



## OPEN ACCESS

EDITED BY  
Masahito Ohue,  
Tokyo Institute of Technology, Japan

REVIEWED BY  
Amitava Roy,  
National Institute of Allergy and  
Infectious Diseases (NIH), United States  
Yuichi Togashi,  
Ritsumeikan University, Japan

\*CORRESPONDENCE  
Adam Liwo,  
✉ adam.liwo@ug.edu.pl

SPECIALTY SECTION  
This article was submitted to Biological  
Modeling and Simulation,  
a section of the journal  
Frontiers in Molecular Biosciences

RECEIVED 16 October 2022  
ACCEPTED 29 November 2022  
PUBLISHED 14 December 2022

CITATION  
Ślusarz R, Lubecka EA, Czaplowski C and  
Liwo A (2022), Improvements and new  
functionalities of UNRES server for  
coarse-grained modeling of protein  
structure, dynamics, and interactions.  
*Front. Mol. Biosci.* 9:1071428.  
doi: 10.3389/fmolb.2022.1071428

COPYRIGHT  
© 2022 Ślusarz, Lubecka, Czaplowski  
and Liwo. This is an open-access article  
distributed under the terms of the  
[Creative Commons Attribution License  
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or  
reproduction in other forums is  
permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original  
publication in this journal is cited, in  
accordance with accepted academic  
practice. No use, distribution or  
reproduction is permitted which does  
not comply with these terms.

# Improvements and new functionalities of UNRES server for coarse-grained modeling of protein structure, dynamics, and interactions

Rafał Ślusarz<sup>1</sup>, Emilia A. Lubecka<sup>2</sup>, Cezary Czaplowski<sup>1</sup> and Adam Liwo<sup>1\*</sup>

<sup>1</sup>Faculty of Chemistry, University of Gdańsk, Fahrenheit Union of Universities in Gdańsk, Gdańsk, Poland, <sup>2</sup>Faculty of Electronics, Telecommunication and Informatics, Gdańsk University of Technology, Fahrenheit Union of Universities in Gdańsk, Gdańsk, Poland

In this paper we report the improvements and extensions of the UNRES server (<https://unres-server.chem.ug.edu.pl>) for physics-based simulations with the coarse-grained UNRES model of polypeptide chains. The improvements include the replacement of the old code with the recently optimized one and adding the recent scale-consistent variant of the UNRES force field, which performs better in the modeling of proteins with the  $\beta$  and the  $\alpha+\beta$  structures. The scope of applications of the package was extended to data-assisted simulations with restraints from nuclear magnetic resonance (NMR) and chemical crosslink mass-spectroscopy (XL-MS) measurements. NMR restraints can be input in the NMR Exchange Format (NEF), which has become a standard. Ambiguous NMR restraints are handled without expert intervention owing to a specially designed penalty function. The server can be used to run smaller jobs directly or to prepare input data to run larger production jobs by using standalone installations of UNRES.

## KEYWORDS

protein-structure modeling, coarse graining, UNRES model of polypeptide chains, molecular dynamics, data-assisted simulations

## 1 Introduction

Coarse-grained simulations are now a well established methodology with which to study large systems at large time scales. Compared to all-atom simulations, they offer a 1,000-fold or greater extension of simulation time scale (Khalili et al., 2005a; Kmiecik et al., 2016), enabling us to simulate biologically important events in real time (Kmiecik et al., 2016; Liwo et al., 2020). This gain, however, is achieved at the expense of accuracy compared to using the all-atom representation. One reason for decreasing the accuracy is a lower resolution of coarse-grained models compared to that of all-atom models. However, a more important reason is that most of the coarse-grained force fields are constructed by analogy to the all-atom ones, which is not correct from the physics point of

view. The physical origin of a coarse-grained force field is the potential of mean force of a given system, in which the degrees of freedom not explicitly considered in the model are averaged out (Liwo et al., 2001; Ayton et al., 2007). Consequently, the interaction potentials have a lower symmetry than the spherical symmetry and taking into account multibody terms is a necessity, as opposed to all-atom force fields (Liwo et al., 2001; Liwo et al., 2020). Moreover, many coarse-grained force fields are derived from structural data rather than in a bottom-up manner from all-atom energy surfaces (Kmieciak et al., 2016; Liwo et al., 2020), which makes the derived energy terms more difficult to understand and control. On the other hand, including even sparse restraints from experiments such as, e.g., nuclear magnetic resonance (NMR), chemical cross link mass spectroscopy (XL-MS), and small angle X-ray diffraction data (SAXS) can reduce the impact of force-field inaccuracy on the results. Therefore, data-assisted simulations are often carried out (Bonomi et al., 2016; Brodie et al., 2017; Karczyńska et al., 2018; Fajardo et al., 2019; Bottaro et al., 2020; Czaplewski et al., 2021).

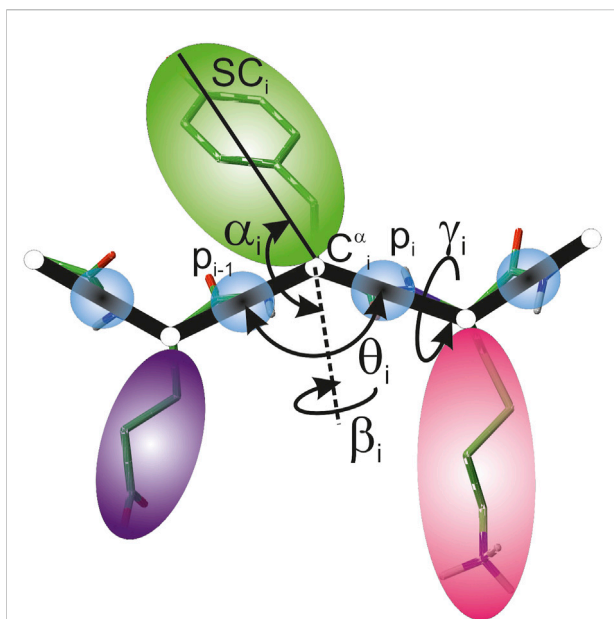
For proteins, a number of coarse-grained models are available, namely AWSEM (Davtyan et al., 2012), OpenAWSEM (Lu et al., 2021), CABS (Kolinski, 2004), MARTINI (Marrink and Tieleman, 2013; Marrink et al., 2022), SIRAH (Darré et al., 2015), and UNRES which, under the name UNICORN was upgraded to treat proteins, nucleic acids, and polysaccharides (Liwo et al., 2014; Sieradzan et al., 2022a). The MARTINI model is the most general one and the package enables automatic coarse-graining of any system with no or little user intervention. On the other hand, only AWSEM, CABS, and UNRES are capable of folding proteins. For peptides and proteins the OPEP model (Chebaro et al., 2012) is also used, which has the folding capacity. For protein-structure prediction, the ROSETTA model (Rohl et al., 2004) has been used with great success. However, this model makes extensive use of bioinformatics-based filters and is designed to locate the candidate predictions as structure with lowest potential energies. The above models differ in the degree of coarse graining, type of potential and methods of conformational search. AWSEM, ROSETTA, and OPEP use all-atom backbone, while the backbone is coarse-grained in CABS, MARTINI, SIRAH, and UNRES. CABS is entirely based on statistical potentials, while the other models mentioned here have physics-based components.

The structure-based coarse-grained models constitute another class of models, in which the native structure is the global minimum. There are two types of these models, namely the  $G\ddot{o}$ -like and elastic-network models. In the  $G\ddot{o}$ -like models (Taketomi et al., 1975; Hills and Brooks, 2009), long-range residue-residue contacts are assigned a potential with energy minimum, while non-native contacts are assigned all-repulsive potentials. In the elastic-network models harmonic or anharmonic potentials (sometimes double-well potentials) are imposed on all residue pairs (Sinitskiy and Voth, 2008; Trylska, 2010; Kmieciak et al., 2018).

Compared to established all-atom packages, the installation of the respective software, preparing the input data, running calculations, and processing the results is more difficult. To facilitate job data preparation and running simulations, web servers were created for the AWSEM (Jin et al., 2020), CABS (CABS-fold) (Błaszczuk et al., 2013), OPEP (PEP-FOLD3) (Lamiabile et al., 2016) and UNRES/UNICORN (UNRES web server) (Czaplewski et al., 2018) models. Protein structure prediction with the ROSETTA model can be accomplished by using the ROBETTA web server (Kim et al., 2004). However, these servers do not run data-assisted simulations except for the UNRES server.

The UNRES model of polypeptide chains (Liwo et al., 2014; Sieradzan et al., 2022a) developed in our laboratory is a heavily coarse-grained model, with only two interaction sites per amino-acid residue, namely a united peptide group and a united side chain. The effective energy function has been developed on the physical basis, by expressing the potential of mean force of a system in terms of Kubo cluster cumulant functions (Kubo, 1962), which are approximated analytically (Liwo et al., 2001; Sieradzan et al., 2017) by Kubo cluster cumulants. Owing to this method of derivation, the respective interaction potentials are dependent on both site distance and site orientation and the expressions for multibody terms, which are essential to reproduce regular secondary structures (Kolinski and Skolnick, 1992; Liwo et al., 2001), have been derived. UNRES has been successful in protein-structure prediction, in studying protein-folding dynamics and thermodynamics, and in solving biological problems (Sieradzan et al., 2022a).

