Postprint of: Antoniak A., Biskupek I., Bojarski K. B., Czaplewski C., Giełdoń A., Kogut M., Kogut M.M., Krupa P., Lipska A.G., Liwo A., Lubecka E.A., Marcisz M., Maszota-Zieleniak M., Samsonov S.A., Sieradzan A.K., Ślusarz M.J., Ślusarz R., Wesołowski P.A., Zięba K., Modeling protein structures with the coarse-grained UNRES force field in the CASP14 experiment, Journal of Molecular Graphics and Modelling, Vol. 108 (2021), 108008, DOI: 10.1016/j.jmgm.2021.108008

Modeling protein structures with the coarse-grained UNRES force field in the CASP14 experiment

Anna Antoniak[a], Iga Biskupek[a], Krzysztof K. Bojarski[a], Cezary Czaplewski[a], Artur Giełdoń[a], Mateusz Kogut[a], Małgorzata M. Kogut[a], Paweł Krupa[b], Agnieszka G. Lipska[a], Adam Liwo[a,c,*], Emilia A. Lubecka[d], Mateusz Marcisz[a,e], Martyna Maszota-Zieleniak[a], Sergey A. Samsonov[a], Adam K. Sieradzan[a], Magdalena J. Ślusarz[a], Rafał Ślusarz[a], Patryk A. Wesołowski[a,e], Karolina Zięba[a]

[a] *Faculty of Chemistry, University of Gdańsk, Wita Stwosza 63, 80-308 Gdańsk, Poland*
[b] *Institute of Physics, Polish Academy of Sciences, Aleja Lotników 32/46, Warsaw, PL-02668, Poland*
[c] *School of Computational Sciences, Korea Institute for Advanced Study, 87 Hoegiro, Dongdaemun-gu, 130-722 Seoul, Republic of Korea*
[d] *Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, G. Narutowicza 11/12, 80-233 Gdańsk, Poland*
[e] *Intercollegiate Faculty of Biotechnology, University of Gdańsk and Medical University of Gdańsk, ul. Abrahama 58, 80-307 Gdańsk, Poland*

Abstract

The UNited RESidue (UNRES) force field was tested in the 14th Community Wide Experiment on the Critical Assessment of Techniques for Protein Structure Prediction (CASP14), in which larger oligomeric and multimeric targets were present compared to previous editions. Three prediction modes were tested (i) *ab initio* (the UNRES group), (ii) contact-assisted (the UNRES-contact group), and (iii) template-assisted (the UNRES-template group). For most of the targets, the contact restraints were derived from the server models top-ranked by the DeepQA method, while the DNCON2 method was used for 11 targets. Our consensus-fragment procedure was used to run template-assisted predictions. Each group also processed the Nuclear Magnetic Resonance (NMR)- and Small Angle X-Ray Scattering (SAXS)-data assisted targets. The average Global Distance Test Total Score (GDT_TS) of the 'Model 1' predictions were 29.17, 39.32, and 56.37 for the UNRES, UNRES-contact, and UNRES-template predictions, respectively, increasing by 0.53, 2.24, and 3.76, respectively, compared to CASP13. It was also found that the GDT_TS of the UNRES models obtained in *ab initio* mode and in the contact-assisted mode decreases with the square root of chain length, while the exponent in this relationship is 0.20 for the UNRES-template group models and 0.11 for the best performing AlphaFold2 models, which suggests that incorporation of database information, which stems from protein evolution, brings in long-range correlations, thus enabling the correction of force-field inaccuracies.

Keywords: protein structure prediction, multiscale modeling, UNRES force field, free and bioinformatics-assisted modeling, replica-exchange molecular dynamics

1. Introduction

Computer simulations of proteins and their complexes play an increasing role in biophysics, biochemistry, and biomedicine, having such practical applications as drug design [1, 2, 3, 4]. Despite the development of high-speed computers, especially the ANTON supercomputer dedicated to run Molecular Dynamics (MD) simulations [5, 6]

*Corresponding author
Email address: adam.liwo@ug.edu.pl (Adam Liwo)

and Graphics Processing Unit (GPU)-based computers [7, 8], long-time all-atom simulations are possible only for relatively small proteins, comprising about 100 residues. Therefore, multiscale modeling is usually the method of choice, in which different parts of the system are treated at appropriate resolution. An important part of the multiscale approach is coarse-grained modeling, in which groups of atoms are merged into extended interaction sites [9]. Owing to the elimination of the fast degrees of freedom, coarse graining enables us to extend the time scale of simulations by several orders of magnitude [10].

The quality of the force field is a key issue, especially if long-term protein dynamics and protein-protein or protein-ligand binding is concerned. This is especially important when using coarse-grained models, because reduction of representation usually implies more sophisticated interaction potentials [11, 12], which are not so easy to parameterize. In particular, the site-site interaction potentials of most coarse-grained force fields are usually too "sticky", this resulting in too compact modeled structures [13, 14, 15, 16, 17]. Assessment of the force field is, therefore, necessary. One of the hardest tests is that of the ability of a force field to reproduce the native structures of proteins. The Community Wide Experiments on the Critical Assessment of Techniques for Protein Structure Prediction (CASP), conducted since 1992 every other year, enable protein-structure modelers to test their approaches with the proteins, the structures of which had not been solved at prediction time and, consequently, provide an impartial and unbiased test. Knowledge-based approaches, especially the AlphaFold approach developed by DeepMind [18, 19], which scored tremendous success in CASP13 and effectively solved the problem of protein-structure prediction for single-domain proteins in CASP14, have the upper hand in the protein-structure prediction as such. Nevertheless, the CASP experiments are an ideal means to test the force fields.

In the last years, we have been developing the physics-based UNited RESidue (UNRES) force field for studying the structure, dynamics, and thermodynamics of proteins and protein complexes [20, 21, 22, 23, 24, 25, 26]. UNRES is a highly-reduced physics-based model with only two interaction sites per residue. Its effective energy function is defined as a cluster-cumulant expansion of the potential of mean force of a protein in water [21, 25]. The solvent is implicit in UNRES and interactions with it are included in the effective potentials. Recently, we developed a scale-consistent theory of force-field derivation, owing to which the atomic details of a system are implicitly included in the resulting coarse-grained effective energy function [12, 25]. This feature solves part of the problem of force-field "stickiness" pointed out in ref. 17, because the force field contains explicit terms that couple the backbone-local and backbone-electrostatic conformational states, which prevent too compact local chain fragments [26]. On the other hand, part of the "stickiness" problem in UNRES, as well as in other coarse-grained force fields, is likely to be caused by the absence of explicit terms that account for the transfer of an interaction site from the solvent to protein inside. This problem is now being addressed in our laboratory. The upgraded UNRES force field was calibrated with 9 training proteins of various structural classes [26]; this recent version has been termed the NEWCT-9P force field. This force field has already been tested in CASP13, demonstrating significant improvement over the previous versions of UNRES in the *ab initio*, as well as bioinformatics- and data-assisted prediction [27, 28, 29, 30, 31]. We used our prediction protocol [27], which is based on Multiplexed Replica Exchange Molecular Dynamics (MREMD) [32] simulations with UNRES [33]. The results of CASP13 demonstrated that UNRES already performs reasonably well, the main problem, compared to knowledge-based approaches, being its coarser resolution. Therefore, before CASP14, we focused on developing UNRES to treat very large systems and not on upgrading the force field. A concise description of UNRES and its recent modifications is provided in section S-1 of the Supplementary Material.

In this paper we report the performance of UNRES extended to large protein systems with the NEWCT-9P force field in the CASP14 blind-prediction experiment in the *ab initio*, contact-assisted, and template-assisted modes. In each mode, we also carried out Nuclear Magnetic Resonance (NMR)- and Small Angle X-Ray Scattering (SAXS)-data-assisted predictions, which were very limited this time. In what follows we describe the prediction methodology that we used in CASP14 (section 2) and the results obtained (section 3). The conclusions from the study are summarized in section 4.

2. Methods

2.1. Prediction protocol

We used the protocols for protein-structure prediction without bioinformatics input [27, 30] and with using the consensus fragments derived from server predictions [28, 34]. The procedure consists of five stages. In stage 1,

restraints to assist prediction are prepared. The restraint-penalty functions have been described in detail in our earlier work [30, 31, 35] and are also summarized in section S-2 of the Supplementary Material. In stage 2, the UNRES implementation [22, 33] of MREMD [36], with geometry restraints defined in stage 1, if applicable, is used to search the conformational space of the target protein. The simulation procedure used to handle the CASP14 targets is briefly described in subsection 2.2.

In stage 3, the conformations from the last section of the MREMD run are processed by using the Weighted Histogram Analysis Method (WHAM) [37] and their relative free energies are calculated. The purpose of these calculations is to enable us to estimate the probabilities of the obtained conformations to occur at a desired temperature (note that a conformational ensemble resulting from MREMD simulations contains conformation simulated at all replica temperatures). The information from WHAM is carried over to stage 4, in which, first, the probabilities of the conformations from the simulated ensemble are calculated at a selected temperature (usually 260 K or 280 K). The ensemble is subsequently subjected to Ward's minimum-variance clustering [38] into 5 families. For each cluster (family), the conformation of the cluster that is closest to cluster center is selected as a candidate prediction. The models are ranked according to the probabilities of the clusters they are selected from. The probability of a cluster is defined as the sum of the probabilities of all conformations that belong to it. Consequently, the free energies of the respective clusters are included in candidate-prediction ranking. The procedure of WHAM calculations and cluster analysis has been described in detail in our earlier work [22].

