High Performance Computing with Coarse Grained Model of Biological Macromolecules

Emilia Lubecka[1,2], *Adam Sieradzan*[2], *Cezary Czaplewski*[2], *Paweł Krupa*[3], *Adam Liwo*[2]

The Unified Coarse Grained Model of biological macromolecules (UCGM) that is being developed in our laboratory is a model designed to carry large-scale simulations of biological macromolecules. The simplified chains representations used in the model allows to obtain 3-4 orders of magnitude extention of the time-scale of simulations, compared to that of all-atom simulations. Unlike most of the other coarse-grained force fields, UCGM is a physics-based force field, independent of structural databases and applicable to treat non-standard systems. In this communication, the efficiency and scalability of the new version of UCGM package in Fortran 90, with two parallelization levels: coarse-grained and fine-grained, is reported for systems with various size and oligomeric state. The performance was tested in the canonical- and replica exchange MD mode, with small- and moderate-size proteins and protein complexes (20 to 1636 amino-acid residues), as well as with large systems such as, e.g., human proteosome 20S with size over 6200 amino-acid residues, which show the advantage of using coarse-graining. It is demonstrated that, with using massively parallel architectures, and owing to the physics-based nature of UCGM, real-time simulations of the behavior of subcellular systems is feasible.

Keywords: Unified Coarse Grained Model, UNRES force field, molecular dynamics simulations, fine-grained parallelization, coarse-grained parallelization.

Introduction

The knowledge of the three-dimensional structure of macrobiomolecules is crucial for understanding many biological processes. Most biologically important molecules are too large to handle for the classical simulation tools. Therefore, coarse-grained models and force fields have become useful in these studies [36]. Reduced representations of molecules allow saving computational cost of a few orders of magnitude in comparison with full-atomistic simulations. Various types of coarse-grained models such as the SICHO and CABS models have been developed [12, 13]. These knowledge-based protein models, developed in the Kolinski group, use only three degrees of freedom per residue. Compared to coarse-grained models of proteins [11], coarse-grained models of saccharides have a relatively short history. Most of them use three or four interaction centers per sugar ring [25, 27, 33], including the polysaccharide component of the well-known MARTINI force field [23]. The MARTINI force field allows to simulate also other important biomolecules: lipids, proteins [28], DNA etc. [26]

The Unified Coarse Grained Model of biological macromolecules (UCGM) [18] that is being developed in our laboratory is physics-based model, designed to carry large-scale simulations of biological macromolecules: proteins (the UNRES model [19]), nucleic acids (the NARES-2P model [6]), polysaccharides (the SUGRES-1P model), and lipid membranes. The chain representation in UCGM is reduced to two interaction sites (united peptide groups and united side chains in UNRES, united phosphate groups and united sugar-base groups in NARES-2P) or one interaction site (united sugar rings in SUGRES-1P). In UCGM the effective energy function was derived via generalized-cumulant expansion [16] of the potential of mean force of polypep-

[1]Institute of Informatics, Faculty of Mathematics, Physics and Informatics, University of Gdańsk, Gdańsk, Poland
[2]Faculty of Chemistry, University of Gdańsk, Gdańsk, Poland
[3]Institute of Physics, Polish Academy of Sciences, Warsaw, Poland

tide chains [20, 35]. Therefore, it reproduces the structure, dynamics, and thermodynamics of biopolymers very well despite the significant reduction of the number of interaction sites.

UNRES, which is the most advanced component of UCGM and has the longest history of development, has succeeded in physics-based protein-structure predictions [7, 14], investigation of protein-folding [38] and studying biological processes such as amyloid formation [32], and iron-sulfur cluster biogenesis [29]. NARES-2P [6] can reproduce the structures and thermodynamics of small DNA and RNA molecules in free molecular-dynamics simulations and has also recently been applied with success to model the stability of telomere sequences of DNA [34]. The preliminary version of SUGRES-1P [18] can reproduce the helical structure of amylose and the sheet-like structure of cellulose.

The article is organized as follows. Section 1 describes the UNRES model and force field, and coarse-grained molecular dynamics used in our study. In Section 2, we present details of implementation UNRES with Fortran 90. In Section 3, we report the results and its analysis. Conclusion summarizes our study.

