of a system, not only the compatibility between the CG effective energy surfaces with the AA potential energy surfaces but also the compatibility of the CG forces with the average forces computed at the AA level is very important. To achieve this, the force-matching method can be used, which has been developed for the purpose of parametrizing the CG force field by the Voth group [27,39,103] as the Multi-Scale Coarse-Grained (MS-CG) method. In this method, the CG energy function is parametrized to minimize the sum of the squares of the differences between the CG and the corresponding average AA forces. It is assumed that the centers of the masses of the sites are coarse-grained centers, all interactions have radial symmetry, and all effective interactions are pairwise. Consequently, the distance profiles of the site–site potentials can be determined without assuming any specific functional form. The multibody terms are assumed to be embedded into the effective potentials. We extended the method to non-spherical and multibody potentials (however, assuming their functional forms) and recently used it to parametrize the UNRES force field, obtaining its variant that produces forces compatible with those computed with the all-atom approach; it is transferable and is able to fold proteins [26].

Machine learning is playing an increasing role in molecular modeling, the most spectacular success being achieved by Google Deep Mind in the 14th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP14), in which the group predicted the structures of most of the target proteins with crystallographic accuracy [104,105]. It is also increasingly used in force field optimization, regarding both the all-atom [106–108] and coarse-grained force fields [109–111].

Selected CG force fields used in MD simulations are summarized in Sections 2.3.1–2.3.7. Section 2.3.7 is devoted to glycosoaminoglycans, which, despite their significance in cell functioning, seem to be underrepresented in the literature. Apart from these force fields, MD simulations are carried out with the structure-based or Gō-like models [112,113], which are designed so that the interactions between the residues that are in contact in the experimental structure possess energy minima, while the other interactions are all repulsive, and the elastic network models, in which harmonic, anharmonic or double-well potentials are employed to keep the structure in the neighborhood of the native structure [114–116]. The Gō-like models have been used to study protein folding [117–119], including the folding and unfolding of knotted proteins [120], and to study the mechanostability of virus capsids [121]. The elastic network models are used in investigations of the fluctuations around the native structures, particularly in finding the functionally important motions [114–116].

2.3.1. AWSEM

The Associative memory, Water mediated, Structure and Energy Model (AWSEM) [122] has been developed in the Wolynes and Papoian labs to handle protein systems. The geometry of the chain is defined in terms of C^α, C^β and backbone carboxyl-oxygen atoms, which are interaction sites; the positions of backbone N and H atoms are calculated assuming the ideal trans-peptide group geometry. The energy function is a sum of the backbone potential, which consists of the connectivity, Ramachandran, chirality and excluded volume terms, the contact potential that consists of the pairwise $C^\beta \cdots C^\beta$ (C^β replaced with C^α for Gly) and the water-mediated potentials, the burial potential, the hydrogen-bonding potential that depends on the oxygen–nitrogen and oxygen–hydrogen distances, peptide group orientation and an explicit helical term, the associative memory term, which encodes the alignment to proteins with known structures, and the desolvation term. AWSEM has been integrated into the LAMMPS [123] general MD package. It has been used successfully in protein structure prediction and simulations of protein folding and assembly [124], including protein aggregation [125].

2.3.2. MARTINI

MARTINI [69,76,77] is the most popular CG force field. Originally developed to simulate lipid systems [66], it has been extended to proteins [67], polysaccharides [68], nucleic acids [70] and materials science [126]. A major advantage of MARTINI is a standardized coarse-graining scheme, in which chain fragments are merged into sites comprising four non-hydrogen atoms on average, while rings are divided into three-atom fragments. Depending on the character, each CG particle is assigned a polar (P), nonpolar (N), apolar (C) or charged (Q) type, with standardized parameters. This scheme enables automatic coarse-graining without user intervention. Water and lipid molecules are explicit, except for the variant termed dry MARTINI [127]. Four water molecules are assembled into an extended solvent site. The energy function is of neoclassical type (Equation (12)). MARTINI was originally developed to work with the GROMACS MD suite [128] but is also used with other standard MD packages.

MARTINI has been used worldwide to simulate a variety of biological systems [69]. It has also been successfully applied in scoring the docked protein structures [129] and in protein–RNA docking [130]. However, the force field cannot predict protein secondary structure and, consequently, secondary structure restraints have to be imposed when simulating systems containing proteins and nucleic acids.

2.3.3. OPEP and HiRe-RNA

Optimized Potential for Efficient protein structure Prediction (OPEP) [82,83,131] is a CG force field, which was initially developed to perform simulations of polypeptides. As in AWSEM, the all-atom representation of the polypeptide backbone is used, while each sidechain is represented by a single spherical bead. The effective energy function consists of local (bond stretching, bond angle, torsional) and long-range terms, which include the sidechain–sidechain contact potentials and a sophisticated backbone-hydrogen-bonding term, which reflect the nearly linear arrangement of the N-H$\cdots$O groups as observed in experimental structures. The H-bond potentials also include four-body terms that promote the formation of regular hydrogen bond patterns. The force field has been parametrized using a combination of bottom-up and top-down approaches. Recently [132], the model has been extended to run constant pH simulations, which is accomplished by Monte Carlo sampling of protonation states. Owing to careful design and parametrization, unrestrained folding simulations can be performed with OPEP. Based on OPEP, the PEP-FOLD3 server [133] was created, which enables the user to run peptide and small protein folding simulations.

Based on the concept of OPEP, the HiRe-RNA model of nucleic acids [134] was developed, in which four interaction sites (P, O5', C5', C4' and C1') are assigned per backbone unit, a pyrimidine base is represented by one bead and a purine base is represented by two beads. A special orientation-dependent potential has been designed to keep paired nucleic acid bases at appropriate geometry. OPEP/HiRe-RNA have been used in investigating protein dynamics, including the effect of hydrodynamics interactions, small protein and RNA folding, and in investigating protein aggregation, including amyloid formation [131]. The OPEP force field, after training, was able to score structures of protein–protein complexes with higher accuracy than that of ZDOCK [135].

2.3.4. oxDNA and oxRNA

oxDNA/oxRNA, developed by Ouldridge, Louis and Doye [136–141], is a low-resolution DNA and RNA model with three beads per nucleotide. These sites are the backbone repulsion site, the base repulsion site and the base hydrogen-bonding/stacking site, respectively, and they form a linear fragment perpendicular to the direction of the chain. The interaction potentials have been engineered to reproduce base pairing but depend on the kinds of interacting bases and not on their position in the sequence. Salt effects have been included [140]. The model has been parametrized to reproduce the geometry and thermodynamics of DNA hybridization and has been used with success to model the formation and rearrangements of DNA and RNA nanostructures [141]. The model has been included in the oxDNA package [140].

2.3.5. SIRAH

The Southamerican Initiative for a Rapid and Accurate Hamiltonian (SIRAH) force field [142,143] is a CG force field to treat proteins, later extended to nucleic acids [144], lipids and polysaccharides. There are three interaction sites per residue for the polypeptide backbone, located at the amide-N (GN), C$^\alpha$ (GC) and carbonyl-O (GO) atoms, while each sidechain is represented by 1 (for Gly) to 7 (for Arg and Trp) beads. Water is explicit in the model. Each "water" particle (WT4) represents 11 tetrahedrally coordinated water molecules and comprises four beads. Hydrated ions Na$^+$, K$^+$ and Cl$^-$ are considered explicitly, each extended ion particle comprising the actual ion and the six water molecules constituting the first hydration shell. As in OPEP, the interaction potential consists of local and long-range terms. Coulombic interactions are calculated explicitly and partial charges are present on the beads of the WT4 particles, which results in the appropriate handling of hydration. The parameters have been determined to reproduce the structural and thermodynamic properties of model systems. SIRAH has been ported to GROMACS [128] and AMBER [145]. It has been used in the simulation of natively unfolded proteins and protein aggregation.

2.3.6. UNICORN

The UNIfied COarse gRaiNed model (UNICORN) [26,47] is the most heavily coarse-grained model of those described in this section. It has been created by merging the models specific for proteins (UNRES) [24,47,80,146], nucleic acids (NARES-2P) [147] and for polysaccharides (SUGRES-1P) [47,148,149]. Recently, its extension to treat protein–nucleic acid complexes has been developed [150]. For each macromolecule type, one bead is placed in the middle of the backbone virtual bond to represent a peptide group, a phosphate group or a sugar ring, respectively, and one on the side group, if applicable. The respective potentials have been derived by using the recently developed scale-consistent theory [25,26], which was evolving together with UNRES [24,25,151]. The energy function consists of local (virtual bond stretching, virtual bond angle bending, virtual bond torsional, sidechain chirality) and long-range terms, including the multibody terms. The torsional terms depend both on virtual bond dihedral angles and on the virtual bond angles, which is a consequence of the application of the scale-consistent theory [25]. Owing to careful physics-based derivation of the energy function, UNICORN can model regular secondary structure patterns without engineering specific potentials. Solvent is implicit in the interaction potentials. The force field has recently been extended to the lipid bilayer environment, which is treated as a continuous medium [152]. The force fields have been calibrated with a combination of the bottom-up and top-down approaches [25,47,80,102,147]. UNICORN is probably the only coarse-grained model in which the effective energy function depends on temperature [79].

UNRES is applicable to the energy-based prediction of protein structures [153–158], to study protein folding [159] and free-energy landscapes [160] and to solve a variety of biological problems [47,161,162], including the formation of oligomers and fibrils of amyloidogenic peptides [163–166]. Some of these studies are briefly described in Section 3. UNRES also enables the researchers to run simulations of proteins that include D-amino-acid [80,167] and phosphorylated residues [168] and to consider the dynamic breaking and formation of disulfide bonds [169]. An extension to treat the binding of proteins to nanoparticles has also been developed [170]. The package is available in standalone form at www.unres.pl (accessed on 10 September 2021) and also through the web server [171], which has recently been extended to add the peptide–protein and protein–protein docking functionality [172].

NARES-2P not only reproduces the double-helix structure and folding thermodynamics of DNA and RNA molecules but also the pre-melting transition in DNA [147]. With limited restraints, NARES-2P is capable of modeling complex DNA and RNA folds [173]. NARES-2P was used to investigate telomere stability [174] and the influence of single-strand breaks on the mechanical stability of various DNA chains [175].

SUGRES-1P is still under development [148,149]. It is being extended [149] to include glycosaminoglycans (GAGs), for which the CG force fields are scarce despite their high biological significance. The other force fields developed for glycosoaminoglycans are described in Section 2.3.7.

2.3.7. Coarse-Grained Potentials for Glycosoaminoglycans

Glycosaminoglycans (GAGs) are a special class of linear anionic polysaccharides composed of disaccharide periodic units that contain a uronic acid and an N-acetylated aminosugar [176] residue, with some of the hydroxyl groups sulfated. The types of particular monosaccharide building blocks, glycosidic linkages and positions of sulfation ("sulfation code" [177]) contribute to the high chemical and structural variety of this family of carbohydrates, which are traditionally classified into several groups: heparin, heparan sulfate, chondrotin sulfate, dermatan sulfate, keratan sulfate. The GAGs are located in the extracellular matrix of the cell, where they participate in a number of biologically relevant processes, such as cell signaling, cell proliferation, cell adhesion, anticoagulation and antiogenesis, by establishing interactions, their targets being collagen [178], growth factors [179], chemokines [180] and cathepsins [181], among others. While all-atom ap-

proaches are sufficient to model short GAGs (6–8 units) and their interactions [182], CG models are required for modeling very long GAGs, which contain hundreds and thousands of monosaccharide blocks. Such long GAGs are present in the extracellular matrix [183] and guide processes such as the creation of the protein gradients [184] or establishing collagen networks [180].

The first CG model of GAGs, along with the respective force field, for use in MC simulations, was developed by Bathe et al. [185]. The model contains five centers per GAG unit: the mass center, the charge center and one oxygen and two carbon atoms, forming a glycosidic linkage. It was parametrized based on all-atom MD simulations of model systems. The first model applicable in MD simulations was developed by the Almond group [186,187], which covers heparan sulfate, hyaluronic acid, chodroitin sulfate and dermatan sulfate. In this model, a GAG unit consists of two centers, which are the sugar-ring center and the glycosidic-oxygen atom, respectively. Sugar-ring puckering is also accounted for. The parameters were derived from microsecond all-atom simulations using the GLYCAM06 force field [188]. The model was applied in studies of heparan sulfate polysaccharides consisting of 50–200 units [186] and of proteoglycans consisting of GAGs as a glyco-part at a CG level [187]. The resulting force field is compatible with the GLYCAM06 and AMBER FF03 [189] force fields, thus enabling multiscale simulations of glycans to be carried out.

A more detailed CG representation developed to model 17 different GAGs, including the natural ones (heparin, desulfated heparan sulfate, chondroitin sulfate, dermatan sulfate and hyaluronic acid) and their artificially sulfated derivatives, was proposed by Samsonov et al. [190]. In addition to the sugar-ring center and glycosidic-oxygen atom, N-acetyl, N-sulfate, 4-, 6-sulfates (for aminosugars) and 2-, 3-sulfates (for uronic acids) as well as carboxyl groups (distinguishable for glucuronic and iduronic acids) were introduced as separate centers. The model has been designed for use with the AMBER package and the parameters of the respective force field were obtained from the results of all-atom MD simulations. The model shows good agreement among end-to-end distance and radii of gyration for GAGs with a length of up to around 32 units.

