J Mol Model (2014) 20:2306
DOI 10.1007/s00894-014-2306-5

ORIGINAL PAPER

A unified coarse-grained model of biological macromolecules based on mean-field multipole–multipole interactions

Adam Liwo · Maciej Baranowski · Cezary Czaplewski · Ewa Gołaś · Yi He · Dawid Jagieła · Paweł Krupa · Maciej Maciejczyk · Mariusz Makowski · Magdalena A. Mozolewska · Andrei Niadzvedtski · Stanisław Ołdziej · Harold A. Scheraga · Adam K. Sieradzan · Rafał Ślusarz · Tomasz Wirecki · Yanping Yin · Bartłomiej Zaborowski

Received: 28 February 2014 / Accepted: 12 May 2014 / Published online: 15 July 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract A unified coarse-grained model of three major classes of biological molecules—proteins, nucleic acids, and polysaccharides—has been developed. It is based on the observations that the repeated units of biopolymers (peptide groups, nucleic acid bases, sugar rings) are highly polar and their charge distributions can be represented crudely as point multipoles. The model is an extension of the united residue (UNRES) coarse-grained model of proteins developed previously in our laboratory. The respective force fields are defined as the potentials of mean force of biomacromolecules immersed in water, where all degrees of freedom not considered in the model have been averaged out. Reducing the representation to one center per polar interaction site leads to the representation of average site–site interactions as mean-field dipole–dipole interactions. Further expansion of the potentials of mean force of biopolymer chains into Kubo's cluster-cumulant series leads to the appearance of mean-field dipole–dipole interactions, averaged in the context of local interactions within a biopolymer unit. These mean-field interactions account for the formation of regular structures encountered in biomacromolecules, e.g., α-helices and β-sheets in proteins, double helices in nucleic acids, and helicoidally packed structures in polysaccharides, which enables us to use a greatly reduced number of interacting sites without sacrificing the ability to reproduce the correct architecture. This reduction results in an extension of the simulation timescale by more than four orders of magnitude compared to the all-atom representation. Examples of the performance of the model are presented.

Keywords Coarse-graining · Mean-field approach · Multipole–multipole interactions · Proteins · Nucleic acids · Polysaccharides

This paper belongs to Topical Collection 9th European Conference on Computational Chemistry (EuCo-CC9)

A. Liwo (✉) · C. Czaplewski · E. Gołaś · D. Jagieła · P. Krupa · M. Makowski · M. A. Mozolewska · A. Niadzvedtski · A. K. Sieradzan · R. Ślusarz · T. Wirecki · B. Zaborowski
Faculty of Chemistry, University of Gdańsk, ul. Wita Stwosza 63, 80-308 Gdańsk, Poland
e-mail: adam@sun1.chem.univ.gda.pl

M. Baranowski · S. Ołdziej
Laboratory of Biopolymer Structure, Intercollegiate Faculty of Biotechnology, University of Gdansk and Medical University of Gdansk, ul. Kładki 24, 80-922 Gdańsk, Poland

E. Gołaś · Y. He · P. Krupa · M. A. Mozolewska · H. A. Scheraga · T. Wirecki · Y. Yin · B. Zaborowski
Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, NY 14853-1301, USA

M. Maciejczyk
Department of Physics and Biophysics, Faculty of Food Sciences, University of Warmia and Mazury in Olsztyn, Michała Oczapowskiego 4, 10-719 Olsztyn, Poland

Introduction

Coarse-graining is the method of choice when simulating large systems [1–3]. Particular research effort has been directed toward the development of coarse-grained models of biological macromolecules such as proteins [3–17], nucleic acids [3, 18–33], carbohydrates [34–36], and biological assemblies, such as lipid bilayers [37, 38]. In this approach, a number of atoms are merged into single interaction sites, and the solvent surrounding the system is usually treated at the mean-field level in the form of a continuous medium. The main purpose of such an approach is to enable us to run simulations at time and size scales that are orders of magnitude greater than possible using the all-atom approach [39]. This is a great advantage despite the exponential growth of computing power

in recent years (especially that due to the introduction of graphical processor units [40], which capitalized on parallel computations at a scale unknown before, and the very recent construction of the ANTON supercomputer by Shaw and coworkers [41], which made ab initio simulations of the folding of small proteins at the detailed atomistic scale possible [42]). On the other hand, recent work suggests that coarse graining can also be used as a means to understand the rules behind the formation of macromolecular structure and macromolecular dynamics [43–45].

Constructing coarse-grained force fields is a much greater challenge than constructing all-atom force fields; the physical foundations of coarse-grained force fields were discovered only relatively recently [1, 46]. These force fields are divided into two main categories: knowledge-based and physics-based. Knowledge-based force fields are derived based on statistics determined from structural databases [4], while physics-based force fields relate all-atom energy surfaces to effective coarse-grained energy surfaces [13]. Physics-based force fields can, in turn, be divided into neoclassical force fields, in which the functional form is copied from that of all-atom force fields (e.g., the very widely applied MARTINI force field [38]), and those that are based on the understanding of a coarse-grained force field as a potential of mean force in which the degrees of freedom that are not omitted from the model have been integrated out [1, 46].

Based on our understanding of coarse-grained force fields as potentials of mean force, over the last 20 years we have been developing our physics-based united residue (UNRES) model of polypeptide chains [46–54]. To derive the force field in a systematic and consistent way, we developed a method of factorizing the PMF in contributions arising from smaller fragments of the system (thereby making it computable and transferable). These factors can also be expanded into the Kubo cluster-cumulant series [55], thereby enabling us to obtain analytical expressions for the respective terms, especially multibody terms, which are derived in other force fields in a heuristic manner [4]. Another very important feature of the UNRES model is that it emphasizes the role of electrostatic interactions involving polar peptide groups, which are represented as the mean-field interactions between peptide-group dipoles. These mean-field interactions are the main factors responsible for the formation of regular α-helical and β-sheet structure in proteins [44, 46].

The success of the UNRES model prompted us to extend the philosophy of constructing coarse-grained models to other biological macromolecules, namely nucleic acids and polysaccharides, and to produce the unified coarse-grained model (UCGM) for all these three classes of macromolecules that occur in all living organisms as building materials and perform a variety of functions. Very recently, using the very same concept, we extended the UNRES model to the nucleic acid united residue two-point model (NARES-2P), in which one interaction site per nucleotide is the phosphate group and the second is the nucleic acid base merged with its sugar ring. These sites serve as the polar units which interact via mean-field dipole–dipole interactions. Despite its simplicity, the NARES-2P model reproduces the double-helical structures of small DNA and RNA molecules and the melting thermodynamics of small DNA molecules surprisingly well [56]. We have also extended the treatment to polysaccharides, to produce the sugar united residue one-point model (SUGRES-1P).

In this paper, the theory behind the unified coarse-grained model is presented, and its components—the UNRES, NARES-2P, and the as-yet unpublished SUGRES-1P models—are described. Results of simulations performed using the three force fields are presented, and perspectives on their unification into one system—which will be able to treat not only the structures and dynamics of the isolated components but also interactions and composites of them, such as glycans—are outlined.

Methods

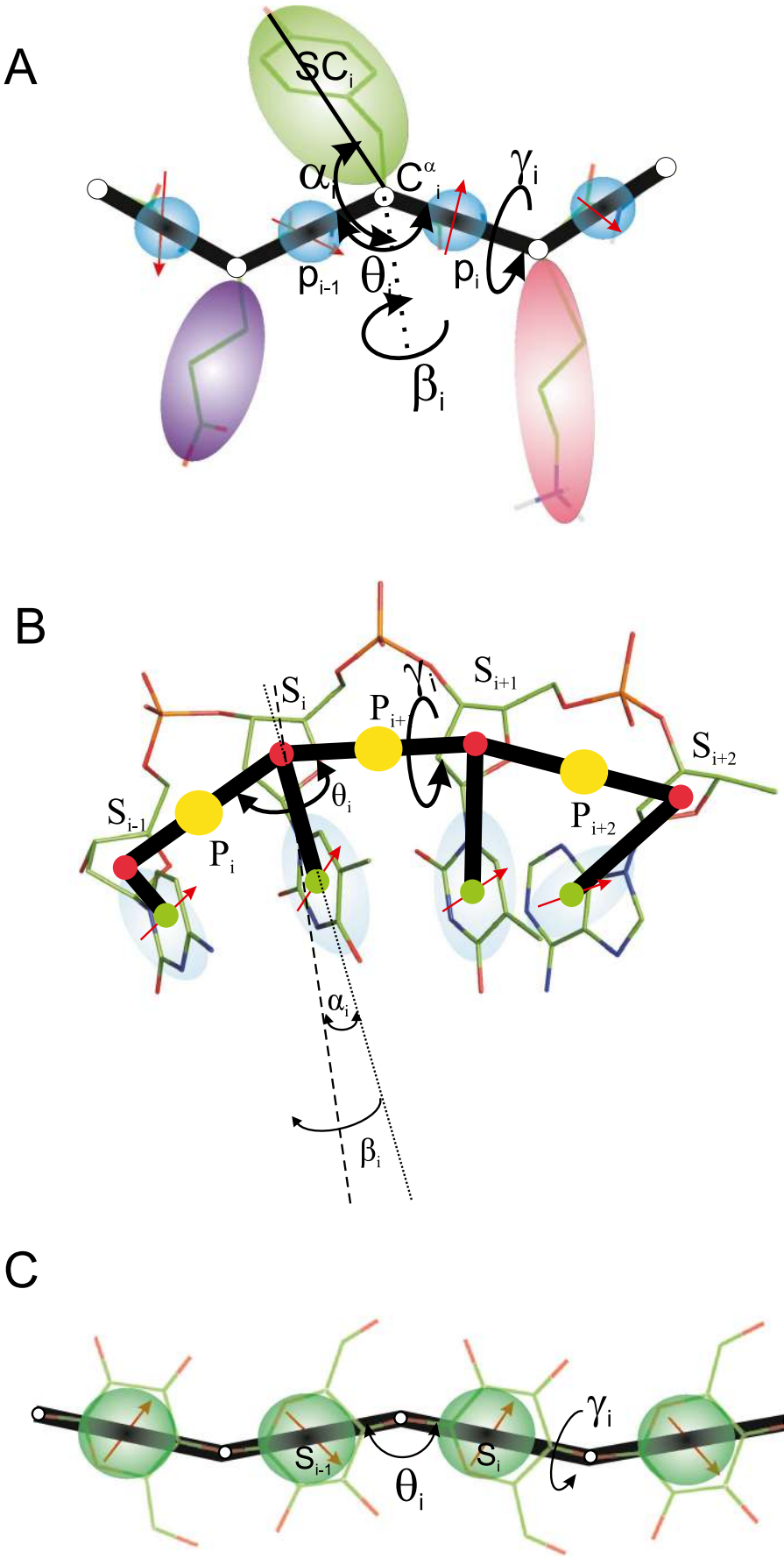
The unified coarse-grained model of biological macromolecules

As mentioned in the "Introduction," the unified coarse-grained model of biological macromolecules is a generalization of the approach taken when designing the UNRES model for proteins [46–54]. It assumes that (i) a biopolymer unit has an easily distinguishable polar site with a charge distribution represented by a point multipole and that (ii) the mean-field interactions between the polar sites, averaged in the context of local interactions, determine the symmetry of regular structures. The components of the model—the UNRES, NARES-2P, and SUGRES-1P models for proteins, nucleic acids, and polysaccharides, respectively—are visualized in Fig. 1.