1. Theory

1.1. UNRES Model and Force Field

In the UNRES model (Fig. 1) [18], a polypeptide chain is represented as a sequence of α-carbon (C^α) atoms connected by virtual bonds (with different $C^\alpha \cdots SC$ bond lengths, d_{sc}), with united side chains (SC) attached to them and united peptide groups (p) positioned in the middle between the two consecutive C^αs. The SCs and ps are the only interacting sites, while the C^αs only assist in geometry definition. Backbone geometry of the simplified polypeptide chain is defined by the $C^\alpha \cdots C^\alpha \cdots C^\alpha$ virtual-bond angles θ (θ_i has the vertex at C_i^α) and the $C^\alpha \cdots C^\alpha \cdots C^\alpha \cdots C^\alpha$ virtual-bond-dihedral angles γ (γ_i has the axis passing trough C_i^α and C_{i+1}^α). The local geometry of the ith side-chain center is defined by the zenith angle α_i (the angle between the bisector of the respective angle θ_i and the $C_i^\alpha \cdots SC_i$ vector) and the azimuth angle β_i (the angle of counter-clockwise rotation of the $C_i^\alpha \cdots SC_i$ vector about the bisector from the $C_{i-1}^\alpha \cdots C_i^\alpha \cdots C_{i+1}^\alpha$ plane, starting from C_{i-1}^α). The energy function consists of local (including virtual-bond-deformation, virtual-bond-angle bending, virtual-bond-torsional and double torsional, and virtual-side-chain rotamer potentials), pairwise site-site interaction potentials, and correlation (multibody) potentials [19]. The solvent is implicit in the force field. UNRES has been derived based on the potential of mean force (PMF) of polypeptide chains in water, which has been factorized into Kubo cluster-cumulant functions [16], which are, in turn, identified with specific effective energy terms [20, 35]. Expansion of the cluster-cumulant functions into Kubo cluster cumulants [16] yielded approximate analytical expressions for the energy terms, including the correlation terms [20, 35]. The force field was calibrated with solution structural data of 7 traning proteins with various types of structure, by using the maximum-likelihood method recently developed in our laboratory [15].

Because of PMF-based origin of UNRES, the force field has the sense of free energy restricted to a given coarse-grained conformation rather than the potential energy. The temperature dependence was implemented in our earlier work [21]; it affects mostly the correlation terms, thereby contributing to weakening the regular structure elements with increasing temperature.
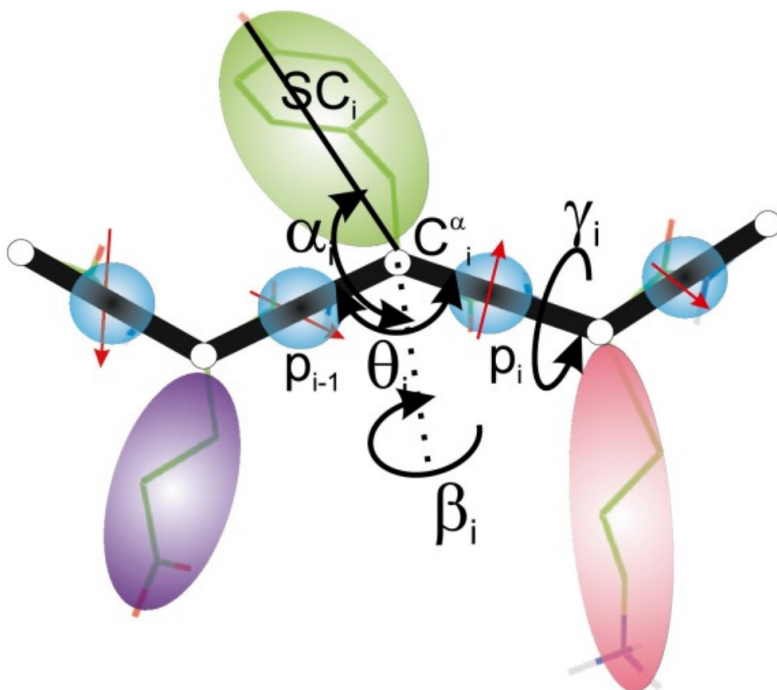
Figure 1. The UNRES model of polypeptide chains, with united peptide-groups (p)(blue circles), united side chains (SC)(ellipsoids) and α carbons (C^α)(small open circles)

1.2. Molecular Dynamics and Replica Exchange Molecular Dynamics with UNRES

Molecular dynamics is the core of the main conformational-search engine with UNRES. The equations of motion are Langevin equations because of implicit solvent treatment; however, for faster search, the simple Berendsen thermostat [2] (which originates from application of the Langevin equations to a system) is often used [8, 9]; the Nosé–Hoover and Nosé–Poincaré thermostats that enable more strict control of temperature conservation have also been implemented [10]. Because the virtual bonds are treated as stretchable rods, it was natural to select the $C^\alpha \cdots C^\alpha$ and $C^\alpha \cdots SC$ virtual-bond vectors as variables. With such a representation, the inertia matrix is a full, albeit constant, matrix. Recently, by selecting the C^α and SC coordinates as variables, we reduced it to a pentadiagonal form (J. Sans and A. Liwo, to be submitted), this resulting in linear and not quadratic scaling the memory requirements with system size. A version of adaptive multiple-time-step velocity-Verlet algorithm has been developed to integrate the equations of motion [8, 30].

