

UNRES server for physics-based coarse-grained simulations and prediction of protein structure, dynamics and thermodynamics

Cezary Czaplewski*, Agnieszka Karczyńska, Adam K. Sieradzan and Adam Liwo

Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland

Received January 30, 2018; Revised April 12, 2018; Editorial Decision April 16, 2018; Accepted April 18, 2018

ABSTRACT

A server implementation of the UNRES package (<http://www.unres.pl>) for coarse-grained simulations of protein structures with the physics-based UNRES model, coined a name UNRES server, is presented. In contrast to most of the protein coarse-grained models, owing to its physics-based origin, the UNRES force field can be used in simulations, including those aimed at protein-structure prediction, without ancillary information from structural databases; however, the implementation includes the possibility of using restraints. Local energy minimization, canonical molecular dynamics simulations, replica exchange and multiplexed replica exchange molecular dynamics simulations can be run with the current UNRES server; the latter are suitable for protein-structure prediction. The user-supplied input includes protein sequence and, optionally, restraints from secondary-structure prediction or small x-ray scattering data, and simulation type and parameters which are selected or typed in. Oligomeric proteins, as well as those containing D-amino-acid residues and disulfide links can be treated. The output is displayed graphically (minimized structures, trajectories, final models, analysis of trajectory/ensembles); however, all output files can be downloaded by the user. The UNRES server can be freely accessed at <http://unres-server.chem.ug.edu.pl>.

INTRODUCTION

The biological functions of proteins depend on their structure and dynamics, therefore, research on these subjects is central in molecular biology and medicinal chemistry, including drug design. However, experimental techniques (x-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and, since very recently, cryoelectron microscopy), provide structures of only a small fraction of pro-

tein sequences discovered yearly (1) and only fragmentary information is available by experimental techniques regarding protein dynamics. Computer modeling, aided by experimental and database information is, therefore, routinely used for the prediction of unknown protein structures and simulation of protein dynamics. For protein-structure prediction comparative modeling has proved to be the most successful method nowadays (1); however, it fails when a protein represents a completely new fold. Protein dynamics is usually simulated by using all-atom molecular dynamics (MD). However, despite the continuous development of computer technology, including the construction of machines dedicated to run MD simulations (2), the discrepancy between the simulation and biological time scale restricts the applicability of this technique to solve concrete biological problems. Coarse-grained models of proteins, in which an amino-acid residue is represented by a few extended atoms and, consequently, the simulation time- and size-scale is extended by orders of magnitude are, therefore, a subject of intense development (3). Generally, these models can be divided into those that use the knowledge- (derived from database statistics) and physics-based (derived by translating the all-atom energy function to that corresponding to a reduced model) potentials.

The coarse-grained UNRES model developed in our laboratory (4,5) is a highly-reduced physics-based model of proteins, in which only two interaction sites per residue (united side chains and united peptide groups) are kept. The α -carbon (C^α) atoms are present in the model to assist in the definition of chain geometry. The effective energy function has been defined as a potential of mean force of polypeptide chains in water, which has been subsequently expanded into Kubo cluster cumulant functions, identified with the respective energy terms (6). MD (7,8) and its replica-exchange (REMD) (9) and multiplexed replica exchange (MREMD) (10) extensions have been implemented in UNRES (11). In contrast to most of the protein coarse-grained models, owing to its physics-based origin, the UNRES force field can be used in simulations, including those aimed at protein-structure prediction, without ancillary information from structural databases (12,13); however, the implementation

*To whom correspondence should be addressed. Tel: +48 58 523126; Fax: +48 58 5235012; Email: cezary.czaplewski@ug.edu.pl

includes the possibility of using restraints (14). UNRES has been used with success in protein-structure prediction (12–14), studying protein-folding kinetics and free-energy landscapes as well as to solve biological problems (4,5,15).

MATERIALS AND METHODS

Data processing

Oligomeric proteins (16), proteins that include D-amino-acid residues (17) and those with disulfide links (18) can be treated with UNRES. For general information on UNRES the reader is referred to the pertinent book chapter (4) and review article (5), while the theory behind the model is described in detail in our recent work (6).