In the following subsections, we will outline the method used to derive the coarse-grained force field through cluster-cumulant-function expansion of the potential of mean force of the system developed in our earlier work [46, 49, 51]. We will then provide short descriptions of the components of the model.

Potential of mean force of a coarse-grained system and its expansion into Kubo cluster-cumulant functions

In our approach for polypeptide chains [46], we assume that the effective energy function of a system is the potential of mean force (PMF), also termed the restricted free energy function (RFE), with all degrees of freedom that are lost when passing from the all-atom to the coarse-grained model averaged out. These neglected degrees of freedom include solvent degrees of freedom, side-chain rotation angles, and the dihedral angles λ for rotation of the peptide groups about the


A
SC$_i$
α_i
C$^{\alpha}_i$
γ_i
p$_{i-1}$
θ_i
p$_i$
β_i
B
S$_i$
γ_i
S$_{i+1}$
P$_{i+1}$
S$_{i+2}$
θ_i
S$_{i-1}$
P$_i$
P$_{i+2}$
α_i
β_i
C
S$_{i-1}$
θ_i
S$_i$
γ_i

◀ **Fig. 1** **a**–**c** The components of the unified coarse-grained model of biological macromolecules (UCGM). **a** UNRES (proteins), **b** NARES-2P (nucleic acids), **c** SUGRES-1P (polysaccharides). The polar interaction sites bearing point dipoles (depicted as *red arrows*) are colored *blue* and the virtual bonds are shown as *thick black lines*. It should be noted that the dipoles are not fixed but rotate about the virtual-bond axes to give average potentials. For nucleic acids (panel **b**), the dipole moments of the purine bases are approximately parallel to the S···B virtual-bond (rotation) axis; they are approximately perpendicular for pyrimidine bases, which explains the geometry of Watson–Crick pairing [56]. All-atom chains are superposed on coarse-grained representations for better illustration. In the UNRES model of polypeptide chains (**a**), the interaction sites are side chains, represented as ellipsoids of revolution of different sizes (*SC*) attached to the corresponding α-carbon atoms (represented by *small open circles*), and peptide-bond centers (*p*). The equilibrium length of a peptide bond is 3.8 Å for the *trans* and 2.8 Å for the *cis* configuration. For the *i*th residue, the geometry of the respective chain fragment can be described using virtual-bond angles θ_i, virtual-bond dihedral angles γ_i, and the polar angles α_i and β_i. In the NARES-2P model of the nucleotide chain, the interaction sites are phosphate groups (*P*), represented by *yellow circles*, and the nucleic acid bases fused with sugar rings, represented by *ellipsoids*, with their geometric centers at the Bs (*green circles*). The Ps are located halfway between two consecutive sugar atoms. The backbone virtual-bond angles θ and the virtual-bond dihedral angles γ, as well as the polar angles α and β that define the orientation of the sugar-base vector with respect to the backbone, are also shown. *Small red circles* represent the sugar-ring centers (*S*) which serve as geometric points. In the SUGRES-1P model of the polysaccharide chains, the interaction sites are the sugar rings (*S*), represented by *blue transparent spheres*, while *white circles* represent the glycosidic oxygen atoms (*O*) that serve as anchor points

$C^{\alpha}\cdots C^{\alpha}$ virtual bonds. The solvent degrees of freedom are usually averaged out explicitly using Monte Carlo (MC) or molecular dynamics (MD) simulations, or implicitly using data from the PDB [57]. Thus, the variables describing the geometry of the macromolecule–water system are divided into two sets: the *primary* variables (**X**), which describe the coarse-grained degrees of freedom, and the less important or *secondary* variables (**Y**) that are averaged out. In general, the RFE [$F(\mathbf{X})$] is expressed as

$$F(\mathbf{X}) = -RT\ln\left\{\frac{1}{V_{\mathbf{Y}}}\int_{\Omega_{\mathbf{Y}}}\exp\left[-E(\mathbf{X};\mathbf{Y})\Big/RT\right]\mathrm{d}V_{\mathbf{Y}}\right\}, \tag{1}$$

where $V_{\mathbf{Y}} = \int_{\Omega_{\mathbf{Y}}}\mathrm{d}V_{\mathbf{Y}}$, $E(\mathbf{X};\mathbf{Y})$ is the original (all-atom) energy function, R is the universal gas constant, T is the absolute temperature, $\Omega_{\mathbf{Y}}$ is the region of the **Y** subspace of variables over which the integration is carried out, and $V_{\mathbf{Y}}$ is the volume of this region.

To identify the effective energy terms, the all-atom energy $E(\mathbf{X};\mathbf{Y})$ is expressed as a sum of component energies, each of which is either the sum of energies within a given unit (the *local interaction energies*) or between given units (the *long-range interaction energies*), as given by Eq. 2 below. The RFE (Eq. 1) is decomposed into factors, each of which is a *Kubo cluster-cumulant function* [55], as expressed by Eq. 3 below and illustrated in Fig. 2.

$$E(\mathbf{X};\mathbf{Y}) = \sum_{i=1}^{n}\varepsilon_i(\mathbf{X};\mathbf{z_i}), \tag{2}$$

where $\varepsilon_i(\mathbf{X};\mathbf{z_i})$ is the ith component energy, $\mathbf{z_i}$ contains the secondary degrees of freedom on which ε_i depends, and n is the number of energy components.

$$F(\mathbf{X}) = \sum_i f_i^{(1)}(\mathbf{X}) + \sum_{i<j} f_{ij}^{(2)}(\mathbf{X}) + \sum_{i<j<k} f_{ijk}^{(3)}(\mathbf{X}) + \ldots + \sum_{i_1<2\ldots<i_n} f_{i_1 i_2\ldots i_n}^{(n)}(\mathbf{X}) \tag{3}$$

The factors are expressed as

$$f_{i_1 i_2\ldots i_k}^{(k)} = \langle\langle\varepsilon_{i_1}\varepsilon_{i_2}\ldots\varepsilon_{i_k}\rangle\rangle_f = \sum_{l=1}^{k}\sum_{\substack{i_{m_1}<i_{m_2}<\ldots<i_{m_l}\\ m_i\in[1..k]}}(-1)^{k-l}F_{i_{m_1}i_{m_2}\ldots i_{m_l}}^{(l)} = \sum_{l=1}^{k}\sum_{\substack{i_{m_1}<i_{m_2}<\ldots<i_{m_l}\\ m_i\in[1..k]}}(-1)^{k-l}\left\langle\left\langle\varepsilon_{i_{m_1}}\varepsilon_{i_{m_2}}\ldots\varepsilon_{i_{m_l}}\right\rangle\right\rangle \tag{4}$$

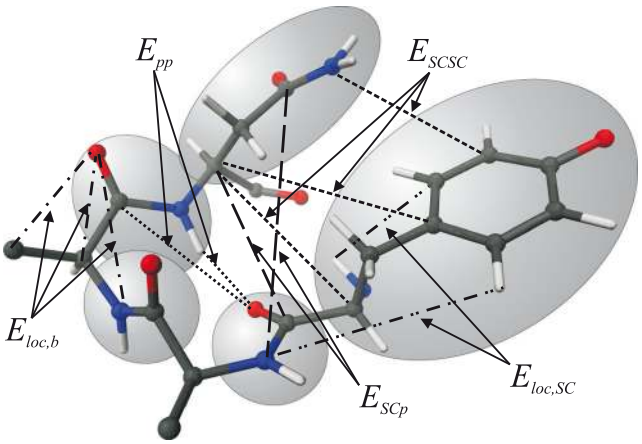


Fig. 2 The splitting of the interaction energy into component energies, as illustrated using a fragment of a polypeptide chain. *SC* denotes a side chain and *p* denotes a peptide group. The atoms of two side chains and three peptide groups of the portion of the polypeptide chain shown in the picture are embedded in *shaded ellipsoids*. For the sake of clarity, only some of the interactions are shown and the water molecules are not included. The terms E_{AB} denote the interaction energies between the atoms of sites A and B (e.g., E_{SCp} denotes the interactions between a side chain and a peptide group), while $E_{loc,b}$ and $E_{loc,SC}$ denote the energies that contribute to local-interaction energies within the backbone and the side-chain part of a given residue, respectively. Reproduced with permission from Figure 2 of [48]

where

$$F^{(k)}_{i_1,i_2,\ldots i_k}(\mathbf{X}) \equiv \langle\langle \varepsilon_{i_1}\varepsilon_{i_2}\ldots\varepsilon_{i_k}\rangle\rangle = -\frac{1}{\beta}\ln\left\{\frac{1}{V_{\mathbf{y}_I}}\int_{\Omega_I}\exp\left[-\beta\sum_{l=1}^{k}\varepsilon_{i_l}(\mathbf{X};\mathbf{z}_{i_l})\right]\mathrm{dV}_{\mathbf{y}_I}\right\} \quad (5)$$

is the RFE containing only a subset of component interactions (here, V_{y_I} is the volume of the subspace spanned by variables $y_{i_1}, y_{i_2}, \ldots, y_{i_k}$).

