Journal of Molecular Graphics and Modelling 79 (2018) 118–132

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Shielding effect in protein folding

Adam K. Sieradzan[a,*], Agnieszka G. Lipska[a], Emilia A. Lubecka[a,b]

[a] Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland
[b] Institute of Informatics, University of Gdańsk, Wita Stwosza 57, 80-308 Gdańsk, Poland

ARTICLE INFO

Article history:
Received 30 June 2017
Received in revised form 19 October 2017
Accepted 20 October 2017
Available online 1 November 2017

Keywords:
Proteins
UNRES force field
Local interactions
Physicochemical properties
Potentials of mean force

ABSTRACT

One of the most important interactions responsible for protein folding and stability are hydrogen bonds between peptide groups. There is a constant competition between the water molecules and peptide groups in a hydrogen bond formation. Also side-chains take part in this process by reducing hydration of peptide group (shielding effect) that promotes the protein folding. In this paper, a new approach to take into account a shielding effect is presented. A modification of the energy function is derived and incorporated into the UNited RESidue (UNRES) force field. Canonical Molecular Dynamics and Replica Exchange Molecular Dynamics with UNRES force field is applied to study the influence of this effect on protein structure, folding kinetics and free energy landscapes. The results of test calculations suggest that even small contribution of this effect into energy function changes force field behavior as well as speeds up the folding process significantly.

1. Introduction

The importance of hydration of peptide group in proteins was recognized over five decades ago [1–4]. The importance of this phenomenon stems from the fact that the formation of a hydrogen bond between two peptide groups is crucial for protein folding and stability and there is a competition between water molecules and peptide groups in the formation of hydrogen bonds [5,6]. It was also found that the reduced hydration of a peptide group by side-chains (shielding effect) promotes protein folding [7].

In vivo, on a macro scale, the screening from solvent is performed by chaperones, which allow only a limited amount of water molecules to be in a chaperone cavity leading to dehydratation of interior of protein [8].

In an all-atom force field with an explicit solvent this type of interaction is directly incorporated, especially if polarizability is considered. Hence, the importance of the shielding effect can be investigated by comparing the differences in the folding of small proteins and peptides with the use of different water models. Examples are investigation of the free energy landscape of alanine oligopeptides [9], comparison of hydration behavior and conformation preference of the Trp-cage in different rigid body models [10,11] and hydration of beta-hairpin [12,13]. It was also suggested that the shielding effect is a stabilizing factor of α-helix in solution [12,14–16] and can be even one of factors keeping proteins stable in a high temperature [17].

However, in a coarse-grained force fields with implicit solvent, the shielding effect has to be incorporated into the effective Hamiltonian. In many force fields those interactions are omitted and side-chain peptide group interaction is incorporated only in a form of Lennard-Jones potential [18–24].

However, reduced hydration of backbone not only leads to fewer interactions with water molecules but also enhances the interactions between two peptide groups [7]. The shielding function of side-chain promotes the formation of peptide-group hydrogen bond and those spots may serve as nucleation sites in protein folding [7].

The analytical solution of the shielding effect phenomenon for micro-scale has been derived [25]. The electrostatic potential from point charge at a given position in space in solvent modified by the particle with different dielectric constant (Fig. 1) can be described by (Eq. (1))

$$\Phi_0 = \frac{q}{4\pi\epsilon_1 r_q} + \sum_{n=0}^{\infty} \frac{A_n P_n(\cos\theta)}{r_p^{n+1}} \tag{1}$$

where Φ_0 is electrostatic potential, q is point charge, ϵ_1 is dielectric constant of solvent, r_q is distance between the point charge and where the potential is measured, P_n is Legendre polynomial, θ is angle between vector connecting the point where electrostatic potential is computed and the vector connecting point charge and center of particle, r_p is distance between where the center of parti-

* Corresponding author.
E-mail address: adasko@sun1.chem.univ.gda.pl (A.K. Sieradzan).

https://doi.org/10.1016/j.jmgm.2017.10.018
1093-3263/

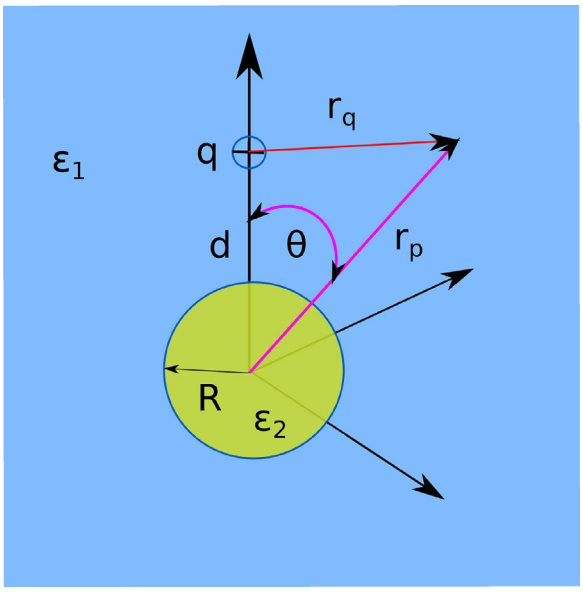

Fig. 1. Schematic representation of point charge in solution with dielectric constant ϵ_1 interacting with a spherical particle with dielectric constant ϵ_2.

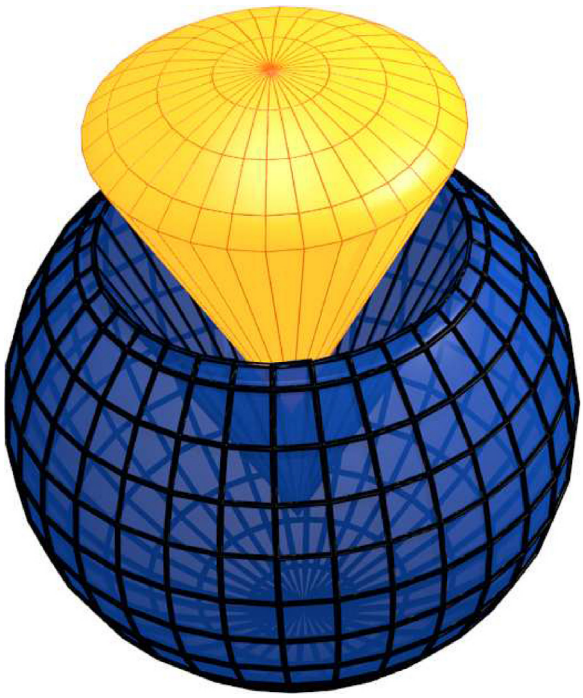

Fig. 2. Illustration of approximation of shielding effect in coarse-grain force field. The blue sphere denotes the first hydration sphere of peptide group, yellow represents the spherical sector occupied by side-chain disallowing water to approach the peptide group. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

cle and the point where the potential is computed, and A_n is given by (Eq. (2))

$$A_n = \frac{-q}{4\pi\epsilon_1}\frac{n(\epsilon_2-\epsilon_1)}{n\epsilon_2+(n+1)\epsilon_1}\frac{R^{2n+1}}{r_p^{n+1}} \tag{2}$$

where ϵ_2 is dielectric constant of particle and R is radius of particle. However, application of this equation due to computational reasons is limited.

There is a constant interest in developing more accurate electrostatic approximations for coarse-grain force fields [26]. In this article we show a new approach to treat the shielding effect of side chains. We implemented this new approach into the UNited RESidue (UNRES) force field. We also report the results of initial tests of the shielding effect on protein structure, thermodynamics and kinetics.

2. Methods

2.1. Theoretical model of shielding effect

The dipole–dipole interaction between two peptide is expressed by:

$$V(r) = -\frac{\mu_1\mu_2}{4\pi\epsilon\epsilon_0 r_{12}^3}(\cos\theta_{12} - 3\cos\theta_1\cos\theta_2) \tag{3}$$

where μ_1 (μ_2) is the dipole moment of the interacting particle 1 (2), ϵ is the dielectric constant of the environment, ϵ_0 is dielectric constant of the vacuum, r_{12} is distance between center of interacting particles, θ_{12} is angle between the dipole vectors, and θ_1 (θ_2) is angle between the vector connecting the interacting centers and the dipole vector of 1 (2) molecule.

Therefore the dielectric constant (ϵ) depends on the environment. The interior part of a protein is estimated to have ϵ value of ≈2–3 [27]. In this paper we approximated effective dielectric constant as a function (f) of the volume of hydration spheres (V_i, V_j) of the interacting peptide groups i and j, occupied by side-chains.

$$\epsilon = f(V_i; V_j) \tag{4}$$

As this would require simulations with a three-body potential it would result in $O(n^3)$ problem therefore further approximation is used.

$$\frac{1}{\epsilon} = f(V_i)f(V_j) \tag{5}$$

Therefore to incorporate the shielding effect, in the first step we approximated how the first hydration sphere of the interacting sites changes when a third body is in their vicinity. A spherical sector was used to calculate what fraction of first hydration sphere volume cannot be accessed by water molecules (Fig. 2).

The effective energy of interaction between peptide group i and j is therefore modified as given by (Eq. (6)):

$$U_{new} = U_{org} \times fac(i) \times fac(j) \tag{6}$$

where U_{new} is energy function after taking into consideration the shielding effect, U_{org} is original energy function describing electrostatic interaction between i and j, $fac(i)$ [$fac(j)$] is a factor proportional to fraction of the first hydration sphere of site i [j] occupied by side-chains and is given by (Eq. (7)):

$$fac(i) = 1 - w_{shield} + w_{shield} \times \sum_{SC} frac_V \tag{7}$$

where w_{shield} is susceptibility factor of the force field to shielding effect, $frac_V$ is a fraction of volume of the first hydration sphere occupied by a side-chain.

2.2. Introduction of shielding effect into UNRES force field

In the UNRES [28–36] model, the polypeptide chain is described by two geometric points per residue. The first one is C^α atom and the second one is the center of geometry of side chain (denoted as SC). Moreover, there are two interaction sites: the peptide group, which is positioned halfway between the two consecutive C^α atoms and the side-chain (SC). The peptide groups are modeled with spheres

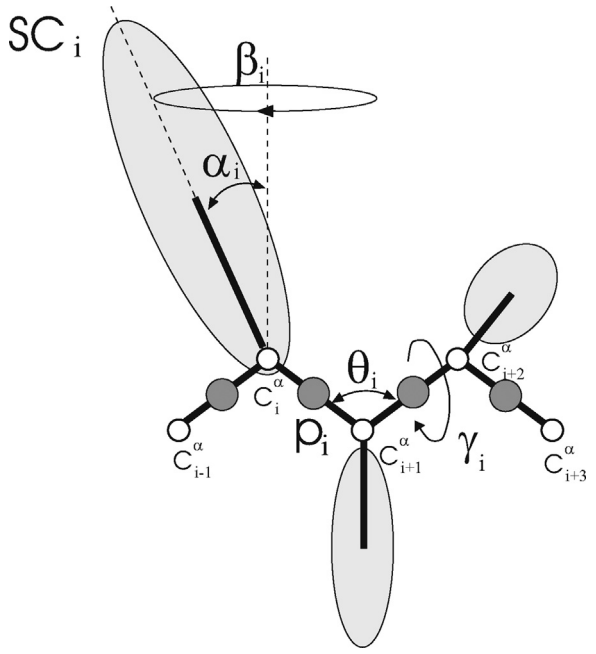

Fig. 3. The UNRES model of polypeptide chains. The interaction sites are peptide-group centers (p), and side-chain centers (SC) attached to the corresponding α-carbons with different $C^{\alpha}\cdots SC$ bond lengths, d_{SC}. The peptide groups are represented as dark gray circles and the side chains are represented as light gray ellipsoids of different size. The α-carbon atoms are represented by small open circles. The geometry of the chain can be described either by the virtual-bond vectors $\mathbf{dC}_i$ (from Ci_{α} to $Ci+1_{\alpha}$), $i=1,2,\ldots,n-1$, and $\mathbf{dX}_i$ (from Ci_{α} to SC_i), $i=2,\ldots,n-1$, represented by thick lines, where n is the number of residues, or in terms of virtual-bond lengths, backbone virtual-bond angles θ_i, $i=1,2,\ldots,n-2$, backbone virtual-bond-dihedral angles γ_i, $i=1,2,\ldots,n-3$, and the angles α_i and β_i, $i=2,3,\ldots,n-1$ that describe the location of a side chain with respect to the coordinate frame defined by $Ci-1_{\alpha}$, Ci_{α}, and Ci_{α}, $Ci+1_{\alpha}$.