In stage 5, the five coarse-grained models selected and ranked in stage 4 are converted to all-atom structures, by using the Protein Chain Reconstruction Algorithm (PULCHRA) [39] and Side-Chains with Rotamer Library (SCWRL) [40] algorithms and refined by running a sequence of 500 minimization steps, a short (0.3 ps) canonical MD run, and 500 final minimization steps, by using the Assisted Model Building with Energy Refinement, version 2014 (AMBER14) package [41] with the ff14SB force field and Generalized Born Surface Area (GBSA) implicit-solvent model. The resulting refined all-atom structures are submitted to CASP. The protocol was applied to both monomeric and oligomeric targets, the difference being only in the generation of the initial structures.

Prediction type is defined by the restraints determined in stage 1. In the CASP14 exercise, we ran unassisted, contact-assisted and server-model-assisted predictions and, consequently, participated as three groups: the UNRES group (group 360), the UNRES-contact group (group 096), and the UNRES-template group (group 018), respectively. The UNRES group used only weak secondary-structure restraints determined by the Predict Secondary Structure (PSIPRED) server [42] as implemented in our earlier work [30], the UNRES-contact group used the secondary-structure restraints and the residue-residue contact restraints, which were determined from selected server models and, for 11 targets, by the DNCON2 method [43], and the UNRES-template group used the restraints from the consensus fragments derived from the server models. For this purpose, the server models were selected based on their quality assessed by the by means of the DeepQA server [44] and fragments of similar structure were picked to derive the distance and angle restraints [28, 34]. Additionally, the knowledge-based Dynamic Fragment Assembly (DFA) pseudopotentials [45, 46] were also determined, which were added to the energy function. There were only three data-assisted targets, of which only the structure of one NMR-assisted target (T1088) was solved by the completion of CASP14, the other two targets cancelled. All three UNRES-based groups treated the data-assisted targets. The respective experimental data are available from https://predictioncenter.org/download_area/CASP14/extra_experiments/.

A scheme illustrating the above-described protocol is shown in Figure 1. The UNRES model and other elements of the prediction protocol are described in detail in the references cited above and are also briefly described in sections S-1 and S-2 of the Supplementary Information, respectively.

2.2. Simulation details (stage 2 of the prediction protocol)

For all targets, except the three data-assisted targets (S1063, N1077, and N1088), we ran temperature MREMD [36]. For H1081 and T1099, canonical MD simulations of 1,000,000 and 200,000 steps, respectively, were run for the UNRES and UNRES-contact groups due to time and resource constraints. MREMD is an extension of REMD [47]. In REMD, a number of independent trajectories are run at different temperatures and, after a given number of steps (every 10,000 MD steps in this work), the temperatures are exchanged between trajectories according to the Boltzmann criterion. The exchange occurs if a high-energy structure has a low bath temperature and a low-energy structure has a higher bath temperature, this enabling the high-energy structure to get out of the region it is stuck in, and enabling the low-energy structure to explore finer details of the low-energy region of the conformational space. In MREMD several replicas are run at each temperature to enable efficient exploration of the conformational space [36].

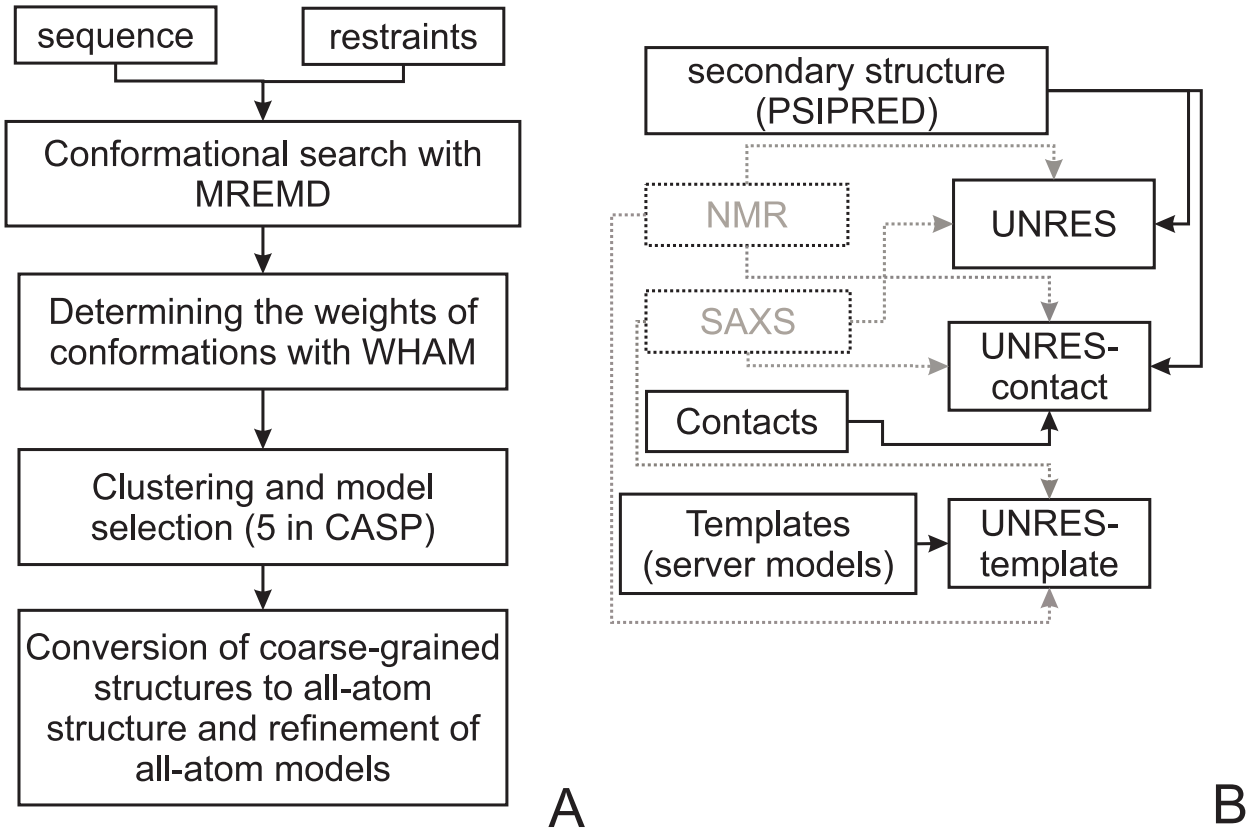

Figure 1. Scheme of prediction of CASP14 targets by the UNRES, UNRES-contact, and UNRES-template groups. Panel A summarizes the calculation flow and panel B shows the restraints used by the respective UNRES-based groups. The arrows in panel B connect the restraints to the groups which applied them. The experimental restraints (SAXS and NMR) are printed in grey and the corresponding boxes and arrows are dashed lines, because these restraints were provided for three CASP14 targets only.

In this work, the Langevin-dynamics implementation with UNRES [10, 48] with the variable times step (VTS) [48] and adaptive multiple time step algorithm (A-MTS) [49] was used to run MD trajectories. The replicas were run at 12 temperatures (260, 272, 279, 284, 288, 291, 294, 298, 308, 322, 341, and 370 K), 4 trajectories per temperature (48 trajectories total) per target. Replicas were exchanged every 10,000 MD steps. The spacing of replica temperatures was determined by using the Hansmann algorithm [50], which maximizes the walk of replicas in the temperature space and was found to be optimal for most of the proteins. Each trajectory consisted of 20,000,000 steps, 4.89 fs step length (97.8 ns UNRES time, amounting to about 0.1 ms real time [10]). Water friction was scaled by the factor of 0.01 in the computation of the friction and stochastic forces, as in our earlier work [10]. Snapshots were collected every 10,000 steps and the snapshots from the second half of simulations were taken for analysis, which assured convergence [22, 26].

The energy function was a sum of the UNRES energy and appropriate penalty terms discussed in section S-2 of the Supplementary Material, as given by eq. (1).

$$U = U_{UNRES} + w_{dih}V_{dih} + w_{cont}V_{cont} + w_{template}V_{template} + e_{DFA} + w_{SAXS}V_{SAXS} + w_{NMR}V_{NMR} \quad (1)$$

where U_{UNRES} is the UNRES energy, V_{dih} is the energy of backbone-virtual-bond-dihedral-angle restraints (that control the secondary structure), V_{cont} is the contact-restraint energy, $V_{template}$ is the template-restraint energy, e_{DFA} is the energy of the DFA pseudo-potentials, V_{SAXS} is the SAXS-restraint energy, V_{NMR} is the NMR-restraint energy, and the ws are the weights of the respective terms. For the definition of the restraint terms, see section S-2 of the Supplementary Material. The weights $w_{template} = 0$ and $e_{DFA} = 0$ for the UNRES and UNRES-contact groups, additionally, $w_{cont} = 0$ for the UNRES groups, $w_{dih} = w_{cont} = 0$ for the the UNRES-template group, $w_{SAXS} = 0$ for all targets but S1063 and $w_{NMR} = 0$ for all targets but N1077 and N1088.

For the data-assisted targets, Hamiltonian replica exchange was run, with replicas differing in the weight of the NMR or the SAXS term [eq. (1)]. In these simulations, in addition to running replicas at different temperatures to enable the conformations heat up or cool down, the weight(s) of the restraint(s) are varied from replica to replica. Consequently, if a trajectory has a large weight of the restraint potential and is stuck in a conformation that satisfies part of restraints but is very far from the minimum of the restraint function, while another one has a low restraint-potential weight but is close to satisfy all restraints, exchanging the restraint weights enables the first one to get out of the unfavorable region, while the other one to get closer to satisfying the restraints. For S1063, 4 Hamiltonian replicas were assigned to each of the 12 temperatures listed above, with $w_{SAXS} = 0$, 143.62, 379, and 1000, respectively, the statistical weights in WHAM processing being calculated with $w_{SAXS} = 1000$. For N1077 and N1088, 8 Hamiltonian replicas were assigned to each temperature, the weights being $w_{NMR} = 0$, 0.0634678, 0.18963, 0.295432, 0.444649, 0.653766, 0.824766, and 1, respectively, the statistical weights of the conformations being calculated for $w_{NMR} = 1$. The weights were selected to maximize the number of walks in the weight space, by using a variant of the Hansmann algorithm [50], adapted to the weight space.