For improved conformational search, replica exchange (REMD) [5] and multiplexed replica exchange (MREMD) [31] extensions of MD were implemented with UNRES [4]. In the MREMD method, a series of independent MD trajectories are run, each at a different temperature, with synchronization every pre-defined (usually 10,000 or 20,000) number of steps. At each synchronization point, the energies of the conformations from consecutive trajectories are compared and exchange of temperature occurs following the Metropolis criterion. This approach gives a chance for high-energy conformations stuck at low temperatures to escape from high-energy regions because of overcoming energy barriers upon heating and, conversely, to low-energy conformations simulated at high temperatures to cool down and achieve even lower energy. The

MREMD method differs from REMD in that more than one trajectory is simulated at a given temperature.

2. Implementation

2.1. Parallelization

UNRES was parallelized on two levels: coarse-grained and fine-grained [22]. At the coarse-grained level, each MD trajectory is handled by a given task or task group. This means trivial parallelization scheme when many canonical MD trajectories are run or task-farm-type parallelization scheme for REMD/MREMD runs, with the master processor controlling the synchronization and exchange of replicas and the slave processor running only MD trajectories between the synchronization points. The REMD/MREMD UNRES jobs cannot be run in a serial mode and the number of cores per job must be equal to the number of trajectories if coarse-grained-parallelization only is executed or a multiple of the number of trajectories for fine-grained parallelization. Because the management work is a small fraction of trajectory computations and the speed of calculations of all trajectories is virtually the same, the master processor also computes a MD trajectory. The efficiency of coarse-grained parallelization is close to 100 %.

To treat large systems, fine-grained parallelization has been developed [22] in which energy and force evaluation has been split between the tasks handling a given trajectory. The task-group master distributes coordinates to the slave tasks and collects the energy and force components from them; it also does its share of energy and force calculations.

2.2. Implementation with Fortran 90

Originally, UNRES was written in Fortran 77. Because of this, the code started to be more and more obscure as new functions were added to the package. Therefore, recently [24] we rewrote the code in Fortran 90, grouping subroutines and data structures into clearly defined modules, which can also be shared by the programs for post-processing UNRES results (WHAM, which executes the weighted-histogram analysis method [17] given the results of REMD/MREMD simulations in order to comupute ensemble averages and probabilities of conformations and CLUSTER to group the conformations into families [21]). This approach enabled us to reduce the redundancy in the code to a great extent. Moreover, the code takes advantage of dynamic memory allocation, this making memory usage more efficient, especially for 2-,3- and 4-dimensional arrays. The source code in the UNRES package with Fortran 90 has the following directory structure:

- unres (main program modules, source code with all UNRES functionalities);
  - data (modules containing arrays declarations that correspond to COMMON block files in the UNRES package with Fortran 77);
- wham (weighted analysis method source code);
- cluster (cluster analysis source code);
- xdrfpdb (format conversion programs source code).

Because the postprocessing programs require a negligible fraction of compute time compared to that required for simulations with the main UNRES module, only the performance of UNRES will be discussed in this paper.

3. Results and Discussion

We tested the scalability of the Fortran 90 implementation of UNRES with proteins of different size and on various hardware platforms, which included our local Beowulf cluster composed of 20-core Dell nodes Intel Xeon CPU E5-2640 v4 2.40GHz connected with Gigabit Ethernet, at Faculty of Chemistry, University of Gdańsk (piasek4.chem.univ.gda.pl), the 3214-node cluster, each node containing a 12-core Intel Xeon E5 v3 2,3 GHz processor (35568 cores total), with Infiniband connection, at the Centre of Informatics - Tricity Academic Supercomputer & network (CI TASK) in Gdańsk (tryton-ap.gv.kdm.task.gda.pl), the 1084-node Cray supercomputer, each node containing 2 Intel Xeon E5-2690 v3 12-core processors with Cray Aries connection (okeanos.hpc.icm.edu.pl), and the 1024-node IBM Blue Gene/Q supercomputer, each node containig 16 4-thread PowerPC A2 cores with 5D Torus connectivity (nostromo.hpc.icm.edu.pl), at the Interdisciplinary Centre for Mathematical and Computer Modelling, University of Warsaw.

The proteins used for benchmark runs were as follows: tryptophan cage (20 residues, single chain, PDB code: 1L2Y), double-stranded RNA binding domain (68 residues, single chain, PDB code: 1STU), M-domain from GAG polyprotein (87 residues, single chain, PDB code: 1A6S), endonuclease NucB (110 residues, single chain, PDB code: 5OMT), BID domain of Bep6 (137 residues, single chain, PDB code: 4YK1), CDI complex (263 residues, two chains, PDB code: 5HKQ), UDP-glucose glycoprotein glucosyltransferase (1494 residues, single chain, PDB code: 5NV4), STRA6 receptor (819 residues, two chains, and 1636 residues, four chains; PDB code: 5SY1) and human proteasome 20S (3143 residues, 14 chains, and 6286 residues, 28 chains; PDB code: 4R3O).