Although the UNRES package is available for download (at <http://www.unres.pl>) since several years ago, a user needs to install it on its system and run in batch mode. This feature leaves out a large number of potential users who prefer to submit jobs using a web-based interface. Moreover, a parallel compute server is required for most of the functions of the package to run. Therefore, we have recently created a server based on the UNRES package, coined a name ‘UNRES server’, which enables a user to submit jobs using a web-based interface. The present article is devoted to the description of this server.

A scheme of the UNRES server, outlining its function and data pipelines are shown in Figure 1. As shown, three types of calculations are available: (i) single local energy minimization (the MIN button), (ii) single-trajectory canonical MD (the MD button) and (iii) multiplexed replica-exchange molecular dynamics (the MREMD button). The mode of calculations is selected by the user. For the MREMD type of calculation, a production simulation is followed by running weighted histogram analysis method (19) to enable us to compute the probabilities of conformations, thermodynamic quantities and ensemble averages at any temperature (20), and cluster analysis of the ensemble at the desired temperature (usually 10–20 K below that of the heat-capacity peak) to construct the final models. This procedure copies our physics-based protein-structure prediction procedure that has been tested in the CASP exercises (13). The final models are converted into all-atom representation by using the PULCHRA (22) and SCWRL (23) methods and subjected to a short energy minimization with the AMBER force field with the Generalized-Born method to treat hydration (24).

Each calculation can be run in the basic or advanced mode; in the basic mode default parameters which were proved to work in most situations are applied, while in the advanced mode the user can select the parameters of the calculations. In particular, one of the two variants of UNRES force field can be selected, a ‘canonical’ one obtained by extensive search of parameter space (25) and subsequently supplemented with additional torsional potentials accounting for the coupling between backbone-local and sidechain-local conformational states (26), which improves the qual-

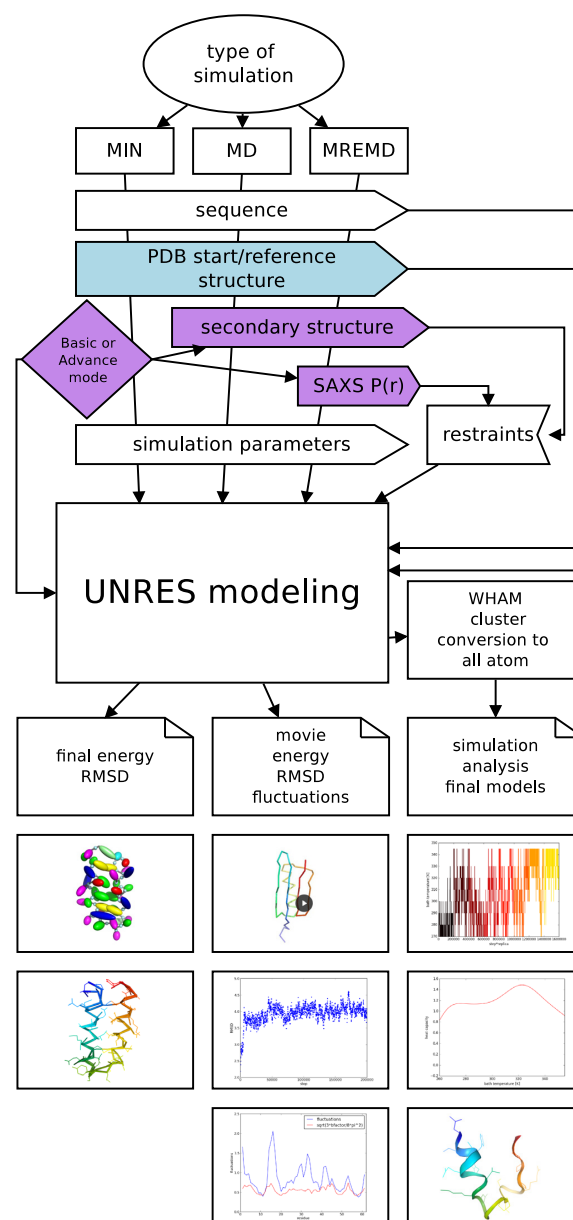


Figure 1. UNRES server operation flow.

ity of local structure, hereafter referred to as FF2, and the UNRES force field recently calibrated with seven proteins of different structural classes, by using the maximum-likelihood approach developed in our laboratory, hereafter referred to as OPT-WTFSA-2 (27). Both variants of UNRES have been validated in our earlier work (12,13,25–27).