2.4. Extensions of MD

Although the original purpose of MD was to study the time evolution of molecular systems, the MD simulations carried out in canonical mode can be used to obtain equilibrium ensembles and, thereby, equilibrium structural averages (e.g., the end-to-end distance), thermodynamics properties (e.g., average energy, heat capacity, etc.) or distributions, which can be compared with the experimental observables. However, because a simulation run at a constant temperature can get stuck in a local minimum region, extensions of MD aimed at enhancing the conformational sampling were developed [46,191–195], which can be divided into three groups. In the methods of the first group, which consist, among others, of umbrella sampling (US) [196,197], thermodynamic integration (TI) [198] and steered molecular dynamics (SMD) [199], the system is guided along a given reaction coordinate. These methods are often combined with the generalized ensemble methods to ensure sufficient sampling. The methods of the second group are metadynamics methods [200,201], in which the system is pushed away from the already visited region(s) of the conformational space, by adding a series of Gaussian terms centered at the encountered values of the selected reaction coordinate. These methods are also often used together with generalized ensemble methods. An approach related to metadynamics, termed Gaussian-accelerated MD, which does not require reaction coordinate selection, was proposed recently [195]. The methods of the third groups are aimed at visiting the whole energy space. This group comprises simulated annealing (SA) [202] and generalized ensemble methods [203–206]. Aside from the methods summarized above, the forward-flux sampling method has been developed [207,208] for the efficient determination of kinetics from MD.

The generalized ensemble methods are the most efficient in searching the conformational space. Among them, the Replica Exchange Molecular Dynamics (REMD) [203,204] and its variants are the most straightforward and, at the same time, very efficient. In REMD, several MD simulations (replicas) are carried out at different temperatures $(T_0, T_1, \cdots, T_{M_T})$. The replicas evolve independently and, after a certain number of MD steps, an exchange of temperatures between neighboring replicas (with indices i and $i + 1$, respectively) is attempted, the acceptance of the exchange following the Metropolis criterion with the exchange probability ω expressed by Equation (13), which reflects the fact that the effective CG energy function can depend on temperature [79,209]:

$$\omega(\mathbf{q}_i \to \mathbf{q}_j) = \min[1, \exp(-\Delta)] \tag{13}$$

with

$$\Delta = \left[\beta_j U(\mathbf{q}_j; \beta_j) - \beta_i U(\mathbf{q}_j; \beta_i)\right] - \left[\beta_j U(\mathbf{q}_i; \beta_j) - \beta_i U(\mathbf{q}_i; \beta_i)\right] \tag{14}$$

where $\beta_i = 1/k_B T_i$, T_i being the absolute temperature corresponding to the ith trajectory, and $\mathbf{q}_i$ denotes the variables of the conformation of the ith trajectory at the attempted exchange point. To enhance the sampling further, several trajectories can be run at a given temperature, this approach being called the Multiplexed Replica Exchange Molecular Dynamics (MREMD) [210,211].

Further enhancement of sampling can be achieved by extending the replica scheme to include the components of the energy function; this extension has been coined the Hamiltonian Replica Exchange Molecular Dynamics (HREMD) [212–214]. Thus, a number (e.g., M) of canonical MD simulations are carried out simultaneously at different temperatures and with different potential energy functions (V), which can differ by the repulsive Lennard–Jones terms, if the purpose is to allow the system to overcome sterical clashes or by the restraint function(s), e.g., umbrella or experimental-restraint potentials, if the purpose is to explore different ranges of order parameters [79] or to perform data-assisted structure modeling [214]. The replicas constitute a two-dimensional (T_i, V_j) grid. For each replica, an exchange is attempted with its neighboring replica in one or two dimensions (up, down or diagonal on the grid, the direction being selected at random). The exchange is accepted based on the probability ω expressed by Equation (15):

$$\omega(\mathbf{q}_{ij} \to \mathbf{q}_{kl}) = \min[1, \exp(-\Delta)], \ \ (k, l) \in \{(i+1, j), (i, j+1), (i+1, j+1)\} \tag{15}$$

with

$$\Delta = \left[\beta_k V_l(\mathbf{q}_{kl}; \beta_k) - \beta_i V_j(\mathbf{q}_{kl}; \beta_i)\right] - \left[\beta_k V_l(\mathbf{q}_{ij}; \beta_k) - \beta_i V_j(\mathbf{q}_{ij}; \beta_i)\right] \tag{16}$$

where V_j is the potential energy function (including the restraining potential) corresponding to the (T_i, V_j) trajectory, and $\mathbf{q}_{ij}$ denotes the variables of the respective conformation of this trajectory at the attempted exchange point. A variant of HREMD was developed recently [215] for mixed-resolution MD simulations with the MARTINI [66] and GROMOS [216] force fields.

Multicanonical algorithms [217,218], also known as entropic sampling [218], are another class of generalized ensemble methods, in which the energy is replaced in the Metropolis criterion by the logarithm of the density of states (the microcanonical entropy). A simulation is converged when all energies are sampled with the same frequency; therefore, this method is well suited to overcome energy barriers. Once the density of states is obtained, all ensemble averages can be computed. The multicanonical algorithms have also been combined with REMD to form the Replica Exchange Multicanonical Algorithm (REMUCA) and Multicanonical Replica Exchange Algorithm (MUCAREM) [204]. These algorithms were implemented in the UNRES CG force field [219].

3. Examples

In this section, examples of CG MD studies with UNRES and NARES-2P are discussed, which pertain to investigating the kinetics of protein folding and ensemble-based structure modeling.

3.1. Investigation of Protein-Folding Kinetics and Pathways

3.1.1. Folding Kinetics of FBP WW Domain and Its Mutants

The most straightforward comparison of the time scale of coarse-grained simulations with the experimental time scale is that of the timing of the events, which can be captured both by experiments and simulations. The quantities that can be obtained from both experiments and simulations are the rate constants of protein folding. We studied the folding of the Formin Binding 28 (FBP28) WW domain by using UNRES MD [159]. The native structure of this protein is an antiparallel three-stranded β-sheet (PDB: 1E0L) [220] and the mechanism, thermodynamics and kinetics of its folding and of the folding of its mutants were the subject of extensive experimental studies [221]. We carried out a canonical simulation study of the wild-type protein and its three single-residue mutants, Y11R, Y19L and W30F, and three mutants with deletions of five N-terminal or four C-terminal residues or both, ΔNY11R, ΔNΔCY11R and ΔNΔCY11R/L26A. These mutants exhibit different folding rates and stability compared to the wild-type protein [221].

For each protein, 512 canonical simulation trajectories were run at a temperature below the folding transition temperature. The duration of each simulation was at least 5 μs of coarse-grained MD time. The fractions of folded and intermediate structures, defined based on the Root Mean Square Deviation (RMSD) cut-off from the respective experimental structures, were calculated every 2000 MD steps (each with length of 4.89 fs) as averages over the 512 trajectories and single- and double-exponential kinetic equations were fitted to them. Except for the Y19L and Y30F mutants, all proteins were found to exhibit double-exponential kinetics, indicating the presence of an intermediate. This result was in fair agreement with the experiment, which suggests that Y19L and ΔNY11R exhibit a single-exponential and the other protein exhibits a double and single-exponential kinetics, respectively. However, an analysis of the free-energy landscapes demonstrated that the intermediate, which has the N-terminal hairpin fully folded and the C-terminal hairpin partially formed (Figure 2), is always present but, for Y19L, the transition from the intermediate state to the native structure corresponds to a low free-energy barrier, making the kinetics effectively single-exponential.

We compared the rate constants corresponding to the fast phase of folding (no accurate experimental data were available for the slow-phase rate constants; this phase probably corresponds to the formation of the C-terminal β-hairpin [222]). A plot of the correlation between the logarithms of the rate constants determined from simulation and the experimental rate constants is shown in Figure 3. It can be seen that the correlation is good, if only the wild-type protein and single-point mutants are considered, but it deteriorates when the mutants with N- and C-terminus deletions are added. Nevertheless, Figure 3 demonstrates that the experimental rate constants are roughly three orders of magnitude smaller compared to those resulting from simulations, this demonstrating the time-scale extension of CG simulations.
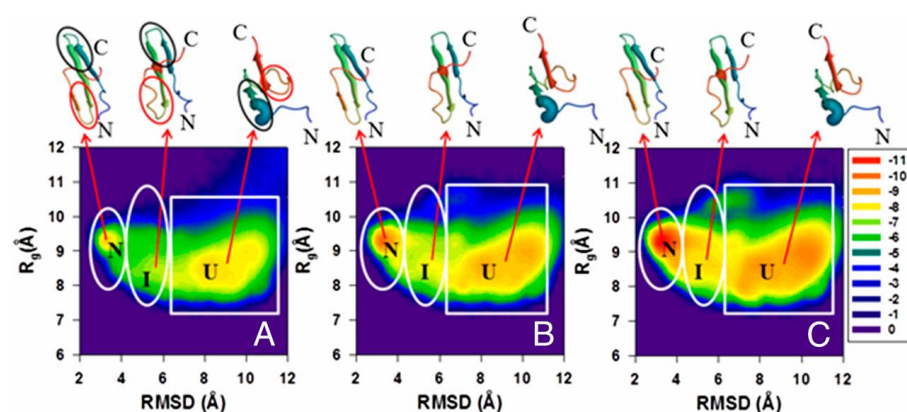
Figure 2. Variation of the distribution of conformational states in terms of FELs (in kcal/mol) along the C^α-RMSD and R_g order parameters for the wild-type FBP28 WW domain. The data have been collected from different sections of all 512 trajectory sets for the molecule (shown in (**A**–**C**), respectively). The FEL corresponding to the initial parts of the trajectories (with the average fraction of the native structures up to 20% of the maximum fraction) is shown in (**A**); the FEL from the middle parts of the trajectories (the fraction of the native structures between 20% and 50% of the maximum fraction) is shown in (**B**); and the FEL from the final parts of the trajectories (the fraction of the native structures exceeds 50% of the maximum fraction) is shown in (**C**), respectively. The letters "U", "I" and "N" correspond to unfolded, intermediate and native states, respectively. The representative structures of unfolded, intermediate and native states are plotted on top of each state. Hairpin 1 and hairpin 2 are circled by black and red lines, correspondingly, in (**A**). Reproduced from R. Zhou et al. *Proc. Natl. Acad. Sci. USA*, 111, 18243–18248 (2014). Copyright 2014 National Academy of Sciences.
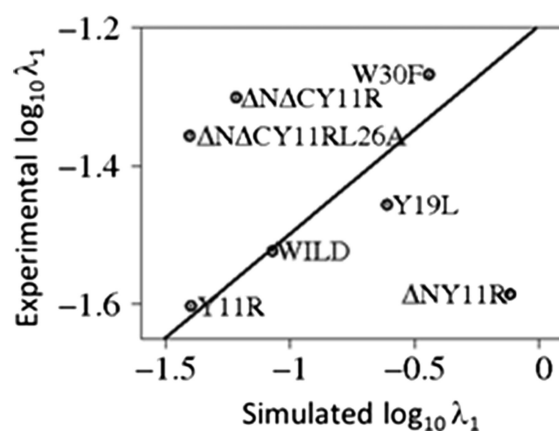

Figure 3. A correlation plot of the experimental and simulated fast cumulative constants with a least-squares fitting line. The experimental rate constants are expressed in μs^{-1}, while those obtained from simulations are expressed in ns^{-1}. Reproduced from R. Zhou et al. *Proc. Natl. Acad. Sci. USA*, 111, 18243–18248 (2014). Copyright 2014 National Academy of Sciences.

3.1.2. Effect of Hydrodynamic Interactions on Folding

Because our study on the FBP WW domain and its mutants demonstrated that protein-folding kinetics can be modeled reasonably by using MD with UNRES, we undertook an investigation of the effect of the non-diagonal friction tensor (see Section 2.1) due to the solvent (which is implicit in UNRES), which gives rise to the so-called hydrodynamic interactions (HIs), on the folding dynamics. The HIs are manifested as the apparent drag of two objects moving through a liquid [40,41,223]. The friction matrix in hydrodynamic interactions is usually expressed by the Rotne–Prager (RP) tensor [43]. The HIs probably play an important role in protein-folding protein–protein association [224], the formation of lipid membranes [225], the sliding motion of protein along DNA [226], etc.