The speed-ups vs. the number of cores for single-trajectory MD runs are displayed in Fig. 2, while the corresponding efficiency plots are shown in Fig. 3. These data pertain to single-trajectory MD runs, i.e., they illustrate the performance of the software in the fine-grained parallel runs. As expected, the speed up and efficiency increase with the number of residues in a system. For the smallest protein, only up to 20-times speed up can be achieved, while for large proteins over 50 % efficiency can be achieved with up to 24 cores or up to 128 cores with Blue Gene/Q. Except for Blue Gene/Q, which has fast interconnect, all cores used for fine-grained parallelization were contained in a single node. Use of cores from more than one node for fine-grained parallelization resulted in performance deterioration. For example, use of 48 fine-grained processors (2 nodes) on the Cray resulted in the same or lower speed up than that with 24 processors (a single node).

The fraction of communication time split into synchronization (barrier) and data exchange is plotted for the 5HKQ protein in Fig. 4 as a function of the number of fine-grained tasks. It can be seen that communication constitutes less than 10 % of the total wall-clock time and that the main fraction of it is spent on synchronization, which is understandable because the data are exchanged between the cores of the same hardware unit.

To test the efficiency of the parallel implementation of the code in full-blown calculation, replica-exchange runs consisting of 2 to 48 trajectories, single-trajectory runs taken as reference, were carried out for the 1A6S protein (87 residues). Each trajectory consisted of 400,000 steps and exchange of replicas occurred every 10,000 steps. The non-setup time (defined as the wall-clock time after subtracting that for data input, initializing variables, and computing the inertia matrix) efficiencies as functions of number of trajectories (coarse-grained tasks) and different numbers of cores per trajectory (fine-grained tasks) are shown in Fig. 5. It can be seen from the figure that the non-setup time and efficiency effectively depend only on the number of fine-
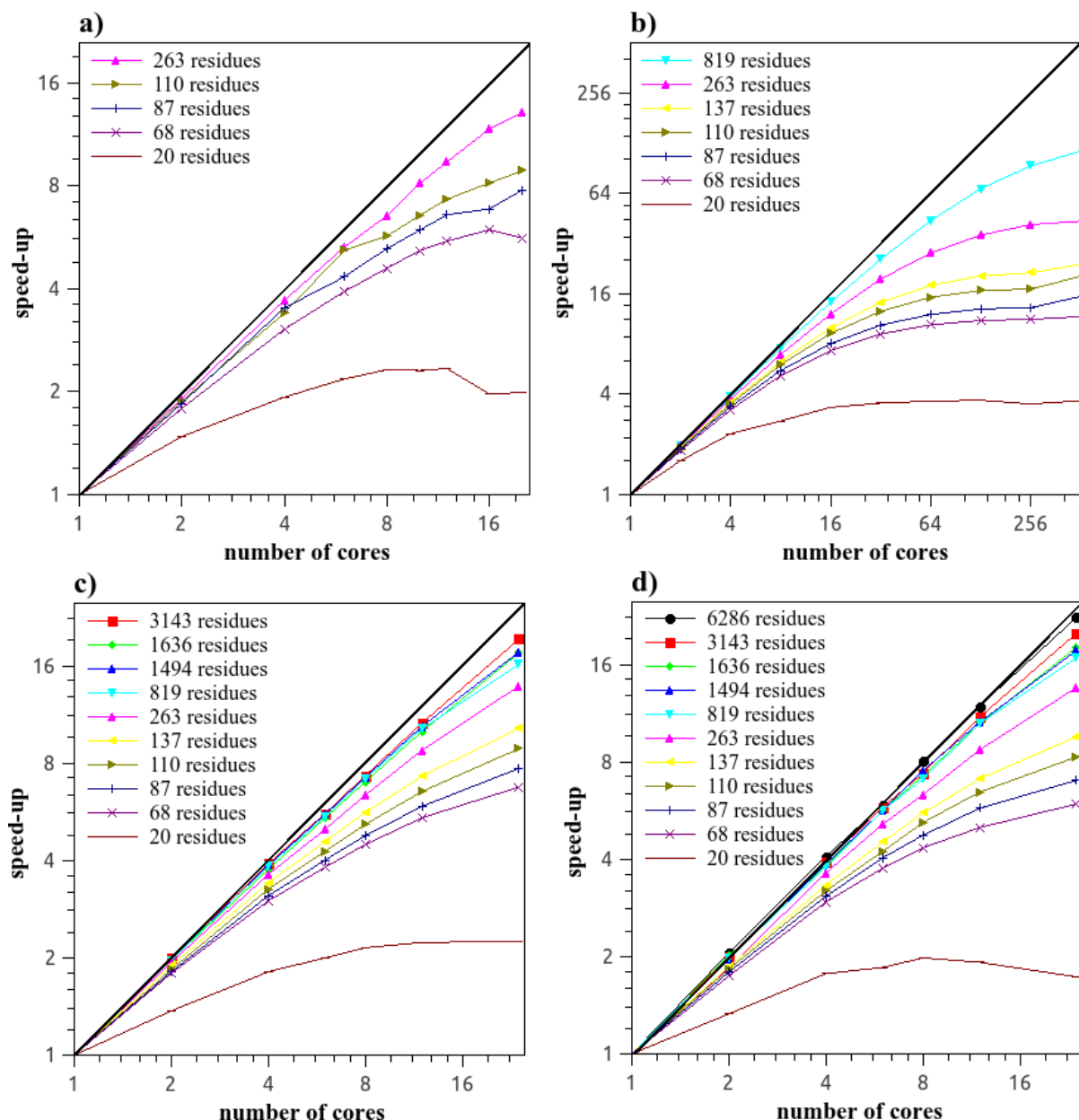
Figure 2. Plots of the speed-ups for canonical single-trajectory UNRES/MD runs obtained for proteins with various chain length vs. the number of processors on (a) piasek4.chem.univ.gda.pl, (b) nostromo.hpc.icm.edu.pl, (c) okeanos.hpc.icm.edu.pl, and (d) tryton-ap.gv.kdm.task.gda.pl. The logarithmic scale with base 2 is used on both axes