The UNRES package is accessible as a standalone version (<https://unres.pl>) and as a web server, with which small production jobs can be run. The first version (Czaplewski et al., 2018) was released 4 years ago. That version ran a then state-of-the-art variant of the UNRES force field, which had a significant predictive power but produced too compressed  $\beta$ -strands (Krupa et al., 2017a). The first version of the server could also handle SAXS restraints. Recently, we developed a scale-consistent theory of coarse-grained force-field derivation (Sieradzan et al., 2017; Liwo et al., 2020) and a new version of UNRES (Liwo et al., 2019), which handles the  $\beta$ -strand and loop geometry much better, producing structures with a higher resolution, as demonstrated in the CASP13 (Lubecka et al., 2019) and CASP14 (Antoniak et al., 2021) community-wide experiments of the assessment of methods for protein-structure prediction (<https://predictioncenter.org>). The UNRES package has also been enhanced with the XL-MS (Fajardo et al., 2019; Kogut et al., 2021) and NMR (Lubecka and Liwo, 2021, 2022) data-assisted-calculation capacities, which have recently been included in the server. Finally, the UNRES code has recently been heavily optimized for memory and speed (Sieradzan et al., 2022b) and the optimized code has been included in the server. The new features of the UNRES server



**FIGURE 1**

UNRES model of polypeptide chains. The interaction sites are united peptide groups located between the consecutive  $\alpha$ -carbon atoms (light-blue spheres) and united side chains attached to the  $\alpha$ -carbon atoms (spheroids with different colors and dimensions). The backbone geometry of the simplified polypeptide chain is defined by the  $C^{\alpha} \cdots C^{\alpha} \cdots C^{\alpha}$  virtual-bond angles  $\theta$  ( $\theta$  has the vertex at  $C_i^{\alpha}$ ) and the  $C^{\alpha} \cdots C^{\alpha} \cdots C^{\alpha} \cdots C^{\alpha}$  virtual-bond-dihedral angles  $\gamma$  ( $\gamma$  has the axis passing through  $C_i^{\alpha}$  and  $C_{i+1}^{\alpha}$ ). The local geometry of the  $i$ th side-chain center is defined by the polar angle  $\alpha_i$  (the angle between the bisector of the respective angle  $\theta_i$  and the  $C_i^{\alpha} \cdots SC_i$  vector) and the azimuth angle  $\beta_i$  (the angle of counter-clockwise rotation of the  $C_i^{\alpha} \cdots SC_i$  vector about the bisector from the  $C_{i-1}^{\alpha} \cdots C_i^{\alpha} \cdots C_{i+1}^{\alpha}$  plane, starting from  $C_{i-1}^{\alpha}$ ). For illustration, the bonds of the all-atom chains, except for those to the hydrogen atoms connected with the carbon atoms, are superposed on the coarse-grained picture. Reproduced with permission from Zaborowski et al., 2015, J. Chem. Inf. Model., 55, 2050 (2015). Copyright 2015 American Chemical Society.

$$\begin{aligned}
 U = & w_{SC} \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_{i \neq j} U_{SC_i p_j} + w_{PP}^{VDW} \sum_{i < j-1} U_{p_i p_j}^{VDW} \\
 & + w_{PP}^{el} f_2(T) \sum_{i < j-1} U_{p_i p_j}^{el} + w_{tor} f_2(T) \sum_i U_{tor}(\gamma_i, \theta_i, \theta_{i+1}) \\
 & + w_b \sum_i U_b(\theta_i) + w_{rot} \sum_i U_{rot}(\theta_i, \hat{r}_{SC_i}) + w_{bond} \sum_i U_{bond}(d_i) \\
 & + w_{ssbond} \sum_i U_{ssbond}(d_i^{SS}) + w_{corr}^{(3)} f_3(T) \sum_{i < j-1} U_{corr;i}^{(3)} \\
 & + w_{turn}^{(3)} f_3(T) \sum_i U_{turn;i}^{(3)} \quad (1)
 \end{aligned}$$

where the terms  $U_{SC_i SC_j}$  are sidechain-sidechain interaction energies represented by the modified Gay-Berne potentials (Liwo et al., 1997),  $U_{SC_i p_j}$  are excluded-volume potentials that prevent the collapse of the united side chains on the backbone,  $U_{p_i p_j}^{VDW}$  (with spherical symmetry) and  $U_{p_i p_j}^{el}$  (with axial symmetry) are the non-bonded and mean-field-electrostatic interaction potentials of united peptide groups,  $U_{bond}$  are the bond-deformation potentials,  $U_b$  and  $U_{tor}$  are the backbone-virtual-bond-angle and the backbone-virtual-bond-torsional potentials, respectively,  $\theta_i$  and  $\gamma_i$  denoting the virtual-bond- and virtual-bond-dihedral angles, respectively (Figure 1),  $U_{rot}$  are the side-chain-rotamer potentials, in which  $\hat{r}_{SC_i}$  denotes the local coordinates of the unit vector pointing from  $C_i^{\alpha}$  to  $SC_i$ , while  $U_{corr}^{(3)}$  and  $U_{turn}^{(3)}$  are multibody terms that account for the coupling of the backbone-local and backbone-electrostatic interactions (Liwo et al., 2001; Sieradzan et al., 2017).  $U_{ssbond}$  denotes the terms that account for the energetics of disulfide bonds, including their formation and breaking (Chinchio et al., 2007; Krupa et al., 2017b). The solvent is implicit in the interaction potentials and the present parameterization corresponds to physiological pH. The  $w$ s are the weights of the energy terms and have been determined, along with some other parameters, by maximum-likelihood calibration of the force field (Liwo et al., 2019).

The factors  $f_n(T)$  account for the dependence of the force-field terms that correspond to higher-order terms in the Kubo cluster-cumulant expansion on temperature (Liwo et al., 2007), as given by Eq. 2.

$$f_n(T) = \frac{\ln[\exp(1) + \exp(-1)]}{\ln\{\exp[(T/T_0)^{n-1}] + \exp[-(T/T_0)^{n-1}]\}} \quad (2)$$

where  $T_0 = 300$  K.

As mentioned in the Introduction, the UNRES effective energy function originates from the potential of mean force of polypeptide chains in water, which is represented (Liwo et al., 2001; Sieradzan et al., 2017) by a truncated series of Kubo cluster-cumulant functions (Kubo, 1962). Consequently, it has the sense of free energy and depends on temperature (Liwo et al., 2007). The energy terms in the present version of UNRES have been derived based on our recently developed scale-consistent theory of coarse graining (Sieradzan et al., 2017), in which the atomistic details are rigorously embedded in the effective interactions potentials. As a result of the application of this theory, the torsional potentials depend not only on the virtual-bond-

are described in this article. Examples are provided to illustrate the new data-assisted-calculation features.

## 2 Materials and methods

### 2.1 UNRES model and force field

In the UNRES model, a polypeptide chain is represented by a sequence of  $\alpha$ -carbon ( $C^{\alpha}$ ) atoms linked with virtual bonds, with peptide groups (p) located halfway between the consecutive  $C^{\alpha}$ s and united side chains (SC) attached to the  $C^{\alpha}$ s with the  $C^{\alpha} \cdots SC$  virtual bonds (Figure 1). Only the united peptide groups and the united side chains are interaction sites, while the  $C^{\alpha}$ s assist in chain-geometry definition.

The UNRES energy function is expressed by Eq. 1.

dihedral angles but also on the adjacent virtual-bond angles, tending to zero when a chain fragment becomes linear. This feature of the torsional potentials eliminates the problem of their indefiniteness in such situations (Sieradzan et al., 2017). Similarly, the terms accounting for the correlation of the backbone-electrostatic and backbone-local interactions, which formerly depended on the distance and orientation of the peptide groups involved and the backbone virtual-bond-dihedral angles located at them, now also depend on the virtual-bond angles. These features greatly improved the accuracy of the modeling of  $\beta$  and  $\alpha+\beta$  proteins (Liwo et al., 2019). The latest NEWCT-9P variant of the UNRES force field implemented in the upgraded UNRES server has been parameterized by a maximum likelihood method (Zaborowski et al., 2015) with nine proteins of all secondary-structure types (Liwo et al., 2019).

The rigorous physics-based derivation of the effective energy terms outlined above distinguishes UNRES from other coarse-grained force fields, in which the energy terms have been constructed by analogy to all-atom energy terms or on a heuristic basis (Kmiciek et al., 2016; Liwo et al., 2020). Owing to this derivation, the UNRES energy terms can be tracked down to elemental atomic interactions. The rigorous derivation also enabled us to capture the dependence of the effective potentials on site orientation and to derive the expressions for correlation terms. Details of the UNRES model and force field are available in the references cited (Liwo et al., 2014; Sieradzan et al., 2017; Liwo et al., 2019; Sieradzan et al., 2022a).

## 2.2 Molecular dynamics and its extensions with UNRES

The engine of the conformational search with UNRES is molecular dynamics, which has been implemented using the Lagrange formalism (Khalili et al., 2005b,a). Due to the axial symmetry of the interacting sites, the equations of motion are more complicated than those of the coarse-grained models that use spherical potentials (AWSEM, MARTINI, and SIRAH). The inertia matrix is not diagonal; however, it is a constant matrix. In our recent work (Sieradzan et al., 2022b), we reduced the inertia matrix to a five-band form, this saving both memory and computing time. MD simulations with UNRES can be run in the microcanonical (NVE) and canonical (NVT) mode, the latter with the Berendsen or the Langevin thermostat. For better conformational search, replica-exchange (REMD) (Hansmann, 1997) and multiplexed replica exchange molecular dynamics (Rhee and Pande, 2003) have been implemented in UNRES (Czaplewski et al., 2009). A binless version of the weighted histogram analysis method (WHAM) (Kumar et al., 1992) has been implemented (Liwo et al., 2007) to process the results of REMD/MREMD simulations in order to calculate ensemble-averaged properties and to determine conformational ensembles at the desired temperatures.