The factors of the first order, $f^{(1)}$, correspond to the PMFs of isolated units (e.g., isolated amino-acid residues) or those between isolated pairs of units (e.g., pairs of interacting side chains), while factors of order 2 and higher correspond to the multibody or correlation terms. All of the factors depend on temperature, and this dependence increases with increasing factor order because of the increasing order of the first term in the generalized-cumulant expansion of this factor [46, 50]. In our approach, as opposed to other coarse-grained force fields, this temperature dependence is explicitly accounted for [50].

The contributions of the correlation terms to the PMF and thus their importance depend on how many secondary variables are shared between the component energies, ε_i, included in a particular factor. If no secondary variables are shared, the corresponding factor is equal to zero. For polypeptide chains, the variables that are strongly shared between the factors are the angles λ for rotation of the peptide groups about the $C^{\alpha}\cdots C^{\alpha}$ virtual bonds.

The factor expansion is truncated [46] to achieve a good compromise between the complexity of the force field and its ability to reproduce the structure and dynamics of the system. We found that the fourth-order expansion is sufficient for the UNRES force field [58]. For the neoclassical force fields, e.g., MARTINI [35, 38, 59], all long-range interactions are approximated by factors of order 1 (i.e., by the potentials of mean force of isolated pairs of sites), while factors of order 2 occur only in the torsional potentials (these factors account for the coupling between the conformational states of the consecutive polymer units [46]). Approximate analytical formulae for factors can be obtained by taking the first nonzero generalized cumulant of its expansion into a generalized-cumulant series (which is very useful for correlation terms) [46] or by adapting the expressions from all-atom force fields (for the first-order factors and torsional potentials). These analytical expressions must be parameterized and the whole force field calibrated to reproduce the structure and physical properties of the system under study.

A general scheme of the construction of coarse-grained force fields based on the cluster-cumulant-expansion approach of the PMF is shown in Scheme 1.

The UNRES model of polypeptide chains

In our UNRES model [46–54] (Fig. 1a), a polypeptide chain is represented by a sequence of α-carbon (C^{α}) atoms linked by virtual bonds with attached united side chains (SC) and united peptide groups (p) located midway between the consecutive α-carbon (C^{α}) atoms (Fig. 1a). Only the united peptide groups and united side chains act as interaction sites. The C^{α} atoms serve only to define the geometry of the backbone trace, and are not interaction sites in the UNRES model.

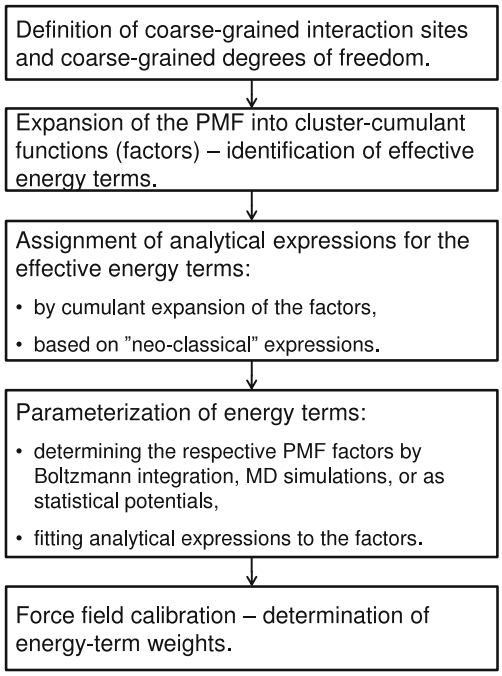
The energy of the virtual-bond polypeptide chain is expressed by

$$\begin{aligned} U = {} & w_{\mathrm{SCSC}}\sum_j\sum_{i<j}U_{\mathrm{SC}_i\mathrm{SC}_j} + w_{\mathrm{SCp}}\sum_j\sum_{i\neq j}U_{\mathrm{SC}_i\mathrm{p}_j} \\ & + f_2(T)w_{\mathrm{el}}\sum_j\sum_{i<j-1}U_{\mathrm{p}_i\mathrm{p}_j} \\ & + f_2(T)w_{\mathrm{tor}}\sum_i U_{\mathrm{tor}}(\gamma_i) \\ & + f_3(T)w_{\mathrm{tord}}\sum_i U_{\mathrm{tord}}(\gamma_i,\gamma_{i+1}) + w_b\sum_i U_{\mathrm{b}}(\theta_i) \\ & + w_{\mathrm{rot}}\sum_i U_{\mathrm{rot}}(\alpha_{\mathrm{SC}_i},\beta_{\mathrm{SC}_i}) \\ & + \sum_{m=2}^{N_{\mathrm{corr}}} f_m(T)w^{(m)}_{\mathrm{corr}}U^{(m)}_{\mathrm{corr}} + f_3(T)w^{(3)}_{\mathrm{turn}}U^{(3)}_{\mathrm{turn}} \\ & + f_4(T)w^{(4)}_{\mathrm{turn}}U^{(4)}_{\mathrm{turn}} + w_{\mathrm{bond}}\sum_i U_{\mathrm{bond}}(d_i) \\ & + w_{\mathrm{SS}}\sum_{\text{disulfide bonds}}U_{\mathrm{SS}_i} + n_{\mathrm{SS}}E_{\mathrm{SS}}, \end{aligned} \quad (6)$$

with

$$f_n(T) = \ln[\exp(1)+\exp(-1)] \Big/ \ln\left\{\exp\left[(T/T_{\circ})^{n-1}\right] + \exp\left[-(T/T_{\circ})^{n-1}\right]\right\} \quad (7)$$

The terms $U_{\mathrm{SC}_i\mathrm{SC}_j}$ correspond to the mean free energy of solvent-mediated interactions between the side chains. The terms $U_{\mathrm{SC}_i\mathrm{p}_j}$ correspond to the excluded-volume potential of the side chain–peptide group interactions. The terms $U_{\mathrm{p}_i\mathrm{p}_j}$ represent the energy of mean-field electrostatic interactions between backbone peptide groups. The terms U_{tor} and U_{tord} are the torsional and double-torsional potentials, respectively, for rotation about a given virtual bond or two consecutive



Scheme 1 General scheme of the construction and parameterization of coarse-grained force fields based on Kubo cluster-cumulant expansion of the PMF

virtual bonds. The terms U_{b} and U_{rot} are the virtual-bond-angle-bending and side-chain-rotamer potentials, respectively, and the term U_{bond} accounts for backbone and side-chain virtual-bond stretching [51, 60]. We recently [61] extended the backbone-virtual-bond stretching term to account for the *trans–cis* transition of peptide groups. The terms $U_{\mathrm{corr}}^{(m)}$ and $U_{\mathrm{turn}}^{(m)}$ correspond to the correlations (of order m) between peptide-group electrostatic and backbone-local interactions [46, 49]. The terms $U_{\mathrm{turn}}^{(m)}$ (the "turn" terms) involve consecutive segments of the chain. The correlation terms are absolutely essential for reproducing regular secondary structures, such as α-helices and β-sheets [46, 62]. We found [58] that correlation terms of order 3 and 4 are sufficient for the force field to reproduce regular protein structures. The terms U_{SS_i} are the energies of distortion of disulfide bonds from their equilibrium configuration, E_{SS} is the energy of formation of an "unstrained" disulfide bond in the chain (relative to the presence of two free cysteine residues), and n_{SS} is the number of disulfide bonds. The $w$terms are the weights of the respective energy terms. The multipliers $f_n(T)$ account for the temperature dependence of the dominant terms corresponding to the generalized-cumulant expansion of the PMF factors (Eq. 4); for a factor with a lowest nonzero cumulant of order m, the multiplier varies as $1/T^{m-1}$ with temperature [50]. For detailed expressions of the respective energy terms, the reader is referred to our earlier work [46–53].

All terms except $U_{\mathrm{SC}_i\mathrm{SC}_j}$ were determined by numerically computing the PMF surfaces of systems representing the corresponding PMF factors from the energy surfaces calculated by ab initio quantum mechanics (for U_{tor}, U_{tord}, U_{pp}, U_{corr} [51], U_{SS} [63]) or semiempirical AM1 (for U_{b} [60] and U_{rot}, and U_{vib} [60]) energy surfaces and fitting the respective analytical expressions to them. We initially [64] determined the side chain–side chain interaction potentials as knowledge-based potentials from the Protein Data Bank (PDB); however, they were recently [53, 65–68] re-determined from the PMFs of models of pairs of amino-acid side chains in water from all-atom MD simulations in explicit water.

To determine the energy-term weights (the w terms in Eq. 6), we developed [50, 58, 69] a hierarchical optimization approach in which the objective is to fit the weights so as to reproduce the order of structure formation and the thermodynamics of thermal folding/unfolding of the proteins selected for calibration.