For most of the monomeric targets, the initial structures for simulations for all groups were randomly picked from the server models with top DeepQA scores selected to generate restraints for the UNRES-template groups (section S-2.4 of the Supplementary Material). The server models were, in turn, selected from the 'stage 2' models generated by the servers participating in CASP14, which were available 72 hours after the release of a target. Additionally, we generated up to 10 models by our in-house installation of the I-TASSER (Iterative Threading ASSEmbly Refinement) knowledge-based protein-structure modeling software from the Zhang group [51]. All models were scored by using the DeepQA server [44] and ranked from the highest to the lowest DeepQA score. The numbers of models used to start simulations are collected in Table S1 of the Supplementary Material. Each initial structure was energy minimized with the Limited Memory Broyden-Fletcher-Goldfarb-Shanno (LBFGS) [52] algorithm. If the final energy was higher than 10^5 kcal/mol, the next model was tried. For T1029 and T1033, randomly-generated structures were used to start MREMD simulations by the UNRES and UNRES-contact groups (Table S1).

For the oligomeric targets, the HHPred server [53] was used to find the oligomers of homologous proteins and the initial structures were built based on these templates, as in our previous work [30, 31]. The oligomers of weakly homologous targets were assembled from monomer models (directly from servers or from UNRES-based predictions), by using the Symmetry Assembler (SAM) program from the Grudinin group [54] or manually, by using PyMol [55], based on optimal packing or external information of the function of a given protein. Two special cases were H1081 and T1099, which were the largest oligomeric targets.

H1081 is a 20-mer, each monomer comprising 758 residues, 15,160 residues total. The monomer sequence is highly homologous to that of acid induced arginine decarboxylase from *E. coli* (PDB: 2VYC), which forms a decamer. The 20-mer initial models of H1081 were built by modeling the decamer first, by means of SwissModeller [56]. Subsequently, two decamers were stacked over each other so as to make contacts between residues and rotated by 5, 10, 15, and 20 degrees. Two additional models were created by positioning the decamers side-by-side. These initial models were energy-minimized and four of them, which had reasonably low energies, were selected for simulations.

The duck hepatitis virus capsid (target T1099) comprises 240 subunits, 262 residues each, 62,880 residues total. The virus capsid has the T=4 icosahedral symmetry (https://viralzone.expasy.org/808), according to the Caspar and Klug system [57]. The smallest subunit is a tetramer. We selected Zhang-server model 1, which generated the least overlaps when assembled into capsid subunits, to construct the initial model of the capsid. In the first step of capsid-model construction, four monomers were superposed on the the respective tetramer from the known structure of the human hepatitis B virus (PDB: 3J2V). The distances between the center of the masses of the monomers were increased to avoid clashes. The structure was then energy minimized with the UNRES force field. Subsequently, a decamer was built by adding neighboring units. Because the C-terminal parts severely overlapped in the decamers, the C-terminal section, from residue 171 on, which was largely disordered in the server models, was randomly re-generated to avoid overlaps and subjected to local energy minimization. Finally, the central tetramer was extracted from the decamer and its 60 copies were superposed on the 3J2V capsid structure to generate the initial capsid model. Two variants of the capsid model, differing in the geometry of the randomly-generated C-terminal section, were constructed. Both models were energy minimized and subjected to conformational search. Because, for this target, the smallest unit containing all interactions had to be submitted, which was a hexamer, the capsid models obtained in simulations were cut into 60 overlapping hexamers, which were together subjected to cluster analysis.

The resources used varied depending on target size. For targets up to 200 amino-acid residues, the 20,000,000 MD steps took up to 24 wall-clock hours with 576 cores (12 cores per trajectories). For targets with up to 500 residues, about 6 wall-clock days were required with these resources. For the largest target, T1099, 200,000 MD steps required about a 2 days of calculations with 24 cores (for this target we ran only canonical MD). The latter timing corresponds to about 1 ns of UNRES time or 1 μs of real time (about 500 ns/day), taking into account the at least 1,000-fold time-scale extension of coarse-grained time-scale compared to all-atom time-scale [10]. It should be noted that all-atom simulations of a related hepatitis virus took 30 ns/day with 2016 cores of CRAY X6E [58]. We used Cray XC40 of the Interdisciplinary Center of Mathematical and Computer Modeling of the University of Warsaw, ICM, the Intel Xeon E5 v3 cluster at the Centre of Informatics – Tricity Academic Supercomputer & Network (CI TASK) in Gdansk, and the Intel Xeon cluster at the Academic Computer Centre Cyfronet AGH in Krakow.

2.3. Comparison of the models with the experimental structures

The Global Distance Test Score (GDT_TS) [59, 60] is a primary measure used in CASP to compare the structures of the models of proteins or their domains with the respective experimental structures. It is the average of the percentage of residues in the computed structure that are within 1 Å, 2 Å, 4 Å, and 8 Å distance, respectively, from their counterparts in the experimental structure, as given by eq. (2). The GDT_TS measure was used throughout this paper to discuss the quality of UNRES group models.

$$GDT\_TS = \frac{GDT_{P1} + GDT_{P2} + GDT_{P4} + GDT_{P8}}{4} \tag{2}$$

where GDT_{Pn} is the percentage of the C^{α} atoms whose distance from the C^{α} atoms of the experimental structure is below the n Å cut-off. The α-carbon Root-Mean-Square Deviation (C^{α}-RMSD or RMSD) was also used for comparing models with the experimental structures; it is defined by eq. (3).

$$RMSD = \min_{\mathbf{t},\mathbf{R}} \sqrt{\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{X}_{Ti} - (\mathbf{R}\mathbf{X}_M + \mathbf{t})_i\|^2} \tag{3}$$

where $\mathbf{X}_M$ and $\mathbf{X}_T$ are the coordinates of the model (M) and the target (T) structures, n being the number of the C^{α} atoms, whereas $\mathbf{R}$ and $\mathbf{t}$ are the rotation matrix and the translation vector of the model with respect to the experimental structure, which are determined to minimize the RMSD.

The primary measures of the similarity of the models of oligomeric proteins to the respective experimental structures, which are used since CASP12 [61, 62], are the Interface Patch Similarity (IPS), and Interface Contact Similarity (ICS) [61], respectively, which are expressed as the Jaccard coefficient and the $F1$ score, respectively [61]. The IPS is the ratio of the number of interface-patch residues common to the model and to the target and that of the number of all interface residues that occur in the model and in the target [eq. (4)] [61].

$$IPS(M,T) = J_C(\mathbf{I}_M, \mathbf{I}_T) = \frac{|\mathbf{I}_M \cap \mathbf{I}_T|}{|\mathbf{I}_M \cup \mathbf{I}_T|} \tag{4}$$

where $\mathbf{I}_M$ and $\mathbf{I}_T$ are the sets of residues in the interface patch of the model and of the target, respectively, and $|\ldots|$ denotes the number of elements in a set. The ICS [eq. (5)] is the harmonic mean of precision (P; the percentage of the correct interchain contacts among all interchain contacts in the model [eq. (6)], and recall [R; the percentage of correctly reproduced native interchain contacts, eq. (7)].

$$ICS(M,T) = F1(P,R) = \frac{2 \times P(M,T) \times R(M,T)}{P(M,T) + R(M,T)} \times 100\% \tag{5}$$

$$P(M,T) = \frac{|\mathbf{C}_M \cap \mathbf{C}_T|}{|\mathbf{C}_M|} \times 100\% \tag{6}$$

$$R(M,T) = \frac{\mathbf{C}_M \cap \mathbf{C}_T}{|\mathbf{C}_T|} \times 100\% \tag{7}$$

where $\mathbf{C}_M$ and $\mathbf{C}_T$ are the sets of interface contacts present in the model and in the target, respectively.

The other measures of comparison of the models of oligomeric proteins to the respective experimental structure are the global (over the whole oligomer) RMSD and GDT_TS, and the interface RMSD (I-RMSD), which is computed from eq. (3) with the set of superposed atoms reduced to those present in the protein-protein interfaces in the experimental structure.

3. Results and Discussion

All three UNRES-based groups submitted models of 63 regular single-chain targets, which were not cancelled. We processed a total 16 (out of 22 total) oligomeric targets, the UNRES group submitting the models of 15, the UNRES-contact group of 10, and the UNRES-template group of 16 of them. All three groups submitted the models of all 3 data-assisted targets, 5 models of each; however, the SAXS assisted target (S1063) was cancelled and the models of the first NMR-assisted target (N1077) were not evaluated due to not resolving the structure in time. The single-chain targets comprised 93 domains, of which 23 were of easy Template Based Modeling (TBM-easy) type, 31 of hard Template Based Modeling (TBM-hard) type, 15 on the border between Free Modeling and Template Based Modeling (FM/TBM), and 23 of Free Modeling (FM) type, respectively. Multidomain proteins (MultiDom) constituted a set of 10 targets. Additionally, 9 more proteins consisted of multiple domains and the submitted models of their whole structures were also evaluated in CASP. Therefore, in addition to analyzing our results obtained for the the MultiDom category, we analyzed the results of all 19 multidomain proteins, grouping them into the D0 category (which also includes the MultiDom targets). The domains and the entire proteins are together denoted by the term Evaluation Units (EUs) [63]. Altogether, our predictions comprised 112 EUs.

In the subsequent subsections of this section we describe the results of the UNRES, UNRES-contact, and UNRES-template groups. In all comparisons, we used the values of the measures of the model-target similarity downloaded from the official CASP14 page (https://predictioncenter.org/casp14/index.cgi) under the "Parseable data" link.