The UNRES molecular-dynamics code has been parallelized (Liwo et al., 2010) and its parallel implementation has been heavily upgraded and optimized recently (Sieradzan et al., 2022b). With the new code, molecular dynamics simulations of protein systems with sizes exceeding 100,000 amino-acid residues, reaching over 1 ns/day with 24 cores are feasible. Compared to the previous version, the new code is 2–50 times faster, depending on protein size and degree of parallelization (Sieradzan et al., 2022b).

## 2.3 Experimental restraints

As mentioned in the Introduction, UNRES can handle the experimental restraints from NMR, XL-MS, and SAXS experiments. These are included by adding the respective penalty functions to the UNRES energy, as given by Eq. 3.

$$V = U_{UNRES} + w_{NMR}^{\theta} V_{NMR}^{\theta} + w_{NMR}^{\gamma} V_{NMR}^{\gamma} + w_{NMR}^{dist} V_{NMR}^{dist} + w_{XL-MS}^{dist} V_{XL-MS}^{dist} + w_{SAXS} V_{SAXS} \quad (3)$$

where  $V_{NMR}^{\theta}$  and  $V_{NMR}^{\gamma}$  are the penalty terms corresponding to virtual-bond angles  $\theta$  and virtual-bond-dihedral angles  $\gamma$  (cf. Figure 1), respectively, derived from NMR data,  $V_{NMR}^{dist}$  is the penalty term corresponding to the interproton distances derived from NMR data,  $V_{XL-MS}^{dist}$  is the distance-penalty term corresponding to crosslink restraints,  $V_{SAXS}$  is the SAXS-restraint term, and the  $w$ s are the weights of the respective penalty terms. The specific terms are described in Sections 2.3.1–2.3.3.

### 2.3.1 NMR restraints

As follows from Eq. 3, the NMR restraints consist of the angular and the distance terms. Chemical shifts (CS) and coupling-constants data provide the restraints on the backbone dihedral angles  $\phi$  and  $\psi$ . These are converted to the restraints on the  $\theta$  and the  $\gamma$  angles (cf. Figure 1) by using the formulas from (Nishikawa et al., 1974). The penalty function is a flat-bottom function as given by Eqs 4–7 (Lubecka and Liwo, 2022).

$$V_{NMR}^{\theta} = g(\theta, \theta_l, \theta_u) \quad (4)$$

$$V_{NMR}^{\gamma} = g(\gamma, \gamma_l, \gamma_u) \quad (5)$$

with

$$g(x, x_l, x_u) = \begin{cases} \frac{1}{4} \left( \delta + \frac{x_u - x_l}{2} \right)^4 & \text{for } \delta < \frac{x_l - x_u}{2} \\ 0 & \text{for } \frac{x_l - x_u}{2} < \delta < \frac{x_u - x_l}{2} \\ \frac{1}{4} \left( \delta - \frac{x_l - x_u}{2} \right)^4 & \text{for } \delta > \frac{x_u - x_l}{2} \end{cases} \quad (6)$$

$$\delta = \left( x - \frac{x_l + x_u}{2} \right) \bmod 2\pi \quad (7)$$



where  $\theta_b$ ,  $\theta_u$ ,  $\gamma_b$ , and  $\gamma_u$  are the lower and upper boundaries on the virtual-bond angles  $\theta$  and virtual-bond-dihedral angles  $\gamma$ , respectively (which are calculated from the boundaries on the  $\phi$  and  $\psi$  backbone dihedral angles).

To include the distance restraints, the positions of the protons are estimated first by using our recently developed ESCASA algorithm (Lubecka and Liwo, 2021), which is based on analytical formulas. This algorithm also provides analytical gradients of the estimated proton positions with respect to coarse-grained coordinates. A flat-bottom penalty function with a mild slope is imposed on each interproton distance estimated by NMR, as given by Eq. 8.

$$V_{cont}(d, d_l, d_u, A) = \begin{cases} A \frac{(d - d_l)^4}{\sigma^4 + (d - d_l)^4} [1 + \kappa \ln \cosh(d - d_l)] & \text{for } d < d_l \\ 0 & \text{for } d_l \leq d \leq d_u \\ A \frac{(d - d_u)^4}{\sigma^4 + (d - d_u)^4} [1 + \kappa \ln \cosh(d - d_u)] & \text{for } d > d_u \end{cases} \quad (8)$$

where  $d$  is a proton-proton distance estimated from an UNRES structure,  $d_l$  and  $d_u$  are the lower and upper distance boundaries, respectively, which are taken from NMR data,  $\sigma$  is the thickness of the transition region between zero and maximum restraint height,  $A$  is the height of the restraint well, and  $\kappa$  is the slope of the restraint at large distances. The default values of  $\sigma$  and  $A$  are 0.5 Å and 1.0 kcal/mol, respectively. The original penalty function from our earlier work (Sierdzan and Jakubowski, 2017; Lubecka and Liwo, 2019) corresponds to  $\kappa = 0$  and quickly approaches the asymptote  $A$ , contributing virtually no force when  $d \gg d_u$ . Thus, the penalty terms do not force incompatible restraints (which usually correspond to wrongly predicted contacts), preventing a simulation from producing non-protein-like structures. With a small  $\kappa > 0$  (default 0.01), the right asymptote is  $A + \kappa(d - d_u)$ , which provides a small gradient at large distances, thus mildly guiding the search towards satisfying the restraint but not forcing it if incompatible with the other restraints.

To treat ambiguous restraints, the restraint-penalties of an ambiguous set are put into a log-exp function, which has the shape of intersecting gorges and, therefore, takes the minimum value regardless of whether one only or more restraints of an ambiguous-restraint sets are satisfied (Eq. 9) (Lubecka and Liwo, 2022).

$$V_{NMR}^{dist}(\{d\}; d_l, d_u, A) = -\frac{1}{\alpha} \ln \left\{ \sum_{i=1}^{n_{amb}} \exp[-\alpha V_{cont}(d_i; d_l, d_u, A)] \right\} \quad (9)$$

where  $\{d\}$  is the set of distances corresponding to a given ambiguous restraint,  $\alpha$  is an arbitrary parameter, and  $V_{cont}(d_i; d_l, d_u, A)$  is defined by Eq. 8. With  $\alpha$  large enough (default 20),  $V_{NMR}(\{d\}; d_l, d_u, A)$  is nearly 0, regardless of whether only one or all restraints of the ambiguous set are satisfied. Thus, the

restraints of an ambiguous set, which are incompatible with the structure are eliminated.

In this work, we optimized the calculation of the ambiguous penalty function given by Eq. 9 by considering only the exponentials for which the distance-penalty is not too big to make them close to zero and parallelized the evaluation of the penalty function to achieve load balance.

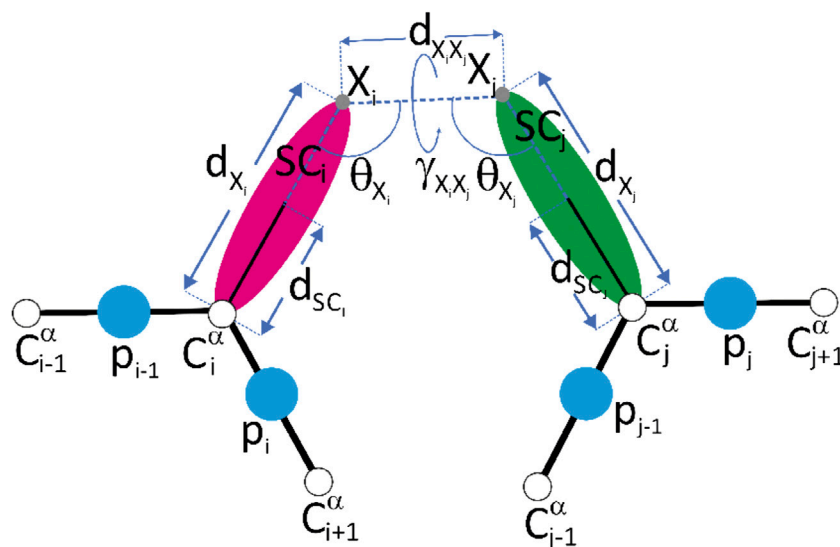
Apart from UNRES, the NMR-data-assisted simulations can be run with the CABS (CABS-NMR) (Latek and Koliński, 2011) and ROSETTA (CS-ROSETTA) (Nerli and Sgourakis, 2019) coarse-grained models. However, ROSETTA uses all-atom backbone, with which the calculation of interproton distances and backbone dihedral angles is straightforward, while the procedure implemented in CABS-NMR requires the conversion from the coarse-grained to the all-atom representation to evaluate the NMR-penalty term Latek and Koliński (2011), which involves additional computational effort and restricts the use of the approach to Monte Carlo simulations. As opposed to this, the ESCASA algorithm expresses the proton coordinates analytically in terms of coarse-grained geometry and, consequently, enables us to compute analytical forces due to the NMR-penalty term which, in turn, makes possible to implement it in molecular dynamics simulations (Lubecka and Liwo, 2021).