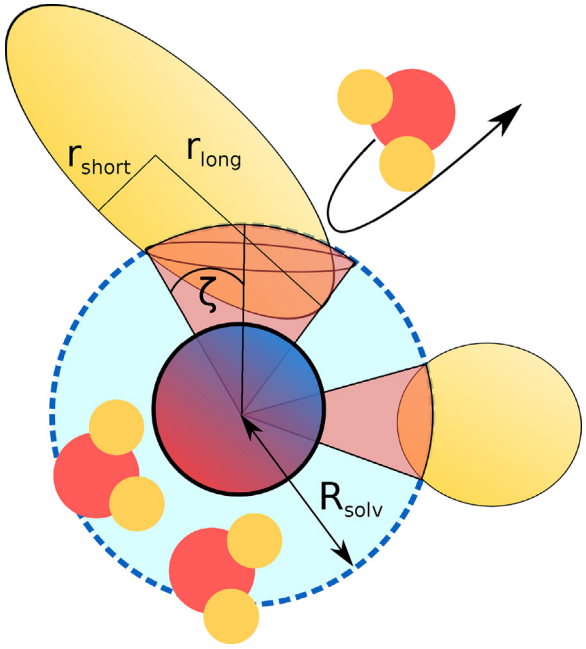

Fig. 4. Schematic illustration of volume of the first hydration sphere of peptide group (circle with dipole moment indicated as gradient from blue to red) occupied by water (blue) and lipidic environment side-chain (orange). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

whereas side-chains are ellipsoids of revolution about their longer axes. It should be noted that the C^{α} atoms are not center of interaction and assist only in geometry definition. The effective energy function originates from the restricted free energy (RFE) or potential of mean force (PMF) of the conformational ensemble restricted to coarse-grained conformation defined by C^{α} and SC atoms and is expressed by (Eq. (8)):

$$\begin{aligned} U = {} & w_{SC}\sum_{i<j} U_{SC_iSC_j} + w_{SCp}\sum_{i\neq j} U_{SC_ip_j} + w_{pp}^{VDW}\sum_{i<j-1} U_{p_ip_j}^{VDW} + w_{pp}^{el} f_2(T)\sum_{i<j-1} U_{p_ip_j}^{el} \\ & + w_{tor} f_2(T)\sum_i U_{tor}(\gamma_i) + w_{tord} f_3(T)\sum_i U_{tord}(\gamma_i,\gamma_{i+1}) \\ & + w_b\sum_i U_b(\theta_i) + w_{rot}\sum_i U_{rot}(\alpha_{SC_i},\beta_{SC_i}) + w_{bond}\sum_i U_{bond}(d_i) \\ & + w_{corr}^{(3)} f_3(T) U_{corr}^{(3)} + w_{corr}^{(4)} f_4(T) U_{corr}^{(4)} + w_{turn}^{(3)} f_3(T) U_{turn}^{(3)} + w_{turn}^{(4)} f_4(T) U_{turn}^{(4)} \\ & + w_{SC-corr} f_2(T)\sum_{m=1}^{3}\sum_i U_{SC-corr}(\tau_i^{(m)}) \end{aligned} \tag{8}$$

where the U's are energy terms, θ_i is the backbone virtual-bond angle between three consecutive C^{α} atoms, γ_i is the backbone virtual-bond-dihedral angle between four consecutive C^{α} atoms, α_i and β_i are the angles defining the location of the center of the united side chain of residue i (Fig. 3) with respect to $Ci-1_{\alpha}$, Ci_{α} and $Ci+1_{\alpha}$ plane, d_i is the length of the ith virtual bond, which is either a $C^{\alpha}\cdots C^{\alpha}$ virtual bond or $C^{\alpha}\cdots SC$ virtual bond, τ is backbone–side-chain virtual-bond-dihedral angle.

Each energy term is multiplied by an appropriate weight, w_x, and the terms corresponding to factors of order higher than 1 are additionally multiplied by the respective temperature factors that were introduced in earlier work [31] and which reflect the dependence of the first generalized-cumulant term in those factors on temperature, as discussed in Refs. [31,37]. The factors f_n are defined by (Eq. (9)).

$$f_n(T) = \frac{\ln\left[\exp(1)+\exp(-1)\right]}{\ln\left\{\exp\left[\left(T/T_{\circ}\right)^{n-1}\right]+\exp\left[-\left(T/T_{\circ}\right)^{n-1}\right]\right\}} \tag{9}$$

where $T_{\circ}=300$ K.

The term $U_{SC_iSC_j}$ represents the mean free energy of the hydrophobic (hydrophilic) interactions between the side chains, which implicitly contains the contributions from the interactions of the side chain with the solvent. The term $U_{SC_ip_j}$ denotes the excluded-volume potential of the side-chain – peptide-group interactions. The peptide-group interaction potential is split into two parts: the Lennard–Jones interaction energy between peptide-group centers ($U_{p_ip_j}^{VDW}$) and the average electrostatic energy between peptide-group dipoles ($U_{p_ip_j}^{el}$); the second of these terms accounts for the tendency to form backbone hydrogen bonds between peptide groups p_i and p_j. The terms U_{tor}, U_{tord}, U_b, U_{rot}, and U_{bond} are the virtual-bond-dihedral angle torsional terms, virtual-bond dihedral angle double-torsional terms, virtual-bond angle bending terms, side-chain rotamer, and virtual-bond-deformation terms; these terms account for the local properties of the polypeptide chain. The terms $U_{corr}^{(m)}$ represent correlation or multibody contributions from the coupling between backbone-local and backbone-electrostatic interactions, and the terms $U_{turn}^{(m)}$ are correlation contributions involving m consecutive peptide groups; they are, therefore, termed turn contributions. The multibody terms are indispensable for reproduction of regular α-helical and β-sheet structures [30,38,39]. One of the physical meaning of the correlation potentials is energetically favoring of multiple dipole alignment [30]. The $U_{SC-corr}$ terms are local physics-based side-chain backbone correlation potentials dependent only on torsional potential τ [40].

In the UNRES force field there are five energy terms which incorporate electrostatic interaction between peptide groups: $U_{p_ip_j}^{el}$,

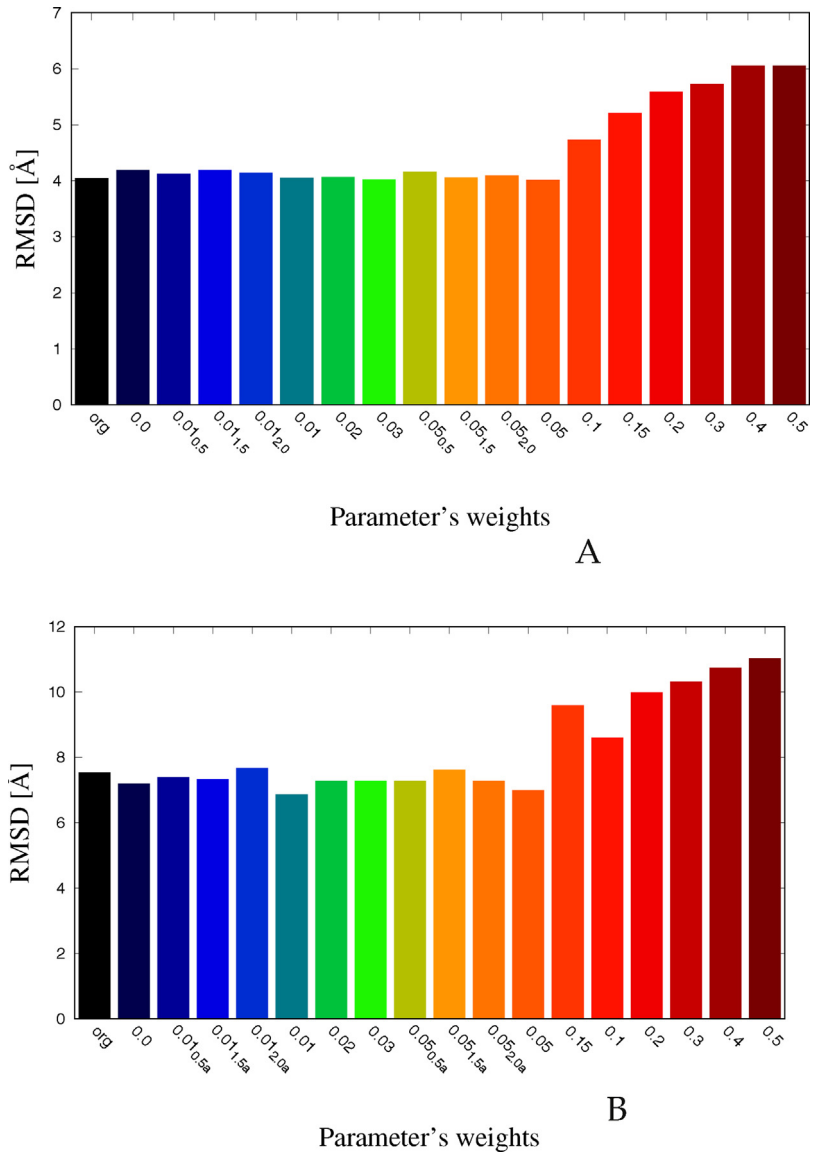

Fig. 5. C^{α} RMSD of top structure (A) and cluster with lowest C^{α} RMSD (B) obtained during MREMD simulation averaged over 8 test proteins. The org denotes the original UNRES force field, shielding weight 0.0 denotes UNRES force field with no shielding effect but with the temperature dependence of $U_{p_i p_j}^{VDW}$, $U_{SC_i SC_j}$, $U_{SC_i p_j}$, the subscript under shielding weight denotes the smoothing function (σ) value, if not stated value of 1.0 Å was used.

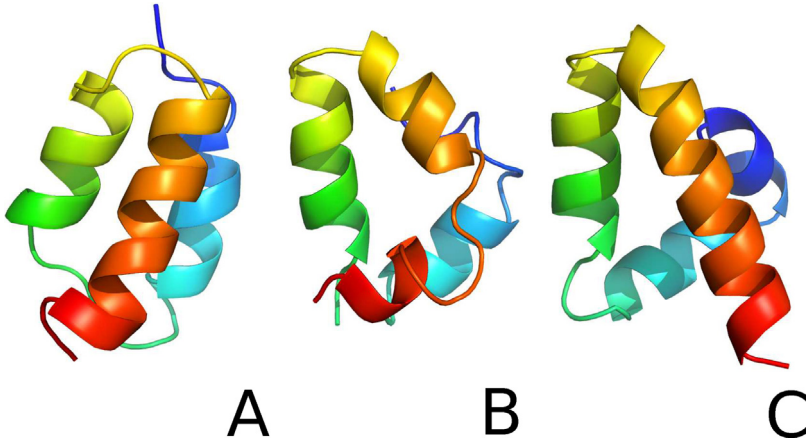

Fig. 6. Structure of 1GAB protein (A) experimental, (B) closest to experimental structure cluster for simulation with shielding weight 0.05 and (C) closest to experimental structure cluster for simulation without shielding effect.