The NARES-2P model of nucleic acids and the model of protein–nucleic acid interactions

In the NARES-2P model, depicted in Fig. 1b, a polynucleotide chain is represented by a sequence of virtual sugar (S) atoms that are located at the geometric centers of the sugar rings and linked by virtual bonds with attached united sugar bases (B) and united phosphate groups (P). These united sugar bases and the united phosphate groups serve as interaction sites. The energy of the virtual-bond chain is expressed by

$$\begin{aligned} U = {} & w_{\mathrm{BB}}^{\mathrm{GB}} \sum_i \sum_{j<i} U_{\mathrm{B}_i\mathrm{B}_j}^{\mathrm{GB}} + w_{\mathrm{BB}}^{\mathrm{el}} \sum_i \sum_{j<i} U_{\mathrm{B}_i\mathrm{B}_j}^{\mathrm{el}} \\ & + w_{\mathrm{PP}} \sum_i \sum_{j<i} U_{\mathrm{P}_i\mathrm{P}_j} + w_{\mathrm{PB}} \sum_i \sum_{j\neq i} U_{\mathrm{P}_i\mathrm{B}_j} + w_{\mathrm{bond}} \sum_i \\ & U_{\mathrm{bond}}(d_i) + w_{\mathrm{b}} i \sum_i U_{\mathrm{b}}(\theta_i) + w_{\mathrm{tor}} f_2(T) \sum_i U_{\mathrm{tor}}(\gamma_i) \\ & + w_{\mathrm{rot}} \sum_i U_{\mathrm{rot}}(\alpha_i, \beta_i), \end{aligned} \tag{8}$$

where $U_{\mathrm{B}_i\mathrm{B}_j}^{\mathrm{GB}}$ denotes the nonbonded interactions between the coarse-grained sugar-base sites, which is described by the Gay–Berne anisotropic potential [70], $U_{\mathrm{B}_i\mathrm{B}_j}^{\mathrm{el}}$ denotes the mean-field interactions between nucleic-acid-base dipoles (similar to that between peptide groups in UNRES [47]), $U_{\mathrm{P}_i\mathrm{P}_j}$ denotes the mean-field potential of phosphate group interactions, which consists of a Debye–Hückel term to account for solvent- and counterion-mediated charge–charge interactions [71], and the Lennard–Jones term $U_{\mathrm{P}_i\mathrm{B}_j}$ denotes the excluded-volume potential of the interactions of phosphate groups with sugar-base centers, the role of which is to prevent the collapse of these sites on each other, while U_{bond}, U_{b}, U_{tor}, and U_{rot} account for virtual-bond stretching, virtual-bond-angle bending, the energetics of rotation about the S···S virtual bonds, and the energetics of the local geometric states

of sugar-base sites. No correlation terms, except for the torsional potentials, are present in the current version.

The terms $U^{GB}_{B_iB_j}$ and $U^{elec}_{B_iB_j}$ were determined by numerical integration of the respective all-atom energy surfaces calculated with the AMBER force field, as done in our early work on the UNRES potential [47], and were fitted to the respective analytical expression. The dominant term was found to account for the mean-field interactions of the dipole-moment component parallel to the axis of one base with the dipole-moment component of the second base, which is perpendicular to its axis. This term has a minimum when the two base axes are perpendicular to each other, which is close to the geometry of the Watson–Crick base pairs. The local terms were determined as knowledge-based potentials from nucleic acid structures. The multipliers $f_n(T)$ are defined by Eq. 7. In the current version, $w^{GB}_{B_iB_j} = 0.5$ and the other weights were set to 2 to achieve the physiological melting temperature.

Coarse-grained model of polysaccharides (SUGRES-1P)

The sugar model developed in this project (depicted schematically in Fig. 1c) is a single-center model in which the anchor points are the glycosidic oxygen atoms (usually 1 and 4), with the sugar interaction site positioned between them. The ignored degrees of freedom are the rotation angles of the sugar rings about the O···O virtual bonds, usually the O(1)···O(4) virtual bonds, as seen in the structures of, e.g., cellulose and starch. Thus, the resulting force field has a component arising from mean-field backbone–dipole interactions that are averaged in the context of local interactions, just as for the UNRES model, and with the same functional forms [46]. Off-1,4 connections (the 1,2, 1,3, 1,6, etc. connections), including chain branching, fix the plane of the sugar ring involved; this rotational restriction is analogous to that imposed by the pyrrolidine ring in proline.

The current version of the SUGRES-1P model was developed for polysaccharides with 1,4-glycosidic bonds and parameterized for α- and β-D-glucose.

$$
\begin{aligned}
U = {} & w^{GB}_{SS}\sum_i\sum_{j<i} U^{GB}_{S_iS_j} + w^{el}_{SS} f_2(T)\sum_i\sum_{j<i} U^{el}_{S_iS_j} \\
& + w^{(3)}_{corr} f_3(T)\sum_i\sum_{j<i} U^{(3)}_{corr;S_iS_j} + w^{(3)}_{turn} f_3(T)\sum_i\sum_i \\
& U^{(3)}_{turn;S_iS_{i+2}} + w^{(4)} U^{(4)}_{corr} + w_{bond}\sum_i U_{bond}(d_i) \\
& + w_b\sum_i U_b(\theta_i) + w_{tor} f_2(T)\sum_i U_{tor}(\gamma_i),
\end{aligned}
\tag{9}
$$

where the terms $U^{GB}_{S_iS_j}$ represent the mean-field van der Waals and solvent-mediated interactions between sugar rings, which are represented by the anisotropic Gay–Berne potential [70], $U^{el}_{S_iS_j}$ represent the mean-field interactions of the sugar-ring dipoles outside of the context of local interactions (the same functional form as used for backbone peptide groups is applied [47]), $U^{(3)}_{corr;S_iS_j}$ and $U^{(3)}_{turn;S_iS_{i+2}}$ denote the correlation contributions that account for the restricted rotation of sugar-ring dipoles (again, the same functional forms are used as those employed for polypeptide chains [46, 49]), $U^{(4)}_{corr}$ is the sum of fourth-order correlation terms adapted from UNRES [48], U_{bond}, U_b, and U_{tor} denote the virtual-bond-deformation, virtual-bond-angle-deformation, and virtual-bond-torsional energies, respectively, and the w terms are the weights of the energy terms.

In the current preliminary version of SUGRES-1P, the parameters of $U^{GB}_{S_iS_j}$ and $U^{el}_{S_iS_j}$ were determined by calculating the potential energy surfaces as functions of the distance between sugar-ring centers and their orientation using the AM1 method of molecular quantum mechanics. These potential energy surfaces were then used to compute the potentials of mean force, by averaging out the rotation about the O(4)···O(4) virtual-bond axes, as done in our earlier work on the derivation of the $U_{p_ip_j}$ potentials for polypeptide chains [47, 49]. Therefore, the present version of SUGRES-1P can treat fibrillar polysaccharides which may contain only solitary water molecules inside. To include water, long-range interaction potentials were determined from molecular dynamics simulations using the same procedure as employed to determine the side chain–side chain interaction potentials [53]. The local-interaction parameters were determined from the PMFs of trisugars composed of all possible combinations of α- and β-D-glucose; these energy surfaces were subsequently used to compute the virtual-bond-torsional and virtual-bond-valence potentials using the procedures developed for the parameterization of UNRES [49, 72]. Two-dimensional Fourier series were also fitted to the energy surfaces of trisugars to derive the initial approximations of the parameters of the $U^{(3)}_{corr;S_iS_j}$ and $U^{(3)}_{turn;S_iS_{i+2}}$ correlation terms, as done in our earlier work on UNRES [46]. No further refinement of these parameters has been carried out so far.

Implementation of the components of UCGM

The UNRES model was initially used with the conformational space annealing (CSA) method of global optimization [73] to predict protein structures as global minima of the potential energy function. To extend its applications, we later implemented Langevin dynamics with UNRES [39, 74]. The equations of motion for the UNRES chain are Langevin dynamics equations because the solvent is implicit in UNRES. Consequently, it contributes to conservative forces (through the RFE) and gives rise to nonconservative forces which originate in the energy exchange of the polypeptide

chain with the solvent (the stochastic and friction forces). Because the geometry of an UNRES chain is not uniquely defined by the Cartesian coordinates of the interacting sites, we chose the virtual-bond vectors ($C^{\alpha}\cdots C^{\alpha}$ and $C^{\alpha}\cdots SC$) as generalized coordinates **q** and implemented the Lagrange approach to derive the equations of motion [39, 75, 76].

To enable larger MD steps (up to 20 fs, compared to the 1–2 fs time-step size applied in typical MD programs such as AMBER [77]), we have also designed the adaptive multiple time-step integration algorithm (A-MTS) [74].

To sample the conformational space more efficiently than achievable by canonical MD, we extended [78, 79] the UNRES/MD approach to the multiplexed replica-exchange molecular dynamics method (MREMD) [80].

The reader is referred to our earlier works on MD [39, 74–76] and REMD/MREMD [78, 79] implementations of UNRES.

We recently [81] parallelized the energy and force evaluations, which enabled us to run calculations of >500-residue proteins in a few days with massively parallel systems. To compute the averages from the results of simulations carried out at different temperatures, we adapted [50] the histogram-reweighting technique known as the weighted histogram analysis method (WHAM) [82]. With these extensions, we were able to calculate thermodynamic and ensemble-averaged structural characteristics of protein folding [50] and develop a physics-based protocol for protein-structure prediction in which the candidate predictions are conformations averaged over subensembles of structures with the highest probability below the folding-transition temperature [50].

The NARES-2P and SUGRES-1P models were built into the UNRES/MD platform and thus enabled us to carry out canonical [39] and replica-exchange [79] simulations of nucleic acids and polysaccharides, respectively.

The UNRES package, with full documentation, is available to the academic community at http://www.unres.pl. It will be extended to incorporate NARES-2P and SUGRES-1P as soon as these components are fully developed and parameterized. The current versions of NARES-2P and SUGRES-1P can be obtained from the authors on request.

Results

In this section, we briefly summarize the results obtained with UNRES and the results of initial test calculations obtained with NARES-2P and SUGRES-1P.