The studies carried out with the Gō-like models [40,41] suggested that HIs accelerate folding, owing to the faster initial collapse that takes place if they are introduced. However, the Gō-like models are biased towards native interactions. Therefore, we studied the effect of hydrodynamic interactions on the folding kinetics with UNRES, taking the N-terminal domain of staphylococcal protein A (PDB: 1BDD) and the FBP 28 WW domain [42] as examples. The first protein is a 46-residue three-α-helix bundle and, as mentioned above, the second one is a 37-residue three-stranded anti-parallel β-sheet, PDB: 1E0L). For both proteins, we ran two series of 200 independent canonical MD trajectories, one with plain-Langevin dynamics and another one with Langevin dynamics including hydrodynamic interactions, at $T = 300$ K and $T = 335$ K for 1BDD and 1E0L, respectively, which are below the respective transition midpoints equal to 320 K and 339 K for 1BDD and 1E0L, respectively. The progress of folding was monitored as the variation of the fraction of native-like conformations and that of the conformation of the respective intermediate with time, computed from all trajectories at a given snapshot [42].

As opposed to the results of earlier studies [40,41], we found that the folding becomes slower upon introducing hydrodynamic interactions [42] of both proteins. The reason for this is the presence of non-native intermediates (Figure 4), which are kinetic traps. For 1BDD, the intermediate is a mirror image of the native three-helix bundle, and for 1E0L, it has a partially helical conformation. The intermediates persist for a much longer time when hydrodynamic interactions are included and, for 1E0L, even a steady state is reached.
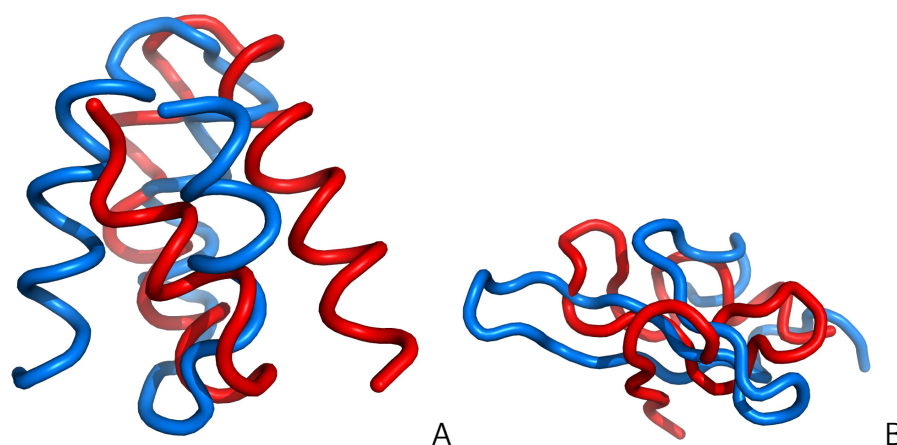

Figure 4. Superpositions of the experimental (blue) and intermediate (red) structures of (**A**) 1BDD and (**B**) 1E0L proteins. Reproduced from A. Lipska et al., *J. Chem. Phys.* 144, 184110 (2016), with the permission of AIP Publishing.

The kinetics was found to be biexponential, with the fast route corresponding to direct folding and the slow one involving an intermediate [42]. The fast route was found to be faster in the presence of hydrodynamic interactions, which led to faster collapse, consistent with the results obtained with the Gō-like models [40,41]. However, the collapse can also lead to an intermediate. The intermediates, especially that in the simulated folding of the FBP WW domain with HIs included, are clearly far from the corresponding native structures, while having a well-defined secondary and tertiary structure, this resulting in a high free-energy barrier of their conversion to the native structures. It should be noted that non-native interactions are not present in the Gō-like Hamiltonians, which is the reason that studies with these models suggested that hydrodynamic interactions accelerate folding.

3.1.3. Phosphorylation-Induced Folding of the Intrinsically Disordered eIF4E-Binding Protein Isoform 2

The Eukaryotic Translation Initiation Factor 4E (eIF4E)-binding protein isoform 2 (4E-BP2) is a natively unfolded protein that can, however, form a marginally stable four-stranded β-sheet upon post-translational phosphorylation of two threonine residues, the respective variant being pT37pT46 4E-BP2 (PDB: 2MX4) [227]. Phosphorylation also results

in an approximately 100-fold decrease in the binding activity. In order to determine the origin of the formation of the structure and the respective folding pathways, we used the recently developed variant of UNRES that includes phosphorylated residues [168].

We ran canonical simulations (80 trajectories) for both the wild type of 4E-BP2$_{18-62}$ and its pT37pT46 variant. The wild type formed a three-stranded anti-parallel β-sheet in part of the structure but the N-terminal strand remained loose (Figure 5). It turned out that the formation of electrostatic contacts involving the phosphorylated residues is crucial for directing the N-terminal strand to close the complete β-structure. Two types of folding trajectories were found, which are shown in Figure 6. The results of this study demonstrated that coarse-grained modeling with UNRES is able to handle well phosphorylated proteins, which is very important in view of the fact that such proteins are involved in signaling.
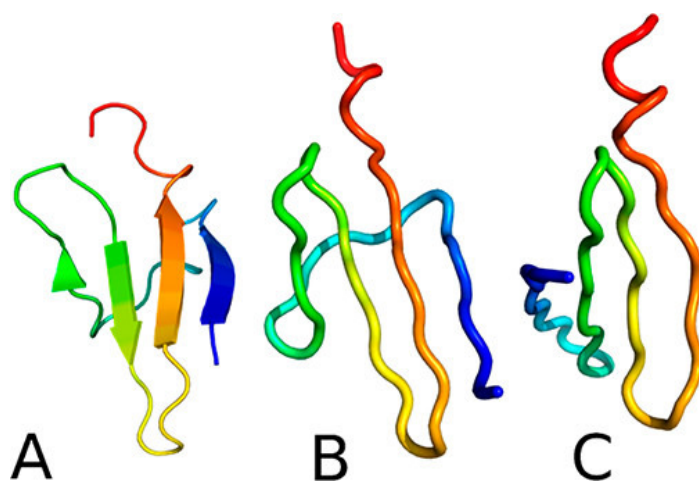

Figure 5. Experimental structure of pT37pT46 4E-BP2$_{18-62}$ (**A**) compared with the most similar structures obtained after simulations and clustering of pT37pT46 4E-BP2$_{18-62}$ (**B**) and WT 4E-BP2$_{18-62}$ (**C**). Reproduced with permission from A.K. Sieradzan et al., *J. Chem. Theory Comput.* 17, 3203–3220 (2021). Copyright 2021 American Chemical Society.
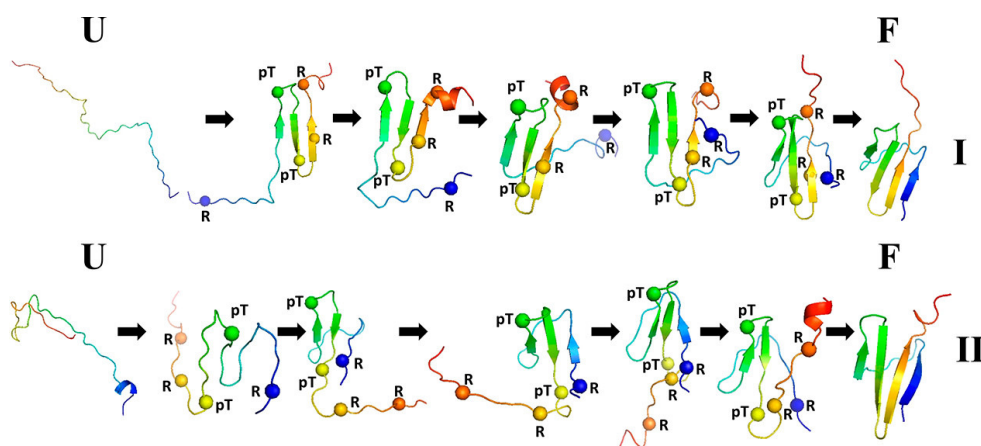

Figure 6. Two folding pathways of pT37pT46 4E-BP2$_{18-62}$ found by UNRES MD simulations. Reproduced with permission from A.K. Sieradzan et al., *J. Chem. Theory Comput.* 17, 3203–3220 (2021). Copyright 2021 American Chemical Society.

3.2. Ensemble-Based Modeling of Protein Structures

The structure of a biomolecule or that of a biomolecular system can be understood as the most probable conformational ensemble at a given temperature and other condition. This definition is consistent with Anfinsen's thermodynamic hypothesis [228] and covers both proteins and other biological macromolecules with a well-defined structure, in which case the ensemble is tight, and natively unfolded proteins. All-atom molecular dynamics

is often used to refine all-atom structures [229] and canonical or extended MD and MC sampling is used with various degrees of success in structural modeling [28,47,82,133,230]. With canonical simulations, it is straightforward to perform cluster analysis of the resulting ensemble and select the mean structures from the clusters as candidate models, ranking them according to decreasing cluster populations; this approach has also been carried over to extended canonical ensemble simulations [231]. We have developed a fully energy-based method of protein structure modeling [79,156], which we used in the CASP exercises [155–158,232,233]. The method consists of four basic stages. In stage 1, an extensive conformational search is carried out with MREMD [211]. In stage 2, our coarse-grained implementation [79] of the Weighted Histogram Analysis Method (WHAM) [234] is used to determine the probabilities of the conformations. In stage 3, cluster analysis is carried out at a temperature below the folding transition temperature, the cluster being ranked in descending order of the sums of the probabilities of the conformations constituting a cluster. The conformation closest to the cluster average is selected as the candidate prediction in the coarse-grained representation. In stage 4, the coarse-grained models obtained in stage 3 are converted to all-atom representation and refined by means of short all-atom MD with the AMBER force field. Modeling can be carried out in the free mode or with auxiliary information from templates produced by servers [235], secondary structure or contact restraints [236], as well as low-resolution experiment restraints [237]. An example of UNRES prediction of an oligomeric protein in the CASP13 exercise is shown in Figure 7.
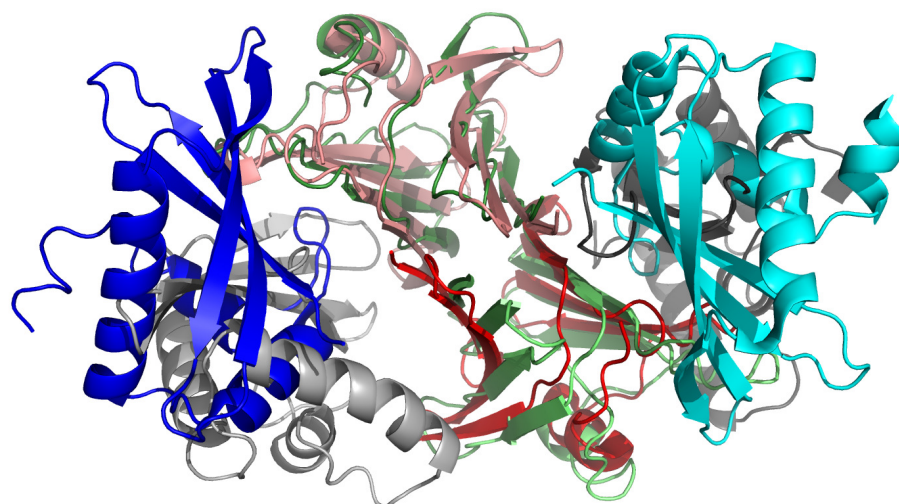

Figure 7. Cartoon representation the UNRES model of CASP13 oligomeric target H0968 obtained by the UNRES-based KIAS-Gdansk group (gray) superposed on the corresponding experimental structures (rainbow-colored). The interface RMSD is 1.33 Å and the model was ranked #1 among the models obtained by all groups. Reproduced with permission from A. Karczyńska et al., *J. Chem. Inf. Model.*, 60, 1844–1864 (2020). Copyright 2020 American Chemical Society.

Another example, in which the structure of the Herpes Virus Entry Mediator (HVEM) peptide complex B- and T-lymphocyte Attenuator (BTLA) was predicted in agreement with the experimental structure [238], is shown in Figure 8.
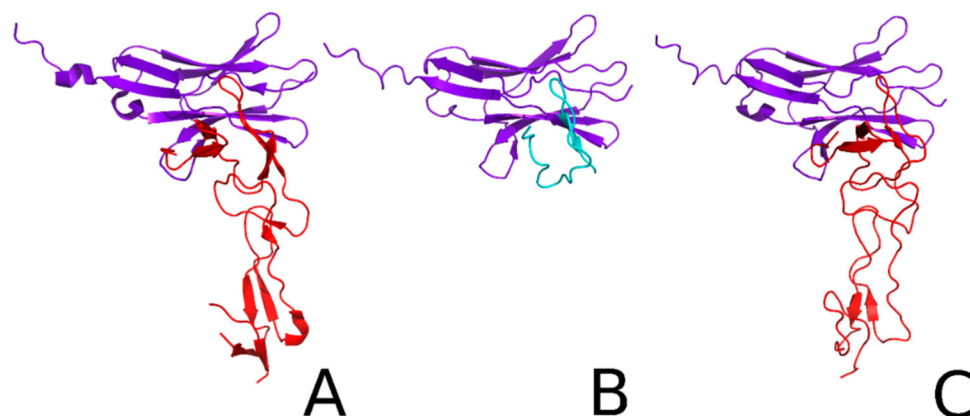
Figure 8. (**A**) Crystal structure of the Herpes Virus Entry Mediator (HVEM) peptide complex B-and T-lymphocyte Attenuator (BTLA) (PDB code: 2AW2), in which the purple color indicates BTLA and red indicates HVEM; (**B**) The dominant cluster of the HVEM(14–39) peptide–BTLA complex obtained from UNRES simulation, in which the purple color indicates BTLA and cyan blue indicates the peptide; (**C**) The fourth cluster of the BTLA–HVEM complex obtained from UNRES simulation, in which HVEM and BTLA are highlighted in red and purple, respectively. Reproduced with permission from M. Spodzieja et al., *Int. J. Mol. Sci.*, 21, 636 (2020) under Creative Common CC BY license.