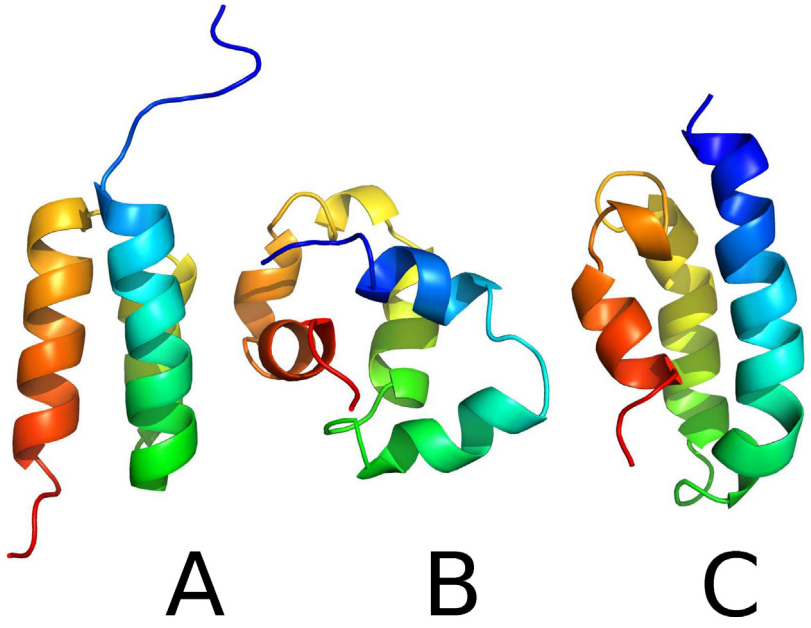

Fig. 7. Structure of 2JWS protein (A) experimental, (B) closest to experimental structure cluster for simulation with shielding weight 0.05 and (C) closest to experimental structure cluster for simulation without shielding effect.

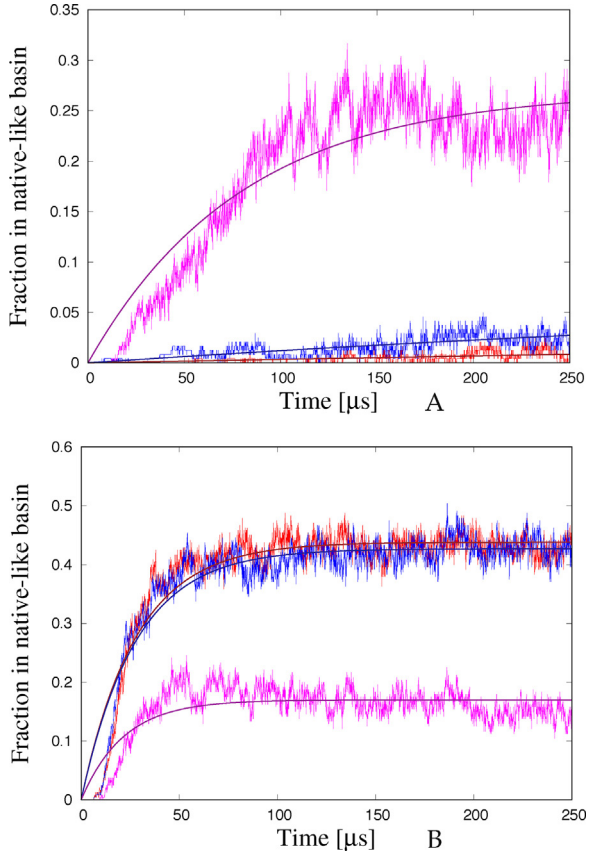

Fig. 8. Plots of fraction of native-like structures against time for (A) 1GAB and (B) 1BDD for different weights of w_{shield}: 0.0 (light red), 0.01 (light blue) and 0.05 (light purple) used, and first-order kinetic plots for fitted parameters 2 for different weights of w_{shield}: 0.0 (dark red), 0.01 (dark blue) and 0.05 (dark purple) used. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
C^{α} RMSD for test proteins obtained with original UNRES force field (org), with w_{shield} 0.0 – no shielding effect but with temperature dependence of energy terms $U_{p_ip_j}^{VDW}$, $U_{SC_iSC_j}$, $U_{SC_ip_j}$, w_{shield} 0.05 and C^{α} RMSD difference between no shielding effect and w_{shield} 0.05. Δ is difference between C^{α} RMSD obtained for best cluster for shielding $w_{shield} = 0.05$ and $w_{shield} = 0.0$.

Protein PDB code	org	0.0	0.05	Δ
1BDD	4.19	4.20	4.22	0.02
1CLB	9.07	6.51	6.48	−0.03
1E0G	7.63	7.29	6.31	−0.98
1E0L	8.96	8.54	7.06	−1.48
1GAB	6.33	6.18	4.84	−1.34
1KOY	6.57	6.60	6.67	0.07
2JWS	6.76	6.14	8.44	2.30
2LGN	10.73	12.11	11.92	−0.19

Table 2
Kinetics parameters, A_0 equilibrium fraction and k reaction rate constant for 1BDD and 1GAB with different shielding weights. In brackets fitting errors are shown.

	1BDD		1GAB	
w_{shield}	A_0	k [ms]	A_0	k [ms]
0.00	0.438 (0.001)	35.76 (0.013)	0.261 (0.893)	0.13 (0.023)
0.01	0.428 (0.001)	35.92 (0.013)	0.094 (0.015)	1.37 (0.012)
0.05	0.170 (0.001)	46.38 (0.041)	0.270 (0.001)	12.49 (0.008)

$U_{corr}^{(3)}$, $U_{corr}^{(4)}$, $U_{turn}^{(3)}$, $U_{turn}^{(4)}$. It should be noted that $U_{p_ip_j}^{el}$ and $U_{corr}^{(4)}$ are proportional to squares of the dipole moments of the interacting peptide groups and, therefore, the factors [$fac(i)$ and $fac(j)$] in equation (6) are squared.

Because all shielding-effect dependent terms are also temperature dependent, their introduction destroys the balance between the terms of the force field and, consequently, requires the introduction of the temperature dependence of the $U_{p_ip_j}^{VDW}$, $U_{SC_iSC_j}$ and $U_{SC_ip_j}$ energy terms. Those terms had to be scaled with $f_2(T)$ to prevent the collapse of a side-chain on the peptide group which makes a contact with it. Therefore two reference force field were used, without the temperature dependence of the $U_{p_ip_j}^{VDW}$, $U_{SC_iSC_j}$, $U_{SC_ip_j}$ terms (termed org) and with the temperature dependence but without shielding effect ($w_{shield} = 0.0$).

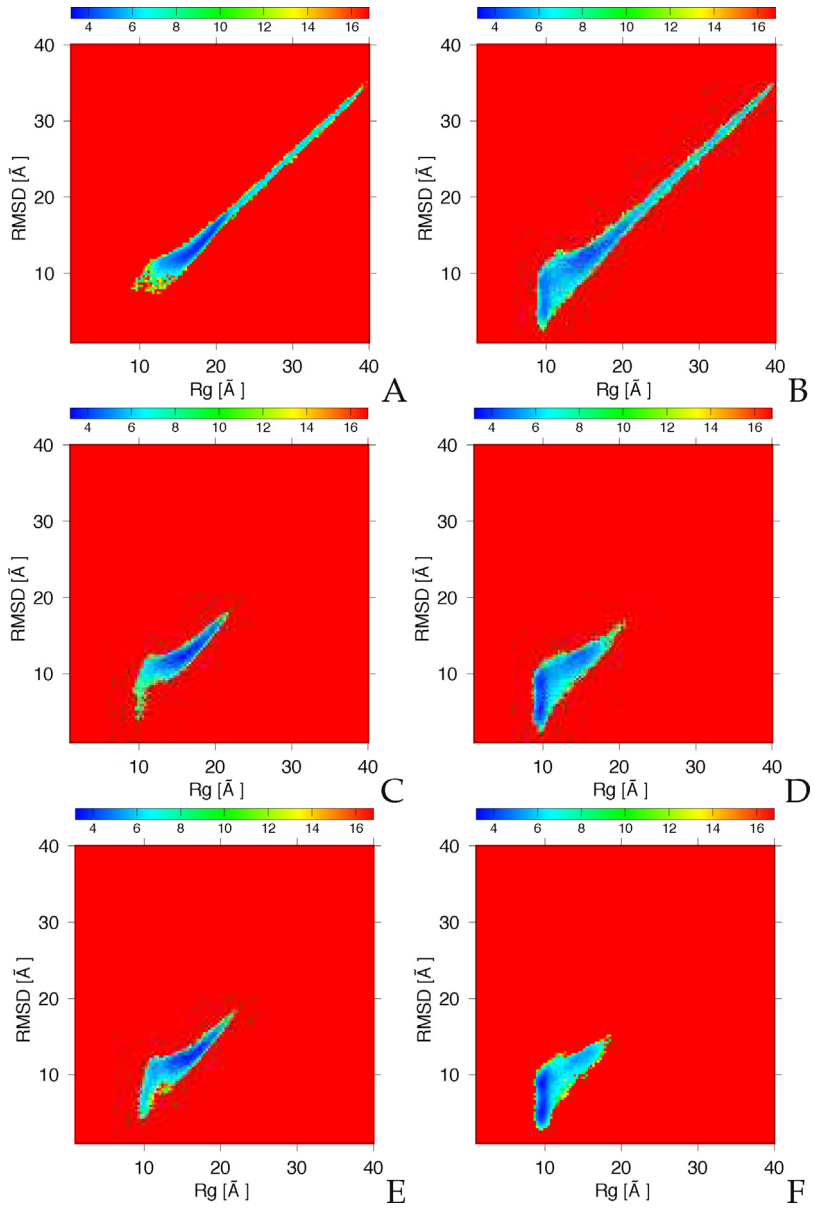

Fig. 9. Non-equilibrium free energy (kcal/mol, color bar above each panel) landscape changes over time for 1GAB protein. (A) and (B) correspond to 0-1mln molecular dynamics time steps, (C) and (D) correspond to 1–2 mln molecular dynamics time steps, (E) and (F) correspond to 2–3 mln molecular dynamics time steps. (A), (C) and (E) free energy landscape for simulations without shielding effect; (B), (D) and (F) free energy landscape for simulations with shielding effect included, $w_{shield} = 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To calculate the fraction of the first hydration sphere occupied by the ellipsoid side-chain ($frac_V$) a spherical sector (Fig. 4) approximation was used (Eq. (10)):

$$frac_V = \frac{V_{SC}}{V_{sphere}} \tag{10}$$

where V_{sphere} is the volume of the first hydration sphere of a peptide group from which the volume of the peptide group has been subtracted $\left(V_{sphere} = \frac{4}{3}\pi(R_{solv}^3 - R_{pep}^3)\right)$ where R_{pep} is radius of peptide group and R_{solv} is radius of hydration sphere of peptide group [41]. V_{SC} is the volume of a spherical sector given by (Eq. (11)):

$$V_{SC} = \frac{2}{3}\pi R_{solv}^3(1 - \cos\zeta)\left(scale\frac{R_{solv} - h + \sigma}{\sigma}\right) \tag{11}$$

where ζ is half cone angle (Fig. 4); $\tan\zeta$ is approximated

$$\tan\zeta = \frac{\sqrt{r_{short}(\sin\alpha(r_{long} - r_{short}) + r_{short})}}{h} \tag{12}$$

where r_{short} is the short axis of the ellipsoid, r_{long} is the long axis of the ellipsoid, α is angle between the vector connecting the side-chain center of mass with the peptide group center and the vector of the long axis of the ellipsoid, h is the distance between the side-chain center of mass and the peptide group center. The scaling function (*scale*) is used to ensure that side-chains that are too far from the peptide group do not contribute to the shielding effect as well as to smooth the gradient of the function. The scale function is given by (Eq. (13)):

$$scale = \begin{cases} 0 & \text{if } h > R_{solv} + \sigma \\ 3\lambda^2 - 2\lambda^3 & \text{if } R_{solv} < h < R_{solv} + \sigma \\ 1 & \text{if } h < R_{solv} \end{cases} \tag{13}$$

where σ is the user defined smoothing function quickness and $\lambda = \frac{R_{solv} - h + \sigma}{\sigma}$