As mentioned in the "Methods" section, the initial application of UNRES was to make energy-based predictions of protein structures, in which the native structure was sought as the global minimum in the effective energy surface [73]. Using this approach, we scored the best prediction of target T0063 (HDEA) [83] in the Third Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP3) (see http://www.predictioncenter.org for more information about the CASP exercises). After implementing MD [39, 74, 76] and its extensions [79] in UNRES, we used a much better justified ensemble-based approach to prediction in which candidate predictions are sought as ensembles of geometrically similar structures [50]. Using this approach, we predicted correctly, as one of the only two groups, domain packing for the CASP10 target T0063 [84]; our prediction was featured by the CASP10 assessors. Based on sequence similarity, T0063 was a template-based modeling target but template-based methods failed to

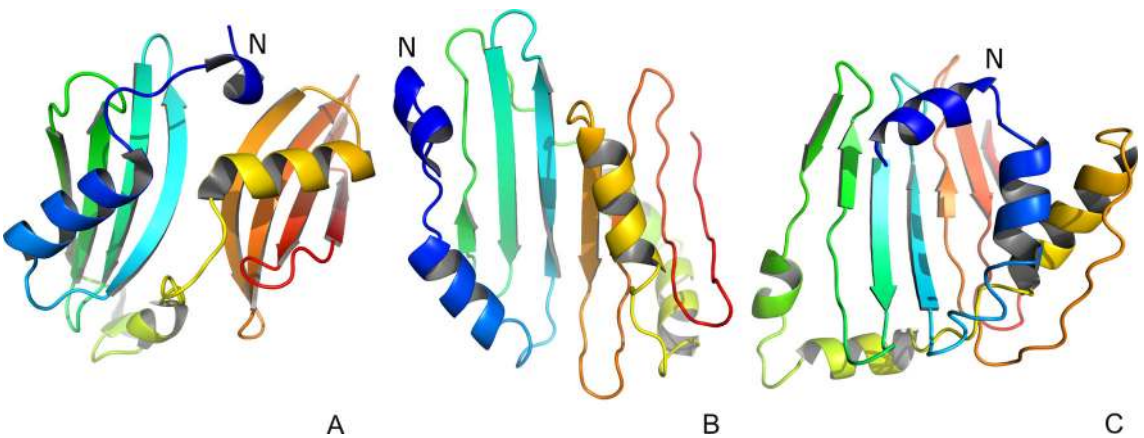


Fig. 3 **a** The experimental 4EXR structure of CASP10 target T0663. **b** Our model 1. **c** Our model 4. The N-termini in panels **a–c** are labeled *N*. The values of GDT_TS are 23.19, 31.98, and 42.80 for model 1 of the whole protein and its domains D1 and D2, respectively, and 22.04, 31.98, and 40.15 for model 4 of the whole protein and its domains D1 and D2, respectively. The respective GDT_TS values of the models with the highest GDT_TS values submitted to CASP are 42.93 (model 1 from group 27), 68.61 (model 3 from group 27), and 98.20 (model 4 from group 27). The drawings of the structures were produced with PYMOL (http://www.pymol.org). Reproduced with permission from Figure 1A-C of [84]

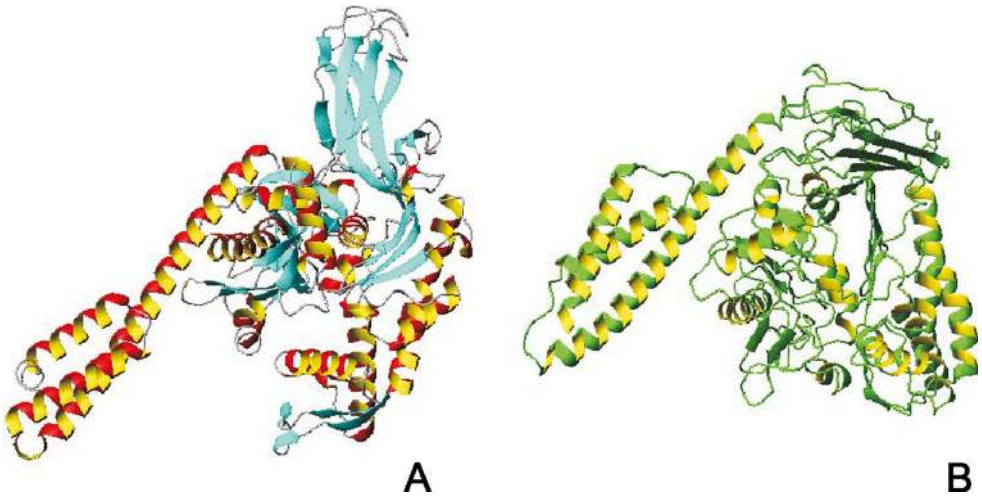


Fig. 4 **a** The experimental structure [94] of the open conformation of DnaK (a bacterial chaperone, PDB: 1BQ9) and **b** the structure simulated [93] with UNRES/MD, starting from the closed (substrate-binding) conformation of the chaperone, before the experimental structure was determined

predict correct domain packing. Our predicted structure of this target is compared with the experimental structure in Fig. 3.

With the MD implementation of UNRES, we carried out extensive studies of protein folding, including a simulation of the kinetics of the folding of the B-domain of staphylococcal protein A [85], a description of the folding pathway of protein A obtained through network analysis [86], and free-energy landscapes of protein A and the FBP28 WW domain and its variants [87–89]. We also applied the UNRES/MD approach to determine the mechanisms of biophysical processes, including amyloid formation and growth [90, 91] as well as signaling [92], and to investigate the Hsp70 chaperone cycle [93]. In particular, with UNRES, we simulated the transition between the substrate-binding (closed) and ATP-bound (open) conformations of DnaK, a bacterial Hsp70 chaperone [93]. The open structure calculated by UNRES turned out to be very similar to the ATP-bound structure of DnaK solved one year later [94]. Our calculated structure is compared with the experimental structure in Fig. 4.

For small proteins, UNRES/MREMD calculations require only several hours to achieve the convergence of ensemble averages; for example, for the 46-residue fragment of the B-domain of staphylococcal protein A (a three-α-helix-bundle structure; this is one of the benchmark systems for UNRES calculations), 20 million MD steps per trajectory are run in about 7 h with Intel Pentium processors, with one core handling one trajectory. For larger systems (200–300 residue proteins), the same number of steps require about 24 CPU hours, with 4–16 cores handling one trajectory. A detailed study of the speed of UNRES and its parallel efficiency can be found in our earlier work [81].

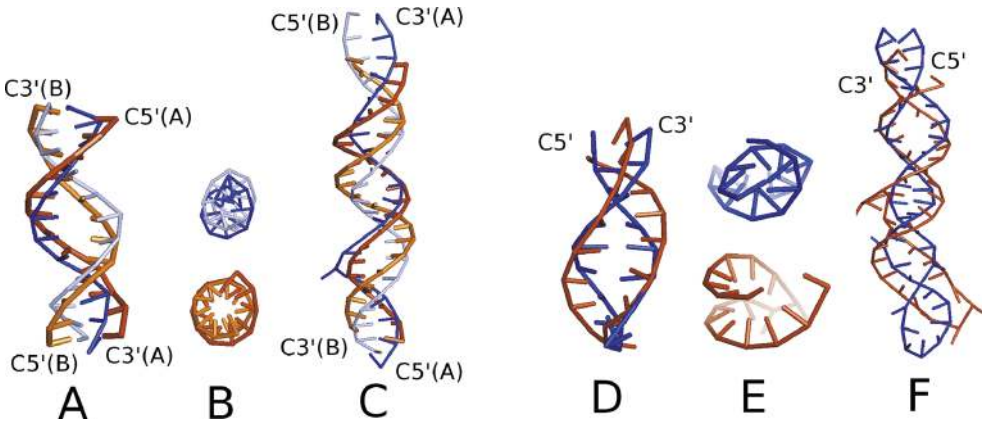


Fig. 5 **a**–**f** Calculated ensemble-averaged structures at T=300 K obtained in MREMD simulations (*thin blue sticks*) of the two small DNA molecules 9BNA (**a** and **b**) and 2JYK (**c**) and of the two small RNA molecules 2KPC (**d** and **e**) and 2KX8 (**f**), as compared to the respective experimental structures (*thick brown sticks*). For the *side views* (presented for all molecules), the calculated structures are superposed on the experimental structures, while the experimental structures are shown below the calculated structures for the *top views* (presented for 9BNA and 2JYK). The root-mean-square deviations (RMSDs) over the S centers averaged over all native-like clusters are 4.5 Å, 8.1 Å, and 5.7 Å for 9BNA, 2JYK, and 2KPC, respectively, and 9.8 Å for 2KX8 with respect to each experimental structure. The lowest RMSD values obtained in the respective MREMD runs are 2.9 Å, 5.6 Å, 1.6 Å, and 6.9 Å for 9BNA, 2JYK, 2KPC, and 2KX8, respectively. Reproduced with permission from Figure 2 of [56]

Among the other physics-based force fields, the optimized potential for protein structure prediction (OPEP) from the Derreumaux group, which uses a detailed all-atom representation of the protein backbone and united side chains, was applied in ab initio folding. The latest version of the force field succeeded in folding the tryptophan zipper and the FBP28 WW domain (a three-stranded antiparallel β-sheet protein); the root-mean-square deviation (RMSD) of the most populated cluster was 3.8 Å [15, 16, 95]. This resolution, when scaled by protein size, is comparable to the resolution of the UNRES force field, although UNRES has been tested with larger sets of small proteins [50, 69, 96] and was also tested with larger proteins in the CASP experiments [52, 56, 83, 97, 98]. Just like UNRES [90, 91], OPEP was successfully used to simulate the aggregation of amyloidogenic peptides [3, 10, 11, 15].