3.3. Investigation of Telomere Stability

Telomeres are repetitive nucleotide sequences that occur at linear DNA termini. They form unusually stable structures, in view of the fact that their sequences contain many A and T bases, which usually results in reduced stability of the double helix. These structures play a variety of roles in cell biology. Shortening telomeres keeps track of the previous cell divisions and, therefore, controls the ageing process [239]. Moreover, the loop-like or lasso-like structures of telomeres play an important role in chromosome stability [240]. Furthermore, the telomeres also adjust the cellular response to stress and growth stimulation [241].

By using steered molecular dynamics (SMD) [242] with the NARES-2P model of nucleic acids [102,147], as well as all-atom molecular dynamics with the AMBER ffSB14 force field [145,243], we attempted to explain the origin of the unusual stability of telomere structures [174]. We carried out the simulation of human-like (TTAGGG), plant (TTTAGG), insect (TTAGG) and *Candida guilermondi* (GGTGTAC) telomere repeats with various lengths, and other non-telomere repeats, AT, CG and TTTTTCCCC, for comparison [174]. The double helices were pulled either longitudinally, each strand from the other end of the helix, or pulled apart vertically, both strands at the same end of the double helix.

The force strength profiles obtained in NARES-2P simulations are shown in Figure 9A. It can be seen from the figure that all profiles exhibit a peak, with position dependent on sequence length. The profiles obtained from the all-atom simulations were largely independent of sequence, because a fast pulling force rate had to be applied, this resulting in the formation of a flat double ladder. Conversely, the faster dynamics with NARES-2P enabled us to apply a slower pulling force rate, resulting in sequence dependence of stability. As could be expected, the force peak was the highest for the CG repeats, which form the strongest Watson–Crick pairs, and the lowest for the AT repeats, which form weaker pairs (Figure 9A).
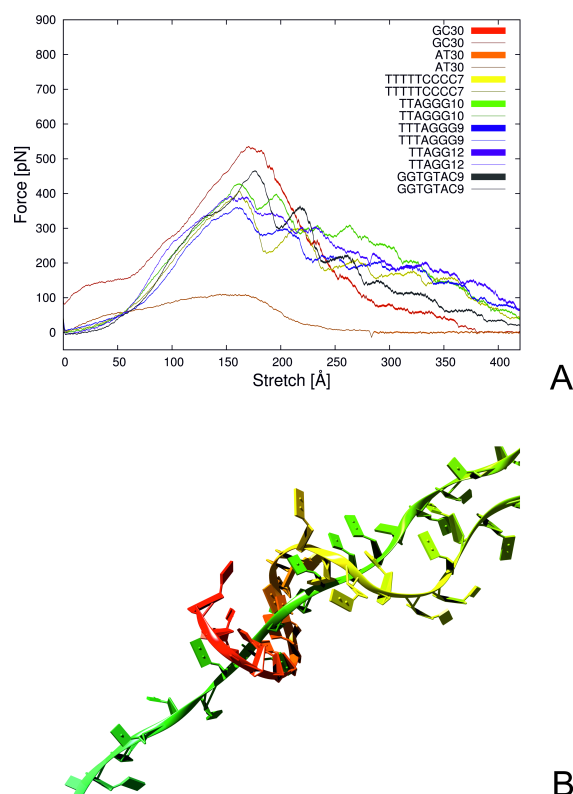
Figure 9. (**A**) Plot of the force needed to stretch the DNA perpendicular to the hydrogen bonds between bases, as a function of sequence; Thin lines indicate averages over 64 trajectories and the colored areas indicate one standard deviation. (**B**) Structure of telomere after regrabbing. The free end (red) wraps around the second DNA chain. Reproduced with permission from A.K. Sieradzan et al., *J. Phys. Chem. B*, 121, 2207–2219 (2017). Copyright 2015 American Chemical Society.

Nevertheless, it can be seen from Figure 9A that the force peaks corresponding to the telomere sequences as well as for the TTTTCCCC repeat are almost as high as that for the CG repeat. Moreover, the pulling force decays slower with stretch extent than that for the CG repeat and the profiles have a fine structure, which results from cycles of dissociation of parts of one strand from another one and subsequent "regrabbing", as shown in Figure 9. The "regrabbing" probably is the main factor preventing chain separation and seems to be characteristic of the telomere (and the TTTTTCCCC) repeats, since it is not observed for the CG repeats. It can, therefore, be the reason for the ability of telomeres to resist damage and, thus, to protect the chromatin structure. As also found from our study, telomere resistance to forces increases significantly with repeat length. Thus, shortening the telomere upon ageing reduces its ability to protect the chromatin, finally leading to apoptosis in the cell.

4. Discussion

The field of coarse-grained description of the structures and dynamics of biomolecular systems is growing rapidly, both regarding the theoretical development and the scope of applications, and has advanced considerably since the seminal work by Levitt and Warshel [11]. A solid theoretical foundation of this approach has been provided by the Mori–Zwanzig theory of the projection operator [15,32,33], with which the effective equations of motion of the coarse-grained degrees of freedom could be derived [16–18,34] (Equation (1)). With this derivation, it is clear that the PMF of the system should be considered as the potential energy in the equations of motion, which has pave the way towards the physics-based systematic development of CG force fields, similar to the Born–Oppenheimer approximation, which enables us to express the Potential Energy Surface (PES) of a system solely in terms of the coordinates of the nuclei [6]; this has been the basis of the development of all-atom force fields. By using the factor expansion of the PMF [25,26], it is,

in turn, possible to split the PMF into components that can be interpreted as specific energy terms, transferable between systems. This decomposition also enables us to derive the analytical formulas for the effective energy terms, including the multibody terms that are crucial for the correct, unassisted modeling of biomolecule structures and are not easy to obtain in other ways. The examples presented in Section 3 demonstrate that the force fields derived in such a way can model reliably the structures of proteins and the kinetics and mechanisms of protein folding, including those of post-translationally modified proteins, and the stability of DNA, including that of telomere sequences, for which the all-atom approach failed to detect the sequence dependence of stability.

There is still a long way to go until the coarse-grained force fields (and the all-atom ones as well) become so reliable as to model the structure and dynamics with experimental accuracy. The machine learning approaches [106–108,110,111] can be of great help in this regard, and their recent tremendous success in the knowledge-based prediction of protein structures [104,105] is a strong reason to believe that they can also lead to tremendous advancements in the development of physics-based force fields. Nevertheless, many reasonably performing coarse-grained force fields are available even at present, including the most popular MARTINI, which can be used for virtually all biologically important systems [66–70], Examples are UNICORN [47] and SIRAH [142,143], which also have the same scope of application, AWSEM [122] and OPEP [82,83,131] for proteins and HiRe-RNA [134] and oxDNA/oxRNA [136–141] for nucleic acids. With carefully chosen restraints, these approaches yield reliable results. Moreover, a number of force fields are available for use with MC simulations, such as, the very well performing CABS knowledge-based force field for proteins [28,31] or the SimRNA knowledge-based force fields for nucleic acids [29]. Furthermore, valuable results are also obtained with simple structure-based Gō-like models [117–121] or the elastic network models [114–116], which, although not transferable, are easy to construct for particular systems.

Another very important issue in CG dynamics is the relationship between the CG time scale and the actual time scale. Equation (1) provides the basis for this through the presence of the friction kernel and random forces. Both terms contain the whole history of the dynamics of a system and are, therefore, very difficult to handle rigorously, although research in this direction is gaining momentum [17,18]. The example presented in Section 3.1.2 demonstrates that introducing a non-diagonal solvent friction tensor (the hydrodynamic interactions) has a different effect on simulated folding depending on whether the Gō-like or physic-based CG force fields are used [42]. Furthermore, considering the non-diagonal form of the inertia tensor, which naturally arises while coarse graining biopolymers chains [21,26], as opposed to the commonly applied representation of the interaction sites as point masses [27,66], appears essential for the correct representation of the dynamics of the coarse-grained degrees of freedom.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AMBER	Assisted Model Building with Energy Refinement
AWSEM	Associative memory, Water mediated, Structure and Energy Model
CASP	Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction
CG	Coarse-Grained
CHARMM	Chemistry at Harvard Molecular Mechanics
FEL	Free Energy Landscape
GAG	Glycosoaminoglycan
MC	Monte Carlo
MD	Molecular Dynamics
MREMD	Multiplexed Replica Exchange Molecular Dynamics
NARES-2P	Nucleic Acid united Residue 2-Points
OPEP	Optimized Potential for Efficient protein structure Prediction
PMF	Potential of Mean Force
REMD	Replica Exchange Molecular Dynamics
SIRAH	Southamerican Initiative for a Rapid and Accurate Hamiltonian
SUGRES-1P	Sugar united Residue 1-Point
UNRES	United Residue
UNICORN	Unified Coarse-Grained Model

References

1. Alder, B.J.; Wainwright, B.E. Molecular dynamics by electronic computers. In *Proceedings of the International Symposium on Statistical Mechanical Theory of Transport Processes*; Prigogine, I., Ed.; Wiley: New York, NY, USA, 1958; pp. 97–131.
2. van Gunsteren, W.F. Molecular dynamics and stochastic dynamics: A primer. In *Computer Simulation of Biomolecular Systems*; van Gunsteren, W.F., Weiner, P.K., Wilkinson, A.J., Eds.; ESCOM: Leiden, The Netherlands, 1993; pp. 3–36.
3. Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: New York, NY, USA, 2000; p. 2.
4. Scheraga, H.A.; Khalili, M.; Liwo, A. Protein-folding dynamics: Overview of molecular simulation techniques. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57–83. [CrossRef]
5. Durrant, J.D.; McCammon, J.A. Molecular dynamics simulations and drug discovery. *BMC Biol.* **2011**, *9*, 71. [CrossRef]
6. Atkins, P.; Friedman, R. *Molecular Quantum Mechanics*; Oxford University Press: Oxford, UK, 2010.
7. Leach, A.R. *Molecular Modeling: Principles and Applications*; Pearson Education: Harlow, UK, 2010.
8. Shaw, D.E.; Deneroff, M.M.; Dror, R.O.; Kuskin, J.S.; Larson, R.H.; Salmon, J.K.; Young, C.; Batson, B.; Bowers, K.J.; Chao, J.C.; et al. Anton, a special-purpose machine for molecular dynamics simulation. *Commun. ACM* **2008**, *51*, 91–97. [CrossRef]
9. Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D.E. Structure and dynamics of an unfolded protein examined by molecular dynamics simulation. *J. Am. Chem. Soc.* **2012**, *134*, 3787–3791. [CrossRef] [PubMed]
10. Larsson, D.S.D.; Liljas, L.; van der Spoel, D. Virus capsid dissolution studied by microsecond molecular dynamics simulations. *PLoS Comput. Biol.* **2012**, *8*, e1002502. [CrossRef]
11. Levitt, M.; Warshell, A. Computer simulation of protein folding. *Nature* **1975**, *253*, 694–698. [CrossRef]
12. Voth, G. *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 1st ed.; CRC Press: Boca Raton, FL, USA; Taylor & Francis Group: Abingdon, UK, 2008.
13. Kolinski, A. *Multiscale Approaches to Protein Folding*; Springer: New York, NY, USA, 2011.
14. Papoian, G.A. *Coarse-Grained Modeling of Biomolecules*; CRC Press: Boca Raton, FL, USA, 2017.
15. Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press: New York, NY, USA, 2001; Chapter 8.
16. Kinjo, T.; Hyodo, S. Equation of motion for coarse-grained simulation based on microscopic description. *Phys. Rev. E* **2007**, *75*, 051109. [CrossRef] [PubMed]

17. Klippenstein, V.; Tripathy, M.; Jung, G.; Schmid, F.; van der Vegt, N.F.A. Introducing memory in coarse-grained molecular simulations. *J. Phys. Chem. B* **2021**, *125*, 4931–4954. [CrossRef]

18. Han, Y.; Jin, J.; Voth, G.A. Constructing many-body dissipative particle dynamics models of fluids from bottom-up coarse-graining. *J. Chem. Phys.* **2021**, *154*, 084122. [CrossRef] [PubMed]

19. Rudnicki, W.R.; Bakalarski, G.; Lesyng, B. A mezoscopic model of nucleic acids. Part 1. Lagrangian and quaternion molecular dynamics. *J. Biomol. Struct. Dyn.* **2000**, *17*, 1097–1108. [CrossRef]

20. Alvarado, C.; Kazerounian, K. On the rotational operators in protein structure simulations. *Prot. Eng.* **2003**, *16*, 717–720. [CrossRef] [PubMed]

21. Khalili, M.; Liwo, A.; Rakowski, F.; Grochowski, P.; Scheraga, H.A. Molecular dynamics with the united-residue (UNRES) model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode. *J. Phys. Chem. B* **2005**, *109*, 13785–13797. [CrossRef] [PubMed]

22. Tozzini, V. Coarse-grained models for proteins. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144–150. [CrossRef] [PubMed]

23. Tozzini, V. Minimalist models for proteins: A comparative analysis. *Quart. Rev. Biophys.* **2010**, *43*, 333–371. [CrossRef] [PubMed]

24. Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H.A. Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. *J. Chem. Phys.* **2001**, *115*, 2323–2347. [CrossRef]

25. Sieradzan, A.K.; Makowski, M.; Augustynowicz, A.; Liwo, A. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. I. Backbone potentials of coarse-grained polypeptide chains. *J. Chem. Phys.* **2017**, *146*, 124106. [CrossRef]

26. Liwo, A.; Czaplewski, C.; Sieradzan, A.K.; Lubecka, E.A.; Lipska, A.G.; Golon, Ł.; Karczyńka, A.; Krupa, P.; Mozolewska, M.A.; Makowski, M.; et al. Scale-consistent approach to the derivation of coarse-grained force fields for simulating structure, dynamics, and thermodynamics of biopolymers. In *Progress in Molecular Biology and Translational Science. Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly*; Strodel, B., Barz, B., Eds.; Academic Press: London, UK, 2020; Volume 170, Chapter 2, pp. 73–122.