Input data

Amino-acid sequence is required for all types of simulations; it can be input by the user as one-letter code or read from a PDB file (only the first model is taken), which can be uploaded by the user or downloaded from the PDB database given the PDB code. The reference/initial structure, if the user chooses so, is also read from a PDB file. In

the current implementation, no missing residues in a structure read from a PDB file are allowed, an error message is displayed if this occurs. For oligomeric proteins, different chains are automatically recognized while reading a PDB file; it is also possible to request calculations for a selected chain only. For example, by supplying the string 5G3Q the user specifies that a calculation should be run for all chains present in the PDB file, while 5G3Q:B means that chain B has been selected. User-supplied sequence of an oligomeric protein must contain the 'XX' chain separator; for example the string AAGGAAXXAAGGAA means that the sequence is a dimer consisting of two AAGGAA chains. D-residues are recognized in PDB files; in the user-supplied sequence they are marked by lowercase letters. For example, the sequence AAaAA means that the chain contains D-alanine in the third position. The positions of the disulfide links are read from the SSBOND records of the PDB file.

For the energy-minimization type, the sequence is always read from a PDB file, along with the starting structure. Energy minimization is carried out with the quasi-Newton Secant Unconstrained Minimization Solver (SUMSL) algorithm (28). In the advanced mode, the user can select the number of minimization steps and the number of maximum function evaluations, and decide whether to use an initial Monte Carlo search of the local geometry of side-chain centroids to remove overlaps.

For the MD type of calculations, a single canonical MD trajectory can be run. The basic mode includes selection of the starting structure (extended chain, randomly generated structure, or the structure read from a PDB file), number of MD steps, temperature and the seed to initialize the random-number generator. Initial energy minimization is turned on if the starting structure is read from a PDB file or randomly generated. Calculations are run with the Langevin thermostat, by using an adaptive multiple-time-step (A-MTS) quasi-symplectic algorithm developed in our laboratory (29), which is similar to the RESPA algorithm (30). In the advanced mode, the thermostat (Langevin or Berendsen) can be selected and secondary-structure restraints can be input [in the PSIPRED (31) format]. The 'molecular time unit' (mtu) used in UNRES MD amounts to 48.9 fs (7). However, it should be noted that, because of averaging over the secondary degrees of freedom, the time scale of UNRES MD is extended by 1000–10 000 times compared to the all-atom time scale (8). Simulation parameters can be changed in the advanced mode.

The MREMD type of calculations implies running multiple trajectories in a parallel run. In the basic input mode, an REMD calculation (no multiplexing) is carried out with eight replicas run at temperatures from 270 K to 345 K. In the advanced mode, the user can choose the number of replica temperatures and the values of the temperatures, as well as multiplexing of each replica and replica-exchange frequency. In addition to this, the distance distribution from small x-ray scattering (SAXS) measurements can be input to run SAXS-restrained simulations (32). MD parameters can be selected in the advanced mode, as for MD-type runs. At last, the user can also select the temperature at which to cluster the final conformational ensemble.

Output

For each simulation type, the output is displayed graphically and can be accessed up to 2 weeks from the time of job completion provided that the user has saved the web address of the job. Moreover, the UNRES input file(s) and all output files can also be downloaded from the server; see <http://www.unres.pl/docs> for the description of input/output files.

For the respective types of calculations, the output is the following:

Minimization: UNRES representation of minimized structure, minimized structure superposed on the starting (reference) structure, C^α-root-mean square deviation (C^α-rmsd) from the starting structure, percentage of native contacts.