NARES-2P has not yet been applied to solve practical problems; however, we carried out extensive tests of this approach [56]. To assess the predictive power of NARES-2P, unrestricted multiplexed replica exchange simulations,

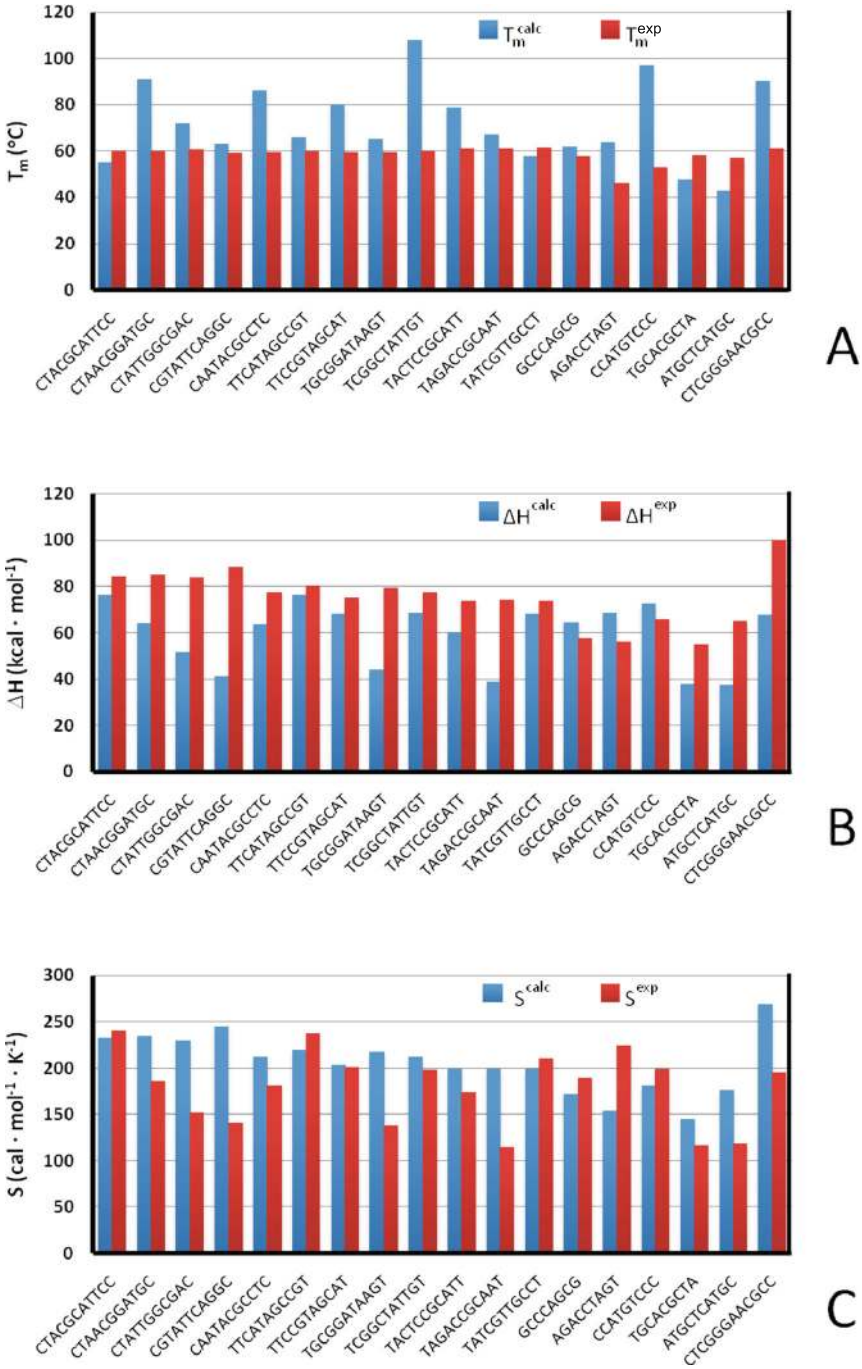


Fig. 6 **a**–**c** Comparison of the experimental (*red bars*) and calculated (*blue bars*) temperatures (**a**), enthalpies (**b**), and entropies (**c**) of melting of model small DNA molecules. Data from [56]

started from extended unpaired chains, were carried out with two small DNA molecules (9BNA, 2×12 nucleotides; and 2JYK, 2×21 nucleotides) and two RNA molecules (2KPC, 17 nucleotides; 2KX8, 44 nucleotides) molecules. The conformational ensembles below the melting temperature consisted almost exclusively of native right-handed double-helical structures. Example results are shown in Fig. 5.

The coarse-grained DNA model from the Ouldridge group, which uses Morse-like potentials to reproduce base pairing and stacking with base-pair-type specific parameters [25, 30], and the HiRe-RNA model from the Derreumaux group [3, 28, 32], which uses Gaussian-type multibody terms to account for base pairing, can also fold nucleic acid molecules. However, both of these contain more interaction sites per nucleotide unit, and the functional forms of the potentials have been constructed to reproduce base pairing and stacking, while these features arise in NARES-2P from the mean-field electrostatic nature of the dominant base–base interaction terms. In addition, a number of statistical potentials [23, 24] reproduce the experimental RNA structures in ab initio folding simulations.

It is very interesting that removing or reducing the $U_{B_iB_j}^{elec}$ component destroyed the folding capability of the NARES-2P force field, while removing local interactions (even the virtual-bond-angle terms) did not impair the ability of the force field to form double helices. Only right- and left-handed double helices appeared in comparable amounts due to the absence of the torsional potential that defines chain chirality [56]. These results suggest that the mean-field dipole–dipole interactions help to form structure. Unlike for proteins, the related correlation interactions do not appear to be required to reproduce double-helical structure.

We have also tested the ability of NARES-2P to reproduce the thermodynamic parameters associated with DNA melting. To accomplish this, we ran [56] MREMD simulations of a number of small DNA molecules for which the thermodynamics of melting were studied by calorimetry [99, 100]. As shown in Fig. 6, the agreement between the calculated and experimental melting temperatures, enthalpies, and entropies of melting is reasonable.

Because the NARES-2P energy function is less computationally expensive than the UNRES energy function (it does not have correlation terms), NARES-2P requires less time for a given number of MD steps. For example, for the 2KX8 RNA molecule (44 nucleotides), 20,000,000 MD steps take only 3 h. On the other hand, because the bases are usually mispaired in the initial folding stages and have to rearrange, it takes three- to fourfold more MD steps to obtain converged conformational averages as compared to the UNRES simulations for proteins.

The SUGRES-1P force field is at the initial development stage. Nevertheless, the limited tests carried out so far are encouraging. In Fig. 7, the average structure of the most populated cluster of conformations of a helical section of cyclic amylose, and that of a dimer of two 12-residue α-D-glucose chains (a unit of amylose), obtained in unrestricted MREMD simulations using the SUGRES-1P force field, are compared with the respective experimental data. As shown, the force field is able to reproduce the double-helical fold of both systems.

Conclusions and outlook

The examples illustrated in the "Results" section have shown that it is possible to construct a unified coarse-grained model with a very small number of interaction sites per unit that describes the structure and energetics of proteins, nucleic acids, and polysaccharides surprisingly well. The success of

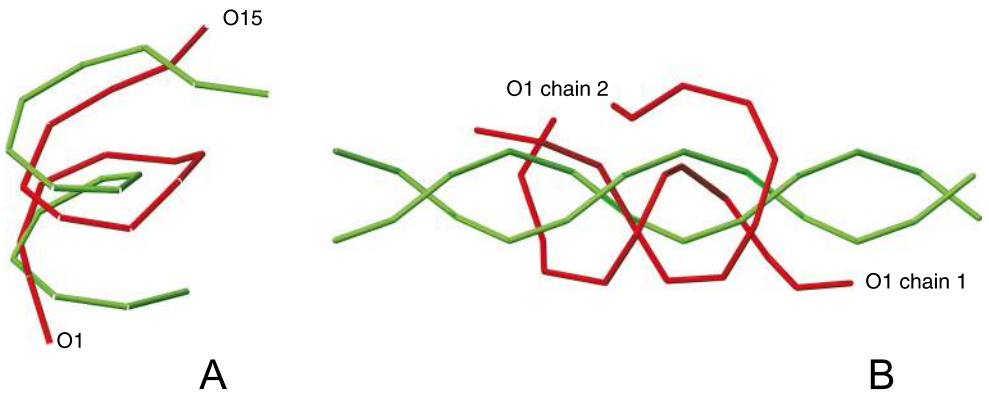


Fig. 7 **a** Superposition of the experimental (*green*) and calculated (using the SUGRES-1P model; *red*) O4 traces of part of cyclic amylose. The RMSD over the O4 atoms is 5.5 Å. **b** Superposition of the experimental (*green*) and calculated (using the SUGRES-1P model; *red*) O4 traces of amylose A (structure from the PolySac3DB database [101]; http://polysac3db.cermav.cnrs.fr/polysacdb/amylose-a/AmyA_double.pdb). Calculations were carried out on the dimer composed of two 12-residue monomers. The RMSD over the O4 atoms is 7 Å. It can be seen that, despite considerable distortion of the calculated structure at the ends, the right-handed-twisted double-helical structure and parallel packing of the chains is preserved

the UCGM most probably results from two principles of its design: (i) the origin of the effective energy function in the potential of mean force, which is then split into factors, enabling us to extract pure components pertaining to a given part of the system under consideration without the danger of counting the same contributions multiple times, and (ii) focusing on electrostatic and local interactions between polar units, the interactions of which seem to determine biopolymer architecture. Moreover, all three components of the model are based on the same geometric design: placing backbone sites between two anchor points and attaching branches (side chains, nucleic acid bases) to the same anchor points. Therefore, merging all of the components of UCGM into one system is a relatively simple task. In particular, it is feasible to interface the oligosaccharide part to a protein to form the respective glycan. At present, we are also extending the model to protein–nucleic acid interactions. We have already developed the potentials of interactions between protein side chains and nucleic acid bases (Yin Y, Sieradzan AK, Liwo A, He Y, Scheraga HA, manuscript in preparation). Using these extensions, the model will become a tool with which it will be possible to study the energetics and dynamics of biochemical processes using a small fraction of the computational effort required by all-atom simulations, while still being able to keep track of the physics of the respective phenomena.

The transferability and universality resulting from maintaining close connections of the effective UNRES, NARES-2P, and SUGRES-1P energy functions with the physics of interactions in these types of macromolecules is the greatest advantage of the unified coarse-grained model.

On the other hand, the resolution of the components of the model, even the most advanced UNRES model for proteins, is only moderate (about 5 Å for an approx. 50-residue protein). Fortunately, this feature does not seem to be inherent in the coarse-grained approach because some of our test calculation resulted in average RMSDs of about 2 Å for a 67-residue protein [96]. The force fields constituting UCGM are probably still missing details of local interactions. Work on improving the representation of local interactions is in progress in our laboratory. Very recently [54], we introduced torsional potentials involving the virtual $C^{\alpha}\cdots SC$ bonds. This modification improved the resolution of the force field by about 0.5 Å on average [54].