grained tasks per trajectory. This is because all MD trajectories are running at virtually the same compute speed, this providing good load balancing and data transmission occurs rarely (in the reported calculations only 40 times).

To compare UNRES with the all-atom simulations we have chosen one of the most popular molecular dynamics software packages, GROMACS [1, 3]. We have used the all-atom GROMACS 5.1.2 program run on tryton-ap.gv.kdm.task.gda.pl for 1L2Y, 1A6S, and 5OMT proteins, and 10,000 steps of molecular dynamics with amber03 force field [37]. The non-setup times for the above-mentioned proteins, with a various numbers of processors are plotted in figure 6. As can
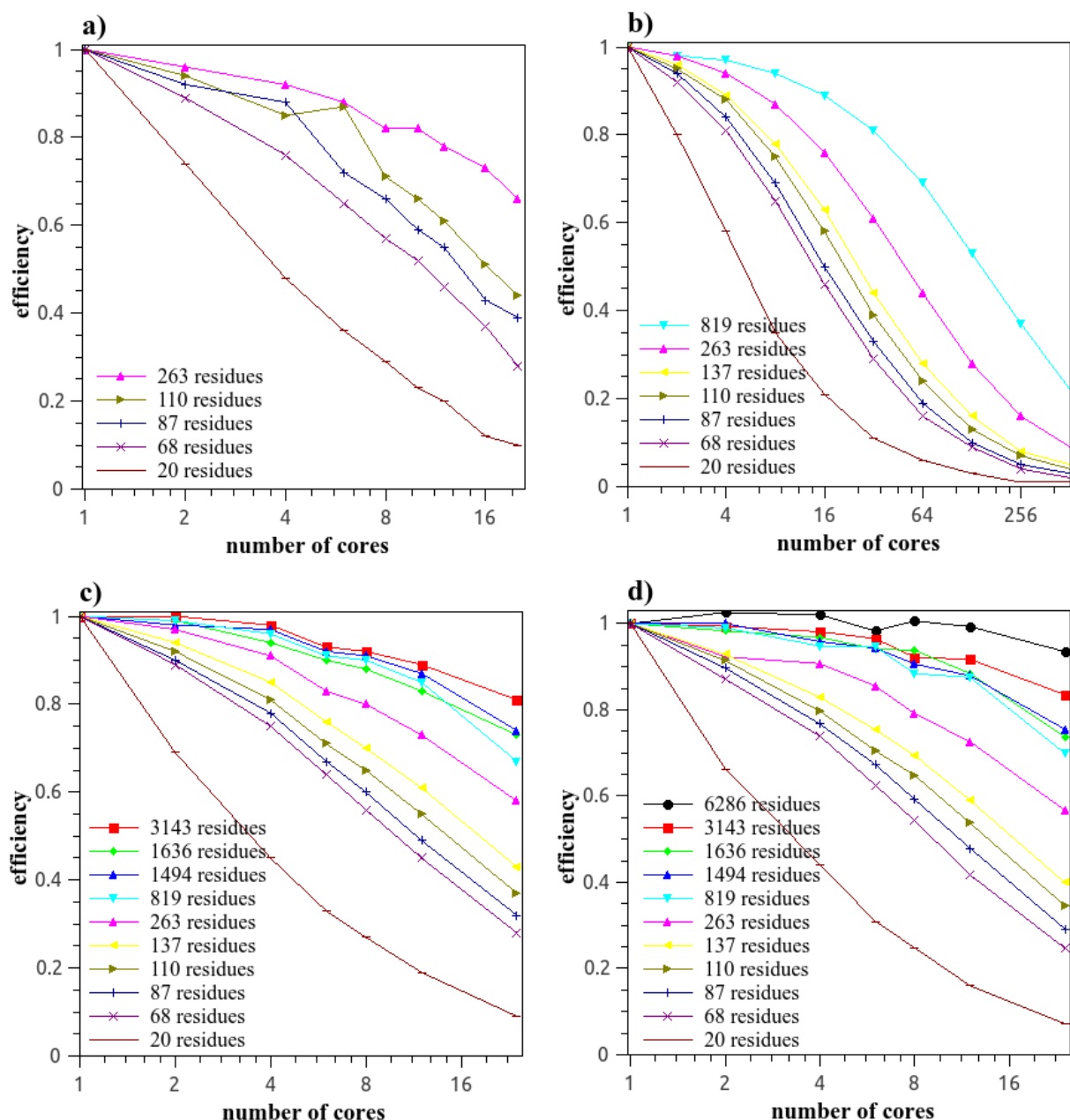
Figure 3. Plots of the efficiencies for canonical single-trajectory UNRES/MD runs obtained for proteins with various numbers of residues vs. the number of processors on (a)piasek4.chem.univ.gda.pl, (b) nostromo.hpc.icm.edu.pl, (c) okeanos.hpc.icm.edu.pl, and (d) tryton-ap.gv.kdm.task.gda.pl. The logarithmic scale with base 2 is used on x-axis