Canonical MD: Temperature distribution, plots of potential energy, radius of gyration, C^α-rmsd, fraction of native contacts, fluctuations versus time, movie of the trajectory.

MREMD: Plots of heat capacity and ensemble-average C^α-rmsd versus temperature, plots of walk in temperature space, C^α-rmsd from the reference structure (if present) versus energy, C^α-rmsd versus time, representative conformations of the five families obtained after clustering (cf. section *Data processing*). These structures, in all-atom form, obtained after the conversion of the UNRES structures, can be downloaded by clicking on the respective button. For each model, its probability (fraction in the ensemble), average cluster C^α-rmsd, as well as the C^α-rmsd, TMscore and GDT_TS of the respective model are displayed.

RESULTS AND DISCUSSION

Examples included in server tutorial

The tutorial includes examples of running server jobs for all calculation types in the basic and in the advanced mode. Each calculation corresponding to a basic-mode example can be run within minutes, while those corresponding to advanced mode require up to a couple of hours. By pressing the 'Load example data' button all input data are loaded and parameters are set. The results of pre-run computations are displayed by selecting the 'Tutorial' item from the top navigation bar.

The following basic-mode examples include (i) minimization of the energy of the experimental structure of the N-terminal portion of the B-domain of staphylococcal protein A (PDB code: 1BDD); (ii) canonical MD simulation of the IGG binding domain of streptococcal protein G (PDB code: 1IGD), starting from the experimental structure; a fluctuation plot in residue index is displayed and compared with that of the B-factor; (iii) REMD simulation of the tryptophan cage mini-protein (PDB code: 1L2Y), starting from the extended chain; this is a full-blown structure-prediction run.

Advanced-mode examples include: (i) minimization of the P8MTCP1 disulfide-bonded α -helical hairpin mini-protein (PDB code: 1EI0); (ii) canonical MD run of the tryptophan-cage mini-protein (PDB code: 1L2Y) starting from the extended structure; (iii) MREMD run of CASP12 target T0882 also discussed in section '*Selected test cases*'; (iv) REMD run of the Bacteriocin CbnXY mini-protein (PDB code: 5UJG) starting from extended chain with simu-

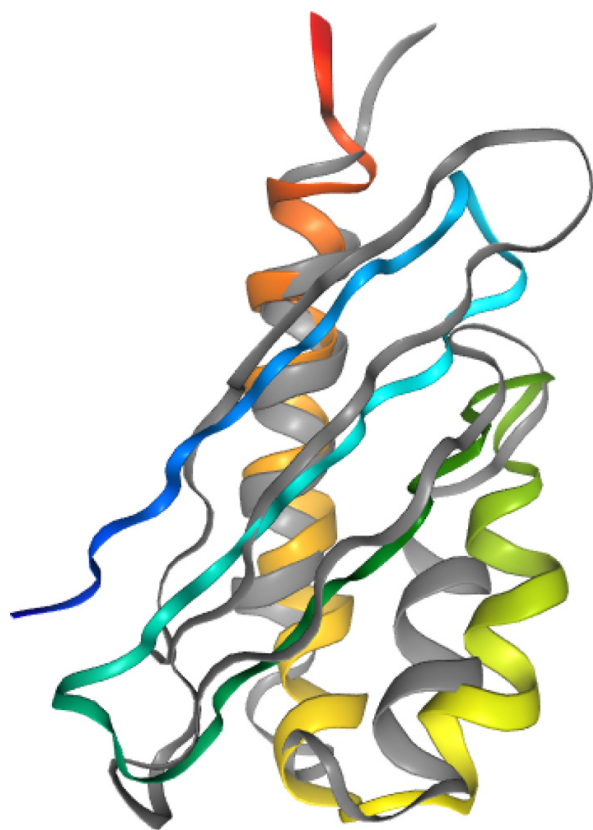


Figure 2. Overlap of the experimental 5G3Q:B structure (gray) of CASP12 target T0882 with UNRES server Model 1, rainbow-colored from the N-terminus (blue) to the C-terminus (red).

lated SAXS data in the form of the Gaussian-smoothed distance distribution calculated from the experimental structure showing that very good agreement between the distance distribution calculated from the models and the input distribution is achieved; (v) the central portion of Factor H, modules 10–15 starting from the NMR structure (PDB code: 2KMS structure) with including the experimental distance distribution from SAXS [the SASDA25 entry of the SASBDB database (33)]. Examples (iii) and (v) are discussed in more detail in section ‘Selected test cases’.