3.1. Regular single-chain targets

3.1.1. General results

The violin plots of the GDT_TS distributions of the first models, where the maximum, the minimum, and the mean values, as well as the approximate distributions are displayed, for all EU categories (TBM-easy, TBM-hard, FM/TBM, FM, MultiDom, and D0), are shown in Figure 2. The respective plots for the best models and all models are shown in Figure S2A and S2B and detailed results are summarized in Table S2 of the Supplementary Material,

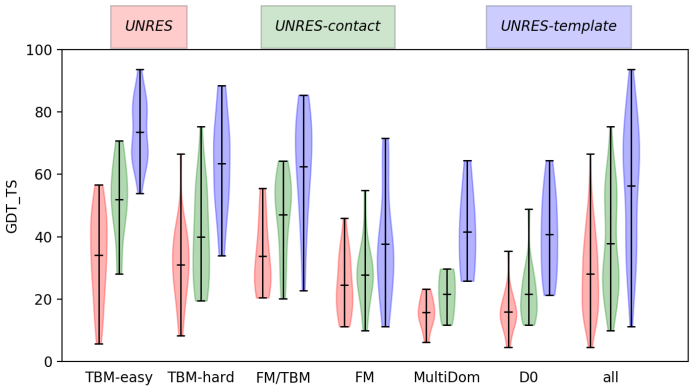

Figure 2. Violin plots of the GDT_TS of the first models of the respective CASP14 EU types and all EUs for the UNRES, UNRES-contact, and UNRES-template groups. The mean, minimum, and the maximum values in the violin plots are shown as black horizontal lines, the GDT_TS ranges are shown as vertical lines, and approximate distributions are shown as color-filled shapes.

respectively. Because the UNRES group models of targets T1046s1 and T1046s2 were erroneously submitted twice as the UNRES and as the UNRES-template models, respectively, these two targets are omitted from the analysis of the UNRES-template group results.

As could be expected, the distributions are shifted towards higher GDT_TS in the order $\text{GDT\_TS}_{\text{UNRES}} < \text{GDT\_TS}_{\text{UNRES-contact}} < \text{GDT\_TS}_{\text{UNRES-template}}$. It can, therefore, be concluded that the model quality increases with increasing the content of knowledge-based information, including the case of FM-modeling targets. It can also be noted that, for all groups, the GDT_TS decreases in the order $\text{GDT\_TS}_{\text{TBM-easy}} > \text{GDT\_TS}_{\text{TBM-hard}} \approx \text{GDT\_TS}_{\text{FM/TBM}}$. For the UNRES and UNRES-contact groups, $\text{GDT\_TS}_{\text{FM}} > \text{GDT\_TS}_{\text{MultDom}} \approx \text{GDT\_TS}_{\text{D0}}$, this meaning that the model quality decreases for multidomain targets compared to single-domain ones. Thus, model accuracy decreases with increasing target difficulty for these two groups. Conversely, for the UNRES-template group, the GDT_TS is slightly greater for the MultiDom and D0 targets, compared to the FM targets.

3.1.2. Comparison with CASP13

The differences between the average GDT_TS values of the first, best, and all models in CASP14 from those obtained in CASP13 by the three groups [the present UNRES group corresponding to the CASP13 UNRES group (group 288), the UNRES-contact group to the CASP13 wf-BAKER-UNRES group (group 492) and the UNRES-template group to the CASP13 KIAS-Gdansk group (group 208)] are shown in Figure S3A – S3C of the Supplementary Material. Because there were only 2 multidomain CASP13 targets, which were assigned to a separate category (FM-sp), and could be matched to the CASP14 MultiDom category, we compared the average GDT_TS of all multidomain CASP14 targets (the D0 category assigned in this work) with their CASP13 counterparts. Even though the CASP13 targets were different and had presumably different degree of difficulty than those of CASP14, comparing the mean GDT_TS values from both experiments gives some insight on the progress of the prediction procedure.

It can be seen from Figure S3 that there is no major change in the UNRES group model quality except for the 'Model 1' predictions of the FM/TBM domains ($\Delta_{\text{GDT\_TS}}$=3.12, 77% significance by Student's test). Overall, $\Delta_{\text{GDT\_TS}}$ is 0.53 (63 % significance) and 0.94 (70 % significance) for the first and the best UNRES models, respectively, and 0.09 (55 % significance) averaged over all models. This small overall GDT_TS increase, is understandable, because the UNRES force field has not been upgraded since CASP13. The increase was likely caused by starting the conformational search (which was carried out by MREMD) from multiple server models in CASP14 and from one server model only in CASP13.

For the UNRES-contact group, the average GDT_TS has remarkably increased only for the TBM-easy domains, with $\Delta_{\text{GDT\_TS}}$ of 2.52 (78 % significance), 3.71 (90 % significance) for the first and the best models, respectively, and by 4.02 (100 % significance) for all models. Small increase in GDT_TS is also observed for the TBM-hard

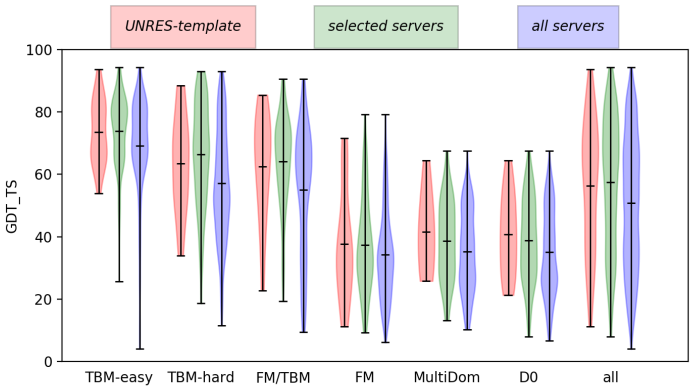

Figure 3. Violin plots of the GDT_TS of the first UNRES-template, selected-server, and all server models of the respective EU types and all EUs of the single-chain proteins. The mean, minimum, and the maximum values are shown as black horizontal lines, the GDT_TS ranges are shown as vertical lines, and approximate distributions are shown as color-filled shapes.

domains (by 0.95, 0.91, and 1.11 for the first and the best models and on average, respectively), while some decrease is observed for the FM/TBM and FM targets except for the all-model-averaged GDT_TS of the FM/TBM domains. Overall, GDT_TS increased by 2.24 (85 % significance), 2.83 (91 % significance) and 3.07 (100 % significance) for the first, best models, and all models, respectively.

The greatest increase of GDT_TS has occurred for the UNRES-template group (by 3.76, 92 % significance, 3.18, 89 % significance, and 3.65, 100 % significance) for the first, best, and all models, respectively. The GDT_TS decreased with respect to CASP13 only for the FM category domains (3.18, 76 % significance, by 2.03, 69 % significance, and by 2.90, 94 % significance for the first, best, and all models, respectively). Still, UNRES-template performs significantly better than UNRES and UNRES-contact for the FM domains (Figures 2 and S2). Consistent increase in average GDT_TS, for the first, the best, and all models, is observed for all remaining target categories.

3.1.3. Comparison of UNRES-template models with server models

Because the UNRES-template predictions had heavy input from the server models, we compared the average GDT_TS values from the UNRES-template models with those from the server models selected to derive restraints and those from all server models. The respective violin plots of GDT_TS are presented, along with the errorbar plot of the differences, with boxes indicating the statistical significances of the differences, in Figure 3 and Figure S4C of the Supplementary Material (first models) and in Figure S4A and S4B of the Supplementary Material (best models and all models), respectively. Detailed plots for selected targets are shown in Figure S6A – S6D (TBM-easy and TBM-hard EUs), S7A – S7C (FM/TBM and FM EUs), and S8A – S8B (multidomain proteins) of the Supplementary Material.

Analyzing the results averaged over all models, which have reasonable statistical significance, it can be seen from Figure 3 and from Figure S4 that the GDT_TS values of the UNRES-template models are smaller than those for the selected server models for the TBM-easy (by 1.14, 85 % significance), TBM-hard (by 1.98, 89 % significance), and FM/TBM domains (by 1.67, 79 % significance), while they are larger for the FM domains (by 1.12, 76 % significance), for the MultiDom (by 2.24, 87 % significance) and for the D0 targets (by 1.68, 74 % significance). The results are qualitatively similar for the 'Model 1' and best-model predictions; however, they have lower statistical significance due to smaller sample sizes. The net improvement over the server models for these two targets categories is consistent with the results of our previous studies [28, 29, 31, 64] and is a consequence of the physics-based character of UNRES, which enables it to correct wrong templates and achieve correct packing. For all categories, the average GDT_TS values are greater for the UNRES-template models than those averaged over all server models, which means that our server-model-selection procedure enables us to pick good models to derive restraints.

The differences of the GDT_TS values in the CASP14 and CASP13 experiments for the UNRES-templates, all server, selected server and for all target categories are shown as errorbar plots in Figure S5A – S5C of the Supplementary Material, the boxes representing their statistical significances superposed. It can be seen that, overall, the

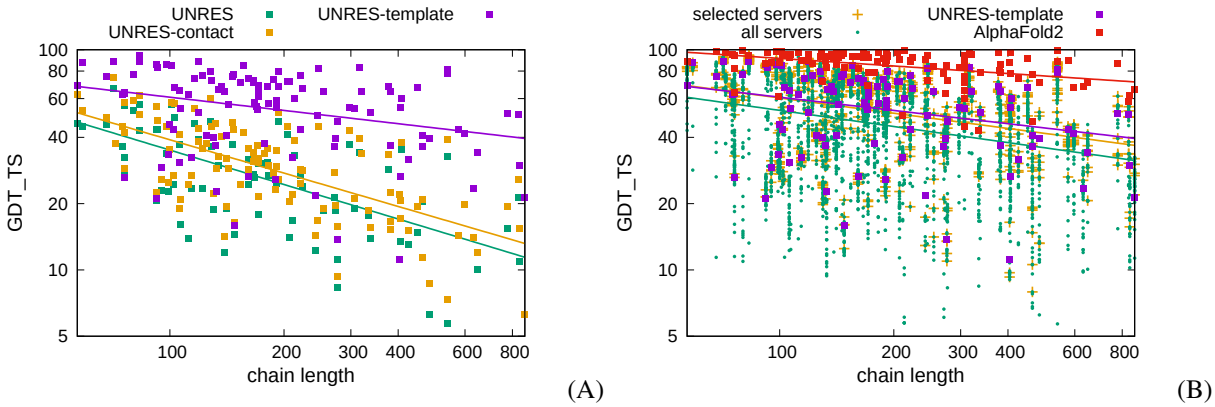

Figure 4. Plots of the GDT_TS of the 'Model 1' predictions of the CASP14 EUs obtained by the UNRES, UNRES-contact, and UNRES-template groups (A) and those from the server models selected to derive the restraints for UNRES-template predictions, all servers, UNRES-template, and AlphaFold2 (the winner group) (B). The logarithmic scale is used on both axes. The straight lines correspond to fitting the power function [eq. (8)] to the respective points.