27. Ayton, G.S.; Noid, W.G.; Voth, G.A. Multiscale modeling of biomolecular systems: In serial and in parallel. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192–198. [CrossRef]

28. Kolinski, A. Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* **2004**, *51*, 349–371. [CrossRef]

29. Boniecki, M.J.; Lach, G.; Dawson, W.K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K.M.; Bujnicki, J.M. SimRNA: A coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **2016**, *44*, e63. [CrossRef]

30. Kmiecik, S.; Kolinski, A. Folding pathway of the B1 domain of protein G explored by multiscale modeling. *Biophys. J.* **2008**, *94*, 726–736. [CrossRef]

31. Kurcinski, M.; Jamroz, M.; Blaszczyk, M.; Kolinski, A.; Kmiecik, S. CABS-dock web server for the flexible docking of peptides to proteins without prior knowledge of the binding site. *Nucleic Acids Res.* **2015**, *43*, W419–W424. [CrossRef]

32. Zwanzig, R. Memory effects in irreversible thermodynamics. *Phys. Rev.* **1961**, *124*, 983–992. [CrossRef]

33. Mori, H. Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.* **1965**, *33*, 423–455. [CrossRef]

34. Rudzinski, J.F. Recent progress towards chemically-specific coarse-grained simulation models with consistent dynamical properties. *Computation* **2019**, *7*, 42. [CrossRef]

35. Langevin, P. Sur le théorie du mouvement brownien. *C. R. Acad. Sci.* **1908**, *146*, 530–533.

36. Liwo, A.; Khalili, M.; Scheraga, H.A. Molecular dynamics with the united-residue (UNRES) model of polypeptide chains; test of the approach on model proteins. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2362–2367. [CrossRef] [PubMed]

37. Khalili, M.; Liwo, A.; Jagielska, A.; Scheraga, H.A. Molecular dynamics with the united-residue (UNRES) model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model *α*-helical systems. *J. Phys. Chem. B* **2005**, *109*, 13798–13810. [CrossRef]

38. Murarka, R.K.; Liwo, A.; Scheraga, H.A. Separation of time scale and coupling in the motion governed by the coarse-grained and fine degrees of freedom in a polypeptide backbone. *J. Chem. Phys.* **2007**, *127*, 155103. [CrossRef] [PubMed]

39. Saunders, M.G.; Voth, G.A. Coarse-graining methods for computational biology. *Annu. Rev. Biophys.* **2013**, *42*, 73–93. [CrossRef]

40. Frembgen-Kesner, T.; Elcock, A.H. Striking effects of hydrodynamic interactions on the simulated diffusion and folding of proteins. *J. Chem. Theory Comput.* **2009**, *5*, 242–256. [CrossRef] [PubMed]

41. Cieplak, M.; Niewieczerzał, S. Hydrodynamic interactions in protein folding. *J. Chem. Phys.* **2009**, *130*, 124906. [CrossRef] [PubMed]

42. Lipska, A.G.; Seidman, S.R.; Sieradzan, A.K.; Giełdoń, A.; Liwo, A.; Scheraga, H.A. Molecular dynamics of protein A and a WW domain with a united-residue model including hydrodynamic interaction. *J. Chem. Phys.* **2016**, *144*, 184110. [CrossRef]

43. Rotne, J.; Prager, S. Variational treatment of hydrodynamic interaction in polymers. *J. Chem. Phys.* **1969**, *50*, 4831–4837. [CrossRef]

44. Levy, R.M.; Karplus, M.; McCammon, J.A. Diffusive Langevin dynamics of model alkanes. *Chem. Phys. Lett.* **1979**, *65*, 4–11. [CrossRef]

45. Davidchack, R.L.; Ouldridge, T.E.; Tretyakov, M.V. New Langevin and gradient thermostats for rigid body dynamics. *J. Chem. Phys.* **2015**, *142*, 144114. [CrossRef]

46. Liwo, A.; Czaplewski, C.; Ołdziej, S.; Rojas, A.V.; Kaźmierkiewicz, R.; Makowski, M.; Murarka, R.K.; Scheraga, H.A. Simulation of protein structure and dynamics with the coarse-grained UNRES force field. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G., Ed.; CRC Press: Boca Raton, MA, USA, 2008; Chapter 8, pp. 1391–1411.

47. Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M.A.; et al. A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* **2014**, *20*, 2306. [CrossRef] [PubMed]

48. Sieradzan, A.K.; Czaplewski, C.; Lubecka, E.A.; Lipska, A.G.; Karczyńska, A.S.; Giełdoń, A.P.; Ślusarz, R.; Makowski, M.; Krupa, P.; Kogut, M.; et al. Extension of UNRES package for physics-based coarse-grained simulations of proteins and protein complexes to very large systems. *Biophys. J.* **2021**, *120*, 83a–84a. [CrossRef]

49. Kleinerman, D.S.; Czaplewski, C.; Liwo, A.; Scheraga, H.A. Implementations of Nosé – Hoover and Nosé – Poincaré thermostats in mesoscopic dynamic simulations with the united-residue model of a polypeptide chain. *J. Chem. Phys.* **2008**, *128*, 245103. [CrossRef] [PubMed]

50. Paterlini, M.G.; Ferguson, D.M. Constant temperature simulations using the Langevin equation with velocity Verlet integration. *Chem. Phys.* **1998**, *236*, 243–252. [CrossRef]

51. Ricci, A.; Ciccotti, G. Algorithms for Brownian dynamics. *Mol. Phys.* **2003**, *101*, 1927–1931. [CrossRef]

52. Ciccotti, G.; Kalibaeva, G. Deterministic and stochastic algorithms for mechanical systems under constraints. *Philos. Trans. R. Soc. Lond. A* **2004**, *362*, 1583–1594. [CrossRef]

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

54. Goga, N.; Rzepiela, J.; de Vries, A.H.; Marrink, S.J.; Berendsen, H.J.C. Efficient algorithms for Langevin and DPD dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 3637–3649. [CrossRef] [PubMed]

55. Nosé, S. A molecular dynamics method for simulation in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268. [CrossRef]

56. Nosé, S. An improved symplectic integrator for Nosé-Poincaré themostat. *J. Phys. Soc. Jpn.* **2001**, *70*, 75–77. [CrossRef]

57. Smith, A.V.; Hall, C.K. α-Helix formation: Discontinuous molecular dynamics on an intermediate-resolution protein model. *Proteins* **2001**, *44*, 344–360. [CrossRef] [PubMed]

58. Cheon, M.; Chang, I.; Hall, C.K. Extending the PRIME model for protein aggregation to all 20 amino acids. *Proteins* **2010**, *78*, 2950–2960. [CrossRef]

59. Sippl, M.J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.* **1990**, *213*, 859–883. [CrossRef]

60. Kolinski, A.; Godzik, A.; Skolnick, J. A general method for the prediction of the three-dimensional structure and folding pathway of globular proteins: Application to designed helical proteins. *J. Chem. Phys.* **1993**, *98*, 7420–7433. [CrossRef]

61. Kmiecik, S.; Gront, D.; Koliński, M.; Wieteska, L.; Dawid, A.; Koliński, A. Coarse-grained protein models and their applications. *Chem. Rev.* **2016**, *116*, 7898–7936. [CrossRef]

62. Singh, N.; Li, W. Recent advances in coarse-grained models for biomolecules and their applications. *Int. J. Mol. Sci.* **2019**, *20*, 3774. [CrossRef]

63. Sun, T.; Minhas, V.; Korolev, N.; Mirzoev, A.; Lyubartsev, A.P.; Nordensiöld, L. Bottom-up coarse-grained modeling of DNA. *Front. Biomol. Sci.* **2021**, *8*, 645527.

64. Giulini, M.; Rigoli, M.; Mattioti, G.; Menichetti, R.; Tarenzi, T.; Fiorentini, R; Potestio, R.F. From system modeling to system analysis: The impact of resolution level and resolution distribution in the computer-aided investigation of biomolecules. *Front. Mol. Biosci.* **2021**, *8*, 676976. [CrossRef]

65. Gay, J.G.; Berne, B.J. Modification of the overlap potential to mimic a linear site-site potential. *J. Chem. Phys.* **1981**, *74*, 3316–3319. [CrossRef]

66. Marrink, S.J.; Risselada, H.J.; Yefimov, S.; Tieleman, D.P.; de Vries, A.H. The MARTINI force field: Coarse Rgained model for biomolecular simulations. *J. Phys. Chem. B* **2007**, *111*, 7812–7824. [CrossRef]

67. Monticelli, L.; Kandasamy, S.K.; Periole, X.; Larson, R.G.; Tieleman, D.P.; Marrink, S.J. The MARTINI coarse-grained force field. *J. Chem. Theory Comput.* **2008**, *4*, 819–834. [CrossRef]

68. Lopez, C.A.; Rzepiela, A.; de Vries, A.H.; Dijkhuizen, L.; Hunenberger, P.H.; Marrink, S.J. Martini coarse-grained force field: Extension to carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210. [CrossRef] [PubMed]

69. Marrink, S.J.; Tieleman, D.P. Perspective on the Martini model. *Chem. Soc. Rev.* **2013**, *42*, 6801–6822. [CrossRef]

70. Uusitalo, J.J.; Ingólfsson, H.I.; Akhshi, P.; Tieleman, D.P.; Marrink, S.J. Martini coarse-grained force field: Extension to DNA. *J. Chem. Theory Comput.* **2015**, *11*, 3932–3945. [CrossRef] [PubMed]

71. Stark, A.C.; Andrews, C.T.; Elcock, A.H. Toward optimized potential functions for protein-protein interactions in aqueous solutions: Osmotic second virial coefficient calculations using the MARTINI coarse-grained force field. *J. Chem. Theory Comput.* **2013**, *9*, 4176–4185. [CrossRef] [PubMed]

72. Schmalhorst, P.S.; Deluweit, F.; Scherrers, R.; Heisenberg, C.P.; Sikora, M.J. Overcoming the limitations of the MARTINI force field in simulations of polysaccharides. *J. Chem. Theory Comput.* **2017**, *13*, 5039–5053. [CrossRef]

73. Fornasier, F.; de Souza, L.; Souza, F.; Reynaud, F.; Pimentel, A. The lipophilicity of coarse-grained cholesterol models. *J. Chem. Inf. Model.* **2020**, *60*, 569–577. [CrossRef]

74. Souza, L.M.P.; Souza, F.R.; Reynaud, F.; Pimentel, A.S. Tuning the hydrophobicity of a coarse grained model of 1,2-dipalmitoyl-Sn-glycero-3-phosphatidylcholine using the experimental octanol-water partition coefficient. *J. Mol. Liq.* **2020**, *319*, 114132. [CrossRef]

75. Souza, F.R.; Pereira Souza, L.M.; Pimentel, A.S. Recent open issues in coarse grained force fields. *J. Chem. Inf. Model.* **2020**, *60*, 5881–5884. [CrossRef]

76. Alessandrini, R.; Souza, P.C.T.; Thallmair, S.; Melo, M.N.; de Vries, A.H.; Marrink, S.J. Pitfalls of the Martini model. *J. Chem. Theory Comput.* **2019**, *15*, 5448–5460. [CrossRef] [PubMed]

77. Souza, P.C.T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B.M.H.; Wassenaar, T.A.; et al. Martini 3: A general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18*, 382–388. [CrossRef]

78. Kubo, R. Generalized cumulant expansion method. *J. Phys. Soc. Jpn.* **1962**, *17*, 1100–1120. [CrossRef]

79. Liwo, A.; Khalili, M.; Czaplewski, C.; Kalinowski, S.; Ołdziej, S.; Wachucik, K.; Scheraga, H. Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins. *J. Phys. Chem. B* **2007**, *111*, 260–285. [CrossRef] [PubMed]