Selected test cases

CASP12 target T0882. The calculation has been run with secondary-structure restraints from PSIPRED (34) obtained at the CASP12 time when the experimental 5G3Q structure was not included in the PDB database. The FF2 variant of the UNRES force field (25,26) was used. Figure 2 displays the overlap of UNRES server model 1 and the experimental 5G3Q:B structure. The C^α-RMSD of that model from the experimental structure is 3.6 Å and the GDT.TS is 58.8%.

Example of including experimental SAXS restraints in simulations. This example pertains to the central portion of Factor H, which has been solved by NMR (35) (PDB code: 2KMS). This is a relatively small two-domain β-sheet protein. The distance distribution calculated from the exper-

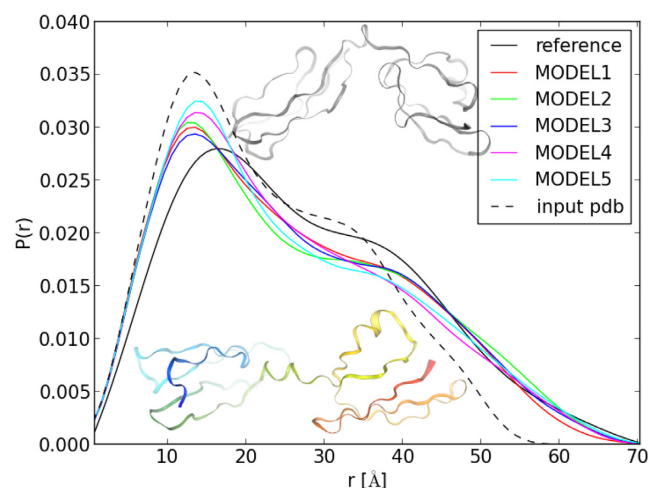


Figure 3. Comparison of the distance distribution from SAXS measurements of the central portion of Factor H, modules 10–15 with those calculated from the experimental 2KMS structure and from the five UNRES server models. The 2KMS experimental structure (top, gray) and UNRES model 1 [bottom, rainbow-colored from the N- (blue) to the C-terminus (red)] are superposed on the graph.

imental structure does not fully overlap with that from SAXS; in particular, the plot calculated from the 2KMS structure decays to 0 quicker than that from SAXS. By running REMD with UNRES and with the SAXS distance-distribution and secondary-structure restraints, much better agreement, in particular in the long-distance part is obtained (Figure 3). The reason for this better agreement is that, in the UNRES-calculated structure, the domains are at a larger angle than those in the experimental structure (Figure 3).

SERVER ARCHITECTURE

The UNRES server is equipped with the web interface written using Django Python Web framework. At the time of submission the server checks the correctness of data provided by the user and reports possible errors. For accepted jobs, the server prepares UNRES input files and submits jobs to the queue on our local cluster. A total of 42 nodes of the cluster are assigned to run server jobs. Upon pressing the ‘Refresh’ button or every 30 s, the server checks if the job is finished and reports the percentage of accomplishment. At last, the results are post-processed (which includes making a movie of an MD trajectory, constructing the plots, etc.) and displayed graphically.

Third-party software employed in the server are (i) pymol (36), (ii) convvpdb.pl from the MMTSB Tool Set (37), (iii) AMBER tools (24), (iv) PULCHRA (22), (v) SCWRL (23), NGL Viewer (38), which is employed as an interactive molecular viewer on the web page.

DATA AVAILABILITY

The UNRES server is available at <http://unres-server.chem.ug.edu.pl>. This website is free and open to all users and there is no login requirement. However, there is optional registration; registered users can access old jobs from their

homepage in the server website without having to copy the link corresponding to a particular job. The source code of the server is available from group GIT repository at mmka.chem.univ.gda.pl/repo/django_unres.