### 2.3.2 XL-MS restraints

Three types of XL-MS restraints have been implemented in UNRES. The restraints of the first type have the form of a flat-bottom bounded function imposed on the respective  $C^\alpha \cdots C^\alpha$  or  $SC \cdots SC$  distances (Eq. 8 in which  $\kappa = 0$ ). These were designed for the non-specific crosslinks proposed by Rappsilber (2011). Because the right boundary of the distance between the crosslinked residues is 25 Å, they do not perform well in data-assisted simulations (Fajardo et al., 2019). The crosslinks of the second type are statistical pseudopotentials (Fajardo et al., 2019; Kogut et al., 2021) that are based on  $C^\alpha \cdots C^\alpha$ -distance distributions of crosslinked residue pairs from known proteins determined by Leitner et al. (2014) (Eq. 10). The cross linking agents are the adipic-acid (ADH) and the pimelic-acid (PDH) hydrazides linking the acidic side chains and disuccinimidyl suberate (DSS), which links the side chains of lysine residues. Zero-length (ZL) crosslinks that occur between basic and acidic side chains or involve the N-terminal amino groups or the C-terminal carboxyl groups are also included. These restraints perform well in the data-assisted modeling of protein structure with UNRES (Leitner et al., 2014; Kogut et al., 2021).

$$W(d) = -ART \ln \left\{ \left[ a + b \left( \frac{d}{\sigma} \right)^4 \right] \exp \left( -\frac{d^2}{2\sigma^2} \right) + c \right\} \quad (10)$$

where  $d$  is the distance between the  $C^\alpha$ -atoms of the crosslinked residues,  $a$ ,  $b$ ,  $c$ , and  $\sigma$  are crosslink-specific parameters,  $R$  is the universal gas constant,  $T$  is the absolute temperature; we assumed



**FIGURE 2**

A scheme of the representation of crosslink restraints between residues with indices  $i$  and  $j$ , respectively, in the UNRES model. The  $C^\alpha$  atoms are shown as white spheres, the united side chains (SC) are shown as colored spheroids, and the united peptide groups (p) are shown as blue spheres. The crosslinkable side chains (aspartic acid, glutamic acid or lysine) are linked with adipic/pimelic acid (ADH or PDH) dihydrazide or by the glutaric/suberic acid (BS<sup>2</sup>G or BS<sup>3</sup>), respectively. The link is anchored in (approximately) the positions of the side-chain carbonyl-carbon atoms of Asp or Glu (for the ADH or PDH crosslinks) or in the positions of the lysine side-chain nitrogen atoms (for the BS<sup>2</sup>G or BS<sup>3</sup> crosslinks), respectively. The anchor points (indicated with "X" and light-gray spheres) are located on the  $C^\alpha \cdots SC$  axes of the UNRES residues. The geometric parameters on which the respective pseudopotentials depend [Eqs 11–14] are also shown in the Figure. Reproduced with permission from Kogut et al. (2021), J. Comput. Chem. 42, 2054 (2021). Copyright 2021 John Wiley and Sons.

$T = 298$  K, hence  $RT = 0.591$  kcal/mol, and  $A$  (default 15) is the weight of the potential, which is assigned the confidence of the crosslink.

Restraints of the third type are the pseudopotentials imposed on the distances between the ends of the acidic residue side chains linked by ADH or PDH or the ends of the lysine side chains linked by glutaric acid dipentylamide (BS<sup>2</sup>) or suberic acid dipentylamide (BS<sup>3</sup>) and the virtual-bond angles and the virtual-bond-dihedral angle of the crosslinked  $C^\alpha \cdots X \cdots X \cdots C^\alpha$  moiety, where  $X$  means the end of the respective side chain (Kogut et al., 2021), as given by Eqs 11–14. These restraint potentials performed well in data-assisted UNRES simulations (Kogut et al., 2021).

$$V(d_{X_i}, d_{X_j}, d_{X_i X_j}, \theta_{X_i}, \theta_{X_j}, \gamma_{X_i X_j}) = V_d(d_{X_i}) + V_d(d_{X_j}) + V_d(d_{X_i X_j}) + V_\theta(\theta_{X_i}) + V_\theta(\theta_{X_j}) + V_\gamma(\gamma_{X_i X_j}, \theta_{X_i}, \theta_{X_j}) \quad (11)$$

$$V_d(d) = \frac{\prod_{j=1}^{N_d} \left[ a_j + \frac{1}{2} k_j (d - d_j^0)^2 \right]}{\sum_{j=1}^{N_d} \prod_{j \neq i} \left[ a_j + \frac{1}{2} k_j (d - d_j^0)^2 \right]} \quad (12)$$

$$V_\theta(\theta) = a_\theta + \sum_{j=1}^{N_\theta} a_j (\cos \theta)^j + b_j (\sin \theta)^j \quad (13)$$

$$V_\gamma(\gamma, \theta_1, \theta_2) = V_\circ + \sum_{j=1}^{N_\gamma} c_j (\sin \theta_1)^j (\sin \theta_2)^j \cos(j\gamma) \quad (14)$$

where  $N_d$ ,  $N_\theta$ , and  $N_\gamma$  are the numbers of terms in the expressions for the virtual-bond-length, virtual-bond-angle, and virtual-bond-dihedral-angle potentials and the other symbols except for geometric variables ( $d_{X_i}$ ,  $\theta_{X_i}$ ,  $\theta_{X_j}$ , and  $\gamma_{X_i X_j}$ ) are adjustable parameters, which have been determined in (Kogut et al., 2021). The geometric variables are illustrated in Figure 2.

A full menu of crosslink-restraint potentials, which include the statistical potentials given by Eq. 10 and the MD-derived potentials given by Eqs 12–14 is available only with UNRES. Implementations with other coarse-grained models such as, e.g., ROSETTA (Nerli and Sgourakis, 2019) and MEDUSA (Brodie et al., 2017) include only the square-well type contact potentials similar to that given by Eq. 8.

### 2.3.3 SAXS restraints

In our approach (Karczyńska et al., 2018), the SAXS restraint-penalty function is based on the experimental distance distribution,  $P^{exp}(r)$ , determined by the Fourier transform of the intensity,  $I(q)$ , where  $r$  is the distance and  $q = 4\pi \sin \theta / \lambda$ ,  $\theta$  being the scatter angle and  $\lambda$  being the

wavelength respective restraint function. The penalty function,  $V_{SAXS}$ , is a maximum-likelihood function, as given by Eq. 15.

$$V_{SAXS} = - \int_0^{d_{max}} P^{SAXS}(r) \ln P^{calc}(r) dr \quad (15)$$

$$\approx - \Delta r \sum_{k=1}^M P^{SAXS}(r_k) \ln P^{calc}(r_k)$$

where  $r_k$  is the distance at the center of the  $k$ th bin of the histogram of the distance distribution from SAXS measurements,  $M$  is the number of bins,  $P^{SAXS}(r)$  is the value of the probability distribution determined by SAXS at  $r$ ,  $P^{calc}(r)$  is the value of the probability distribution calculated from simulations at  $r$ , and  $d_{max}$  is the maximum distance in the molecule, and  $\Delta r$  is the bin size.

In our earlier work (Karczyńska et al., 2018),  $P^{calc}$  was a sum of Gaussians, each centered on a given  $C^\alpha \cdots C^\alpha$  distance, with a fixed standard deviation. As a result, the calculated  $P(r)$  curves were slightly shifted to the right and the parts corresponding to the small distances were too small. Moreover, the solvation shell was not taken into account. Later, we revised the formula to replace the Gaussians with log-normal functions and to introduce an estimate of solvation shell, as given by Eq. 16.

$$P^{calc}(r_k) = \frac{1}{A} \sum_i \sum_{j < i} \exp \left[ - \frac{(\ln r_{ij} - \ln r_k)^2}{2\sigma_{ij}^2} \right] \quad (16)$$

$$A = \Delta r \sum_{k=1}^M \sum_i \sum_{j < i} \exp \left[ - \frac{(\ln r_{ij} - \ln r_k)^2}{2\sigma_{ij}^2} \right] \quad (17)$$

$$\sigma_{ij} = \frac{r_i^o + r_j^o}{r_{ij}} \quad (18)$$

$$r_i^o = \frac{\rho_i}{s} + \sigma_{max} + \frac{(\sigma_{min} - \sigma_{max})(x_i + 1)}{2} \quad (19)$$

where  $r_{ij}$  is the distance between the  $C^\alpha$  atoms of residues  $i$  and  $j$  in the calculated conformation,  $\sigma_{ij}$  is the standard deviation of the respective Gaussian,  $\rho_i$  is the Stokes' radius of residue  $i$ ,  $\sigma_{min}$  and  $(\sigma_{min} + \sigma_{max})/2$  are minimum and maximum size of the solvation shell of residue  $i$  and  $x_i$  is equal to 0 if residue  $i$  has no neighbors and 1 if its solvation shell is maximally filled with neighboring residues.  $A$  is the factor normalizing the calculated probability to 1.

The above implementation of SAXS restraints to run data-assisted simulations seems to be the only one for a coarse-grained model. Grudin et al. (2021) have recently reported an implementation of their Pepsi-SAXS/SANS method to run SAXS-data-assisted simulations; however, they use all-atom representation.