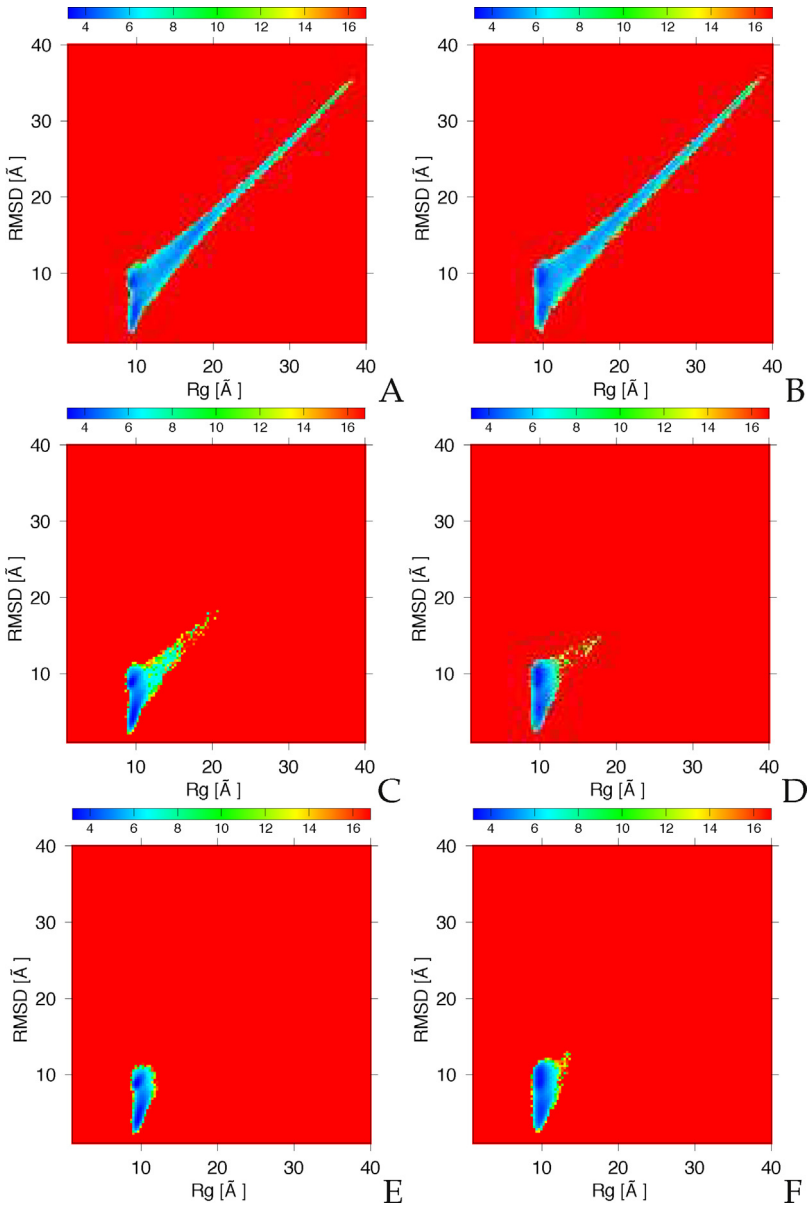

Fig. 10. Non-equilibrium free energy (kcal/mol, color bar above each panel) landscape changes over time for 1BDD protein. (A) and (B) correspond to 0–1 mln molecular dynamics time steps, (C) and (D) correspond to 1–2 mln molecular dynamics time steps, (E) and (F) correspond to 2–3 mln molecular dynamics time steps. (A), (C) and (E) free energy landscape for simulations without shielding effect; (B), (D) and (F) free energy landscape for simulations with shielding effect included, $w_{shield} = 0.05$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The set of energy-term weights was determined by force-field calibration to reproduce the structure and folding thermodynamics of selected training proteins [42]: LysM domain from *E. coli* (PDB code: 1E0G) [43], the Fbp28 WW domain from *Mus musculus* (PDB code: 1E0L) [44], the albumin-binding GA module (PDB code: 1GAB) [45], the IgG-binding domain from streptococcal protein G (PDB code: 1IGD) [46]. Moreover, two peptides were also used to force-field calibration, namely the tryptophan cage (PDB code: 1L2Y) [47] and the tryptophan zipper 2 (PDB code: 1LE1) [48].

2.3. Effect of the shielding term on protein folding

To study the influence of the shielding effect on the force-field ability to fold proteins a set of 50 mln step simulations with a molecular dynamics method [49,50] and its multiplexed replica exchange extension [51] (MREMD) were performed. MREMD simulations were performed for a set temperatures: 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 385 and 400 K. The Berendsen thermostat [52] with the coupling parameter τ = 48.9 fs was used to maintain the constant temperature. In each temperature two trajectories were performed. A similar to previous [53] set of test proteins: 1BDD, 1CLB, 1E0G, 1E0L, 1GAB, 1KOY, 2JWS, 2LGN was used. All of these test proteins are single-chain proteins, however the UNRES force field with shielding effect can be also applied to study protein-protein interactions. Afterward, the results of obtained MREMD simulations were processed with the Weighted Histogram Analysis Method (WHAM), which was implemented in UNRES and described in the earlier work [31], to determine the heat-capacity profiles, conformational ensembles and average properties of the ensemble at any desired temperature. Finally, the cluster analysis was carried out at 280 K, by means of Ward's minimum variance method [54] to obtain 5 clusters. A set of different shielding weights was tested: 0.01, 0.02, 0.03, 0.05, 0.1,

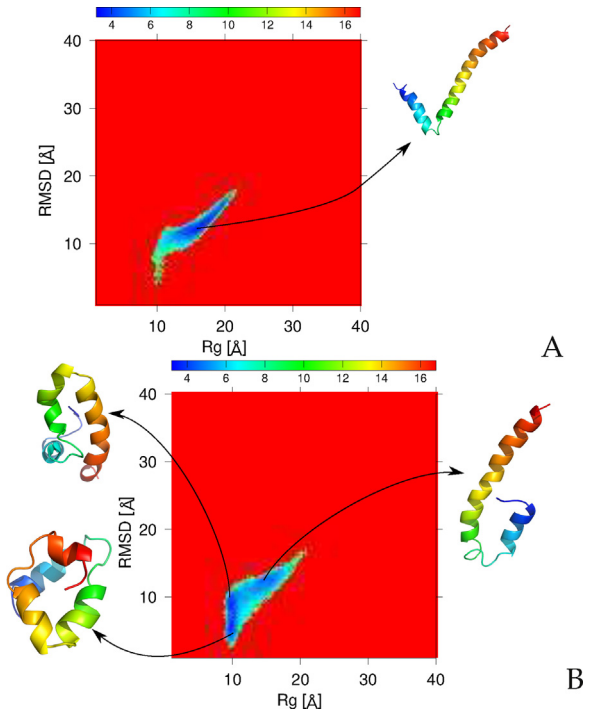

Fig. 11. Free-energy (kcal/mol, color bar above each panel) landscape for time frame 1–2 mln molecular dynamics time steps for 1GAB protein for simulations (A) without shielding effect; (B) with shielding effect included, $w_{shield} = 0.05$. On the side the most frequently occurring structures are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

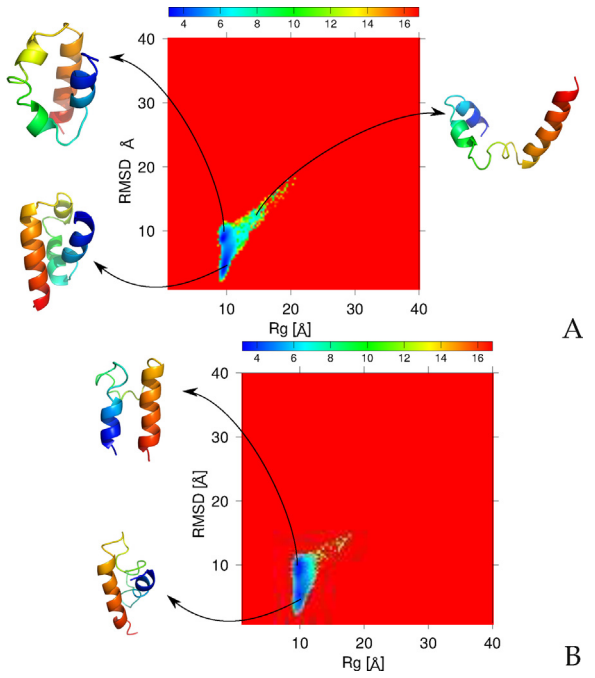

Fig. 12. Free-energy (kcal/mol, color bar above each panel) landscape for time frame 1–2 mln molecular dynamics time steps for 1BDD protein for simulations (A) without shielding effect; (B) with shielding effect included, $w_{shield} = 0.05$. On the side the most frequently occurring structures are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

0.15, 0.2, 0.3, 0.4, 0.5 with a smoothing factor (σ) set to 1 Å. Additionally, for weights 0.01 and 0.05 the influence of the smoothing factor (σ): 0.5, 1.5 and 2.0 Å on the foldability was tested.

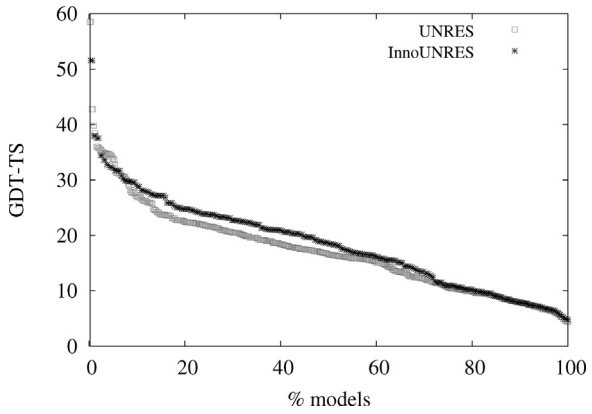

Fig. 13. Comparison of InnoUNRES (black) with UNRES (gray) CASP12 performance as a function of GDT_TS score against fraction of structures.

2.4. Influence of shielding effect on protein folding kinetics

To study the influence of the reduced hydration on folding of proteins the albumin-binding GA module (PDB code: 1GAB), and *Staphylococcus aureus* protein A, immunoglobulin-binding B domain (PDB code: 1BDD) were used. For each system three w_{shield} values were used: 0.01, 0.05 and 0.00 which was used as a reference. For each system 240 canonical molecular dynamics trajectories each 250 μs long (real time after taking the UNRES speed up into consideration) were performed. Simulations were performed in 300 K with the Langevin thermostat. Structures were divided into a native like and a non-native like basin with C^{α} the root-mean-square deviation (C^{α} RMSD) criterion previously used [55] (5.5 Å). Afterward, the first order kinetics equation was fitted to the obtained results (Eq. (14)):

$$f(t) = A_0 * (1 - \exp(-kt)) \tag{14}$$

where $f(t)$ is the fraction of structures in the native-like basin at a given time t, A_0 the value of fractions in the native like basin in the equilibrium, k is the reaction rate constant.

For 0.0 and 0.05 shielding weight the free energy landscape analysis was performed. For the each simulation, structures were binned with regards to C^{α} RMSD and radius of gyration (Rg). The simulations were split into time frames and the analysis was performed for 0–1 mln (0–4.9 μs), 1–2 mln (4.9–9.8 μs) and 2–3 (9.8–14.7 μs) molecular dynamics time steps. The 0.1 Å bin size was used for both variables. The free energy was calculated with $\Delta G_{ij} = -RT\, ln p_{ij}$ where R is the universal gas constant T is the temperature of simulation and p_{ij} is the probability of the occurrence of the structure in the ith C^{α} RMSD and the jth Rg bin. Additionally, for the time frame 1–2 mln molecular dynamics time steps a clustering analysis was performed, with use of the Ward's minimum variance method [54].