80. Liwo, A.; Sieradzan, A.K.; Lipska, A.G.; Czaplewski, C.; Joung, I.; Żmudzińska, W.; Hałabis, A.; Ołdziej, S. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force field and force-field calibration and validation. *J. Chem. Phys.* **2019**, *150*, 155104. [PubMed]

81. Kolinski, A.; Skolnick, J. Discretized model of proteins. I. Monte Carlo study of cooperativity in homopolypeptides. *J. Chem. Phys.* **1992**, *97*, 9412–9426. [CrossRef]

82. Maupetit, J.; Tuffery, P.; Derreumaux, P. A coarse-grained protein force field for folding and structure prediction. *Proteins* **2007**, *69*, 394–408. [CrossRef] [PubMed]

83. Shen, Y.; Maupetit, J.; Derreumaux, P.; Tufféry, P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theory Comput.* **2014**, *10*, 4745–4758. [CrossRef] [PubMed]

84. Levitt, M.; Chothia, C. Structural patterns in globular proteins. *Nature* **1976**, *261*, 552–558. [CrossRef]

85. Liwo, A.; Ołdziej, S.; Czaplewski, C.; Kozłowska, U.; Scheraga, H.A. Parameterization of backbone-electrostatic and multibody contributions to the UNRES force field for protein-structure prediction from *ab initio* energy surfaces of model systems. *J. Phys. Chem. B* **2004**, *108*, 9421–9438. [CrossRef]

86. Yin, Y.; Sieradzan, A.K.; Liwo, A.; He, Y.; Scheraga, H.A. Physics-based potentials for coarse-grained modeling of protein DNA interactions. *J. Chem. Theory Comput.* **2015**, *11*, 1792–1808. [CrossRef]

87. Makowski, M.; Sobolewski, E.; Czaplewski, C.; Liwo, A.; Ołdziej, S.; No, J.H.; Scheraga, H.A. Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. III. Calculation and parameterization of the potentials of mean force of pairs of identical hydrophobic side chains. *J. Phys. Chem. B* **2007**, *111*, 2925–2931. [CrossRef]

88. Makowski, M.; Liwo, A.; Scheraga, H.A. Simple physics-based analytical formulas for the potentials of mean force of the interaction of amino acid side chains in water. VII. Charged-hydrophobic/polar and polar-hydrophobic/polar side chains. *J. Phys. Chem. B* **2017**, *121*, 379–390. [CrossRef]

89. Makowski, M. Physics-based modeling of side chain-side chain interactions in the UNRES force field. In *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes. From Bioinformatics to Molecular Quantum Mechanics*; Liwo, A., Ed.; Springer: Cham, Switzerland, 2018; pp. 89–115.

90. Soper, A.K. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.* **1996**, *202*, 295–306. [CrossRef]

91. Reith, D.; Püt, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624–1636. [CrossRef] [PubMed]

92. Lyubartsev, A.P.; Laaksonen, A. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E* **1995**, *52*, 3730–3737. [CrossRef] [PubMed]

93. Lyubartsev, A.P.; Naômé, A.; Vercauteren, D.P.; Laaksonen, A. Systematic hierarchical coarse-graining with the inverse Monte Carlo method. *J. Chem. Phys.* **2015**, *143*, 243120. [CrossRef]

94. Mashayak, S.Y.; Jochum, M.N.; Koschke, K.; Aluru, N.R.; Rühle, V.; Junghans, C. Relative entropy and optimization-driven coarse-graining methods in VOTCA. *PLoS ONE* **2016**, *10*, e0131754. [CrossRef]

95. Mereghetti, P.; Maccari, G.; Spampinato, G.L.B.; Tozzini, V. Optimization of analytical potentials for coarse-grained biopolymers. *J. Phys. Chem. B* **2016**, *120*, 8571–8579. [CrossRef]

96. Sanyal, T.; Shell, M.S. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *J. Chem. Phys.* **2016**, *145*, 034109. [CrossRef]

97. Latham, A.P.; Zhang, B. Maximum entropy optimized force field for intrinsically disordered proteins. *J. Chem. Theory Comput.* **2020**, *16*, 773–781. [CrossRef]

98. Kullback, S.; Leibler, R.A. On information and sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86. [CrossRef]

99. Zaborowski, B.; Jagieła, D.; Czaplewski, C.; Hałabis, A.; Lewandowska, A.; Żmudzińska, W.; Ołdziej, S.; Karczyńska, A.; Omieczynski, C.; Wirecki, T.; Liwo, A. A maximum-likelihood approach to force-field calibration. *J. Chem. Inf. Model.* **2015**, *55*, 2050–2070. [CrossRef]

100. Hałabis, A.; Żmudzińska, W.; Liwo, A.; Ołdziej, S. Conformational dynamics of the Trp-cage miniprotein at its folding temperature. *J. Phys. Chem. B* **2012**, *116*, 6898–6907. [CrossRef]

101. Krupa, P.; Hałabis, A.; Żmudzińska, W.; Ołdziej, S.; Scheraga, H.A.; Liwo, A. Maximum likelihood calibration of the UNRES force field for simulation of protein structure and dynamics. *J. Chem. Inf. Model.* **2017**, *57*, 2364–2377. [CrossRef]

102. He, Y.; Liwo, A.; Scheraga, H.A. Optimization of a Nucleic Acids united-RESidue 2-Point model (NARES-2P) with a maximum-likelihood approach. *J. Chem. Phys.* **2015**, *143*, 243111. [CrossRef]

103. Izvekov, S.; Voth, G.A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473. [CrossRef] [PubMed]

104. Senior, A.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A.; Bridgland, A.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706–710. [CrossRef] [PubMed]

105. Callaway, E. 'It will change averything': AI makes gigantic leap in solving protein structures. *Nature* **2020**, *588*, 203–204. [CrossRef] [PubMed]

106. Behler, J. Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901. [CrossRef] [PubMed]

107. Huan, T.D.; Batra, R.; Chapman, J.; Krishnan, S.; Chen, L.; Ramprasad, R. A universal strategy for the creation of machine learning-based atomistic force fields. *NPJ Comput. Mater.* **2017**, *3*, 37. [CrossRef]

108. Friederich, P.; Häse, F.; Proppe, J.; Aspuru-Guzik, A. Machine-learned potentials for next-generation matter simulations. *Nat. Mater.* **2021**, *20*, 750–761. [CrossRef]

109. Wang, J.; Chmiela, S.; Müller, K.R.; Noé, F.; Clementi, C. Ensemble learning of coarse-grained molecular dynamics force fields with a kernel approach. *J. Chem. Phys.* **2020**, *152*, 194106. [CrossRef]

110. Doerr, S.; Majewski, M.; Pérez, A.; Krämer, A.; Clementi, C.; Noe, F.; Giorgino, T.; De Fabritiis, G. TorchMD: A deep learning framework for molecular simulations. *J. Chem. Theory Comput.* **2021**, *17*, 2355–2363. [CrossRef]

111. Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N.E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent. Sci.* **2019**, *5*, 755–767. [CrossRef]

112. Taketomi, H.; Ueda, Y.; Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation. 1. Effect of specific amino-acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **1975**, *7*, 445–459. [CrossRef]

113. Hills, R.D.; Brooks, C.L. Insights from coarse-grained Gō models for protein folding and dynamics. *Int. J. Mol. Sci.* **2009**, *10*, 889–905. [CrossRef] [PubMed]

114. Sinitskiy, A.V.; Voth, G.A. Coarse-graining of proteins based on elastic network models. *Chem. Phys.* **2008**, *422*, 165–174. [CrossRef]

115. Trylska, J. Coarse-grained models to study dynamics of nanoscale biomolecules and their applications to the ribosome. *J. Phys. Cond. Matter* **2010**, *22*, 453101. [CrossRef]

116. Kmiecik, S.; Kouza, M.; Badaczewska-Dawid, A.; Kloczkowski, A.; Kolinski, A. Modeling of protein structural flexibility and large-scale dynamics: Coarse-grained simulations and elastic network models. *Int. J. Mol. Sci.* **2018**, *19*, 3496. [CrossRef]

117. Yang, S.C.; Onuchic, J.N.; Levine, H. Effective stochastic dynamics on a protein folding energy landscape. *J. Chem. Phys.* **2004**, *125*, 054910. [CrossRef] [PubMed]

118. Schug, A.; Hyeon, C.; Onuchic, J. Coarse-grained structure-based simulations of proteins and RNA. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G., Ed.; CRC Press: Boca Raton, FL, USA, 2008; Chapter 9, pp. 123–140.

119. Hoang, T.X.; Cieplak, M. Molecular dynamics of folding of secondary structures in Go-like models of proteins. *J. Chem. Phys.* **2000**, *112*, 6851–6862. [CrossRef]

120. Sułkowska, J.I.; Sułkowski, P.; Szymczak, P.; Cieplak, M. Untying knots in proteins. *J. Am. Chem. Soc.* **2010**, *132*, 13954–13956. [CrossRef]

121. Cieplak, M. Mechanostability of virus capsids and their proteins in structure-based models. In *Computational Methods to Study the Structure and Dynamics of Biomolecules and Biomolecular Processes. From Bioinformatics to Molecular Quantum Mechanics*; Liwo, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2018; pp. 295–315.

122. Davtyan, A.; Schafer, N.P.; Zheng, W.; Clementi, C.; Wolynes, P.G.; Papoian, G.A. AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* **2012**, *116*, 8494–8503. [CrossRef]

123. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19. [CrossRef]

124. Chen, M.; Lin, X.; Lu, W.; Onuchic, J.N.; Wolynes, P.G. Protein folding and structure prediction from the ground up II: AAWSEM for α/β proteins. *J. Phys. Chem. B* **2017**, *121*, 3473–3482. [CrossRef] [PubMed]

125. Chen, X.; Chen, M.; Schafer, N.P.; Wolynes, P.G. Exploring the interplay between fibrillization and amorphous aggregation channels on the energy landscapes of tau repeat isoforms. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 4125–4130. [CrossRef]

126. Alessandri, R.; Grünewald, F.; Marrink, S.J. The Martini model in materials science. *Adv. Mater.* **2021**, *33*, 2008635. [CrossRef]

127. Arnarez, C.; Uusitalo, J.J.; Masman, M.F.; Ingólfsson, H.I.; de Jong, D.H.; Melo, M.N.; Periole, X.; de Vries, A.H.; Marrink, S.J. Dry Martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *J. Chem. Theory Comput.* **2015**, *11*, 260–275. [CrossRef]

128. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J.C. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718. [CrossRef]

129. Hou, Q.; Lensink, M.F.; Heringa, J.; Feenstra, K.A. CLUB-MARTINI: Selecting favourable interactions amongst available candidates, a coarse-grained simulation approach to scoring docking decoys. *PLoS ONE* **2016**, *11*, e0155251. [CrossRef]

130. Honorato, R.V.; Roel-Touris, J.; Bonvin, A.M.J.J. MARTINI-based protein-DNA coarse-grained HADDOCKing. *Front. Mol. Biosci.* **2016**, *6*, 102. [CrossRef] [PubMed]

131. Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J.F.; Kalimeri, M.; Barducci, A.; et al. The OPEP protein model: From single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **2014**, *43*, 4871–4893. [CrossRef]

132. Barroso da Silva, F.L.; Sterpone, F.; Derreumaux, P. OPEP6: A new constant-pH molecular dynamics simulation scheme with OPEP coarse-grained force field. *J. Chem. Theory Comput.* **2019**, *15*, 3875–3888. [CrossRef] [PubMed]

133. Lamiable, A.; Thevenet, P.; Rey, J.; Vavrusa, M.; Derreumaux, P.; Tuffery, P. PEP-FOLD3: Faster denovo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* **2016**, *44*, W449–W454. [CrossRef]

134. Cragnolini, T.; Laurin, Y.; Derreumaux, P.; Pasquali, S. Coarse-grained HiRE-RNA model for ab initio RNA folding beyond simple molecules, including noncanonical and multiple base pairings. *J. Chem. Theory Comput.* **2015**, *11*, 3510–3522. [CrossRef] [PubMed]

135. Kynast, P.; Derreumaux, P.; Strodel, B. Evaluation of the coarse-grained OPEP force field for protein-protein docking. *BMC Biophys.* **2016**, *9*, 4. [CrossRef] [PubMed]

136. Ouldridge, T.E.; Louis, A.A.; Doye, J.P.K. DNA nanotweezers studied with a coarse-grained model of DNA. *Phys. Rev. Lett.* **2010**, *104*, 178101. [CrossRef] [PubMed]

137. Ouldridge, T.E.; Louis, A.A.; Doye, J.P.K. Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. *J. Chem. Phys.* **2011**, *134*, 085101. [CrossRef]

138. Šulc, P.; Romano, F.; Ouldridge, T.E.; Rovigatti, L.; Doye, J.P.K.; Louis, A.A. Sequence-dependent thermodynamics of a coarse-grained DNA model. *J. Chem. Phys.* **2012**, *137*, 135101. [CrossRef] [PubMed]