GDT_TS of the first, best, and all models from the selected and from all servers increased significantly compared to CASP13, by about 5 and 15 units, respectively, the statistical significance exceeding 90 %. The greater increase for all servers compared to that of selected servers demonstrates that server-model quality has overall improved since CASP13. For all servers, the average GDT_TS also increased in all categories, independent of whether the first, the best or all models are considered. For selected servers, the GDT_TS increased for all target categories except for FM and D0; for the last category UNRES-template has slight GDT_TS increase. Together with higher average GDT_TS of the UNRES-template models of the MultiDom- and D0-type targets, compared to that of selected-server models (Figure S4C of the Supplementary Material), this suggests that UNRES contributes the added value mostly to multidomain targets.

3.1.4. Dependence of model quality on chain length

To find out how the model quality changes with chain length and how does this dependence relates to that of the server models and the AlphaFold2 group, which outperformed all other groups in CASP14, we analyzed the plots of GDT_TS of the first models vs. chain length, shown in Figure 4A (the three UNRES-based groups) and B (AlphaFold2, servers, and UNRES-template). The relationships are similar for other models. The GDT_TS values corresponding to all EUs were used to make these plots. It can be seen from Figure 4A that, albeit the points are scattered, they are grouped around straight lines in the log-log scale, which suggests a power relationship, as given by eq. (8).

$$GDT\_TS_X(N) = b_X \times N^{-a_X} \tag{8}$$

where GDT_TS_X is the GDT_TS value for an EU model of CASP14 group X, N is the chain length, and a_X and b_X are coefficients. The coefficients obtained after least-squares fitting, together with their standard deviations, are summarized in Table S3 of the Supplementary Material.

For the UNRES and for the UNRES-contact group the weighted percentage of structure within the three RMSD cut-offs included in GDT_TS (1, 4, and 8 Å) drops approximately with the square root of the chain length, this corresponding to a random-walk behavior. This relationship suggests that the errors in structure propagate without any correction and the introduction of contact information only uniformly increases the GDT_TS, compared to UNRES. Conversely, the double-logarithmic plots of $GDT\_TS(N)$ for the UNRES-template group and server models have more than twice as low slopes than those of UNRES and UNRES-contact, which suggests that the evolution information inherent in templates results in error correction. As could be expected, the least slope is observed for the AlphaFold2 models. However, even in this case, GDT_TS drops with chain length. On the other hand, because the accuracy of

experimental structures is also usually lower for larger protein, the non-zero slope could result from the error inherent in the determination of the experimental structures.

3.1.5. Rankings of UNRES-based groups

From section 3.1.2 it follows that of the three UNRES-based groups the UNRES-template group progressed most significantly (in terms of GDT_TS) with respect to CASP13, except for the FM targets. Less significant progress of the UNRES and UNRES-contact groups resulted from the fact that the UNRES force field was not developed since CASP13, due to focusing our effort on upgrading UNRES to handle large targets. Because the other groups used machine-learning methods, the ranking of the UNRES group is only 120 (82 % of all the groups) for the first models of all targets and 126 (84 % of all groups) for the best models. The UNRES-contact group ranks 111 (76 %) and 114 (85 %) for the first and the best models, respectively. For the FM EUs, the rankings are better: 89 (64 %) and 97 (69 %) for the first and the best UNRES models and 90 (64 %) and 94 (67 %) for the first and the best UNRES-contact models, respectively. For the UNRES-template group, the rankings are 31 (21 %) and 34 (23 %) for the first and the best models overall, and 32 (23 %) for both the first and the best FM EU models, respectively. As could be expected from the GDT_TS analysis presented in section 3.1.3, the UNRES-template models of the MultiDom EUs have the highest ranks, which are 22 (16 %) and 29 (21 %) for the first and the best model, respectively.

3.1.6. Prediction examples

In this section we discuss three examples of UNRES-based predictions of three regular targets, whose experimental structure have already been released: T1026 (TBM-hard), T1038 (TBM-easy), and T1090 (FM).

Target T1026 (a monomer of the Faba bean necrotic stunt virus; PDB: 6S44) is a β-sandwich single-domain protein. The T1026-D1 evaluation unit contains residues from 27 to 172. For this target, all UNRES-template models have higher GDT_TS compared to the parent templates (Figure S6C). The UNRES-template model 3, with GDT_TS=78.60, is ranked 7th among all 517 submitted models. The GDT_TS plots of the UNRES-template, UNRES-contact, UNRES, and other group models and the best models of the UNRES-based groups are shown in Figure S9A of the Supplementary Material. Although the best UNRES-contact and UNRES models do not rank that well (219 and 294, respectively with GDT_TS of 52.40 and 46.58, respectively), they exhibit the correct topology of the β-sandwich, only the curvature or packing of its two halves being slightly different than in the experimental structure, this resulting in lower a GDT_TS and a higher RMSD (Figure 5A).

Target T1034, the BIL2 domain from *T. thermophila* BUBL1 locus (PDB: 6TMM) is a 156-residue β-structure protein. The T1043-D1 evaluation unit contains all 156 residues of this protein. Although the UNRES-template models do not have the highest GDT_TS compared to the selected-server models (Figure S6A), the average is still above that of selected-server models. The GDT_TS of UNRES-template model 2 is 85.74 and it ranks 19th among all 559 models of this target. The UNRES-template model 1 ranks 13th with GDT_TS=84.94 among 122 'Model 1' predictions. The best UNRES-contact model (model 1) ranks 44th with GDT_TS=62.34 and the best UNRES model (model 1) ranks 505th with GDT_TS=42.31. The GDT_TS plots and the experimental structure and the best UNRES-based groups models are shown in Figure S9B and in Figure 5B, respectively. It can be seen from Figure 5B that the β-sheet topology is gradually distorted when passing from the UNRES-template to the UNRES model, to become largely different from that in the experimental structure, which is reflected in high RMSD in the UNRES model.

The free-modeling target T1090 (the *S. pombe* chromatin remodelling protein, PDB: 7K7W) has the α/β-barrel fold. The T1090-D1 evaluation unit comprises residues from 2 to 192. As in the case of T1034, some of the selected server models have higher GDT_TS than the UNRES-template models; however, the average over UNRES-template model is higher than that over selected-server models (Figure S7C). The GDT_TS plots and best UNRES-based models are shown in Figure S9C and in Figure 5C, respectively. The best UNRES-template model (model 4, GDT_TS=60.71, RMSD=3.75 Å) ranks 39th out of 517 models submitted for this target, model 1 of this group ranks 16th among all 113 'Model 1' predictions. The best UNRES-contact model (model 1, GDT_TS=32.13) ranks 374th and the best UNRES model 1, (GDT_TS=34.79) ranks 368th. It can be seen that the barrel topology is destroyed in both UNRES and UNRES-contact models.

3.2. Regular oligomeric targets

The composition, difficulty, and method of initial-model generation of the oligomeric targets processed by the UNRES-based groups are summarized in Table S4 of the Supplementary Material. These targets were divided into 23