2.5. Performance of the modified UNRES force field including shielding effect in CASP12 experiment

UNRES force field was tested in many editions of The Critical Assessment of Protein Structure Prediction (CASP) experiments [56–58], including in the latest CASP12 edition. Consequently, we assessed the performance of the modified version of UNRES developed in this work. It was compared with that of the original force field.

The prediction procedure followed by our group during the CASP12 exercise consisted of the following four steps [58].

In the first step, MREMD [59,60] simulations of the target considered were carried out. Each of the MREMD simulations was run

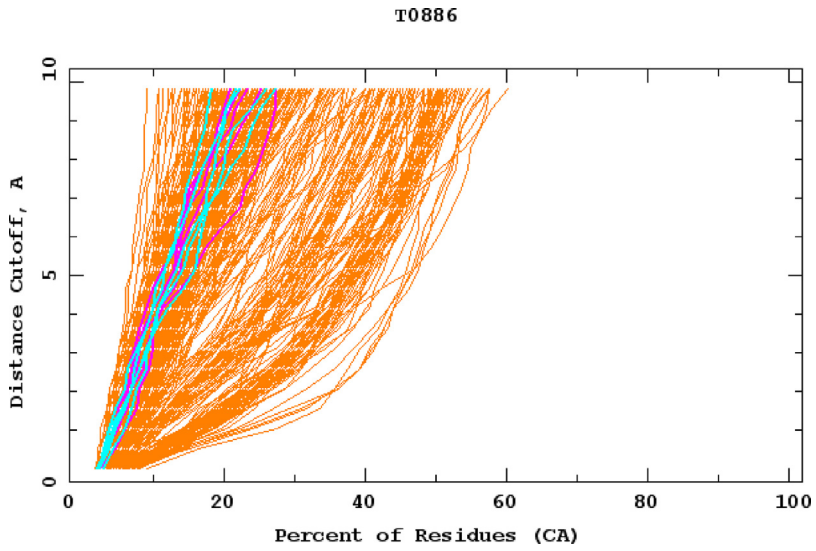

Fig. 14. The GDT_TS plots of T0886 target. Blue lines corresponds to models from Unres group, magenta lines to models from InnoUnres group and the other groups are shown as orange lines. The diagram was taken from the official CASP12 website at http://www.predictioncenter.org/casp12. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

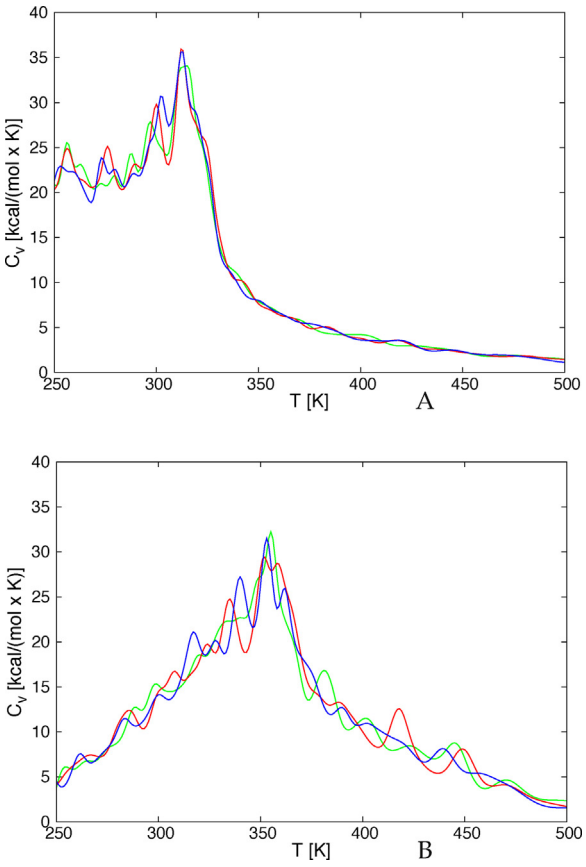

Fig. 15. Plots of heat capacity vs temperature for simulations of T0886 target in (a) Unres group and (b) InnoUnres group.

at 36 temperatures with two trajectories per temperature, providing a total of 72 trajectories. Weak restraints were imposed on the virtual-bond-dihedral angle γ and virtual-valence-bond angle θ, based on the results of secondary-structure predictions by PSIPRED [61], to speed-up the calculations. Typically, 20 million to 30 million MD steps with length of 4.89 fs [49] (0.1–0.14 μs total UNRES simulation time per trajectory, which corresponds to about 0.1–0.14 ms because of UNRES time-scale extension resulting, in turn, from averaging out the fast degrees of freedom [50,62]) were run for the targets considered. Replicas were exchanged every 10,000 MD time steps. The Berendsen thermostat [52] with the coupling parameter $\tau = 48.9$ fs was used to maintain constant temperature. The adaptive multiple time step (A-MTS) algorithm [63] was used to integrate the equations of motion. The simulations were run with the parallelized UNRES code [64] available at www.unres.pl.

In the second step of the procedure, the results of MREMD simulations were processed with WHAM, to determine the heat-capacity profiles and conformational ensembles at any desired temperature. For each system, the heat-capacity profile was analyzed and a temperature of $T \approx T_m - 10$ K was selected to analyze conformational ensembles, where T_m is the temperature of the main heat-capacity peak (the "melting temperature").

In the third step, cluster analysis was carried out at the selected temperatures, by means of Ward's minimum variance method [54]. The C^α RMSD was used as a measure of the distance between conformations. For each protein, the C^α RMSD cutoff was selected as a compromise between a small number of families (ideally five, which was the number of models that could have been submitted for each target) and grouping similar conformations in a given family. The clusters were ranked according to decreasing probabilities, which were computed from the probabilities of their component conformations calculated by WHAM [31]. For each cluster, the conformation closest to the average over the cluster was considered a representative of the whole cluster. The representatives of the five top clusters were selected as prediction candidates and ranked according to decreasing probabilities of the clusters. In the last step, the final conformations were converted to all-atom structures by physics-based method [65,66]. After conversion to the CASP format, the models were submitted.

The results were analyzed with use of the Global Distance Test-Total Score (GDT_TS) scoring function [67] defined as:

$$GDT\_TS = \frac{d_1 + d_2 + d_4 + d_8}{4} \tag{15}$$

where d_x is fraction of C^α atoms within x Å from native structure after superposition.

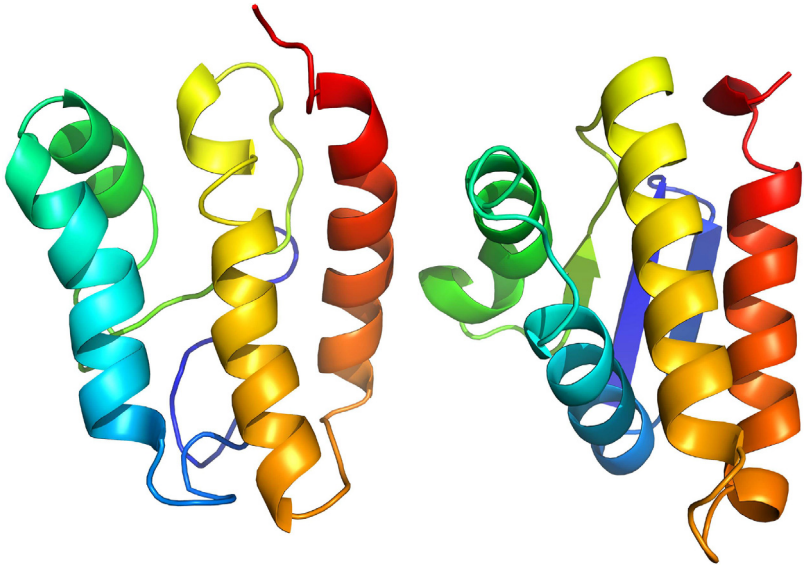

Fig. 16. The best InnoUNRES model (model 3) of target T0869_D1 (left structure) and the experimental structure (right structure; PDB code: 5J4A); in cartoon representation color-coded from N-terminus (blue) to C-terminus (red). The GDT_TS plot for this target is shown in Fig. 17A. The drawings of the structures were produced with PYMOL (www.pymol.org). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.6. Performance of maximum-likelihood optimized UNRES force field with shielding effect

Finally, the influence of the shielding effect was studied with the recently published maximum-likelihood calibrated UNRES force field [68]. The MREMD simulations details and set of proteins were the same as earlier: 1BDD, 1CLB, 1E0G, 1E0L, 1GAB, 1KOY, 2JWS, 2LGN. After MREMD simulations WHAM analysis was performed. For each protein and each w_{shield} RMSD as a function of temperature analysis was performed. For clustering, temperature was chosen where average RMSD(T) function obtained minimum value. Cut-off was set to always obtain five clusters.

Additionally, structures from simulations were sorted with the respect to RMSD from the native structure. The structures with lowest RMSD, 100th RMSD and 10000th RMSD from simulation were analyzed.

3. Results and discussion

3.1. Study of shielding effect on protein folding

As can be seen from Fig. 5B, the introduction of the shielding effect results in decreasing the RMSD for w_{shield} = 0.01 and w_{shield} = 0.05 and smoothing function (Eq. (13)) σ = 1 Å. This improvement, however, is very small. On the other hand, the best structure obtained during simulation have higher RMSD when the shielding effect is included (Fig. 5A). Hence, the shielding effect narrow random search of structures during simulations. This indicates that shielding effect drives toward native-like structures, however narrows conformational space search due to favoring of collapsed structure. This might suggest that the shielding effect leads to overemphasis of the collapsed structure. On average radius of gyration at 300 K for the force field with shielding weight 0.05 is 0.55 Å smaller then without shielding effect. Additional analysis show that the biggest improvement is in case of β-sheet structure or in structures with large fraction of *beta*-sheet (Table 1). For helical proteins some of the calculated structures are worse and some become better compared to those obtained with the original UNRES (Table 1 and Figs. 6 and 7). For example, the calculated structure of 1GAB is improved but for 2JWS, when shielding effect is introduced, a too compact structure is obtained (7B). A long unstructured fragment collapse leading to distrurbed structure.

3.2. Influence of shielding effect on protein folding kinetics

Kinetic plots for 1GAB and 1BDD folding are shown in Fig. 8 and the fitted parameters for first-order kinetic equation are shown in Table 2. As can be seen the introduction of shielding effect speeds up folding of proteins (k), which is not related to the influence on the folded fraction (A_0). For the 1BDD protein the speed up is by 30%, however for the 1GAB protein the speed up is by two order of magnitude. The folding of 1GAB is much slower without including the shielding effect ($w_{shield} = 0$) and the fitted fraction of folded protein (A_0) is charged with a significant error (Table 2).

To analyze the influence of the shielding effect on the protein folding, free-energy landscapes were analyzed. The differences between the force field with and without including the shielding effect become more pronounced when the time evolution of the free-energy landscapes is analyzed (Figs. 9 and 10). For 1GAB Rg decreases significantly faster with the shielding effect for the most frequently occurring structure, while for 1BDD the rate of Rg decreasing (collapsing) is similar with and without shielding effect but still is faster with the shielding effect (Figs. 10C and D).

The obtained clusters of conformations for time frame 4.9–9.8 μs (1–2 mln molecular dynamics time steps) revealed that, for the 1GAB without shielding effect, the dominant structure is helix-turn helix (Fig. 11A), whereas when shielding effect is incorporated in Hamiltonian this structure still can be observed but folded structure is dominant one (Fig. 11B). Even more interesting results are observed for 1BBD protein (Fig. 12). When no shielding effect is present in simulation three major cluster for time frame 4.9–9.8 μs can be observed, namely the folded structure, the mirror image and the open structure (Fig. 12A). However, when shielding effect is taken into account the open and the mirror image structures are no longer observed. An additional cluster with structures corresponding to improper C- and N-terminal helices alignment arise. This suggests that C- and N-terminal helices interact at very early stage of 1BDD folding, and depending on their alignment, structure fold either to native or non-native one. Therefore, for the 1BDD, smaller fraction of native structure (A_0) can be observed (Table 2).