139. Šulc, P.; Romano, F.; Ouldridge, T.E.; Doye, J.P.K.; Louis, A.A. A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.* **2014**, *140*, 235102. [CrossRef]

140. Snodin, B.E.K.; Randisi, F.; Mosayebi, M.; Šulc, P.; Schreck, J.S.; Romano, F.; Ouldridge, T.E.; Tsukanov, R.; Nir, E.; Louis, A.A.; et al. Introducing improved structural properties and salt dependence into a coarse-grained model of DNA. *J. Chem. Phys.* **2015**, *142*, 234901. [CrossRef]

141. Snodin, B.E.K.; Romano, F.; Rovigatti, L.; Ouldridge, T.E.; Louis, A.A.; Doye, J.P.K. Direct Simulation of the Self-Assembly of a Small DNA Origami. *ACS Nano* **2016**, *10*, 1724–1737. [CrossRef]

142. Darré, L.; Machado, M.R.; Brandner, A.F.; González, H.C.; Ferreira, S.; Pantano, S. SIRAH: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* **2015**, *11*, 723–739. [CrossRef]

143. Machado, M.R.; Barrera, E.E.; Klein, F.; Sóñora, M.; Silva, S.; Pantano, S. The SIRAH 2.0 force field: Altius, Fortius, Citius. *J. Chem. Theory Comput.* **2019**, *15*, 2719–2733. [CrossRef]

144. Brandner, A.; Schüller, A.; Melo, F.; Pantano, S. Exploring DNA dynamics within oligonucleosomes with coarse-grained simulations: SIRAH force field extension for protein-DNA complexes. *Biochem. Biophys. Res. Commun.* **2018**, *498*, 319–326. [CrossRef]

145. Pearlman, D.; Case, D.; Caldwell, J.; Ross, W.; Cheatham, T., III; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1–41. [CrossRef]

146. Liwo, A.; Ołdziej, S.; Pincus, M.R.; Wawak, R.J.; Rackovsky, S.; Scheraga, H.A. A united-residue force field for off-lattice protein-structure simulations. I. Functional forms and parameters of long-range side-chain interaction potentials from protein crystal data. *J. Comput. Chem.* **1997**, *18*, 849–873. [CrossRef]

147. He, Y.; Maciejczyk, M.; Ołdziej, S.; Scheraga, H.A.; Liwo, A. Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. *Phys. Rev. Lett.* **2013**, *110*, 098101. [CrossRef]

148. Lubecka, E.A.; Liwo, A. A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. II. Backbone-local potentials of coarse-grained O1→4-bonded polyglucose chains. *J. Chem. Phys.* **2017**, *147*, 115101. [CrossRef]

149. Samsonov, S.A.; Lubecka, E.A.; Bojarski, K.K.; Ganzynkowicz, R.; Liwo, A. Local and long range potentials for heparin-protein systems for coarse-grained simulations. *Biopolymers* **2019**, *110*, e23269. [CrossRef] [PubMed]

150. Sieradzan, A.K.; Giełdoń, A.; Yin, Y.; He, Y.; Scheraga, H.A.; Liwo, A. A new protein nucleic-acid coarse-grained force field based on the UNRES and NARES-2P force fields. *J. Comput. Chem.* **2018**, *39*, 2360–2370. [CrossRef]

151. Liwo, A.; Kaźmierkiewicz, R.; Czaplewski, C.; Groth, M.; Ołdziej, S.; Wawak, R.J.; Rackovsky, S.; Pincus, M.R.; Scheraga, H.A. United-residue force field for off-lattice protein-structure simulations; III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. *J. Comput. Chem.* **1998**, *19*, 259–276. [CrossRef]

152. Zięba, K.; Ślusarz, M.; Ślusarz, R.; Liwo, A.; Czaplewski, C.; Sieradzan, A.K. Extension of the UNRES coarse-grained force field to membrane proteins in the lipid bilayer. *J. Phys. Chem. B* **2019**, *22*, 4758. [CrossRef] [PubMed]

153. Lee, J.; Scheraga, H.A. Conformational space annealing by parallel computations: Extensive conformational search of Met-enkephalin and of the 20-residue membrane-bound portion of melittin. *Int. J. Quant. Chem.* **1999**, *75*, 255–265. [CrossRef]

154. Ołdziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nanias, M.; Vila, J.A.; Khalili, M.; Arnautova, Y.A.; Jagielska, A.; Makowski, M.; et al. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field—Test with CASP5 and CASP6 targets. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 7547–7552. [CrossRef]

155. He, Y.; Mozolewska, M.A.; Krupa, P.; Sieradzan, A.K.; Wirecki, T.K.; Liwo, A.; Kachlishvili, K.; Rackovsky, S.; Jagieła, D.; Ślusarz, R.; et al. Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 14936–14941. [CrossRef]

156. Krupa, P.; Mozolewska, M.; Wiśniewska, M.; Yin, Y.; He, Y.; Sieradzan, A.; Ganzynkowicz, R.; Lipska, A.; Karczyńska, A.; Ślusarz, M.; et al. Performance of protein-structure predictions with the physics-based UNRES force field in CASP11. *Bioinformatics* **2016**, *32*, 3270–3278. [CrossRef]

157. Lubecka, E.A.; Karczyńska, A.S.; Lipska, A.G.; Sieradzan, A.K.; Zięba, K.; Sikorska, C.; Uciechowska, U.; Samsonov, S.A.; Krupa, P.; Mozolewska, M.A.; et al. Evaluation of the scale-consistent UNRES force field in template-free prediction of protein structures in the CASP13 experiment. *J. Mol. Graph. Model.* **2019**, *92*, 154–166. [CrossRef]

158. Karczyńska, A.; Zięba, K.; Uciechowska, U.; Mozolewska, M.A.; Krupa, P.; Lubecka, E.A.; Lipska, A.G.; Sikorska, C.; Samsonov, S.A.; Sieradzan, A.K.; et al. Improved consensus-fragment selection in template-assisted prediction of protein structures with the UNRES force field in CASP13. *J. Chem. Inf. Model.* **2020**, *60*, 1844–1864. [CrossRef]

159. Zhou, R.; Maisuradze, G.G.; Suñol, D.; Todorovski, T.; Macias, M.J.; Xiao, Y.; Scheraga, H.A.; Czaplewski, C.; Liwo, A. Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 18243–18248. [CrossRef]

160. Maisuradze, G.G.; Senet, P.; Czaplewski, C.; Liwo, A.; Scheraga, H.A. Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *J. Phys. Chem. A* **2010**, *114*, 4471–4485. [CrossRef] [PubMed]

161. Golas, E.I.; Maisuradze, G.G.; Senet, P.; Ołdziej, S.; Czaplewski, C.; Scheraga, H.A.; Liwo, A. Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. *J. Chem. Theory Comput.* **2012**, *8*, 1334–1343. [CrossRef]

162. Mozolewska, M.; Krupa, P.; Scheraga, H.; Liwo, A. Molecular modeling of the binding modes of the iron-sulfur protein to the Jac1 co-chaperone from *Saccharomyces cerevisiae* by all-atom and coarse-grained approaches. *Proteins Struct. Funct. Bioinf.* **2015**, *83*, 1414–1426. [CrossRef]

163. Rojas, A.; Liwo, A.; Browne, D.; Scheraga, H.A. Mechanism of fiber assembly; treatment of Aβ-peptide peptide aggregation with a coarse-grained united-residue force field. *J. Mol. Biol.* **2010**, *404*, 537–552. [CrossRef]

164. Rojas, A.; Liwo, A.; Scheraga, H.A. A study of the α-helical intermediate preceding the aggregation of the amino-terminal fragment of the Aβ-amyloid peptide (1–28). *J. Phys. Chem. B* **2011**, *115*, 12978–12983. [CrossRef]

165. Rojas, A.; Maisuradze, G.; Scheraga, H. Dependence of the formation of Tau and A beta peptide mixed aggregates on the secondary structure of the N-terminal region of A beta. *J. Phys. Chem. B* **2018**, *122*, 7049–7056. [CrossRef] [PubMed]

166. Nguyen, H.L.; Krupa, P.; Hai, N.M.; Linh, H.Q.; Li, M.S. Structure and physicochemical properties of the Aβ42 tetramer: Multiscale molecular dynamics simulations. *J. Phys. Chem. B* **2019**, *123*, 7253–7269. [CrossRef]

167. Sieradzan, A.K.; Niadzvedtski, A.; Scheraga, H.A.; Liwo, A. Revised backbone-virtual-bond-angle potentials to Rteat the L- and D-amino acid residues in the coarse-grained united residue (UNRES) force field. *J. Chem. Theory Comput.* **2014**, *10*, 2194–2203. [CrossRef]

168. Sieradzan, A.K.; Bogunia, M.; Mech, P.; Ganzynkowicz, R.; Giełdoń, A.; Liwo, A.; Makowski, M. Introduction of phosphorylated residues into the UNRES coarse-grained model: Toward modeling of signaling processes. *J. Phys. Chem. B* **2019**, *119*, 8526–8534. [CrossRef]

169. Chinchio, M.; Czaplewski, C.; Liwo, A.; Ołdziej, S.; Scheraga, H.A. Dynamic formation and breaking of disulfide bonds in molecular dynamics simulations with the UNRES force field. *J. Chem. Theory Comput.* **2007**, *3*, 1236–1248. [CrossRef]

170. Sieradzan, A.K.; Mozolewska, M.A. Extension of coarse-grained UNRES force field to treat carbon nanotubes. *J. Mol. Model.* **2018**, *24*, 121. [CrossRef] [PubMed]

171. Czaplewski, C.; Karczyńska, A.; Sieradzan, A.K.; Liwo, A. UNRES server for physics-based coarse-grained simulations and prediction of protein structure, dynamics and thermodynamics. *Nucleic Acids Res.* **2018**, *46*, W304–W309. [CrossRef]

172. Krupa, P.; Karczyńska, A.S.; Mozolewska, M.A.; Liwo, A.; Czaplewski, C. UNRES-Dock protein-protein and peptide-protein docking by coarse-grained replica-exchange MD simulations. *Bioinformatics* **2021**, *37*, 1613–1615. [CrossRef]

173. Sieradzan, A.K.; Golon, Ł.; Liwo, A. Prediction of DNA and RNA structure with the NARES-2P force field and conformational space annealing. *Phys. Chem. Chem. Phys.* **2018**, *20*, 19656–19663. [CrossRef]

174. Sieradzan, A.K.; Krupa, P.; Wales, D.J. What makes telomeres unique? *J. Phys. Chem. B* **2017**, *121*, 2207–2219. [CrossRef]

175. Krupa, P.; Wales, D.J.; Sieradzan, A.K. Computational studies of the mechanical stability for single-strand break DNA. *J. Phys. Chem. B* **2018**, *122*, 8166–8173. [CrossRef]

176. Esko, J.D.; Kimata, K.; Lindahl, U. Proteoglycans and sulfated glycosaminoglycans. In *Essentials of Glycobiology*, 2nd ed.; Varki, A., Cummings, R.D., Esko, J.D., Freeze, H.H., Stanley, P., Bertozzi, C.R., Hart, G.W., Etzler, M.E., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring, NY, USA, 2009.

177. Habuchi, H.; Habuchi, O.; Kimata, K. Sulfation pattern in glycosaminoglycan: Does it have a code? *Glycoconj. J.* **2004**, *21*, 47–52. [CrossRef]

178. Peng, Y.; Yu, Y.; Lin, L.; Liu, X.; Zhang, X.; Wang, P.; Hoffman, P.; Kim, S.Y.; Zhang, F.; Linhardt, R.J. Glycosaminoglycans from bovine eye vitreous humour and interaction with collagen type II. *Glycoconj. J.* **2018**, *35*, 119–128. [CrossRef]

179. Shute, J. Glycosaminoglycan and chemokine/growth factor interactions. *Handb. Exp. Pharmacol.* **2012**, *207*, 307–324.