be seen from this figure, the GROMACS computation time scales almost linearly with both number of processors and protein size. This confirmed that GROMACS is currently one of the most efficient and best load-balanced MD software (due, among other things, to the cut-off introduction on non-bonded interactions). As shown, UNRES provides a few times quicker simulations compared to GROMACS. Moreover, because the UNRES event-based time scale is at least 4 times longer than that of the all-atom approach, the effective speed-up with respect to GROMACS is even bigger. When the single-processor times are compared, the UNRES simulation time is 33 times lower for 1L2Y, 10 times lower for 1A6S, and 5 times lower for 5OMT.
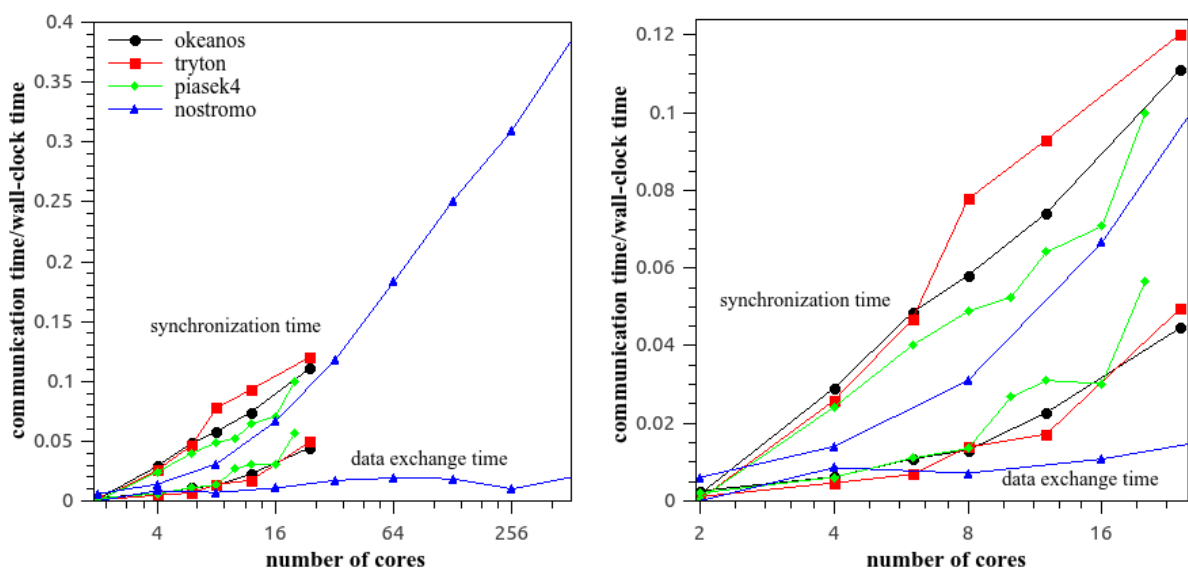
Figure 4. Plots of the fraction of the collective-communication time to the non-setup wall-clock time with a given number of cores in a single-trajectory UNRES/MD run for the 5HKQ protein on okeanos.hpc.icm.edu.pl, tryton-ap.gv.kdm.task.gda.pl, piasek4.chem.univ.gda.pl and nostromo.hpc.icm.edu.pl. The right panel (zoom) shows a higher-magnification view of the bottom left region in the left panel. The logarithmic scale with base 2 is used on x-axis
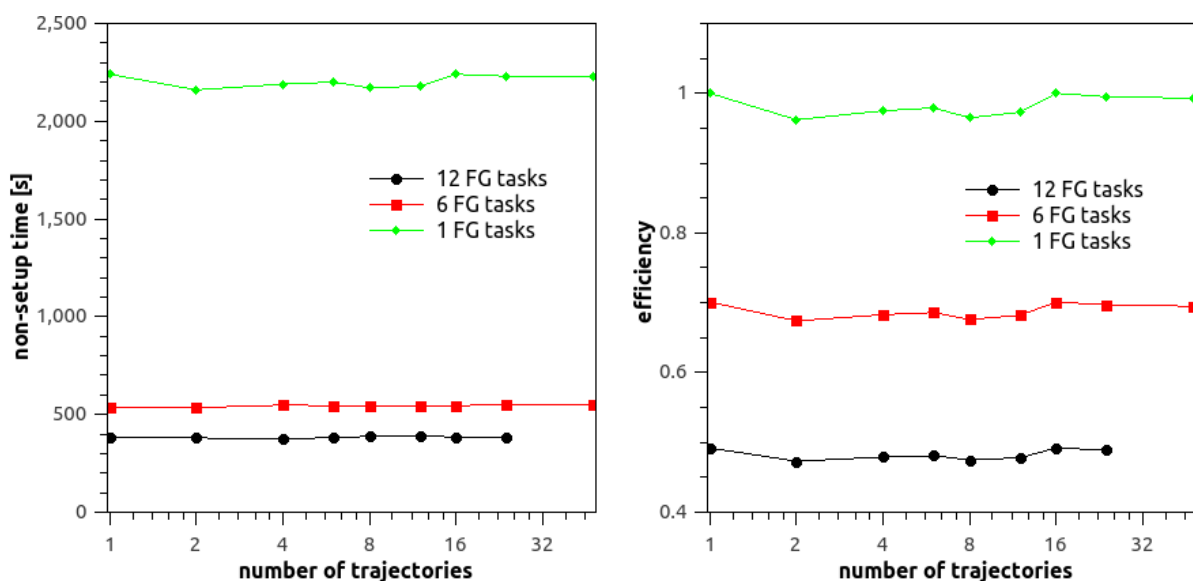