ACKNOWLEDGEMENTS

Computational resources were provided by (i) the Informatics Center of the Metropolitan Academic Network (CI TASK) in Gdańsk, (ii) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM), University of Warsaw (grant GA-71-23), (iii) the Polish Grid Infrastructure (PL-GRID), (iv) our cluster at the Faculty of Chemistry, University of Gdańsk.

FUNDING

National Science Center of Poland (Narodowe Centrum Nauki) (NCN) [UMO-2017/25/B/ST4/01026, UMO-2017/26/M/ST4/00044]. Funding for open access charge: NCN [UMO-2017/25/B/ST4/01026, UMO-2017/26/M/ST4/00044].

Conflict of interest statement. None declared.

REFERENCES

- Tramontano, A. (2005) *The Ten Most Wanted Solutions in Protein Bioinformatics*. Taylor & Francis, Boca Raton.
- Shaw, D.E., Maragakis, P., Lindorff-Larsen, K., Piana, S., Dror, R.O., Eastwood, M.P., Bank, J.A., Jumper, J.M., Salmon, J.K., Shan, Y. *et al.* (2010) Atomic-level characterization of the structural dynamics of proteins. *Science*, **330**, 341–346.
- Kmiecik, 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.
- Liwo, A., Czaplewski, C., Ołdziej, S., Rojas, A.V., Kaźmierkiewicz, R., Makowski, M., Murarka, R.K. and Scheraga, H.A. (2008) Simulation of protein structure and dynamics with the coarse-grained UNRES force field. In: Voth, G., (ed). *Coarse-graining of Condensed Phase and Biomolecular Systems*. CRC Press, Boca Raton, pp. 1391–1411.
- Liwo, A., Baranowski, M., Czaplewski, C., Gołaś, E., He, Y., Jagieła, D., Krupa, P., Maciejczyk, M., Makowski, M., Mozolewska, M.A. *et al.* (2014) A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. *J. Mol. Model.*, **20**, 2306.
- 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.
- Khalili, M., Liwo, A., Rakowski, F., Grochowski, P. and Scheraga, H.A. (2005) Molecular dynamics with the united-residue model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode. *J. Phys. Chem. B*, **109**, 13785–13797.
- Khalili, M., Liwo, A., Jagielska, A. and Scheraga, H.A. (2005) Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model α -helical systems. *J. Phys. Chem. B*, **109**, 13798–13810.
- Hansmann, U.H.E. and Okamoto, Y. (1994) Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. *Physica A*, **212**, 415–437.
- Rhee, Y.M. and Pande, V.S. (2003) Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.*, **84**, 775–786.
- Czaplewski, C., Kalinowski, S., Liwo, A. and Scheraga, H.A. (2009) Application of multiplexing replica exchange molecular dynamics method to the UNRES force field: tests with α and $\alpha + \beta$ proteins. *J. Chem. Theory Comput.*, **5**, 627–640.
- He, Y., Mozolewska, M.A., Krupa, P., Sieradzan, A.K., Wirecki, T.K., Liwo, A., Kachlishvili, K., Rackovsky, S., Jagiela, D., Slusarz, R. *et al.* (2013) Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14936–14941.
- Krupa, P., Mozolewska, M., Wiśniewska, M., Yin, Y., He, Y., Sieradzan, A., Ganzynkiewicz, R., Lipska, A., Karczyńska, A., Ślusarz, M. *et al.* (2016) Performance of protein-structure predictions with the physics-based UNRES force field in CASP11. *Bioinformatics*, **32**, 3270–3278.
- Mozolewska, M., Krupa, P., Zaborowski, B., Liwo, A., Lee, J., Joo, K. and Czaplewski, C. (2016) Use of restraints from consensus fragments of multiple server models to enhance protein-structure prediction capability of the UNRES force field. *J. Chem. Inf. Model.*, **56**, 2263–2279.
- Maisuradze, G.G., Senet, P., Czaplewski, C., Liwo, A. and Scheraga, H.A. (2010) Investigation of protein folding by coarse-grained molecular dynamics with the UNRES force field. *J. Phys. Chem. A*, **114**, 4471–4485.