## 2.4 UNRES web server

The UNRES web server (Czaplewski et al., 2018) is available at <https://unres-server.chem.ug.edu.pl>. No registration is

required to run jobs; however, registered users have access to past jobs. The peptide-protein and protein-protein docking functionality (UNRES-Dock) has recently been added to the server (Krupa et al., 2021). The following types of jobs can be run with the server:

1. Energy minimization of the input structure.
2. Molecular dynamics (MD) simulations. These can be run in both canonical (NVT) and microcanonical (NVE) mode. The Berendsen et al. (1984) and the Langevin thermostats are available to run NVT simulations. A trajectory movie is displayed after the job is completed and fluctuations of  $C^\alpha$  positions are calculated and visualized. If a reference structure has been input, the variation of root mean square deviation (RMSD) from the reference structure with time is calculated and displayed.
3. Replica exchange (REMD) and multiplexed replica-exchange molecular dynamics (MREMD) simulations. Runs of this type are aimed at the modeling of the conformational ensembles, in particular the representative conformations at the selected temperatures. A protocol that consists of a production (M) REMD run, processing the results with WHAM (Kumar et al., 1992), and cluster analysis (Murtagh and Heck, 1987) developed in our earlier work (Krupa et al., 2016) is applied. The final representative structures are converted to all-atom structures by using PULCHRA (Rotkiewicz and Skolnick, 2008) and SCWRL (Wang et al., 2008).
4. Docking-type runs. These runs are aimed at predicting the structures of peptide-protein or protein-protein complexes and are always carried out in the (M)REMD mode. Details of the UNRES docking protocol are described in (Krupa et al., 2021).

The user can select the force field. The input structure can be read from a Protein Data Bank (Berman et al., 2000) (PDB) file (in this case the sequence is not input separately but is taken from the PDB file) or a starting extended or randomly-generated structure can be specified. The present version of the server does not repair incomplete PDB structures; for this purpose, MODELLER (Fiser and Šali, 2003) or other software has to be used. Secondary-structure restraints can also be input. The input can be specified in a simpler way by using the "Basic" options or in a more advanced way by using the "Advanced" options. Details of specifying the input are in the "Input data," "Tutorial," and "UNRES-Dock tutorial" sections of the server. Apart from the visual output, the users (both unregistered and registered) can download all the output files (main output file, run summary files, PDB files, etc.) produced by the server.

In the first version of the UNRES web server (Czaplewski et al., 2018) only SAXS-assisted simulations were enabled, while NMR- and XL-MS-data assisted simulations have been enabled in the present version. These features are described in Section 3.1; for consistency, we have also included a short description of the

SAXS-data assisted feature in that section. The reader is referred to our earlier work (Czaplewski et al., 2018) for the description of the other functions of the UNRES server.

## 3 Results

### 3.1 Implementation of the new features in the UNRES server

The new scale-consistent NEWCT9P force field (Liwo et al., 2019) has been included in the upgraded UNRES server. The old OPT-WTFSA-2 force field (Krupa et al., 2017a) can still be used. It should be noted that the new optimized code runs only the NEWCT-9P force field and selecting OPT-WTFSA-2 means running the slower code.

SAXS restraints in the form of  $P(r)$  as described in Section 2.3.3 are read from the appropriate ASCII text file supplied by the user. The data format is that provided by the CRY SOL program (Svergun et al., 1995).

NMR restraints described in Section 2.3.1 can be read in the plain-text format, in which the restraints for the NMR-assisted targets were provided during the CASP experiments (Sala et al., 2019), in the NMR-star format (Ulrich et al., 2019), in which most of the NMR data are deposited in the PDB, or in the NMR Exchange Format (NEF) (Gutmanas et al., 2015), which has become the standard. If both angular and distance restraints are present in a NMR-data file the user can choose to use only the distance restraints. Details and examples of NMR-restraint input are included in the “Input data” section of the UNRES web server page.

Crosslink restraints (described in Section 2.3.2) are supplied by the user in ASCII files. The crosslink-restraint type (see section Section 2.3.2) is selected from the menu. Details are described in the “Input data” section of the UNRES web server page.

### 3.2 Examples

#### 3.2.1 NMR examples

##### 3.2.1.1 Unambiguous NMR distance and angular restraints

We used the restraint data of the *de novo* designed Foldit3 protein (Koepnick et al., 2019) (PDB: 6msp), with distance restraints taken from the respective PDB entry (NMR restraints v2). The unstructured 17-residue N-terminal tag has been removed and the restraints have been edited accordingly. A total of 1,279 restraints were included in the calculations. The target is an 81-residue  $\alpha+\beta$  protein and was one of the test-set protein used to test the NMR-data-assisted implementation of UNRES (Lubecka and Liwo, 2022). The server example data specify an 8-trajectory REMD run with replica temperatures of 250, 260, 270, 280, 290, 300, 315, and 330 K, respectively. Each trajectory consisted of Langevin-dynamics 2,000,000 steps with a

9.78 fs step length and was started from a randomly-generated conformation. This is a much less resource-demanding run compared to the 144-replica 20,000,000-step Hamiltonian Replica Exchange (HREMD) run for this protein performed to test NMR-data-assisted UNRES (Lubecka and Liwo, 2022). The conformation of the first cluster that fits the NMR data best (model 1) has C $\alpha$  RMSD from the experimental structure of 2.03 Å and Global Distance Test Total Score GDT\_TS (Zemla, 2003; Moult et al., 2013) of 76.88, compared to 1.61 Å and 87.19, respectively in (Lubecka and Liwo, 2022). The superposition of the calculated on the experimental structure is shown in Figure 3A, while the proton-proton contact map corresponding to the model superposed on that resulting from the NMR measurements is shown in Figure 3B.

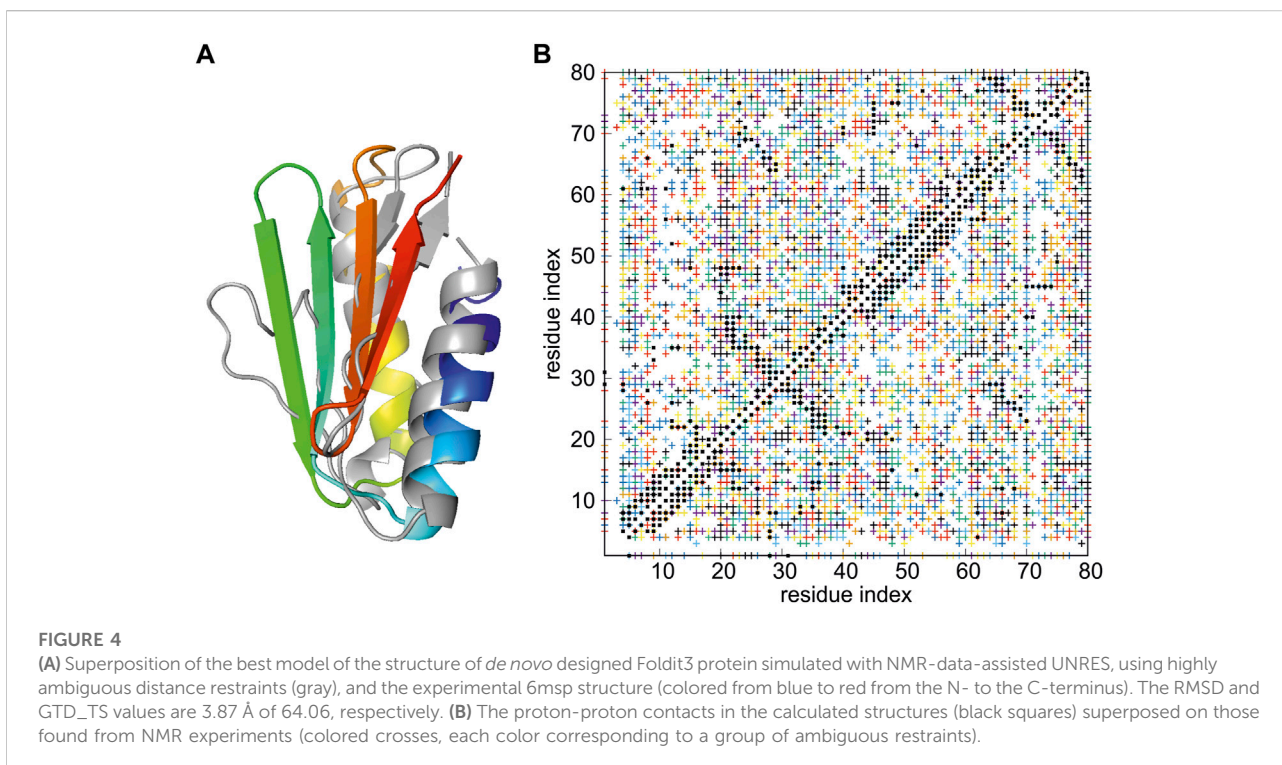
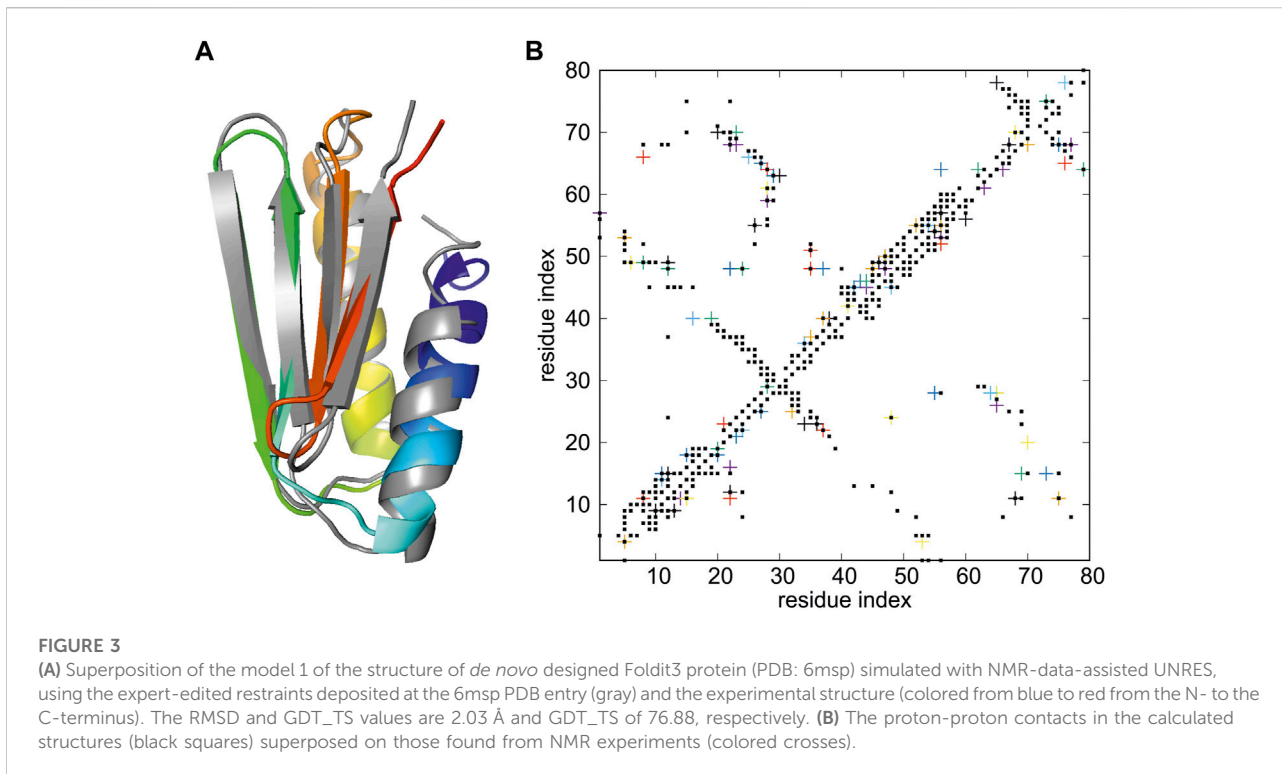
##### 3.2.1.2 Ambiguous NMR distance restraints