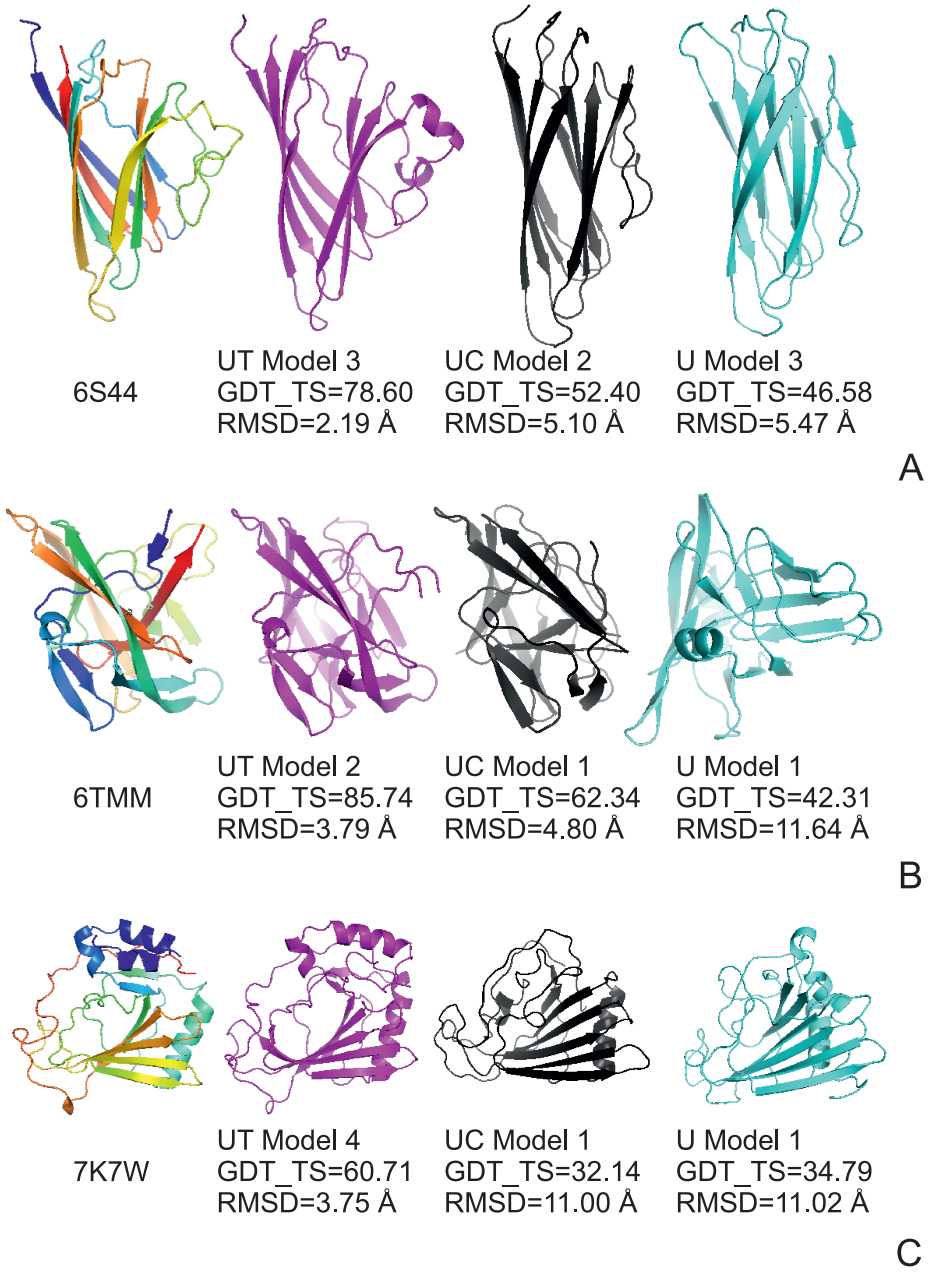

Figure 5. From left to right: the experimental, UNRES-template (UT), UNRES-contact (UC), and UNRES (U) models of targets T1026-D1 (A), T1034-D1 (B), and T1090-D1 (C). The PDB IDs of the experimental structures, the model numbers, the GDT_TS and C^{α}-RMSD values of the models are shown below the respective panels. The experimental structures are rainbow-colored from the N- to the C-terminus, the models are colored according to the corresponding GDT_TS-line colors in Figure S9 of the Supplementary Material.

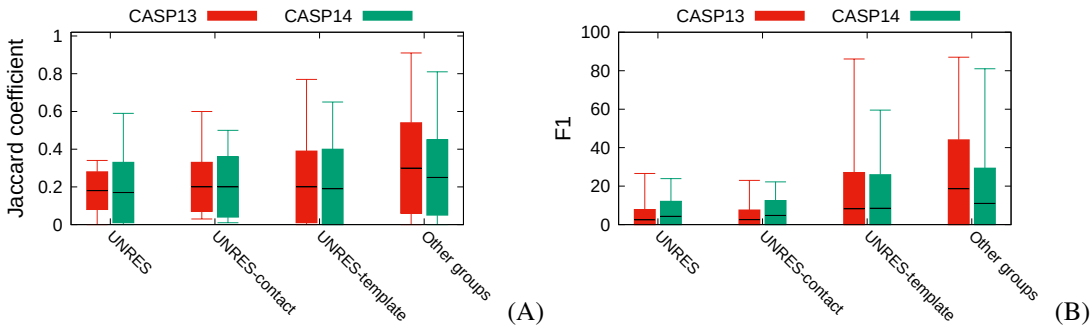

Figure 6. Candlestick plots of the Jaccard coefficient [eq. (4) of section 2.3] (A) and $F1$ score [eq. (5) of section 2.3] (B) of the first models of the oligomeric targets obtained in CASP13 (red) and CASP14 (green) by the UNRES-based groups and all other groups. The lower and upper horizontal black lines correspond to the maximum and the minimum value, the central black lines are the mean values and the boxes are the ranges of the standard deviations.

EUs. The candlesticks plots of the Interface Patch Similarity and Interface Contact Similarity, expressed as the Jaccard coefficient [J_C; eq. (4)] and the $F1$ score [eq. (5)], respectively [61], averaged over all, first, and the best models of the oligomeric targets, respectively, obtained by the UNRES, UNRES-contact, UNRES-template, and other groups, respectively, are shown in Figure 6A and B and in Figure S10A – S10D of the Supplementary Material, respectively. The numerical values of these two measures for the consecutive targets are collected in Table S5 of the Supplementary Material. Because of a small number of targets, we did not do the plots for the respective target categories.

It can be seen from Figure 6 and from Figure S10 that, in both CASP13 and CASP14, the mean values of the J_C and the $F1$ scores are comparable for all UNRES-based groups. The average scores calculated over all groups are also comparable to those calculated over the UNRES-based models. However, the maximum values are greater for the UNRES-template group compared to the UNRES and the UNRES-contact group and the maximum values from other groups are greater than those from the UNRES-template group. It can also be noted that the $F1$ and J_C values corresponding to the UNRES-template group and to other groups are significantly lower compared to those of CASP13, which can be explained by the greater difficulty of the CASP14 target compared to the CASP13 targets, regarding both total number of residues and assembly complexity [65] (Table S5).

Of the UNRES-based groups, UNRES-template ranked best in the oligomeric-target-prediction category. For the first models, the rankings are 8 (33 % of all participating groups) in the easy-, 14 (40 %) in the medium-, 22 (67 %) in the hard-target category, and 15 (38 %) for all targets, respectively. For the best models, the rankings are 11 (46 %), 16 (46 %), 22 (67 %), and 17 (44 %) for easy, medium, hard, and all targets, respectively. The overall ranks of the UNRES group are 24 (62 %) for the first and 23 (59 %) for the best models, and those of the UNRES-contact group are 26 (67 %) for the first and 25 (64 %) for the best models, respectively.

As follows from Table S5, some of the predictions ranked very well. For the medium-difficulty target T1087o, which is a homodimer composed of two antiparallel-packed helical hairpins, each having additionally a long polyproline helix, UNRES and UNRES-template ranked 4 (model 1) and 6 (model 1), respectively. For the homotetrameric target T1073, model 5 of the UNRES group ranked 8. For the heterodimeric target H1047, UNRES model 2 ranked 10. Because the experimental structures of these proteins have not yet been released, we do not show the pictures of our models. Instead, as the illustration of the capacity of the new upgraded UNRES to handle large system, in Figure 7, we show the structure of the UNRES-template model of whole duck hepatitis capsid (target T1099o), compared with the respective experimental structure (PDB: 6VGH). For this target, only the subunits were evaluated and UNRES-template ranked 60 for unit T1099ov2 (model 3), 73 for unit T1099v0 (model 1), and 75 for unit T1099ov1 (model 5), respectively, out of 136 submitted models total (Table S5).

3.3. Data-assisted targets

As mentioned, we treated 3 data assisted targets: S1063 (SAXS-assisted), N1077, and N1088 (both NMR-assisted). The structure of S1063 has not been solved and the target has been cancelled; however, we managed to

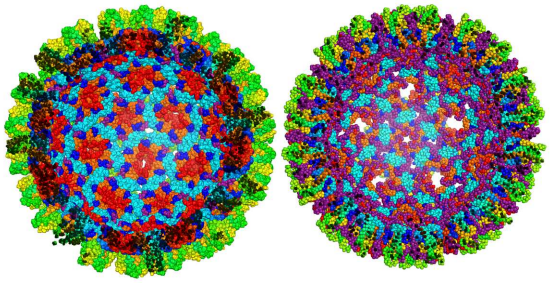

Figure 7. Experimental structure (left) and UNRES group model 1 (right) of target T1099o (duck hepatitis virus capsid, 260 mers). The chains are colored differently to show symmetry elements.

produce models, which are in good agreement with the SAXS data and, therefore, we present these results. The target is a homotetramer and most of the servers return a long α-helix with a small N-terminal, largely β-structure, part for a monomer. The initial structures for UNRES simulations were generated by manual assembly of the helices from tFold-IDT model 1. All structures consistent with the experimental SAXS data consist of long rod-like shapes, in which the pairs of dimers are connected by the N-terminal sections. The plots of the experimental distance distribution and those corresponding to the 5 models from the UNRES-template group are shown in Figure S11A, while the 'Model 1' prediction is shown in Figure S11B of the Supplementary Material.

Of the two NMR-assisted targets, the structure of N1088 has been solved but has not been published yet. A picture of the structure is available from the CASP14 web page under the "Results(Automatic Evaluation)→Assisted→NMR" tab. The structure is a 226-residue β-barrel, part of which seems to exist in two states [66], the N1088-D1 evaluation unit was distinguished, which consists of residues 1-30 and 90-226. The templates to derive restraints for the UNRES-template group and the starting models were taken from BAKER-ROSETTASERVER models 1–5, FEIG-S models 1, 2, 4, and 5, tFold-IDT models 3–5, and tFold models 3 and 5. The UNRES-template model 1 of the entire target ranked 2 (out of 9 groups) among the first models, with GDT_TS=61.06, while the best UNRES-template models (model 3), with GDT_TS=62.61, ranked 3 out of all 43 models. For N1088-D1, the UNRES-template model 1 ranked 1 (out of 9 groups), with GDT_TS=69.13, and the best UNRES-template model 2 ranked 2 out of 43 models, with GDT_TS=70.48. The UNRES-contact and UNRES groups did not fare that well, ranking 6 and 7, respectively, and 18 and 22, respectively, for the first and the best models, respectively. The GDT_TS graphs of the UNRES, UNRES-contact, and UNRES-template models in comparison with those of other groups are shown in Figure S12A of the Supplementary Material, while the best UNRES-template model is shown in panel B of the Figure.