Acknowledgments This work was supported by grants DEC-2012/06/A/ST4/00376 (to AL) and DEC-2013/10/E/ST4/00755 (to MM) from the National Science Center of Poland, by grants Mistrz7.1./2013 (to AL) and START 100.2013 (to AKS) from the Foundation for Polish Science, by grant GM-14312 from the U.S. National Institutes of Health, and by grant MCB10-19767 (to HAS) from the U.S. National Science Foundation (to HAS). This research was supported by an allocation of advanced computing resources provided by the National Science Foundation (http://www.nics.tennessee.edu/), and by the National Science Foundation through TeraGrid resources provided by the Pittsburgh Supercomputing Center. Computational resources were also provided by (a) the supercomputer resources at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, (b) the 624-processor Beowulf cluster at the Baker Laboratory of Chemistry, Cornell University, (c) the 184-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk, and (d) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) of the University of Warsaw, Warsaw, Poland.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Ayton GS, Noid WG, Voth GA (2007) Multiscale modeling of biomolecular systems: in serial and in parallel. Curr Opin Struct Biol 17:192–198
2. Voth G (2008) Introduction. In: Voth G (ed) Coarse-graining of condensed phase and biomolecular systems, 1st edn. CRC (Taylor & Francis Group), Boca Raton, pp 1–4
3. Sterpone F, Melchionna S, Tuffery P, Pasquali S, Mousseau N, Cragnolini T, Chebaro Y, St-Pierre J-F, Kalimeri M, Barducci A, Laurin Y, Tek A, Baaden M, Nguyen PH, Derreumaux P (2014) The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. Chem Soc Rev 43:4871–4893. doi:10.1039/c4cs00048j
4. Koliński A, Skolnick J (2004) Reduced models of proteins and their applications. Polymer 45:511–524
5. Peng S, Ding F, Urbanc B, Buldyrev SV, Cruz L, Stanley HE, Dokholyan NV (2004) Discrete molecular dynamics simulations of peptide aggregation. Phys Rev E Stat Nonlin Soft Matter Phys 4: 041908
6. Tozzini V (2005) Coarse-grained models for proteins. Curr Opin Struct Biol 15:144–150
7. Colombo G, Micheletti C (2006) Protein folding simulations: combining coarse-grained models and all-atom molecular dynamics. Theor Chem Acc 116:75–86
8. Clementi C (2008) Coarse-grained models of protein folding: toy models or predictive tools? Curr Opin Struct Biol 18:10–15
9. Pincus DL, Cho SS, Hyeon HC, Thirumalai D (2008) Minimal models for proteins and RNA: from folding to function. Prog Mol Biol Transl Sci 84:203–250
10. Song W, Wei G, Mousseau N, Derreumaux P (2008) Self-assembly of the β-microglobulin NHVTLSQ peptide using a coarse-grained protein model reveals a β-barrel species. J Phys Chem B 112:4410–4418
11. Lu Y, Derreumaux P, Guo Z, Mousseau N, Wei G (2009) Thermodynamics and dynamics of amyloid peptide oligomerization are sequence dependent. Proteins 75:954–963
12. Thorpe IF, Zhou J, Voth GA (2008) Peptide folding using multiscale coarse-grained models. J Phys Chem B 112:13079–13090
13. Czaplewski C, Liwo A, Makowski M, Ołdziej S, Scheraga HA (2010) Coarse-grained models of proteins: theory and applications (Chapter 3). In: Koliński A (ed) Multiscale approaches to protein modeling. Springer, Berlin, pp 35–83
14. Thorpe IF, Goldenberg DP, Voth GA (2011) Exploration of transferability in multiscale coarse-grained peptide models. J Phys Chem B 115:11911–11926
15. Chebaro Y, Pasquali S, Derreumaux P (2012) The coarse-grained OPEP force field for non-amyloid and amyloid proteins. J Phys Chem B 116:8741–8751
16. Sterpone F, Nguyen PH, Kalimeri M, Derreumaux P (2013) Importance of the ion-pair interactions in the OPEP coarse-

grained force field: parametrization and validation. J Chem Theory Comput 9:4574–4584
17. Zacharias M (2013) Combining coarse-grained nonbonded and atomistic bonded interactions for protein modeling. Proteins 81:81–92
18. Peyrard M, Bishop AR (1989) Statistical-mechanics of a nonlinear model for DNA denaturation. Phys Rev Lett 62:2755–2758
19. Olson WK (1996) Simulating DNA at low resolution. Curr Opin Struct Biol 6:242–256
20. Hyeon C, Thirumalai D (2005) Mechanical unfolding of RNA hairpins. Proc Natl Acad Sci USA 102:6789–6794
21. Knotts T IV, Rathore N, Schwartz DC, de Pablo JJ (2007) A coarse grain model for DNA. J Chem Phys 126:084901
22. Voltz K, Trylska J, Tozzini V, Kurkal-Siebert V, Langowski J, Smith J (2008) Coarse-grained force field for the nucleosome from self-consistent multiscaling. J Comput Chem 29:1429–1439
23. Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV (2008) Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. RNA 14:1164–1173
24. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB (2009) Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. RNA 15:189–199
25. Ouldridge TE, Louis AA, Doye JPK (2010) DNA nanotweezers studied with a coarse-grained model of DNA. Phys Rev Lett 104:178101
26. Maciejczyk M, Spasic A, Liwo A, Scheraga HA (2010) Coarse-grained model of nucleic acid bases. J Comput Chem 31:1644–1655
27. Bernauer J, Huang X, Sim AY, Levitt M (2011) Fully differentiable coarse-grained and all-atom knowledge-based potentials for RNA structure evaluation. RNA 17:1066–1075
28. Pasquali S, Derreumaux P (2010) Hire-RNA: a high resolution coarse-grained energy model for RNA. J Phys Chem B 114: 11957–11966
29. Xia Z, Gardner DP, Gutell RR, Ren P (2010) Coarse-grained model for simulation of RNA three-dimensional structures. J Phys Chem B 114:13497–13506
30. Ouldridge TE, Louis AA, Doye JPK (2011) Structural, mechanical, and thermodynamic properties of a coarse-grained DNA model. J Chem Phys 134:085101
31. Xia Z, Bell DR, Shi Y, Ren P (2013) RNA 3D structure prediction by using a coarse-grained model and experimental data. J Phys Chem B 117:3135–3144
32. Cragnolini T, Derreumaux P, Pasquali S (2013) Coarse-grained simulations of RNA and DNA duplexes. J Phys Chem B 117: 8047–8060
33. Leonarski F, Trylska J (2014) Modeling nucleic acids at the residue-level resolution. In: Liwo A (ed) Computational methods to study the structure and dynamics of biomolecules and biomolecular processes. Springer, Berlin, pp 109–149
34. Molinero V, Goddard III, WA (2006) Molecular modeling of carbohydrates with no charges, no hydrogen bonds, and no atoms. In: Vliegenthar JFG, Woods RJ (eds) NMR spectroscopy and computer modeling of carbohydrates: recent advances. American Chemical Society, Washington, DC, pp 258–270
35. Lopez CA, Rzepiela A, de Vries AH, Dijkhuizen L, Hunenberger PH, Marrink SJ (2009) Martini coarse-grained force field: extension to carbohydrates. J Chem Theory Comput 5:3195–3210
36. Markutsya S, Devarajan A, Baluyut JY, Windus TL, Gordon MS, Lamm MH (2013) Evaluation of coarse-grained mapping schemes for polysaccharide chains in cellulose. J Chem Phys 138:214108
37. Lyubartsev AP (2005) Multiscale modeling of lipids and lipid bilayers. Eur Biophys J 35:53–61
38. Marrink SJ, Risselada JH, Yefimov S, Tieleman DP, de Vries AH (2007) The Martini force field: coarse grained model for biomolecular simulations. J Phys Chem 111:7812–7824
39. Liwo A, Khalili M, Scheraga HA (2005) Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains. Proc Natl Acad Sci USA 102: 2362–2367
40. Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, Beberg AL, Ensign DL, Bruns CM, Pande VS (2009) Accelerating molecular dynamic simulation on graphics processing units. J Comput Chem 30:864–872
41. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossvary I, Klepeis JL, Layman T, Mcleavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC (2008) Anton, a special-purpose machine for molecular dynamics simulation. Commun ACM 51:91–97
42. Shaw DE, Maragakis P, Lindorff-Larsen K, Piana S, Dror RO, Eastwood MP, Bank JA, Jumper JM, Salmon JK, Shan Y, Wriggers W (2010) Atomic-level characterization of the structural dynamics of proteins. Science 330:341–346
43. Krokhotin A, Liwo A, Niemi A, Scheraga HA (2012) Coexistence of phases in a protein heterodimer. J Chem Phys 137:035101
44. Liwo A (2013) Coarse graining: a tool for large-scale simulations or more? Phys Scr 87:058502
45. Krokhotin A, Liwo A, Maisuradze GG, Niemi AJ, Scheraga HA (2014) Kinks, loops, and protein folding, with protein A as an example. J Chem Phys 140:025101
46. Liwo A, Czaplewski C, Pillardy J, Scheraga HA (2001) Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. J Chem Phys 115:2323–2347
47. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA (1993) Prediction of protein conformation on the basis of a search for compact structures; test on avian pancreatic polypeptide. Protein Sci 2:1715–1731
48. Liwo A, Kaźmierkiewicz R, Czaplewski C, Groth M, Ołdziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA (1998) United-residue force field for off-lattice protein-structure simulations. III. Origin of backbone hydrogen-bonding cooperativity in united-residue potentials. J Comput Chem 19:259–276
49. Liwo A, Ołdziej S, Czaplewski C, Kozłowska U, Scheraga HA (2004) 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 108:9421–9438
50. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Ołdziej S, Wachucik K, Scheraga HA (2007) 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 111:260–285
51. Liwo A, Czaplewski C, Ołdziej S, Rojas AV, Kaźmierkiewicz R, Makowski M, Murarka RK, Scheraga HA (2008) Simulation of protein structure and dynamics with the coarse-grained UNRES force field (Chapter 8). In: Voth G (ed) Coarse-graining of condensed phase and biomolecular systems. CRC, Boca Raton, pp 1391–1411
52. Liwo A, He Y, Scheraga HA (2011) Coarse-grained force field: general folding theory. Phys Chem Chem Phys 13:16890–16901
53. Makowski M(2014) Physics-basedmodeling of side chain–side chain interactions in the UNRES force field. In: Liwo A(ed)Computational methods to study the structure and dynamics of biomolecules and biomolecular processes. Springer, Berlin, pp 81–107
54. Krupa P, Sieradzan AK, Rackovsky S, Baranowski M, Ołdziej S, Scheraga HA, Liwo A, Czaplewski C (2013) Improvement of the treatment of loop structures in the UNRES force field by inclusion of coupling between