As in Section 3.2.1.1, the protein is the *de novo* designed protein Foldit3 (Koepnick et al., 2019) (PDB: 6msp), which was the data-assisted CASP13 target n1008. The restraint set, containing a total of 26,626 restraints, was provided to the CASP13 participants at [https://predictioncenter.org/download\\_area/CASP13/extra\\_experiments](https://predictioncenter.org/download_area/CASP13/extra_experiments) by G.T. Montelione. These are raw data prior to expert editing. The number of possible assignments per peak exceeds 100 for some of the peaks; moreover, more than a half of them are violated by the experimental 6msp structure (Lubecka and Liwo, 2022). These data were used to test the NMR-data-assisted functionality of the UNRES package in our previous work (Lubecka and Liwo, 2022). With the server version, 12 REMD trajectories were run at 260, 262, 266, 271, 276, 282, 288, 296, 304, 315, 333, and 370 K, respectively. These temperatures have been determined by using the Hansmann algorithm (Trebst et al., 2006), which maximizes the number of walks in the temperature space. Each trajectory consisted of 2,000,000 steps with a 9.78 fs length and was started from a randomly-generated conformation. The representative conformation of the best model has C $\alpha$  RMSD of 3.87 Å and GDT\_TS of 64.06, compared to 1.66 Å and of 80.00 for the full-blown run with 144 HREMD trajectories, each consisting of 20,000,000 steps (Lubecka and Liwo, 2022). The calculated structure is superposed on the experimental structure in Figure 4A, while the proton-proton contact map corresponding to the model superposed on that resulting from the NMR measurements is shown in Figure 4B. It should be noted that a structure reasonably close to the experimental structure was obtained despite the very high ambiguity of the NMR restraints (Figure 4B) and limited computational resources applied.

### 3.3 XL-MS example

The example is horse heart cytochrome (PDB: 1hrc), a 105-residue  $\alpha$ -protein, which was one of the proteins used in our





previous work (Kogut et al., 2021) to test the XL-MS-data-assisted functionality of the UNRES package. As opposed to that work, no secondary-structure restraints were imposed in the calculations with the UNRES server. The data from lysine-lysine crosslinking experiments were taken from the Xlink Analyzer database (Kosinski et al., 2015), the original source being described in (Seebacher et al., 2006). There are a total of 24 crosslinks. The run consisted of 12 REMD trajectories, 2,000,000 steps per trajectory. The temperatures were set as in Section 3.2.1.2. Each trajectory was started from a randomly-generated conformation. Even with the limited server resources, the GDT\_TS is 26.68, compared to 34.90 with the 48 MREMD trajectories each consisting of 20,000,000 steps and secondary-structure restraint imposed (Kogut et al., 2021).

### 3.3.1 Timing

The calculations of the first NMR example (unambiguous NMR data), take 1.2 ms/MD step with two cores/trajectory of an 20-core Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40 GHz processor when using the optimized code, both running the scale-consistent NEWCT-9P force field. With the old code, the calculations take 3.5 ms/MD step. Thus, code optimization resulted in an about 3-fold reduction of execution time. With the same settings, the calculations of the second NMR example (ambiguous NMR data of the same protein) take 1.7 ms/MD steps when using the optimized and 8.7 ms/MD step with the non-optimized code, respectively. Thus, with the optimized code, the execution time increases only by about 40%, while it increases more than twofold for the non-optimized code. The optimized code is over five times faster than the non-optimized code. The increase of the execution time with optimized code is due to a great total number of ambiguous distance restraints (26,626). With the old code the exponential terms in Eq. 9 corresponding to all these distances were evaluated, while only those that make near-zero contributions to the penalty function are not computed.

For the crosslink example, the calculations took 0.7 ms/MD step with four cores/trajectory of an 20-core Intel(R) Xeon(R) CPU E5-2640 v4 @ 2.40 GHz processor with the optimized code and 2.3 ms/MD step with the old code, giving an over 3-fold speed-up. It should be noted that the examples analyzed here are small proteins, for which full-blown data-assisted calculations are doable with the UNRES web server.

## 4 Conclusion and outlook

We upgraded the UNRES server to include the new scale-consistent variant of the UNRES force field (Liwo et al., 2019) that, owing to the introduction of the dependence of the backbone-virtual-bond torsional and correlation potentials on backbone-virtual-bond angles handles the  $\beta$ -strand and loop geometry and, consequently, that of the  $\beta$ - and  $\alpha+\beta$ -proteins

better than the old version of UNRES. We have also replaced the old code with the recently optimized code (Sieradzan et al., 2022a), this speeding up the calculations at least twice.

The existing SAXS-data-assisted functionality has been upgraded by replacing the old penalty function with one that reproduces the asymmetry of the distance distribution and takes into account the solvation shell in a simple empirical manner (Eqs 15, 16). Two new functionalities recently introduced to the UNRES package were added to the server version, namely NMR- and XL-MS-data-assisted simulations. The NMR penalty function is based on our recently developed ESCASA algorithm (Lubecka and Liwo, 2019) to estimate proton positions from coarse-grained geometry analytically. Highly ambiguous restraints can be handled (Lubecka and Liwo, 2022). The XL-MS restraints include our recently developed pseudopotentials that restrain the distances in a stricter manner than plain distance boundaries or the  $C^\alpha$ -distance based statistical potentials (Kogut et al., 2021).

The introduced modifications have extended the scope of simulations possible to run with the UNRES server. In particular, its capacity to handle contradictory NMR and XL-MS restraints and ambiguous NMR restraints enables the user to run data-assisted simulations of intrinsically-disordered proteins, in which case the restraints can happen to be ambiguous and often do not pertain to a single structure (Bonomi et al., 2016). To our knowledge, the UNRES web server is the only publicly available server for protein simulations that enables the users to run full-blown data-assisted simulations that include the restraints from NMR, XL-MS and SAXS experimental data.

The examples presented in Section 3.2 demonstrate that reasonable results can be obtained, with limited resources, using state-of-the art UNRES and conformational-search methods. However, work on improvement of the UNRES force field and the search method, in particular on the ensemble-oriented conformational search, is underway in our laboratory. These modifications will be gradually introduced to the server version.

## Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## Author contributions

RS and CC implemented the new features in the UNRES server. EL designed and wrote the programs to convert NMR restraints from the NMR-star and the NEF formats to the UNRES format, determined the optimal settings of data-assisted simulations and did test calculations. CC implemented part of the changes and tested the server. AL

designed the manuscript and wrote the text. All authors have read and approved the manuscript.

## Funding

This work was supported by the National Science Centre under grants UMO-2021/40/Q/ST4/00035 (to AL) and UMO-2017/26/M/ST4/00044 (to CC).