4. Conclusions

We tested the ability of the UNRES force field to predict protein structures in unassisted and bioinformatics-assisted mode in the CASP14 experiment. In this CASP experiment, the number of experimental-data-assisted targets was very limited this time (1 target with structure solved by the completion of CASP14) and, therefore, UNRES could not be reasonably tested in this category. UNRES was upgraded to handle very large targets, which enabled us to simulate the 15,160-residue 20-mer (H1081) and whole virus capsid (T1099o; 68,800 residues).

Because the UNRES force field was not upgraded since CASP13, the performance of unassisted UNRES (the UNRES group) remained largely unchanged in terms of GDT_TS for the monomeric (Figures S3 of the Supplementary Material) and in terms of the $F1$ and J_C measures for the oligomeric targets (Figure 6 and Figure S10 of the Supplementary Material). The performance of contact-assisted UNRES (the UNRES-contact group) improved slightly for monomeric targets (Figure S3 of the Supplementary Material), owing to improved method of contact acquisition (from server models ranked according to DeepQA [44] or by using DNCON2 [43]). The largest improvement for single-chain targets was achieved by the server-model-assisted protocol (UNRES-template; Figure S3 of the Supplementary Material). However, this improvement is largely due to the improvement of the server models used to derive restraints (Figure 3 and Figure S4 of the Supplementary Material). The most significant improvement of the UNRES-template

group over the server models selected to derive restraints was achieved for multidomain proteins; for this category, the improvement over CASP13 was also very significant (Figure S3 of the Supplementary Material). Interestingly, a significant improvement over the server models was also achieved for the FM targets, while the performance of server-model-assisted UNRES in this category deteriorated since CASP13. This observation suggests that the FM targets were more difficult in CASP14 than in CASP13. Similarly, the overall performance of the UNRES-template group and of other groups on oligomeric targets was worse in terms of $F1$ and J_C, compared to CASP13 (Figure 6 and Figure S10 of the Supplementary Material), which probably results from larger size and greater complexity of the CASP14 oligomeric targets compared to the respective CASP13 targets [65].

One conclusion from this study is that UNRES needs to be improved in terms of local accuracy to achieve better resolution. The respective work is currently underway in our laboratory. On the other hand, the analysis of the dependence of GDT_TS on protein size shown in Figure 4 suggests that GDT_TS decreases with the square root of the number of residues for the UNRES and the UNRES-contact groups (a random-walk behavior). For the UNRES-template group and the server models, the GDT_TS decreases with the power of 0.2-0.24 of the number of residues, and the exponent is only 0.11 for the best-performing AlphaFold2 group. This observation strongly suggests that the force field alone is missing important long-range correlations, which are empirically included in knowledge-based methods through evolutionary information. Such correlations could emerge if the force field was detailed enough; however, this does not seem viable even if all-atom force fields are considered. Therefore, longer-range correlation terms in the cluster-cumulant expansion of the potential of mean force [21, 25] should probably be included in UNRES and other coarse-grained force fields that work in the *ab initio* mode to improve the quality of the modeled protein structures.

Acknowledgements

This work was supported by grants UMO-2017/25/B/ST4/01026 (to AL), UMO-2017/27/B/ST4/00926 (to AKS), UMO-2017/26/M/ST4/00044 (to CC), UMO-2018/30/E/ST4/00037 (to SAS), and UMO-2018/31/N/ST4/01677 (to KKB) from the National Science Center of Poland (Narodowe Centrum Nauki). Computational resources were provided by (a) the Interdisciplinary Center of Mathematical and Computer Modeling (ICM) the University of Warsaw under grants No. GA76-11, GB71-18, and GA76-17, (b) the Centre of Informatics - Tricity Academic Supercomputer & Network (CI TASK) in Gdańsk, (c) the Academic Computer Centre Cyfronet AGH in Krakow under grants: casp13unres, casp13unres2, and plggagi, and (d) our 796-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/xx.xxxx/j.jmgm.2019.xx.xxx.

[1] J. D. Durrant, J. A. McCammon, Molecular dynamics simulations and drug discovery, BMC Biology 9 (2011) 71.
[2] I. A. Vakser, Protein-protein docking: From interaction to interactome, Biophys. J. 107 (2014) 1785–1793.
[3] X. Lin, X. Li, X. Lin, A review on applications of computational methods in drug screening and design, Molecules 25 (2020) 1375.
[4] J. Wang, Fast identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study, J. Chem. Info. Model. 60 (2020) 3277–3286.
[5] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, C. Mcleavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, S. C. Wang, Anton, a special-purpose machine for molecular dynamics simulation, Commun. ACM 51 (2008) 91–97.
[6] K. Lindorff-Larsen, N. Trbovic, P. Maragakis, S. Piana, D. E. Shaw, Structure and dynamics of an unfolded protein examined by molecular dynamics simulation, J. Am. Chem. Soc. 134 (2012) 3787–3791.
[7] M. S. Friedrichs, P. Eastman, V. Vaidyanathan, M. Houston, S. Legrand, A. L. Beberg, D. L. Ensign, C. M. Bruns, V. S. Pande, Accelerating molecular dynamic simulation on graphics processing units, J. Comput. Chem. 30 (2009) 864–872.
[8] I. Bethune, R. Banisch, E. Breitmoser, A. B. K. Collis, G. Gibb, G. Gobbo, C. Matthews, G. J. Ackland, B. J. Leimkuhler, Mist: A simple and efficient molecular dynamics abstraction library for integrator development, Comp. Phys. Commun. 236 (2019) 224–236.
[9] S. Kmiecik, D. Gront, M. Koliński, L. Wieteska, A. Dawid, A. Koliński, Coarse-grained protein models and their applications, Chem. Rev. 116 (2016) 7898–7936.
[10] M. Khalili, A. Liwo, A. Jagielska, H. A. Scheraga, Molecular dynamics with the united-residue (UNRES) model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model α-helical systems, J. Phys. Chem. B 109 (2005) 13798–13810.
[11] G. Voth, Coarse-Graining of Condensed Phase and Biomolecular Systems, 1st Edition, CRC Press, Taylor & Francis Group, 2008.

[12] A. Liwo, C. Czaplewski, A. K. Sieradzan, E. A. Lubecka, A. G. Lipska, Ł. Golon, A. Karczyńka, P. Krupa, M. A. Mozolewska, M. Makowski, R. Ganzynkowicz, A. Giełdoń, M. Maciejczyk, Scale-consistent approach to the derivation of coarse-grained force fields for simulating structure, dynamics, and thermodynamics of biopolymers, in: B. Strodel, B. Barz (Eds.), Progress in molecular biology and translational science. Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly, Vol. 170, Academic Press, London, 2020, Ch. 2, pp. 73–122.
[13] A. C. Stark, C. T. Andrews, A. H. Elcock, Toward optimized potential functions for protein-protein interactions in aqueous solutions: Osmotic second virial coefficient calculations using the MARTINI coarse-grained force field, J. Chem. Theory Comput. 9 (2013) 4176–4185.
[14] P. S. Schmalhorst, F. Deluweit, R. Scherrers, C. P. Heisenberg, M. J. Sikora, Overcoming the limitations of the MARTINI force field in simulations of polysaccharides, J. Chem. Theory Comput. 13 (2017) 50395053.
[15] F. Fornasier, L. de Souza, F. Souza, F. Reynaud, A. Pimentel, The lipophilicity of coarse-grained cholesterol models, J. Chem. Inf. Model. 60 (2020) 569577.
[16] L. M. P. Souza, F. R. Souza, F. Reynaud, A. S. Pimentel, Tuning the hydrophobicity of a coarse grained model of 1,2-dipalmitoyl-sn-glycero-3-phosphatidylcholine using the experimental octanol-water partition coefficient, J. Mol. Liq. 319 (2020) 114132.
[17] F. R. Souza, L. M. Pereira Souza, A. S. Pimentel, Recent open issues in coarse grained force fields, J. Chem. Inf. Model. 60 (2020) 5881–5884.
[18] A. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Židek, A. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D. Jones, D. Silver, K. Kavukcuoglu, D. Hassabis, Improved protein structure prediction using potentials from deep learning, Nature 577 (2020) 706–710.
[19] E. Callaway, 'it will change averything': Ai makes gigantic leap in solving protein structures, Nature 588 (2020) 203–204.
[20] A. Liwo, S. Ołdziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, H. A. Scheraga, A united-residue force field for off-lattice protein-structure simulations. i. functional forms and parameters of long-range side-chain interaction potentials from protein crystal data, J. Comput. Chem. 18 (1997) 849–873.
[21] A. Liwo, C. Czaplewski, J. Pillardy, H. A. Scheraga, Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field, J. Chem. Phys. 115 (2001) 2323–2347.
[22] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, H. Scheraga, Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins, J. Phys. Chem. B 111 (2007) 260–285.
[23] A. Liwo, C. Czaplewski, S. Ołdziej, A. V. Rojas, R. Kaźmierkiewicz, M. Makowski, R. K. Murarka, H. A. Scheraga, Simulation of protein structure and dynamics with the coarse-grained UNRES force field, in: G. Voth (Ed.), Coarse-Graining of Condensed Phase and Biomolecular Systems, CRC Press, Boca Raton, 2008, Ch. 8, pp. 1391–1411.
[24] A. Liwo, M. Baranowski, C. Czaplewski, E. Gołaś, Y. He, D. Jagieła, P. Krupa, M. Maciejczyk, M. Makowski, M. A. Mozolewska, A. Niadzvedtski, S. Ołdziej, H. A. Scheraga, A. K. Sieradzan, R. Ślusarz, T. Wirecki, Y. Yin, B. Zaborowski, A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions., J. Mol. Model. 20 (2014) 2306.
[25] A. K. Sieradzan, M. Makowski, A. Augustynowicz, A. Liwo, 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 (2017) 124106.
[26] A. Liwo, A. K. Sieradzan, A. G. Lipska, C. Czaplewski, I. Joung, W. Żmudzińska, A. Hałabis, S. Ołdziej, 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 (2019) 155104.
[27] P. Krupa, M. Mozolewska, M. Wiśniewska, Y. Yin, Y. He, A. Sieradzan, R. Ganzynkowicz, A. Lipska, A. Karczyńska, M. Ślusarz, R. Ślusarz, A. Giełdoń, C. Czaplewski, D. Jagieła, B. Zaborowski, H. Scheraga, A. Liwo, Performance of protein-structure predictions with the physics-based unres force field in CASP11, Bioinformatics 32 (2016) 3270–3278.
[28] M. Mozolewska, P. Krupa, B. Zaborowski, A. Liwo, J. Lee, K. Joo, C. Czaplewski, 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 (2016) 2263–2279.
[29] A. Karczyńska, M. A. Mozolewska, P. Krupa, A. Giełdoń, K. K. Bojarski, B. Zaborowski, A. Liwo, R. Ślusarz, M. Ślusarz, J. Lee, K. Joo, C. Czaplewski, Use of the UNRES force field in template-based prediction of protein structures and the refinement of server models: Test with CASP12 targets, J. Mol. Graph. Model. 83 (2018) 92–99.
[30] E. A. Lubecka, A. S. Karczyńska, A. G. Lipska, A. K. Sieradzan, K. Zięba, C. Sikorska, U. Uciechowska, S. A. Samsonov, P. Krupa, M. A. Mozolewska, Ł. Golon, A. Giełdoń, C. Czaplewski, R. Ślusarz, M. Ślusarz, S. N. Crivelli, A. Liwo, Evaluation of the scale-consistent UNRES force field in template-free prediction of protein structures in the CASP13 experiment, J. Mol. Graph. Model. 92 (2019) 154–166.
[31] A. Karczyńska, K. Zięba, U. Uciechowska, M. A. Mozolewska, P. Krupa, E. A. Lubecka, A. G. Lipska, C. Sikorska, S. A. Samsonov, A. K. Sieradzan, A. Giełdoń, A. Liwo, R. Ślusarz, M. Ślusarz, J. Lee, K. Joo, C. Czaplewski, Improved consensus-fragment selection in template-assisted prediction of protein structures with the UNRES force field in CASP13, J. Chem. Inf. Model. 60 (2020) 1844–1864.
[32] V. S. Pande, I. Baker, J. Chapman, S. P. Elmer, S. Khaliq, S. M. Larson, Y. M. Rhee, M. R. Shirts, C. D. Snow, E. J. Sorin, B. Zagrovic, Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing, Biopolymers 68 (2003) 91–109.
[33] C. Czaplewski, S. Kalinowski, A. Liwo, H. A. Scheraga, Application of multiplexing replica exchange molecular dynamics method to the UNRES force field: Tests with α and $\alpha + \beta$ proteins, J. Chem. Theory Comput. 5 (2009) 627–640.
[34] A. S. Karczyńska, M. A. Mozolewska, P. Krupa, A. Giełdoń, A. Liwo, C. Czaplewski, Prediction of protein structure with the coarse-grained UNRES force field assisted by small X-ray scattering data and knowledge-based information, Proteins 86 (2018) 228–239.
[35] E. A. Lubecka, A. Liwo, Introduction of a bounded penalty function in contact-assisted simulations of protein structures to omit false restraints, J. Comput. Chem. 40 (2019) 2164–2178.
[36] Y. M. Rhee, V. S. Pande, Multiplexed-replica exchange molecular dynamics method for protein folding simulation, Biophys J. 84 (2003) 775–786.
[37] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, J. M. Rosenberg, The weighted histogram analysis method for free-energy calculations