backbone- and side-chain-local conformational states. J Chem Theory Comput 9:4620–4632
55. Kubo R (1962) Generalized cumulant expansion method. J Phys Soc Jpn 17:1100–1120
56. He Y, Maciejczyk M, Scheraga HA, Liwo A (2013) Mean-field interactions between nucleic-acid–base dipoles can drive the formation of a double helix. Phys Rev Lett 110:098101
57. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. Nucl Acid Res 28:235–242
58. Ołdziej S, Łagiewka J, Liwo A, Czaplewski C, Chinchio M, Nanias M, Scheraga HA (2004) Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. 3. Use of many proteins in optimization. J Phys Chem B 108: 16950–16959
59. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ (2008) The Martini coarse-grained force field: extension to proteins. J Chem Theory Comput 4:819–834
60. Kozłowska U, Maisuradze GG, Liwo A, Scheraga HA (2010) Determination of side-chain-rotamer and side-chain and backbone virtual-bond-stretching potentials of mean force from AM1 energy surfaces of terminally-blocked amino-acid residues, for coarse-grained simulations of protein structure and folding. 2. Results, comparison with statistical potentials, and implementation in the UNRES force field. J Comput Chem 31:1154–1167
61. Sieradzan AK, Scheraga HA, Liwo A (2012) Determination of effective potentials for the stretching of $C^{\alpha}\cdots C^{\alpha}$ virtual bonds in polypeptide chains for coarse-grained simulations of proteins from ab initio energy surfaces of *N*-methylacetamide and *N*-acetylpyrrolidine. J Chem Theory Comput 8:1334–1343
62. Kolinski A, Godzik A, Skolnick J (1993) 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 98:7420–7433
63. Chinchio M, Czaplewski C, Liwo A, Ołdziej S, Scheraga HA (2007) Dynamic formation and breaking of disulfide bonds in molecular dynamics simulations with the UNRES force field. J Chem Theory Comput 3:1236–1248
64. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Ołdziej S, Scheraga HA (1997) A united-residue force field for off-lattice protein-structure simulations. II. Parameterization of local interactions and determination of the weights of energy terms by Z-score optimization. J Comput Chem 18:874–887
65. Makowski M, Sobolewski E, Czaplewski C, Liwo A, Ołdziej S, No JH, Scheraga HA (2007) Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. 3. Calculation and parameterization of the potentials of mean force of pairs of identical hydrophobic side chains. J Phys Chem B 111:2925–2931
66. Makowski M, Sobolewski E, Czaplewski C, Ołdziej S, Liwo A, Scheraga HA (2008) Simple physics-based analytical formulas for the potentials of mean force for the interaction of amino acid side chains in water. IV. Pairs of different hydrophobic side chains. J Phys Chem B 112:11385–11395
67. Makowski M, Liwo A, Sobolewski E, Scheraga HA (2011) Simple physics-based analytical formulas for the potentials of mean force of the interaction of amino-acid side chains in water. V. Like-charged side chains. J Phys Chem B 115:6119–6129
68. Makowski M, Liwo A, Scheraga HA (2011) Simple physics-based analytical formulas for the potentials of mean force of the interaction of amino-acid side chains in water. VI. Oppositely charged side chains. J Phys Chem B 115:6130–6137
69. He Y, Xiao Y, Liwo A, Scheraga HA (2009) Exploring the parameter space of the coarse-grained UNRES force field by random search: selecting a transferable medium-resolution force field. J Comput Chem 30:2127–2135
70. Gay JG, Berne BJ (1981) Modification of the overlap potential to mimic a linear site–site potential. J Chem Phys 74:3316–3319
71. Kim YC, Hummer G (2008) Coarse-grained models for simulations of multiprotein complexes: application to ubiquitin binding. J Mol Biol 375:1416–1433
72. Kozłowska U, Liwo A, Scheraga HA (2007) Determination of virtual-bond-angle potentials of mean force for coarse-grained simulations of protein structure and folding from ab initio energy surfaces of terminally-blocked glycine, alanine, and proline. J Phys Cond Matter 19:285203
73. Lee J, Scheraga HA (1999) Conformational space annealing by parallel computations: extensive conformational search of Met-enkephalin and of the 20-residue membrane-bound portion of melittin. Int J Quantum Chem 75:255–265
74. Rakowski F, Grochowski P, Lesyng B, Liwo A, Scheraga HA (2006) Implementation of a symplectic multiple-time-step molecular dynamics algorithm, based on the united-residue mesoscopic potential energy function. J Chem Phys 125:204107
75. Khalili M, Liwo A, Rakowski F, Grochowski P, Scheraga HA (2005) Molecular dynamics with the united-residue model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode. J Phys Chem B 109: 13785–13797
76. Khalili M, Liwo A, Jagielska A, Scheraga HA (2005) Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model α-helical systems. J Phys Chem B 109:13798–13810
77. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman P (1995) 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 91:1–41
78. Nanias M, Czaplewski C, Scheraga HA (2006) Replica exchange and multicanonical algorithms with the coarse-grained united-residue (UNRES) force field. J Chem Theory Comput 2:513–528
79. Czaplewski C, Kalinowski S, Liwo A, Scheraga HA (2009) Application of multiplexing replica exchange molecular dynamics method to the UNRES force field: tests with α and α+β proteins. J Chem Theory Comput 5:627–640
80. Mitsutake A, Sugita Y, Okamoto Y (2003) Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test. J Chem Phys 118:6664–6675
81. Liwo A, Ołdziej S, Czaplewski C, Kleinerman DS, Blood P, Scheraga HA (2010) Implementation of molecular dynamics and its extensions with the coarse-grained UNRES force field on massively parallel systems; towards millisecond-scale simulations of protein structure, dynamics, and thermodynamics. J Chem Theory Comput 6:583–595
82. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. J Comput Chem 13: 1011–1021
83. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA (1999) Protein structure prediction by global optimization of a potential energy function. Proc Natl Acad Sci USA 96:5482–5485
84. He Y, Mozolewska MA, Krupa P, Sieradzan AK, Wirecki TK, Liwo A, Kachlishvili K, Rackovsky S, Jagieła D, Ślusarz R, Czaplewski CR, Ołdziej S, Scheraga HA (2013) Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. Proc Natl Acad Sci USA 110:14936–14941
85. Khalili M, Liwo A, Scheraga HA (2006) Kinetic studies of folding of the B-domain of staphylococcal protein A with molecular dynamics and a united-residue (UNRES) model of polypeptide chains. J Mol Biol 355:536–547

86. Yin Y, Maisuradze GG, Liwo A, Scheraga HA (2012) Hidden protein folding pathways in free-energy landscapes uncovered by network analysis. J Chem Theory Comput 8:1176–1189
87. Maisuradze GG, Liwo A, Scheraga HA (2009) How adequate are one- and two-dimensional free energy landscapes for protein folding dynamics? Phys Rev Lett 102:238102
88. Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA (2010) Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. J Phys Chem A 114:4471–4485
89. Maisuradze GG, Zhou R, Liwo A, Xiao Y, Scheraga HA (2012) Effects of mutation, truncation, and temperature on the folding kinetics of a WW domain. J Mol Biol 420:350–365
90. Rojas A, Liwo A, Browne D, Scheraga HA (2010) Mechanism of fiber assembly; treatment of a β-peptide aggregation with a coarse-grained united-residue force field. J Mol Biol 404:537–552
91. Rojas A, Liwo A, Scheraga HA (2011) A study of the α-helical intermediate preceding the aggregation of the amino-terminal fragment of the a β-amyloid peptide (1–28). J Phys Chem B 115: 12978–12983
92. Liwo A, He Y, Weinstein H, Scheraga HA (2011) PDZ binding to the BAR domain of PICK1 is elucidated by coarse-grained molecular dynamics. J Mol Biol 405:298–314
93. Gołaś E, Maisuradze GG, Senet P, Ołdziej S, Czaplewski C, Scheraga HA, Liwo A (2012) Simulation of the opening and closing of Hsp70 chaperones by coarse-grained molecular dynamics. J Chem Theory Comput 8:1750–1764
94. Kityk R, Koop J, Sinning I, Mayer MP (2013) Structure and dynamics of the ATP-bound open conformation of Hsp70 chaperones. Mol Cell 48:863–874
95. Derreumaux P (1999) From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential. J Chem Phys 111:2301–2310
96. Ołdziej S, Liwo A, Czaplewski C, Pillardy J, Scheraga HA (2004) Optimization of the UNRES force field by hierarchical design of the potential-energy landscape. 2. Off-lattice tests of the method with single proteins. J Phys Chem B 108:16934–16949
97. Pillardy J, Czaplewski C, Liwo A, Lee J, Ripoll DR, Kaźmierkiewicz R, Oldziej S, Wedemeyer WJ, Gibson KD, Arnautova YA, Saunders J, Ye Y-J, Scheraga HA (2001) Recent improvements in prediction of protein structure by global optimization of a potential energy function. Proc Natl Acad Sci USA 98: 2329–2333
98. Ołdziej S, Czaplewski C, Liwo A, Chinchio M, Nanias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kaźmierkiewicz R, Ripoll DR, Pillardy J, Saunders JA, Kang YK, Gibson KD, Scheraga HA (2005) Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. Proc Natl Acad Sci USA 102:7547–7552
99. Hughesman CB, Turner RFB, Haynes C (2011) Correcting for heat capacity and 5′-TA type terminal nearest neighbors improves prediction of DNA melting temperatures using nearest-neighbor thermodynamic models. Biochemistry 50:2642–2649
100. Hughesman CB, Turner RFB, Haynes C (2011) Role of the heat capacity change in understanding and modeling melting thermodynamics of complementary duplexes containing standard and nucleobase-modified LNA. Biochemistry 50:5364–5368
101. Sarkar A, Pérez S (2012) PolySac3DB: an annotated data base of 3 dimensional structures of polysaccharides. BMC Bioinforma 13:302