## Acknowledgments

The authors are grateful to the Centre of Informatics–Tricity Academic Supercomputer and Network (CI TASK) in Gdańsk, the Interdisciplinary Centre of Mathematical and Computer Modeling in Warsaw (grant GA71-23), and the Academic Computer Centre Cyfronet AGH in Krakow (grant unres2022) for providing computational resources for the development of the software implemented on the server. Test

## References

- Antoniak, A., Biskupek, I., Bojarski, K. K., Czaplewski, C., Gieldoń, A., Kogut, M., et al. (2021). Modeling protein structures with the coarse-grained UNRES force field in the CASP14 experiment. *J. Mol. Graph. Model.* 108, 108008. doi:10.1016/j.jmgm.2021.108008
- Ayton, G. S., Noid, W. G., and Voth, G. A. (2007). Multiscale modeling of biomolecular systems: In serial and in parallel. *Curr. Opin. Struct. Biol.* 17, 192–198. doi:10.1016/j.sbi.2007.03.004
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., and Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81, 3684–3690. doi:10.1063/1.448118
- Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., et al. (2000). The protein data bank. *Nucleic Acids Res.* 28, 235–242. doi:10.1093/nar/28.1.235
- Błaszczak, M., Jamroz, M., Kmiecik, S., and Kolinski, A. (2013). CABS-Fold: Server for the *de novo* and consensus-based prediction of protein structure. *Nucleic Acids Res.* 41, W406–W411. doi:10.1093/nar/gkt462
- Bonomi, M., Camilloni, C., Cavalli, A., and Vendruscolo, M. (2016). Metainference: A Bayesian inference method for heterogeneous systems. *Sci. Adv.* 2, e1501177. doi:10.1126/sciadv.1501177
- Bottaro, S., Bengtson, T., and Lindorff-Larsen, K. (2020). “Integrating molecular simulation and experimental data: A Bayesian/maximum entropy reweighting approach,” in *Structural bioinformatics methods in molecular biology*. Editor Z. Gáspari (New York, NY: Humana Press), 219–240.
- Brodie, N. I., Popov, K. I., Petrotchenko, E. V., Dokholyan, N. V., and Borchers, C. H. (2017). Solving protein structures using short-distance cross-linking constraints as a guide for discrete molecular dynamics simulations. *Sci. Adv.* 3, e1700479. doi:10.1126/sciadv.1700479
- Chebaro, Y., Pasquali, S., and Derreumaux, P. (2012). The coarse-grained OPEP force field for non-amyloid and amyloid proteins. *J. Phys. Chem. B* 116, 8741–8752. doi:10.1021/jp301665f
- Chinchio, M., Czaplewski, C., Liwo, A., Oldziej, S., and Scheraga, H. A. (2007). Dynamic formation and breaking of disulfide bonds in molecular dynamics simulations with the UNRES force field. *J. Chem. Theory Comput.* 3, 1236–1248. doi:10.1021/ct700084z
- Czaplewski, C., Gong, Z., Lubecka, E. A., Xue, K., Tang, C., and Liwo, A. (2021). Recent developments in data-assisted modeling of flexible proteins. *Front. Mol. Biosci.* 8, 765562. doi:10.3389/fmolb.2021.765562
- Czaplewski, C., Kalinowski, S., Liwo, A., and Scheraga, H. A. (2009). Application of multiplexing replica exchange molecular dynamics method to the unres force

calculations were run on the 796-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk and on the nodes of the UNRES web server.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

field: Tests with  $\alpha$  and  $\alpha+\beta$  proteins. *J. Chem. Theory Comput.* 5, 627–640. doi:10.1021/ct800397z

Czaplewski, C., Karczyńska, A., Sieradzian, A. K., and Liwo, A. (2018). UNRES server for physics-based coarse-grained simulations and prediction of protein structure, dynamics and thermodynamics. *Nucleic Acids Res.* 46, W304–W309. doi:10.1093/nar/gky328

Darré, L., Machado, M. R., Brandner, A. F., González, H. C., Ferreira, S., and Pantano, S. (2015). Sirah: A structurally unbiased coarse-grained force field for proteins with aqueous solvation and long-range electrostatics. *J. Chem. Theory Comput.* 11, 723–739. doi:10.1021/ct5007746

Davtyan, A., Schafer, N. P., Zheng, W., Clementi, C., Wolynes, P. G., and Papoian, G. A. (2012). AWSEM-MD: Protein structure prediction using coarse-grained physical potentials and bioinformatically based local structure biasing. *J. Phys. Chem. B* 116, 8494–8503. doi:10.1021/jp212541y

Fajardo, J. E., Shrestha, R., Gil, N., Belsom, A., Crivelli, S. N., Czaplewski, C., et al. (2019). Assessment of chemical-crosslink-assisted protein structure modeling in CASP13. *Proteins* 87, 1283–1297. doi:10.1002/prot.25816

Fiser, A., and Šali, A. (2003). “Modeller: Generation and refinement of homology-based protein structure models,” in *Methods in enzymology*. Editors C. Carter and R. Sweet (San Diego: Academic Press), 374, 463–493.

Grudinin, S., Martel, A., and Prevost, S. (2021). Pepsi-SAXS/SANS – small-angle scattering guided tools for integrative structural bioinformatics. *Acta Crystallogr. A Found. Adv.* A77, C49. doi:10.1107/s0108767321096288

Gutmanas, A., Adams, P. D., Bardiaux, B., Berman, H. M., Case, D. A., Fogh, R. H., et al. (2015). NMR exchange format: A unified and open standard for representation of NMR restraint data. *Nat. Struct. Mol. Biol.* 22, 433–434. doi:10.1038/nsmb.3041

Hansmann, U. H. (1997). Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* 281, 140–150. doi:10.1016/s0009-2614(97)01198-6

Hills, R. D., and Brooks, C. L. (2009). Insights from coarse-grained Go models for protein folding and dynamics. *Int. J. Mol. Sci.* 10, 889–905. doi:10.3390/ijms10030889

Jin, S., Contessoto, V. G., Chen, M., Schafer, N. P., Lu, W., Chen, X., et al. (2020). AWSEM-suite: A protein structure prediction server based on template-guided, coevolutionary-enhanced optimized folding landscapes. *Nucleic Acids Res.* 48, W25–W30. doi:10.1093/nar/gkaa356

Karczyńska, A. S., Mozolewska, M. A., Krupa, P., Gieldoń, A., Liwo, A., and Czaplewski, C. (2018). Prediction of protein structure with the coarse-grained

UNRES force field assisted by small X-ray scattering data and knowledge-based information. *Proteins* 86, 228–239. doi:10.1002/prot.25421

Khalili, M., Liwo, A., Jagielska, A., and Scheraga, H. A. (2005a). Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model  $\alpha$ -helical systems. *J. Phys. Chem. B* 109, 13798–13810. doi:10.1021/jp058007w

Khalili, M., Liwo, A., Rakowski, F., Grochowski, P., and Scheraga, H. A. (2005b). 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. doi:10.1021/jp058008o

Kim, D. E., Chivian, D., and Baker, D. (2004). Protein structure prediction and analysis using the Robetta Server. *Nucleic Acids Res.* 32, W526–W531. doi:10.1093/nar/gkh468

Kmieciak, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E., and Kolinski, A. (2016). Coarse-grained protein models and their applications. *Chem. Rev.* 116, 7898–7936. doi:10.1021/acs.chemrev.6b00163

Kmieciak, S., Kouza, M., Badaczewska-Dawid, A., Kloczkowski, A., and Kolinski, A. (2018). Modeling of protein structural flexibility and large-scale dynamics: Coarse-grained simulations and elastic network models. *Int. J. Mol. Sci.* 19, 3496. doi:10.3390/ijms19113496

Koepnick, B., Flatten, J., Husain, T., Ford, A., Silva, D.-A., Bick, M. J., et al. (2019). De novo protein design by citizen scientists. *Nature* 570, 390–394. doi:10.1038/d41586-019-1274-4

Kogut, M., Gong, Z., Tang, C., and Liwo, A. (2021). Pseudopotentials for coarse-grained cross-link-assisted modeling of protein structures. *J. Comput. Chem.* 42, 2054–2067. doi:10.1002/jcc.26736

Kolinski, A. (2004). Protein modeling and structure prediction with a reduced representation. *Acta Biochim. Pol.* 51, 349–371. 035001349.

Kolinski, A., and Skolnick, J. (1992). Discretized model of proteins. I. Monte Carlo study of cooperativity in homopolypeptides. *J. Chem. Phys.* 97, 9412–9426. doi:10.1063/1.463317

Kosinski, J., von Appen, A., Ori, A., Karius, K., Müller, C. W., and Beck, M. (2015). Xlink analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol.* 189, 177–183. doi:10.1016/j.jsb.2015.01.014

Krupa, P., Halabis, A., Żmudzińska, W., Ołdziej, S., Scheraga, H. A., and Liwo, A. (2017a). Maximum likelihood calibration of the UNRES force field for simulation of protein structure and dynamics. *J. Chem. Inf. Model.* 57, 2364–2377. doi:10.1021/acs.jcim.7b00254

Krupa, P., Karczyńska, A. S., Mozolewska, M. A., Liwo, A., and Czaplewski, C. (2021). UNRES-Dock protein-protein and peptide-protein docking by coarse-grained replica-exchange MD simulations. *Bioinformatics* 37, 1613–1615. doi:10.1093/bioinformatics/btaa897

Krupa, P., Mozolewska, M. A., Wiśniewska, M., Yin, Y., He, Y., Sieradzan, A. K., et al. (2016). Performance of protein-structure predictions with the physics-based UNRES force field in CASP11. *Bioinformatics* 32, 3270–3278. doi:10.1093/bioinformatics/btw404

Krupa, P., Sieradzan, A. K., Mozolewska, M. A., Li, H., Liwo, A., and Scheraga, H. A. (2017b). Dynamics of disulfide-bond disruption and formation in the thermal unfolding of ribonuclease A. *J. Chem. Theory Comput.* 13, 5721–5730. doi:10.1021/acs.jctc.7b00724

Kubo, R. (1962). Generalized cumulant expansion method. *J. Phys. Soc. Jpn.* 17, 1100–1120. doi:10.1143/jpsj.17.1100

Kumar, S., Bouzida, D., Swendsen, R. H., Kollman, P. A., and Rosenberg, J. M. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 13, 1011–1021. doi:10.1002/jcc.540130812

Lamiabile, A., Thévenet, P., Rey, J., Vavruta, M., Derreumaux, P., and Tufféry, P. (2016). PEP-fold3: Faster de novo structure prediction for linear peptides in solution and in complex. *Nucleic Acids Res.* 44, W449–W454. doi:10.1093/nar/gkw329

Latek, D., and Koliński, A. (2011). CABS-NMR – de Novo tool for rapid global fold determination from chemical shifts, residual dipolar couplings and sparse methyl-methyl NOEs. *J. Comput. Chem.* 32, 536–544. doi:10.1002/jcc.21640

Leitner, A., Joachimiak, L. A., Unverdorben, P., Waltzhoeni, T., Frydman, J., Förster, F., et al. (2014). Chemical cross-linking/mass spectrometry targeting acidic residues in proteins and protein complexes. *Proc. Natl. Acad. Sci. U. S. A.* 111, 9455–9460. doi:10.1073/pnas.1320298111

Liwo, A., Baranowski, M., Czaplewski, C., Golaś, E., He, Y., Jagieła, D., et al. (2014). A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.* 20, 2306. doi:10.1007/s00894-014-2306-5

Liwo, A., Czaplewski, C., Pillardy, J., and Scheraga, H. A. (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. doi:10.1063/1.1383989

Liwo, A., Czaplewski, C., Sieradzan, A. K., Lubecka, E. A., Lipska, A. G., Golon, L., et al. (2020). “Scale-consistent approach to the derivation of coarse-grained force fields for simulating structure, dynamics, and thermodynamics of biopolymers,” in *Progress in molecular biology and translational science. Computational approaches for understanding dynamical systems: Protein folding and assembly*. Editors B. Strodel and B. Barz (London: Academic Press), 170, 73–122. chap. 2.