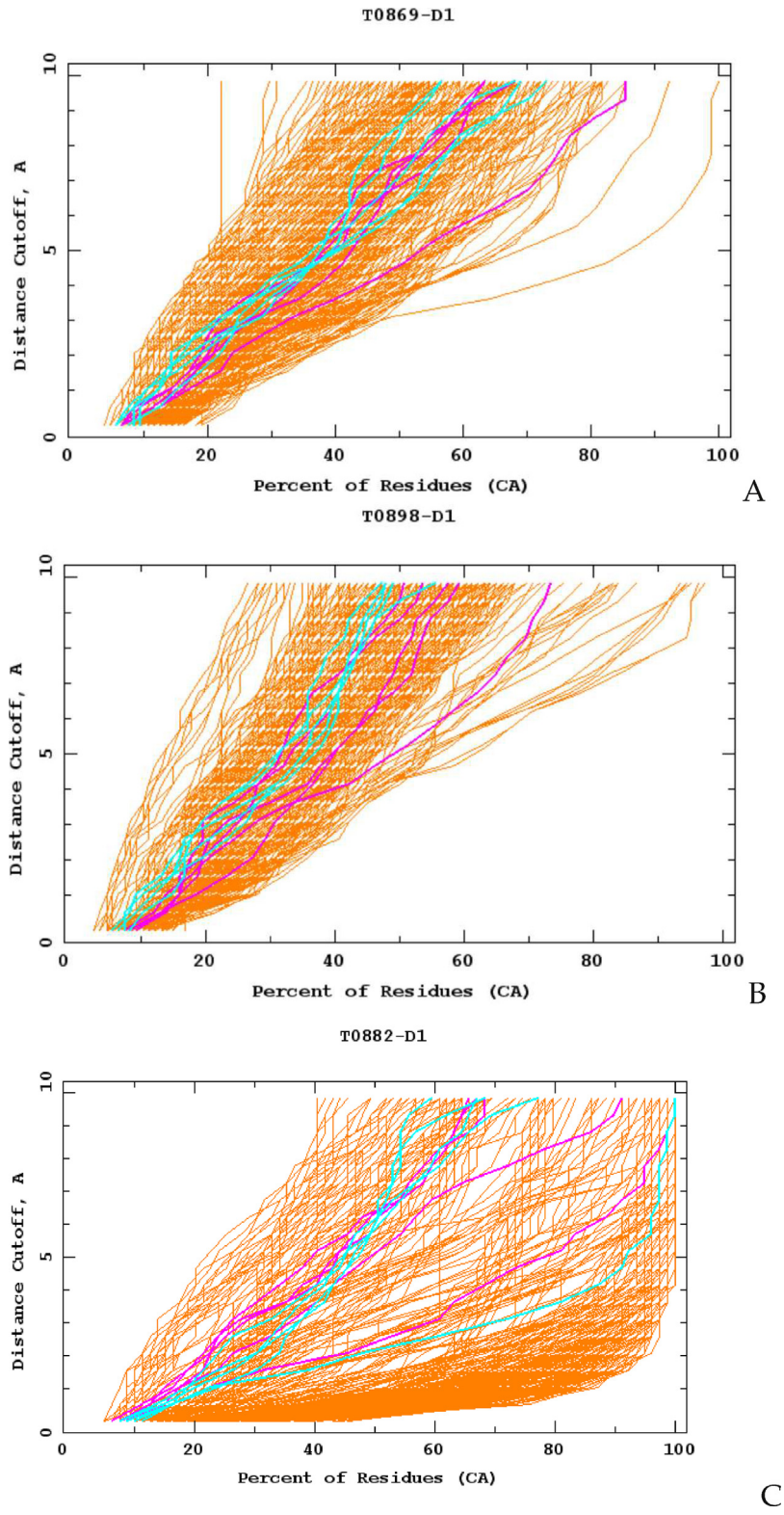

Fig. 17. The GDT_TS plots for targets T0869_D1 (A), T0898_D1 (B) and T0882_D1 (C). UNRES groups' models are marked by cyan lines and those corresponding to InnoUNRES groups' models are marked by magenta lines; models from the other groups shown as orange lines. The diagrams were taken from the official CASP12 website at http://www.predictioncenter.org/casp12. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

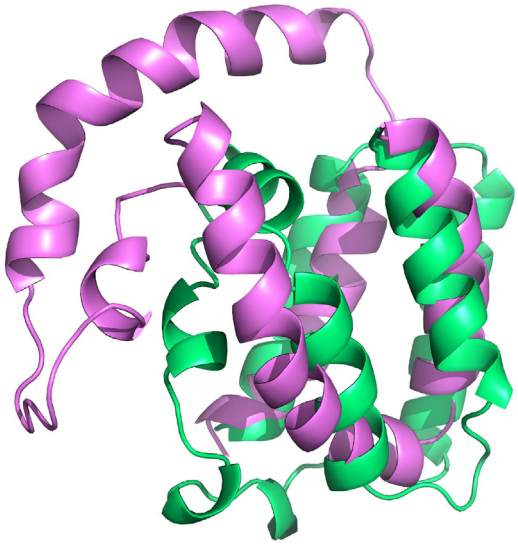

Fig. 18. The best InnoUNRES model (model 5) of target T0898_D1 (violet) and the experimental structure (green). The GDT_TS plot for this target is shown in Fig. 17B. The drawings of the structures were produced with PYMOL (www.pymol.org). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

This leads to conclusion that the shielding effect promotes collapsed structures which lead to faster folding of proteins. Moreover, it may also suggest that is promotes oligomerization.

3.3. Performance of the modified UNRES force field including shielding effect in CASP12 experiment

As the InnoUNRES group (group 145), we have submitted predictions for 35 of 64 targets available to human predictor groups in CASP12 exercises. The best predictions for our group were obtained for T0869_D1 and T0898_D1 targets; while the worst – for T0912 target which, however, is very large (624 residues) and, therefore, hardly tractable by physics-based methods. For the best InnoUNRES group model GDT_TS [67] was 5.43 (compared with 49.08 for the best model submitted to CASP – model 2 from group 384), but it was still better than for the best UNRES group model (GDT_TS = 5.26). While comparing all models from InnoUNRES group with those from the UNRES group, the model with the biggest difference in GDT_TS in favor to the UNRES group was target T0882_D1. In general, the InnoUNRES group have given better results compared with UNRES group predictions, however, this improvement was incremental (Fig. 13).

This improvement in force field performance occurred despite slower convergence. When T0886 is analyzed, revealing small improvement of the GDT_TS (Fig. 14), after 50 mln steps heat capacity curves for UNRES group (Fig. 15A) is almost fully converged. For InnoUNRES group after 50 mln simulation steps a heat capacity curve is not converged (Fig. 15B) as fluctuations occur. Results suggest that the introduction of shielding effect speeds up folding of protein especially at initial stages of MREMD simulations, but they converge slower.

3.3.1. Target T0869_D1

T0869_D1 is a 120 amino-acid residue CdiI immunity protein – a subunit of the CdiA-CT/CdiI complex (heterodimer, PDB code: 5J4A) [69]. From this target only the structure of the 3–106 fragment was solved experimentally and used in the official CASP12 assessment. The CdiI protein consists of a slightly curved three-stranded antiparallel β-sheets with four α-helices (Fig. 16 on right). Our model 3 of this target is close to the experimental structure and

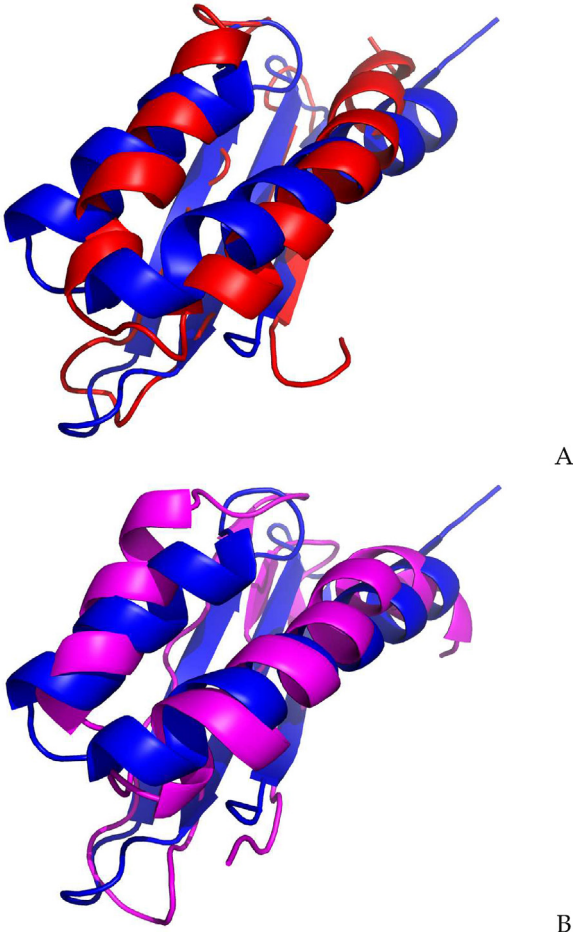

Fig. 19. (A) The best InnoUNRES model (model 5) of target T0882_D1 (red) and the experimental structure (blue). (B) The best UNRES model (model 4) of target T0882_D1 (magenta) and the experimental structure (blue). The GDT_TS plot for this target is shown in Fig. 17C. The drawings of the structures were produced with PYMOL (www.pymol.org). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

secondary-structure elements are correctly oriented in space. The model contains all of the helical elements of the protein (Fig. 16 on left). This structure is also comparable to those obtained by approaches that use knowledge-based input, in particular, to the best models of Zhang (model 4, group 004), Baker (model 3, group 247) and Lee (model 1, group 011) groups. The only noticeable difference is in the placement of the N-terminal β-sheets, which in the best Zhang, Baker and Lee models are placed between the helices, as opposed to our structure and the experimental one in which they are placed next to the helices. However, the β-sheets in our model are of worse quality. The overall C^{α} RMSD of our model 3 from the experimental structure is 8.03 Å compared with 9.86, 12.54 and 11.37 Å for the best Zhang, Baker and Lee models, respectively. These results confirmed our previous conclusions (from CASP11 exercises) [58], that the UNRES force field very often correctly predicts the overall topology of the proteins but lacks details. The GDT_TS value of our model 3 is 37.50, compared with 52.40 for the best model submitted to CASP12 (model 3 from group 173), 27.16 for UNRES group (model 4) (Fig. 17A), and 41.83, 37.50 and 39.90 for the best Zhang, Baker and Lee models, respectively.