180. Li, Z.; Yasuda, Y.; Li, W.; Bogyo, M.; Katz, N.; Gordon, R.E.; Fields, G.B.; Brömme, D. Regulation of collagenase activities of human cathepsins by glycosaminoglycans. *J. Biol. Chem.* **2004**, *279*, 5470–5479. [CrossRef] [PubMed]

181. Sankaranarayanan, N.V.; Nagarajan, B.; Desai, U.R. So you think computational approaches to understanding glycosaminoglycan-protein interactions are too dry and too rigid? Think again! *Curr. Opin. Struct. Biol.* **2018**, *50*, 91–100. [CrossRef] [PubMed]
182. Samsonov, S.A.; Pisabarro, M.T. Computational analysis of interactions in structurally available protein-glycosaminoglycan complexes. *Glycobiology* **2016**, *26*, 850–861. [CrossRef]
183. Rabenstein, D.L. Heparin and heparan sulfate: Structure and function. *Nat. Prod. Rep.* **2002**, *19*, 312–331. [CrossRef]
184. Perrimon, N.; Bernfield, M. Specificities of heparan sulphate proteoglycans in developmental processes. *Nature* **2000**, *404*, 725–728. [CrossRef]
185. Bathe, M.; Rutledge, G.C.; Grodzinsky, A.J.; Tidor, B. A coarse-grained molecular model for glycosaminoglycans: Application to chondroitin, chondroitin sulfate, and hyaluronic acid. *Biophys. J.* **2005**, *88*, 3870–3887. [CrossRef]
186. Sattelle, B.M.; Shakeri, J.; Almond, A. Does microsecond sugar ring flexing encode 3D-shape and bioactivity in the heparanome? *Biomacromolecules* **2013**, *14*, 1149–1159. [CrossRef]
187. Sattelle, B.M.; Shakeri, J.; Cliff, M.J.; Almond, A. Proteoglycans and their heterogeneous glycosaminoglycans at the atomic scale. *Biomacromolecules* **2015**, *16*, 951–961. [CrossRef]
188. Kirschner, K.N.; Yongye, A.B.; Tschampel, S.M.; González-Outeiriño, J.; Daniels, C.R.; Foley, B.L.; Woods, R.J. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29*, 622–655. [CrossRef]
189. Liu, Y.; Palmer, J.C.; Panagiotopoulos, A.Z.; Debenedetti, P.G. Liquid-liquid transition in ST2 water. *J. Chem. Phys.* **2012**, *137*, 214505. [CrossRef]
190. Samsonov, S.A.; Bichmann, L.; Pisabarro, M.T. Coarse-grained model of glycosaminoglycans. *J. Chem. Inf. Model.* **2015**, *55*, 114–124. [CrossRef]
191. Christen, M.; van Gunsteren, W.F. On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review. *J. Comput. Chem.* **2008**, *29*, 157–166. [CrossRef] [PubMed]
192. Barducci, A.; Pfaendtner, J.; Bonomi, M. Tackling sampling challenges in biomolecular simulations. *Meth. Mol. Biol.* **2015**, *1215*, 151–171.
193. Sidky, H.; Colón, Y.J.; Helfferich, J.; Sikora, B.J.; Bezik, C.; Chu, W.; Giberti, F.; Guo, A.Z.; Jiang, X.; Lequieu, J.; et al. SSAGES: Software suite for advanced general ensemble simulations. *J. Chem. Phys.* **2018**, *148*, 044104. [CrossRef]
194. Allison, J.R. Computational methods for exploring protein conformations. *Biochem. Soc. Trans.* **2020**, *48*, 1707–1724. [CrossRef]
195. Wang, J.; Arantes, P.R.; Bhattarai, A.; Hsu, R.V.; Pawnikar, S.; Huang, Y.M.M.; Palermo, G.; Miao, Y. Gaussian accelerated molecular dynamics: Principles and applications. *WIREs Comput. Mol. Sci.* **2021**, *11*, e1521. [CrossRef]
196. Torrie, G.M.; Valleau, J.P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199. [CrossRef]
197. Kästner, J.; Thiel, W. Analysis of the statistical error in umbrella sampling simulations by umbrella integration. *J. Chem. Phys.* **2006**, *124*, 234106. [CrossRef] [PubMed]
198. Pearlman, D.A. Determining the contributions of constraints in free energy calculations: Development, characterization, and recommendations. *J. Chem. Phys.* **1993**, *98*, 8946–8957. [CrossRef]
199. Sieradzan, A.K.; Jakubowski, R. Introduction of steered molecular dynamics into UNRES coarse-grained simulations package. *J. Comput. Chem.* **2017**, *38*, 553–562. [CrossRef]
200. Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566. [CrossRef]
201. Lamprakis, C.; Andreadelis, I.; Manchester, J.; Velez-Vega, C.; Duca, J.S.; Cournia, Z. Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments. *J. Chem. Theory Comput.* **2021**, *17*, 3088–3102. [CrossRef]
202. Kirkpatrick, S.; Gelatt, C.D., Jr.; Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, *220*, 671–680. [CrossRef]
203. Hansmann, U.H.E.; Okamoto, Y. Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. *Physica A* **1994**, *212*, 415–437. [CrossRef]
204. Sugita, Y.; Okamoto, Y. Replica-exchange multicanonical algorithm and multicanonical replica-exchange method for simulating systems with rough energy landscape. *Phys. Rev. Lett.* **2000**, *329*, 261–270. [CrossRef]
205. Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **2001**, *60*, 96–123. [CrossRef]
206. Itoh, S.G.; Okumura, H.; Okamoto, Y. Generalized-ensemble algorithms for molecular dynamics simulations. *Mol. Simul.* **2007**, *33*, 47–56. [CrossRef]
207. van Erp, T.S.; Moroni, D.; Bolhuis, P.G. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* **2003**, *118*, 7762–7774. [CrossRef]
208. Smit, F.X.; Luiken, J.A.; Bolhuis, P.G. Primary fibril nucleation of aggregation prone tau fragments PHF6 and PHF6*. *J. Phys. Chem. B* **2017**, *121*, 3250–3261. [CrossRef]
209. Shen, H.; Liwo, A.; Scheraga, H.A. An improved functional form for the temperature scaling factors of the components of the mesoscopic UNRES force field for simulations of protein structure and dynamics. *J. Phys. Chem. B* **2009**, *113*, 8738–8744. [CrossRef]
210. Rhee, Y.M.; Pande, V.S. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys J.* **2003**, *84*, 775–786. [CrossRef]

211. Czaplewski, C.; Kalinowski, S.; Liwo, A.; Scheraga, H.A. Application of multiplexing replica exchange molecular dynamics method to the UNRES force field: Tests with α and $\alpha + \beta$ proteins. *J. Chem. Theory Comput.* **2009**, *5*, 627–640. [CrossRef]
212. Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian replica exchange method for efficient sampling, of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **2002**, *116*, 9058–9067. [CrossRef]
213. Lee, K.H.; Chen, J. Multiscale enhanced sampling of intrinsically disordered protein conformations. *J. Comput. Chem.* **2015**, *37*, 550–557. [CrossRef] [PubMed]
214. Karczyńska, A.S.; Czaplewski, C.; Krupa, P.; Mozolewska, M.A.; Joo, K.; Lee, J.; Liwo, A. Ergodicity and model quality in template-restrained canonical and temperature/Hamiltonian replica exchange coarse-grained molecular dynamics simulations of proteins. *J. Comput. Chem.* **2017**, *38*, 2730–2746. [CrossRef] [PubMed]
215. Liu, Y.; Pezeshkian, W.; Barnoud, J.; de Vries, A.H.; Marrink, S.J. Coupling coarse-grained to fine-grained models via Hamiltonian Replica Exchange. *J. Chem. Theory Comput.* **2020**, *16*, 5313–5322. [CrossRef]
216. Oostenbrink, C.; Villa, A.; Mark, A.E.; van Gunsteren, W.F.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.* **2004**, *25*, 1656–1676. [CrossRef]
217. Berg, B.A.; Neuhaus, T. Multicanonical ensemble: A new approach to simulate 1st order phase-transitions. *Phys. Rev. Lett.* **1992**, *68*, 9–12. [CrossRef]
218. Lee, J. New Monte Carlo algorithm: Entropic sampling. *Phys. Rev. Lett.* **1993**, *71*, 211–214. [CrossRef]
219. Nanias, M.; Czaplewski, C.; Scheraga, H.A. Replica exchange and multicanonical algorithms with the coarse-grained united-residue (UNRES) force field. *J. Chem. Theory Comput.* **2006**, *2*, 513–528. [CrossRef]
220. Macias, M.J.; Gervais, V.; Civera, C.; Oschkinat, H. Domains and design of a WW prototype. *Nat. Struct. Biol.* **2000**, *7*, 375–379. [CrossRef] [PubMed]
221. Nguyen, H.; Jäger, M.; Moretto, A.; Gruebele, M.; Kelly, J.W. Tuning the free-energy landscape of a WW domain by temperature, mutation, and truncation. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3948–3953. [CrossRef] [PubMed]
222. Karanicolas, J.; Brooks, C.L., III. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 3954–3959. [CrossRef] [PubMed]
223. Tanaka, H. Roles of hydrodynamic interactions in structure formation of soft matter: Protein folding as an example. *J. Phys. Condens. Matter* **2005**, *17*, S2795–S2803. [CrossRef]
224. Frembgen-Kesner, T.; Elcock, A.H. Absolute protein-protein association rate constants from flexible, coarse-grained Brownian dynamics simulations: The role of intermolecular hydrodynamic interactions in Barnase-Barstar association. *Biophys. J.* **2010**, *99*, L75–L77. [CrossRef]
225. Ando, T.; Skolnick, J. On the importance of hydrodynamic interactions in lipid membrane formation. *Biophys. J.* **2013**, *104*, 96–105. [CrossRef]
226. Ando, T.; Skolnick, J. Sliding of proteins non-specifically bound to DNA: Brownian dynamics studies with coarse-grained protein and DNA models. *PLoS Comput. Biol.* **2014**, *10*, e1003990. [CrossRef] [PubMed]
227. Bah, A.; Vernon, R.M.; Soddoqui, Z.; Krzeminski, M.; Munhandiram, R.; Zhao, C.; Senenberg, N.; Kay, L.E.; Forman-Kay, J.D. Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. *Nature* **2015**, *519*, 106–109. [CrossRef]
228. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [CrossRef]
229. Mirjalili, V.; Noyes, K.; Feig, M. Physics-based protein structure refinement through multiple molecular dynamics trajectories and structure averaging. *Proteins* **2014**, *82*, 196–207. [CrossRef]
230. Kolinski, A.; Skolnick, J. Reduced models of proteins and their applications. *Polymer* **2004**, *45*, 511–524. [CrossRef]
231. Gront, D.; Hansmann, U.H.E.; Kolinski, A. Exploring protein energy landscapes with hierarchical clustering. *J. Comput. Chem.* **2005**, *105*, 826–830. [CrossRef]
232. Karczyńska, A.S.; Mozolewska, M.A.; Krupa, P.; Giełdoń, A.; Liwo, A.; Czaplewski, C. Prediction of protein structure with the coarse-grained UNRES force field assisted by small X-Ray scattering data and knowledge-based information. *Proteins* **2018**, *86*, 228–239. [CrossRef]
233. Lubecka, E.; Liwo, A. ESCASA: Analytical estimation of atomic coordinates from coarse-grained geometry for NMR-assisted protein structure modeling. I. Backbone and H^{β} protons. *J. Comput. Chem.* **2021**, *42*, 1579–1589. [CrossRef]
234. Kumar, S.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A.; Rosenberg, J.M. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011–1021. [CrossRef]
235. Mozolewska, M.; Krupa, P.; Zaborowski, B.; Liwo, A.; Lee, J.; Joo, K.; Czaplewski, C. Use of restraints from consensus fragments of multiple server models to enhance protein-structure prediction capability of the UNRES force field. *J. Chem. Inf. Model.* **2016**, *56*, 2263–2279. [CrossRef]
236. Lubecka, E.A.; Liwo, A. Introduction of a bounded penalty function in contact-assisted simulations of protein structures to omit false restraints. *J. Comput. Chem.* **2019**, *40*, 2164–2178. [CrossRef]
237. Karczyńska, A.; Mozolewska, M.A.; Krupa, P.; Giełdoń, A.; Bojarski, K.K.; Zaborowski, B.; Liwo, A.; Ślusarz, R.; Ślusarz, M.; Lee, J.; et al. Use of the UNRES force field in template-based prediction of protein structures and the refinement of server models: Test with CASP12 targets. *J. Mol. Graph. Model.* **2018**, *83*, 92–99. [CrossRef]
238. Spodzieja, M.; Kuncewicz, K.; Sieradzan, A.; Karczyńska, A.; Iwaszkiewicz, J.; Cesson, V.; Węgrzyn, K.; Zhukov, I.; Maszota-Zieleniak, M.; Michielin, O.; et al. Disulfide-linked peptides for blocking. *Int. J. Mol. Sci.* **2020**, *21*, 636. [CrossRef]

239. Wright, W.E.; Piatyszek, M.A.; Rainey, W.E.; Byrd, W.; Shay, J.W. Telomerase activity in human germline and embryonic tissues and cells. *Dev. Genet.* **1996**, *18*, 173–179. [CrossRef]
240. Galati, A.; Micheli, E.; Cacchione, S. Chromatin structure in telomere dynamics. *Front. Oncol.* **2013**, *3*, 46. [CrossRef]
241. von Zglinicki, T.; Saretzki, G.; Döcke, W.; Lotze, C. Mild hyperoxia shortens telomeres and inhibits proliferation of fibroblasts: A model for senescence? *Exp. Cell Res.* **1995**, *220*, 186–193. [CrossRef]
242. Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **1996**, *271*, 997–999. [CrossRef] [PubMed]
243. Galindo-Murillo, R.; Robertson, J.C.; Zgarbová, M.; Šponer, J.; Otyepka, M.; Jurečka, P.; Cheatham, T.E. Assessing the current state of Amber force field modifications for DNA. *J. Chem. Theory Comput.* **2016**, *12*, 4114–4127. [CrossRef] [PubMed]