on biomolecules. I. The method, J. Comput. Chem. 13 (1992) 1011–1021.
[38] F. Murtagh, A. Heck, Multivariate data analysis, Kluwer Academic Publishers, 1987.
[39] P. Rotkiewicz, J. Skolnick, Fast procedure for reconstruction of full-atom protein models from reduced representations, J. Comput. Chem. 29 (2008) 1460–1465.
[40] Q. Wang, A. A. Canutescu, R. L. Dunbrack, SCWRL and MolIDE: Computer programs for side-chain conformation prediction and homology modeling., Nat. Protoc. 3 (2008) 1832–1847.
[41] D. Case, D. Pearlman, J. Caldwell, T. Cheatham III, J. Wang, W. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Siebel, U. C. Singh, P. K. Weiner, P. A. Kollman, AMBER 14, University of California, San Francisco (2014).
[42] L. McGuffin, K. Bryson, D. T. Jones, The PSIPRED protein structure prediction server, Bioinformatics 16 (2000) 404–405.
[43] B. Adhikari, J. Hou, J. Cheng, DNCON2: improved protein contact prediction using two-level deep convolutional neural networks, Bioinformatics 34 (2018) 466–1472.
[44] R. Cao, D. Bhattacharya, J. Hou, J. Cheng, DeepQA: Improving the estimation of single protein model quality with deep belief networks, BMC Bioinformatics 17 (2016) 495.
[45] T. N. Sasaki, M. Sasai, A coarse-grained Langevin molecular dynamics approach to protein structure reproduction, Chem. Phys. Lett. 402 (2005) 102–106.
[46] T. N. Sasaki, H. Cetin, M. Sasai, A coarse-grained Langevin molecular dynamics approach to de novo protein structure prediction., Biochem. Biophys. Res. Commun. 369 (2008) 500–506.
[47] U. H. E. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, Chem. Phys. Lett. 281 (1997) 140–150.
[48] M. Khalili, A. Liwo, F. Rakowski, P. Grochowski, H. A. Scheraga, Molecular dynamics with the united-residue (UNRES) model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode, J. Phys. Chem. B 109 (2005) 13785–13797.
[49] F. Rakowski, P. Grochowski, B. Lesyng, A. Liwo, H. A. Scheraga, Implementation of a symplectic multiple-time-step molecular dynamics algorithm, based on the united-residue mesoscopic potential energy function, J. Chem. Phys. 125 (2006) 204107.
[50] S. Trebst, M. Troyer, U. H. E. Hansmann, Optimized parallel tempering simulations of proteins, J. Chem. Phys. 124 (2006) 174903.
[51] W. Zhang, J. Yang, B. He, S. E. Walker, H. Zhang, B. Govindarajoo, J. Virtanen, Z. Xue, H.-B. Shen, Y. Zhang, Integration of QUARK and I-TASSER for ab initio protein structure prediction in CASP11, Proteins 84 (1, SI) (2016) 76–86.
[52] D. C. Lui, J. Nocedal, On the limited memory BFGS method for large scale optimization, Math. Prog. 45 (1989) 503–528.
[53] L. Zimmermann, A. Stephens, S. Nam, D. Rau, J. Kübler, M. Lozajic, F. Gabler, J. Söding, A. Lupas, V. Alva, A completely reimplemented MPI bioinformatics toolkit with a new HHpred server at its core, J. Mol. Biol. 430 (2018) 2237–2243.
[54] D. W. Ritchie, S. Grudinin, Sperical polar Fourier assembly of protein complexes with arbitrary point groups symmetry, J. Appl. Cryst. 49 (2016) 158–167.
[55] Schrödinger, LLC, The PyMOL molecular graphics system, version 1.3r1, the PyMOL Molecular Graphics System, Version 1.3r1 (August 2010).
[56] A. Waterhouse, M. Bertoni, S. Bienert, D. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, Swiss-model: homology modelling of protein structures and complexes, Nucleic Acids Res. 46(W1) (2018) W296–W303.
[57] D. L. Caspar, A. Klug, Physical principles in the construction of regular viruses, Cold Spring Harb. Symp. Quant. Biol. 27 (1962) 1–24.
[58] D. S. D. Larsson, L. Liljas, D. van der Spoel, Virus capsid dissolution studied by microsecond molecular dynamics simulations, PLoS Comput. Biol. 8 (2012) e1002502.
[59] A. Zemla, LGA: a method for finding 3D similarities in protein structures, Nucleic Acids Res. 13 (2003) 3370–3374.
[60] J. Moult, K. Fidelis, A. Kryshtafovych, T. Schwede, A. Tramontano, Critical assessment of methods of protein structure prediction (CASP) round X, Proteins 82 (Suppl. 2) (2013) 1–6.
[61] A. Lafita, S. Bliven, A. Kryshtafovych, M. Bertoni, B. Monastyrskyy, J. M. Duarte, T. Schwede, G. Capitani, Assessment of protein assembly prediction in CASP12, Proteins 86 (2018) 247–256.
[62] D. Guzenko, A. Lafita, B. Monastyrskyy, A. Kryshtafovych, J. M. Duarte, Assessment of protein assembly prediction in CASP13, Proteins 87 (2019) 1100–1112.
[63] L. N. Kinch, A. Kryshtafovych, B. Monastyrskyy, N. V. Grishin, CASP13 target classification into tertiary structure prediction categories, Proteins 87 (2019) 1021–1036.
[64] G. A. Khoury, A. Liwo, F. Khatib, H. Zhou, G. Chopra, J. Bacardit, L. O. Bortot, R. A. Faccioli, X. Deng, Y. He, P. Krupa, J. Li, M. A. Mozolewska, A. K. Sieradzan, J. Smadbeck, T. Wirecki, S. Cooper, J. Flatten, K. Xu, D. Baker, J. Cheng, A. C. B. Delbem, C. A. Floudas, C. Kesar, M. Levitt, Z. Popović, H. A. Scheraga, J. Skolnick, S. N. Crivelli, Foldit Players, WeFold: A coopetition for protein structure prediction, Proteins 82 (2014) 1850–1868.
[65] Ezgi Karaca's presentation at the CASP14 conference https://www.predictioncenter.org/casp14/doc/presentations/.
[66] Gaetano Montelione's presentation at the CASP14 conference available at https://www.predictioncenter.org/casp14/doc/presentations/.