3.3.2. Target T0898_D1

T0898_D1 is a 169 amino-acid residue target, for which only the section 4-109 was subjected to official CASP12 analysis. The experimental structure of the 4-109-residue section contains six

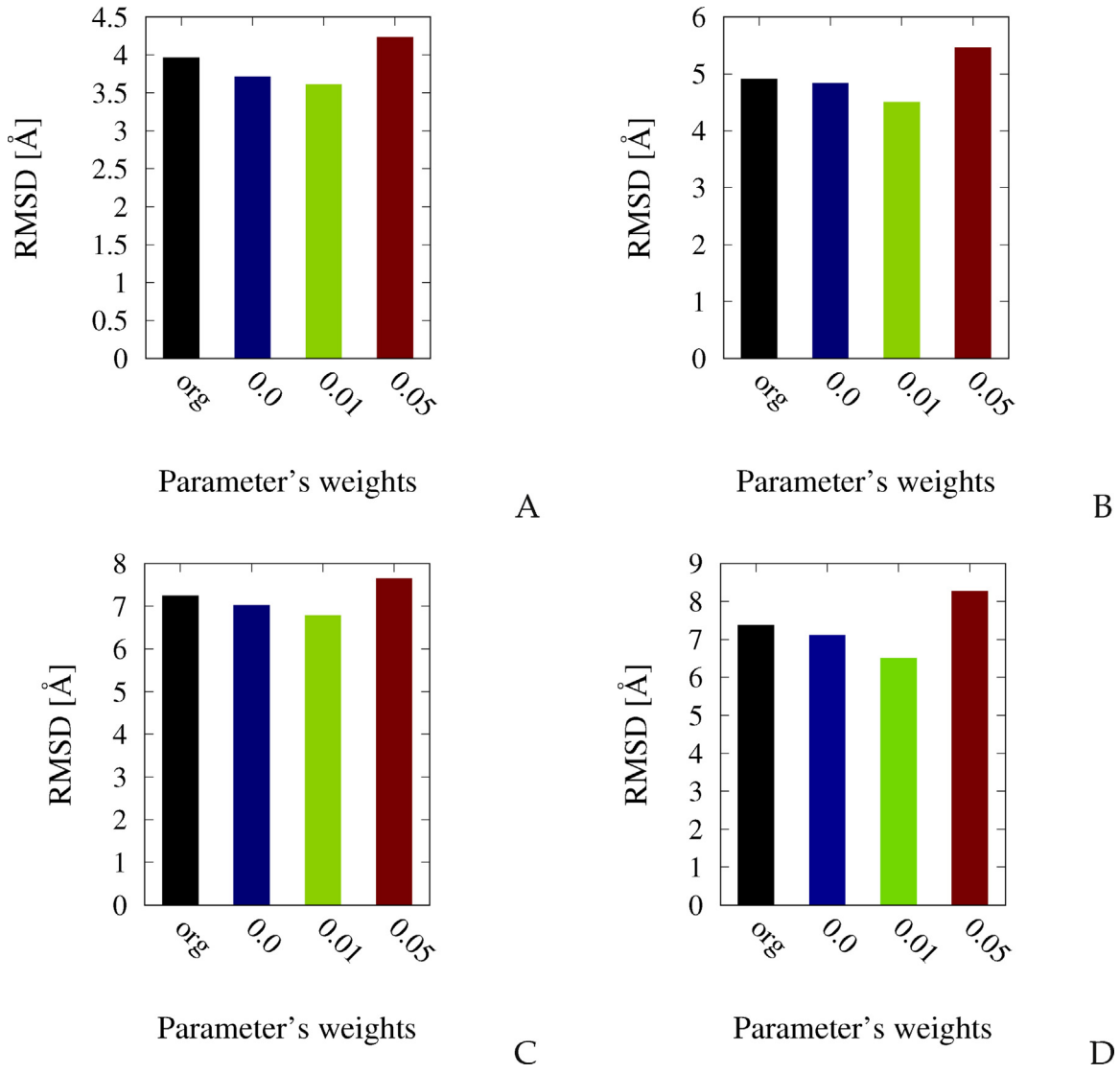

Fig. 20. C^{α} RMSD of top structure (A), top100 structure (B), top10000 structure and cluster with lowest C^{α} RMSD (D) obtained during MREMD simulation averaged over 8 test proteins. The org denotes the maximum-likelihood optimized UNRES force field, shielding weight 0.0 denotes UNRES force field with no shielding effect but with the temperature dependence of $U_{p_i p_j}^{VDW}$, $U_{SC_i SC_j}$, $U_{SC_i p_j}$.

α-helices, as shown in Fig. 18. Our model 5 of this target is close to the experimental structure with C^{α} RMSD equal to 11.75 Å and all helical fragments are correctly determined (Fig. 18). The only noticeable difference is in the placement of the fifth α-helix, which is slightly separated from the other helices. This difference occurs presumably because of a high flexibility of a region preceding the helix. On the contrary, the placement of the first two helices and the last one is in a very good agreement with the experimental structure. The GDT_TS value of this model (Fig. 17B) is 32.78, compared with the highest value of 42.69 obtained by group 498 for model 5 and 23.82 by UNRES group (model 1), and 41.27 (model 4), 33.73 (model 1) and 34.91 (model 4) for the best Zhang, Baker and Lee models, respectively.

3.3.3. Target T0882_D1

T0882_D1 is a 89 amino-acid residue target, for which only the section 5–83 was subjected to official CASP12 analysis. It is a fragment of a hypothetical domain in WNK1 (PBD code: 5G3Q). The experimental structure contains three-stranded antiparallel β-sheets with two α-helices, as shown in Fig. 19. Our model 5 of this target is close to the experimental structure with C^{α} RMSD equal to 4.54 Å. The noticeable difference is in the N-end placement and the location of the loop between the two α-helices. While comparing with the model 4 of the UNRES group, which has C^{α} RMSD equal to 3.74 Å we can observe that those two inaccuracies are better represented here. It should be noted that for InnoUNRES all 5 clusters have correct topology (three-stranded antiparallel β-sheets with two α-helices), wheares in case of UNRES predicted models only 2 have correct general fold. In case of 3 models in UNRES group two helices are merged into one long helix.

3.4. Performance of maximum-likelihood optimized UNRES force field with shielding effect

As can be seen in the Fig. 20 introduction of shielding effect improve RMSD of the ensemble of structures. The introduction of shielding ($w_{shield,maxlik} = 0.01$) effect lead on average to drop of RMSD of the structure with the lowest RMSD by 0.35 Å with respect to original force field (without shielding effect and without temperature dependence of energy terms $U_{p_i p_j}^{VDW}$, $U_{SC_i SC_j}$, $U_{SC_i p_j}$) and by 0.1 Å with respect to force field without shielding effect

Table 3
C^{α} RMSD for test proteins obtained with maximum-likelihood optimized UNRES force field (org), with $w_{shield,maxlik}$ 0.0 – no shielding effect but with temperature dependence of energy terms $U_{p_ip_j}^{VDW}$, $U_{SC_iSC_j}$, $U_{SC_ip_j}$, $w_{shield,maxlik}$ 0.01 and C^{α} RMSD difference between maximum-likelihood optimized force field and $w_{shield,maxlik}$ 0.01. Δ is difference between C^{α} RMSD obtained for best cluster for shielding $w_{shield,maxlik} =$ 0.01 and maximum-likelihood optimized force field.

Protein PDB code	org	0.0	0.05	Δ
1BDD	5.01	5.47	5.10	0.09
1CLB	8.11	6.31	6.04	−2.07
1E0G	7.42	7.88	5.47	−1.95
1E0L	8.49	7.95	6.71	−1.78
1GAB	6.21	4.16	4.35	−1.86
1KOY	6.72	6.73	6.45	−0.27
2JWS	6.52	6.73	8.54	2.02
2LGN	10.45	11.63	9.31	−1.14

(Fig. 20A). The 100th lowest RMSD structure for simulation with shielding effect ($w_{shield,maxlik} = 0.01$) has RMSD to native structure lower by 0.45 Å with respect to original force field (without shielding effect and without temperature dependence of energy terms $U_{p_ip_j}^{VDW}$, $U_{SC_iSC_j}$, $U_{SC_ip_j}$) and by 0.33 Å with respect to force field without shielding effect (Fig. 20B). The 10000th lowest RMSD structure for simulation with shielding effect ($w_{shield,maxlik} = 0.01$) has RMSD to native structure lower by 0.46 Å with respect to original force field (without shielding effect and without temperature dependence of energy terms $U_{p_ip_j}^{VDW}$, $U_{SC_iSC_j}$, $U_{SC_ip_j}$) and by 0.24 Å with respect to force field without shielding effect (Fig. 20C).

Moreover, the largest improvement is observed for structures after clustering. The structure from 5 clusters with the lowest RMSD for simulation with shielding effect ($w_{shield,maxlik} = 0.01$) has RMSD to native structure lower by 0.87 Å on average with respect to original force field (without shielding effect and without temperature dependence of energy terms $U_{p_ip_j}^{VDW}$, $U_{SC_iSC_j}$, $U_{SC_ip_j}$) and by 0.61 Å on average with respect to force without shielding effect (Fig. 20D). This is over 11% improvement of the force field with respect to RMSD of the proteins. This indicate that shielding effect should be introduced to force fields with well-balanced parameter set.

The detailed analysis of the clusters (Table 3) reveal that apart, from 2JWS protein, there is either similar result (1BDD and 1KOY) or a tremendous improvement. The inaccuracy for the 2JWS protein, as in case of original force field, is that the shielding effect drives the long N-terminal unstructured fragment toward helical one resulting in increase of RMSD. This suggests that the shielding effect leads to collapsing of long unstructured fragments.

4. Conclusions

In this work, a new method was proposed to treat peptide group shielding from the solvent, which is based on modifying the effective dielectric constant by the interaction sites which are close to the interacting peptide groups. This dielectric constant is applied to all terms related to the interactions between peptide-group dipoles, this includes the average dipole–dipole-interaction term ($U_{p_ip_j}$) and the correlation terms ($U_{corr}^{(3)}$, $U_{turn}^{(3)}$, $U_{corr}^{(4)}$, and $U_{turn}^{(4)}$). This correction was implemented in the UNited RESidue (UNRES) force field and was tested on set of 8 proteins that revealed that incorporation of the shielding effect improves performance of UNRES force field to fold proteins. Additionally, introduction of the shielding effect speeds up folding kinetics in molecular dynamics. Overall shielding effect lead to more compact structure but may also lead to occurrence of misfolded off-pathway traps. Performance of the modified UNRES force field was tested in blind prediction experiment CASP12 (The Critical Assessment of protein Structure Prediction) showing improved performance. A large improvement was observed for maximum-likelihood optimized force field indicating that shielding effect requires well-balanced force field to be fully functional. However, it was observed that disordered structures are driven to helical structures. To eliminate overstabilization of helical structures which should be unstructured a further recalibration with intrinsically disordered proteins in training set should be performed. Additionally, we are planning to introduce ions into UNRES force field, which would allow to study shielding effect in buffer like conditions. This work is currently underway.

Acknowledgements

This work supported by Sonata UMO-2015/17/D/ST4/00509 (to AKS), Preludium 2016/21/N/ST4/03154 (to AGL) form the National Science Centre (Poland). Computational resources were also provided by (a) the supercomputer resources at the Informatics Center of the Metropolitan Academic Network (IC MAN) in Gdańsk, and (b) computational resources at Interdisciplinary Center for Mathematical and Computer Modeling in Warsaw (ICM), grant GA65-20 (c) our 682-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk.