Figure 5. Plots of non-setup times and efficiencies in 40,0000 step MREMD simulations with replica exchange each 10,000 steps, numbers of trajectories from 1 (reference) to 48, and various numbers of fine-grain tasks per trajectory obtained for the 1A6S (87-residue) protein with tryton-ap.gv.kdm.task.gda.pl. The logarithmic scale with base 2 is used on x-axis. For a given point in the graph, the total number of cores used is obtained by multiplying the number of trajectories by the number of fine-grained tasks

Summarizing, the simplified biopolymer-chain representation in UNRES results in several order of magnitude extension of the time- and size-scale of computations, compared to that of the all-atom simulations.
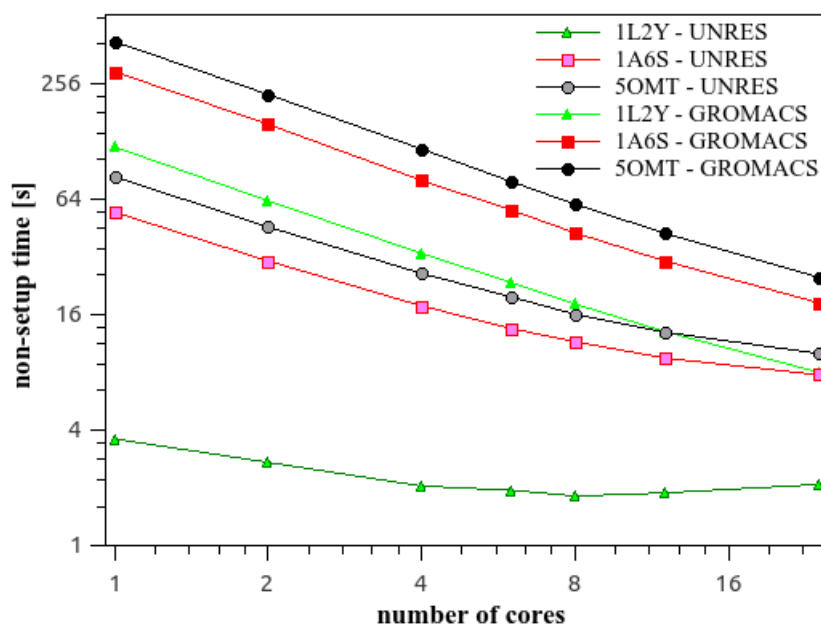
Figure 6. Plots of non-setup times in 10,000 steps MD simulations with UNRES and GROMACS, with various numbers of processors obtained for the 1L2Y (20-residue), 1A6S (87-residue) and 5OMT (110-residue) proteins with tryton-ap.gv.kdm.task.gda.pl. The logarithmic scale with base 2 is used on both axes

Conclusion

In this work, we implemented the Fortran 90 version of the UNRES package for coarse-grained simulations of proteins and protein complexes on a number of parallel platforms and tested its parallel efficiency both in the fine-grained mode (energy and force parallelization in single-trajectory molecular dynamics calculation) and in the full-blown (coarse- and fine-grained) mode, in which multiple trajectories were run in replica exchange calculations with communication between the respective task groups every given number of steps, and parallelization of the computation of each trajectory. For proteins with chain length about 300 amino-acid residues or more over 50 % parallel efficiency is obtained. The efficiency does not deteriorate when replica-exchange calculations are carried out.