Liwo, A., Khalili, M., Czaplewski, C., Kalinowski, S., Ołdziej, S., Wachucik, K., et al. (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. doi:10.1021/jp065380a

Liwo, A., Ołdziej, S., Czaplewski, C., Kleinerman, D. S., Blood, P., and Scheraga, H. A. (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, 890–909. doi:10.1021/ct9004068

Liwo, A., Ołdziej, S., Pincus, M. R., Wawak, R. J., Rackovsky, S., and Scheraga, H. A. (1997). 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, 849–873. doi:10.1002/(sici)1096-987x(199705)18:7<849:aid-jcc1>3.0.co;2-r

Liwo, A., Sieradzan, A. K., Lipska, A. G., Czaplewski, C., Joung, I., Żmudzińska, W., et al. (2019). A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force field and force-field calibration and validation. *J. Chem. Phys.* 150, 155104. doi:10.1063/1.5093015

Lu, W., Bueno, C., Schafer, N. P., Moller, J., Jin, S., Chen, X., et al. (2021). OpenAWSEM with Open3SPN2: A fast, flexible, and accessible framework for largescale coarse-grained biomolecular simulations. *PLoS Comput. Biol.* 17, e1008308. doi:10.1371/journal.pcbi.1008308

Lubecka, E. A., Karczyńska, A. S., Lipska, A. G., Sieradzan, A. K., Zięba, K., Sikorska, C., et al. (2019). Evaluation of the scale-consistent UNRES force field in template-free prediction of protein structures in the CASP13 experiment. *J. Mol. Graph. Model.* 92, 154–166. doi:10.1016/j.jmgm.2019.07.013

Lubecka, E. A., and Liwo, A. (2019). Introduction of a bounded penalty function in contact-assisted simulations of protein structures to omit false restraints. *J. Comput. Chem.* 40, 2164–2178. doi:10.1002/jcc.25847

Lubecka, E., and Liwo, A. (2022). A coarse-grained approach to NMR-data-assisted modeling of protein structures. *J. Comput. Chem.* 43, 2047–2059. doi:10.1002/jcc.27003

Lubecka, E., and Liwo, A. (2021). ESCASA: Analytical estimation of atomic coordinates from coarse-grained geometry for nuclear magnetic resonance-assisted protein structure modeling. I. Backbone and H<sup>β</sup> protons. *J. Comput. Chem.* 42, 1579–1589. doi:10.1002/jcc.26695

Marrink, S. J., Monticelli, L., Melo, M. N., Alessandri, R., Tieleman, D. P., and Souza, P. C. T. (2022). Two decades of Martini: Better beads, broader scope. *WIREs Comput. Mol. Sci.* e1620. doi:10.1002/wcms.1620

Marrink, S. J., and Tieleman, D. P. (2013). Perspective on the martini model. *Chem. Soc. Rev.* 42, 6801–6822. doi:10.1039/c3cs60093a

Moult, J., Fidelis, K., Kryshtafovych, A., Schwede, T., and Tramontano, A. (2013). Critical assessment of methods of protein structure prediction (CASP) round X. *Proteins* 82 (2), 1–6. doi:10.1002/prot.24452

Murtagh, F., and Heck, A. (1987). *Multivariate data analysis*. Kluwer Academic Publishers.

Nerli, S., and Sgourakis, N. G. (2019). CS-ROSETTA. *Methods Enzymol.* 614, 321–362. doi:10.1016/bs.mie.2018.07.005

Nishikawa, K., Momany, F. A., and Scheraga, H. A. (1974). Low-energy structures of two dipeptides and their relationship to bend conformations. *Macromolecules* 7, 797–806. doi:10.1021/ma60042a020

Rappilber, J. (2011). The beginning of a beautiful friendship: Crosslinking/mass spectrometry and modelling of proteins and multi-protein complexes. *J. Struct. Biol.* 173, 530–540. doi:10.1016/j.jsb.2010.10.014

Rhee, Y. M., and Pande, V. S. (2003). Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* 84, 775–786. doi:10.1016/S0006-3495(03)74897-8

Rohl, C. A., Strauss, C. E. M., Misura, K. M. S., and Baker, D. (2004). Protein structure prediction using rosetta. *Methods Enzymol.* 383, 66–93. doi:10.1016/S0076-6879(04)83004-0



- Rotkiewicz, P., and Skolnick, J. (2008). Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.* 29, 1460–1465. doi:10.1002/jcc.20906
- Sala, D., Huang, Y. J., Cole, C. A., Snyder, D. A., Liu, G., Ishida, Y., et al. (2019). Protein structure prediction assisted with sparse NMR data in CASP13. *Proteins* 87, 1315–1332. doi:10.1002/prot.25837
- Seebacher, J., Mallick, P., Zhang, N., Eddes, J. S., Aebersold, R., and Gelb, M. H. (2006). Protein cross-linking analysis using mass spectrometry, isotope-coded cross-linkers, and integrated computational data processing. *J. Proteome Res.* 5, 2270–2282. doi:10.1021/pr060154z
- Sieradzan, A. K., Czaplowski, C., Krupa, P., Mozolewska, M. A., Karczyńska, A. S., Lipska, A. G., et al. (2022a). “Modeling the structure, dynamics, and transformations of proteins with the UNRES force field,” in *Protein folding: Methods and protocols*. Editor V. Muñoz (New York, NY: Springer US), 399–416.
- Sieradzan, A. K., and Jakubowski, R. (2017). Introduction of steered molecular dynamics into UNRES coarse-grained simulations package. *J. Comput. Chem.* 38, 553–562. doi:10.1002/jcc.24685
- Sieradzan, A. K., Makowski, M., Augustynowicz, A., and Liwo, A. (2017). A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. I. Backbone potentials of coarse-grained polypeptide chains. *J. Chem. Phys.* 146, 124106. doi:10.1063/1.4978680
- Sieradzan, A. K., Sans, J., Lubecka, E. A., Czaplowski, C., Lipska, A. G., Leszczynski, H., et al. (2022b). Optimization of parallel implementation of UNRES package for coarse-grained simulations to treat large proteins. *J. Comput. Chem.* (Early View article). doi:10.1002/jcc.27026
- Sinitskiy, A. V., and Voth, G. A. (2008). Coarse-graining of proteins based on elastic network models. *Chem. Phys.* 422, 165–174. doi:10.1016/j.chemphys.2013.01.024
- Svergun, D., Barberato, C., and Koch, M. H. J. (1995). Crysol – A program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* 28, 768–773. doi:10.1107/s0021889895007047
- Taketomi, H., Ueda, Y., and Gō, N. (1975). Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* 7, 445–459. doi:10.1111/j.1399-3011.1975.tb02465.x
- Trebst, S., Troyer, M., and Hansmann, U. H. (2006). Optimized parallel tempering simulations of proteins. *J. Chem. Phys.* 124, 174903. doi:10.1063/1.2186639
- Trylska, J. (2010). Coarse-grained models to study dynamics of nanoscale biomolecules and their applications to the ribosome. *J. Phys. Condens. Matter* 22, 453101. doi:10.1088/0953-8984/22/45/453101
- Ulrich, E. L., Baskaran, K., Dashti, H., Ioannidis, Y. E., Livny, M., Romero, P. R., et al. (2019). Nmr-star: Comprehensive ontology for representing, archiving and exchanging data from nuclear magnetic resonance spectroscopic experiments. *J. Biomol. NMR* 73, 5–9. doi:10.1007/s10858-018-0220-3
- Wang, Q., Canutescu, A. A., and Dunbrack, R. L. (2008). SCWRL and MolIDE: Computer programs for side-chain conformation prediction and homology modeling. *Nat. Protoc.* 3, 1832–1847. doi:10.1038/nprot.2008.184
- Zaborowski, B., Jagieła, D., Czaplowski, C., Hałabis, A., Lewandowska, A., Żmudzińska, W., et al. (2015). A maximum-likelihood approach to force-field calibration. *J. Chem. Inf. Model.* 55, 2050–2070. doi:10.1021/acs.jcim.5b00395
- Zemla, A. (2003). Lga: A method for finding 3D similarities in protein structures. *Nucleic Acids Res.* 31, 3370–3374. doi:10.1093/nar/gkg571