References

[1] G. Némethy, H.A. Scheraga, The structure of water and hydrophobic bonding in proteins. III. The thermodynamic properties of hydrophobic bonds in proteins, J. Chem. Phys. 66 (1962) 1773–1789.
[2] K.A. Dill, Dominant forces in protein folding, Biochemistry 29 (1990) 7133–7155.
[3] G.D. Rose, Hydrogen bonding, hydrophobicity, packing and protein folding, Annu. Rev. Biophys. Biomol. Struct. 22 (1993) 381–415.
[4] M. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu, A.E. García, Water determines the structure and dynamics of proteins, Chem. Rev. 116 (2016) 7673–7697.
[5] D.W. Bolen, G.D. Rose, Structure and energetics of the hydrogen-bonded backbone in protein folding, Annu. Rev. Biochem. 77 (2008) 339–362.
[6] F. Avbelj, R.L. Baldwin, Origin of the change in solvation enthalpy of the peptide group when neighboring peptide groups are added, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 3137–3141.
[7] O. Bignucolo, H.T.A. Leung, S. Grzesiek, S. Bernèche, Backbone hydration determines the folding signature of amino acid residues, J. Am. Chem. Soc. 137 (2015) 4300–4303.
[8] I.A. Kovacs, M.S. Szalay, P. Csermely, Water and molecular chaperones act as weak links of protein folding networks: energy landscape and punctuated equilibrium changes point towards a game theory of proteins, FEBS Lett. 579 (2005) 2254–2260.
[9] D. Nayar, C. Chakravarty, Free energy landscapes of alanine oligopeptides in rigid-body and hybrid water models, J. Phys. Chem. B 119 (2015) 11106–11120.
[10] D. Paschek, R. Day, A.E. García, Influence of water – protein hydrogen bonding on the stability of Trp-cage miniprotein. A comparison between the TIP3P and TIP4P-Ew water models, Phys. Chem. Chem. Phys. 13 (2011) 19840–19847.
[11] M. Gupta, D. Nayar, C. Chakravarty, S. Bandyopadhyay, Comparison of hydration behavior and conformational preferences of the Trp-cage mini-protein in different rigid-body water models, Phys. Chem. Chem. Phys. 18 (2016) 32796.
[12] F. Avbelj, J. Moult, Role of electrostatic screening in determining protein main chain conformational preferences, Biochemistry 34 (1995) 755–764.
[13] D. Nayar, C. Chakravarty, Sensitivity of local hydration behaviour and conformational preferences of peptides to choice of water model, Phys. Chem. Chem. Phys. 16 (2014) 10199–10213.
[14] F. Avbelj, L. Fele, Prediction of the three-dimensional structure of proteins using the electrostatic screening model and hierarchic condensation, Proteins: Struct. Funct. Bioinf. 31 (1998) 74–96.
[15] J.A. Vila, D.R. Ripoll, H.A. Scheraga, Physical reasons for the unusual alfa-helix stabilization afforded by charged or neutral polar residues in alanine-rich peptides, Proc. Natl. Acad. Sci. U. S. A. 97 (2000) 13075–13079.
[16] A.E. García, K.Y. Sanbonmatsu, Alfa-Helical stabilization by side chain shielding of backbone hydrogen bonds, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 2782–2787.
[17] T. Ghosh, S. Garde, A.E. García, Role of backbone hydration and salt-bridge formation in stability of alfa-helix in solution, Biophys. J. 85 (2003) 3187–3193.
[18] L. Monticelli, S.K. Kandasamy, X. Periole, R.G. Larson, D.P. Tieleman, S.J. Marrink, The MARTINI coarse-grained force field: extension to proteins, J. Chem. Theory Comput. 4 (2008) 819–834.
[19] Y. Chebaro, S. Pasquali, P. Derreumaux, The coarse-grained OPEP force field for non-amyloid and amyloid proteins, J. Phys. Chem. B 116 (2012) 8741–8752.
[20] M. Pasi, R. Lavery, N. Ceres, PaLaCe: a coarse-grain protein model for studying mechanical properties, J. Chem. Theory Comput. 9 (2013) 785–793.

[21] P. Kar, S.M. Gopal, Y.M. Cheng, A. Predeus, M. Feig, PRIMO: a transferable coarse-grained force field for proteins, J. Chem. Theory Comput. 9 (2013) 3769–3788.
[22] N. Basdevant, D. Borgist, T. Ha-Duong, Modeling protein–protein recognition in solution using the coarse-grained force field SCORPION, J. Chem. Theory Comput. 9 (2013) 803–813.
[23] A. Liwo, et al., A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions, J. Mol. Model. 20 (2014) 2306.
[24] S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A.E. Dawid, A. Kolinski, Coarse-grained protein models and their applications, Chem. Rev. 116 (2016) 7898–73936.
[25] T.B. Jones, Electromechanics of Particles, Cambridge University Press, 2005.
[26] F. Sterpone, P.H. Nguyen, M. Kalimeri, P. Derreumaux, Importance of the ion-pair interactions in the OPEP coarse-grained force field: parametrization and validation, J. Chem. Theory Comput. 9 (2013) 4574–4584.
[27] T. Simonson, C.L. Brooks, Charge screening and the dielectric constant of proteins: insights from molecular dynamics, J. Am. Chem. Soc. 118 (1996) 8452–8458.
[28] A. Liwo, M.R. Pincus, R.J. Wawak, S. Rackovsky, H.A. Scheraga, Prediction of protein conformation on the basis of a search for compact structures; test on avian pancreatic polypeptide, Protein Sci. 2 (1993) 1715–1731.
[29] A. Liwo, S. Ołdziej, M.R. Pincus, R.J. Wawak, S. Rackovsky, H.A. Scheraga, 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. 18 (1997) 849–873.
[30] A. Liwo, C. Czaplewski, J. Pillardy, H.A. Scheraga, 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 (2001) 2323–2347.
[31] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, H.A. Scheraga, 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 (2007) 260–285.
[32] A. Liwo, C. Czaplewski, S. Ołdziej, A.V. Rojas, R. Kaźmierkiewicz, M. Makowski, R.K. Murarka, H.A. Scheraga, in: G. Voth (Ed.), Coarse-Graining of Condensed Phase and Biomolecular Systems, CRC Press, 2008, pp. 1391–1411 (Chapter 8).
[33] U. Kozłowska, G.G. Maisuradze, A. Liwo, H.A. Scheraga, 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 (2010) 1154–1167.
[34] M. Makowski, A. Liwo, E. Sobolewski, H.A. Scheraga, 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 (2011) 6119–6129.
[35] M. Makowski, A. Liwo, H.A. Scheraga, 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 (2011) 6130–6137.
[36] A.K. Sieradzan, H.A. Scheraga, A. Liwo, 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. Theor. Comput. 8 (2012) 1334–1343.
[37] H. Shen, A. Liwo, H.A. Scheraga, 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 113 (2009) 8738–8744.
[38] A. Kolinski, J. Skolnick, Discretized model of proteins. I. Monte Carlo study of cooperativity in homopolypeptides, J. Chem. Phys. 97 (1992) 9412–9426.
[39] A. Liwo, R. Kaźmierkiewicz, C. Czaplewski, M. Groth, S. Ołdziej, R.J. Wawak, S. Rackovsky, M.R. Pincus, H.A. Scheraga, 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 (1998) 259–276.
[40] P. Krupa, A.K. Sieradzan, S. Rackovsky, M. Baranowski, S. Ołdziej, H.A. Scheraga, A. Liwo, C. Czaplewski, 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 (2013) 4620–4632.
[41] A. Liwo, S. Ołdziej, C. Czaplewski, U. Kozłowska, H.A. Scheraga, 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 (2004) 9421–9438.
[42] Y. He, Y. Xiao, A. Liwo, H.A. Scheraga, 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 (2009) 2127–2135.
[43] A. Bateman, M. Bycroft, The structure of a LysM domain from E-coli membrane-bound lytic murein transglycosylase D (MltD), J. Mol. Biol. 299 (2000) 1113–1119.
[44] M.J. Macias, V. Gervais, C. Civera, H. Oschkinat, Domains and design of a WW prototype, Nat. Struct. Biol. 7 (2000) 375–379.
[45] M.U. Johansson, M. de Château, M. Wikström, S. Forsén, T. Drakenberg, L. Björck, Solution structure of the albumin-binding GA module: a versatile bacterial protein domain, J. Mol. Biol. 266 (1997) 859–865.
[46] J.P. Derrick, D.B. Wigley, The third IGG-binding domain from streptococcal protein G: an analysis by X-ray crystallography of the structure alone and in a complex with FAB, J. Mol. Biol. 243 (1994) 906–918.
[47] J.W. Neidigh, R.M. Fesinmeyer, N.H. Andersen, Designing a 20-residue protein, Nat. Struct. Biol. 9 (2002) 425–430.
[48] A.G. Cochran, N.J. Skelton, M.A. Starovasnik, Tryptophan zippers: stable, monomeric beta-hairpins, Proc. Natl. Acad. Sci. U. S. A. 98 (2001) 5578–5583.
[49] M. Khalili, A. Liwo, F. Rakowski, P. Grochowski, H.A. Scheraga, 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 (2005) 13785–13797.
[50] M. Khalili, A. Liwo, A. Jagielska, H.A. Scheraga, 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 (2005) 13798–13810.
[51] C. Czaplewski, S. Kalinowski, A. Liwo, H.A. Scheraga, Application of multiplexing replica exchange molecular dynamics method to the UNRES force field: tests with α and $\alpha+\beta$ proteins, J. Chem. Theor. Comput. 5 (2009) 627–640.
[52] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[53] A.K. Sieradzan, P. Krupa, H.A. Scheraga, A. Liwo, C. Czaplewski, Physics-based potentials for the coupling between backbone- and side-chain-local conformational states in the united residue (UNRES) force field for protein simulations, J. Chem. Theory Comput. 11 (2015) 817–831.
[54] H. Späth, Cluster Analysis Algorithms, Halsted Press, New York, 1980.
[55] A.G. Lipska, S.R. Seidman, A.K. Sieradzan, A. Giełdoń, A. Liwo, H.A. Scheraga, Molecular dynamics of protein A and a WW domain with a united-residue model including hydrodynamic interaction, J. Chem. Phys. 144 (2016) 184110.
[56] C. Czaplewski, et al., Physics-based protein-structure prediction using the UNRES and ECEPP/3 force fields – tests on CASP5 targets, Fifth Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (2002).
[57] Y. He, M.A. Mozolewska, P. Krupa, A.K. Sieradzan, T.K. Wirecki, A. Liwo, K. Kachlishvili, S. Rackovsky, D. Jagieła, R. Ślusarz, C.R. Czaplewski, S. Ołdziej, H.A. Scheraga, Lessons from application of the UNRES force field to predictions of structures of CASP10 targets, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 14936–14941.
[58] P. Krupa, et al., Performance of protein-structure predictions with the physics-based UNRES force field in CASP11, Bioinformatics 32 (2016) 3270–3278.
[59] V.S. Pande, I. Baker, J. Chapman, S. Elmer, S. Kaliq, S.M. Larson, Y.M. Rhee, M.R. Shirts, C.D. Snow, E.J. Sorin, B. Zagrovic, Atomistic protein folding simulations on the submillisecond timescale using worldwide distributed computing, Biopolymers 68 (2003) 91–109.
[60] U.H. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, Chem. Phys. Lett. 281 (1997) 140–150.
[61] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (2000) 404.
[62] A. Liwo, M. Khalili, H.A. Scheraga, Ab initio simulations of protein-folding pathways by molecular dynamics with the united-residue model of polypeptide chains, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 2362–2367.
[63] F. Rakowski, P. Grochowski, B. Lesyng, A. Liwo, H.A. Scheraga, Implementation of a symplectic multiple-time-step molecular dynamics algorithm, based on the united-residue mesoscopic potential energy function, J. Chem. Phys. 125 (2006) 204107.
[64] A. Liwo, S. Ołdziej, C. Czaplewski, D.S. Kleinerman, P. Blood, H.A. Scheraga, 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. Theor. Comput. 6 (2010) 583–595.
[65] R. Kaźmierkiewicz, A. Liwo, H.A. Scheraga, Energy-based reconstruction of a protein backbone from its α-carbon trace by a Monte Carlo method, J. Comput. Chem. 23 (2002) 715–723.
[66] R. Kaźmierkiewicz, A. Liwo, H.A. Scheraga, Addition of side chains to a known backbone with defined side-chain centroids, Biophys. Chem. (2003) 100 (2003) 261–280, Erratum: Biophys. Chem., 106, 91 (2003).
[67] A. Zemla, C. Venclovas, K. Fidelis, B. Rost, A modified definition of SOV, a segment-based measure for protein secondary structure prediction assessment, Proteins Struct. Funct. Genet. 34 (1999) 220–223.
[68] P. Krupa, A. Hałabis, W. Zmudzinska, S. Ołdziej, H.A. Scheraga, A. Liwo, Maximum likelihood calibration of the UNRES force field for simulation of protein structure and dynamics, J. Chem. Inf. Model. 57 (2017) 2364–2377, PMID: 28809487.
[69] P.M. Johnson, G.C. Gucinski, F. Garza-Sánchez, T. Wong, L.-W. Hung, C.S. Hayes, C.W. Goulding, Functional diversity of cytotoxic tRNase/immunity protein complexes from *Burkholderia pseudomallei*, J. Biol. Chem. 291 (2016) 19387–19400.