- Rojas, A.V., Liwo, A. and Scheraga, H.A. (2007) Molecular dynamics with the united-residue force field: ab initio folding simulations of multichain proteins. *J. Phys. Chem. B*, **111**, 293–309.
- Sieradzan, A., Hansmann, U.H.E., Scheraga, H.A. and Liwo, A. (2012) Extension of UNRES force field to treat polypeptide chains with d-amino acid residues. *J. Chem. Theory Comput.*, **8**, 4746–4757.
- Chinchio, M., Czaplewski, C., Liwo, A., Ołdziej, 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.
- 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.
- Liwo, A., Khalili, M., Czaplewski, C., Kalinowski, S., Ołdziej, S., Wachucik, K. and Scheraga, H.A. (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.
- Ward, J. (1963) Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.*, **58**, 236–244.
- Rotkiewicz, P. and Skolnick, J. (2008) Fast procedure for reconstruction of full-atom protein models from reduced representations. *J. Comput. Chem.*, **29**, 1460–1465.
- 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.
- Case, D.A., Darden, T.A., Cheatham, T.E., Simmerling, C.L., Wang, J., Duke, R.E., Luo, R., Walker, R.C., Zhang, W., Merz, K.M. *et al.* (2012) *AMBER 12*. University of California, San Francisco.
- He, Y., Xiao, Y., Liwo, A. and Scheraga, H.A. (2009) Exploring the parameter space of the coarse-grained UNRES force field by random search: selecting a transferable medium-resolution force field. *J. Comput. Chem.*, **30**, 2127–2135.
- Sieradzan, A.K., Krupa, P., Scheraga, H.A., Liwo, A. and Czaplewski, C. (2015) 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**, 817–831.
- Krupa, P., Hałabis, A., Żmudzińska, W., Ołdziej, S., Scheraga, H.A. and Liwo, A. (2017) Maximum likelihood calibration of the UNRES force field for simulation of protein structure and dynamics. *J. Chem. Inf. Model.*, **57**, 2364–2377.
- Gay, D.M. (1983) Algorithm 611. Subroutines for unconstrained minimization using a model/trust-region approach. *ACM Trans. Math. Softw.*, **9**, 503–524.
- Rakowski, F., Grochowski, P., Lesyng, B., Liwo, A. and Scheraga, H.A. (2006) Implementation of a symplectic multiple-time-step molecular dynamics algorithm, based on the united-residue mesoscopic potential energy function. *J. Chem. Phys.*, **125**, 204107.
- Tuckerman, M., Berne, B.J. and Martyna, G.J. (1992) Reversible multiple time scale molecular dynamics. *J. Chem. Phys.*, **97**, 1990–2001.

31. McGuffin,L., Bryson,K. and Jones,D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics*, **16**, 404–405.
32. Karczynska,A.S., Mozolewska,M.A., Krupa,P., Giełdoń,A., Liwo,A. and Czaplowski,C. (2017) 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.
33. Valentini,E., Kikhney,A.G., Previtali,G., Jeffries,C.M. and Svergun,D.I. (2015) SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.*, **43**, D357–D363.
34. Jones,D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
35. Schmidt,C.Q., Herbert,A.P., Mertens,H.D.T., Guariento,M., Soares,D.C., Uhrin,D., Rowe,A.J., Svergun,D.I. and Barlow,P.N. (2010) The central portion of Factor H (modules 10-15) is compact and contains a structurally deviant CCP module. *J. Mol. Biol.*, **395**, 105–122.
36. Schrödinger LLC (2010) The PyMOL molecular graphics system, version 1.3r1. <https://pymol.org/2/support.html>.
37. Feig,M., Karanicolas,J. and Brooks,C.L. III (2004) MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.*, **22**, 377–395.
38. Rose,A.S. and Hildebrand,P.W. (2015) NGL Viewer: a web application for molecular visualization. *Nucleic Acids Res.*, **43**, W576–W579.