The parallel implementation of UNRES enables us to simulate the dynamics of quite large proteins in real time. For example, for the 263-residue 5HKQ protein, the time required for 10,000 time steps with 24 fine-grained tasks run on the Cray machine is 24 seconds, which translates to about 176 nanoseconds of molecular-dynamics time per (24 hours of computations) with the 4.89 fs time step commonly set in UNRES runs. However, because UNRES simulation time is at least 1000-fold extended with respect to that of all-atom simulations with explicit solvent [8], one day of UNRES simulations for this protein amounts to a 0.176 millisecond molecular dynamics. With this speed up, physics-based simulations of biologically important processes are now at hand.

More informations about UNRES force field and the souce code are available (Description of UNRES program. http://www.unres.pl/unres, accessed: 2018-06-15; Downloads page. http://unres.pl/downloads, accessed: 2018-06-15; UNRES server version. http://unres-server.chem.ug.edu.pl./about, accessed: 2018-06-15).

Acknowledgements

References

1. Abraham, M.J., Murtola, T., Schulz, R., Pll, S., Smith, J.C., Hess, B., Lindahl, E.: Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1-2, 19 – 25 (2015)

2. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R.: Molecular dynamics with coupling to an external bath. J. Chem. Phys. 81, 3684–3690 (1984)

3. Berendsen, H., van der Spoel, D., van Drunen, R.: Gromacs: A message-passing parallel molecular dynamics implementation. Computer Physics Communications 91(1), 43 – 56 (1995)

4. Czaplewski, C., Kalinowski, S., Liwo, A., Scheraga, H.A.: Application of multiplexing replica exchange molecular dynamics method to the UNRES force field: Tests with α and $\alpha + \beta$ proteins. J. Chem. Theor. Comput. 5, 627–640 (2009)

5. Hansmann, U.H.E., Okamoto, Y.: Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem. Physica A 212, 415–437 (1994)

6. He, Y., Maciejczyk, M., Ołdziej, S., Scheraga, H.A., Liwo, A.: Mean-field interactions between nucleic-acid-base dipoles can drive the formation of a double helix. Phys. Rev. Lett. 110, 098101 (2013)

7. He, Y., Mozolewska, M., P., K., Sieradzan, A., Wirecki, T., Liwo, A., Kachlishvili, K., Rackovsky, S., Jagieła, D., Ślusarz, R., Czaplewski, C., Ołdziej, S., Scheraga, H.: Lessons from application of the UNRES force field to predictions of structures of CASP10 targets. Proc. Natl. Acad. Sci. U. S. A. 110(37), 14936–14941 (2013)

8. Khalili, M., Liwo, A., Jagielska, A., Scheraga, H.: 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 (2005)

9. Khalili, M., Liwo, A., Rakowski, F., Grochowski, P., Scheraga, H.: Molecular dynamics with the united-residue model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical model. J. Phys. Chem. B 109, 13785–13797 (2005)

10. Kleinerman, D.S., Czaplewski, C., Liwo, A., Scheraga, H.A.: Implementations of Nosé – Hoover and Nosé –Poincaré thermostats in mesoscopic dynamic simulations with the united-residue model of a polypeptide chain. The Journal of Chemical Physics 128(24), 245103 (2008)

11. Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A.E., Kolinski, A.: Coarse-grained protein models and their applications. Chem. Rev. 116, 7898–7936 (2016)

12. Kolinski, A., Skolnick, J.: Reduced models of proteins and their applications. Polymer 45, 511–524 (2004)

13. Kolinski, A., Skolnick, J.: Assembly of protein structure from sparse experimental data: An efficient monte carlo model. Proteins: Structure, Function, and Bioinformatics 32(4), 475–494 (1998)

14. Krupa, P., Mozolewska, M., Wiśniewska, M., Yin, Y., He, Y., Sieradzan, A., Ganzynkowicz, R., Lipska, A., Karczyńska, A., Ślusarz, M., Ślusarz, R., Giełdoń, A., Czaplewski, C., Jagieła, D., Zaborowski, B., Scheraga, H., Liwo, A.: Performance of protein-structure predictions with the physics-based UNRES force field in CASP11. Bioinformatics 32, 3270–3278 (2016)

15. Krupa, P., Hałabis, A., Żmudzińska, W., Ołdziej, S., Scheraga, H.A., Liwo, A.: Maximum likelihood calibration of the UNRES force field for simulation of protein structure and dynamics. Journal of Chemical Information and Modeling 57(9), 2364–2377 (2017)

16. Kubo, R.: Generalized cumulant expansion method. J. Phys. Soc. Japan 17, 1100–1120 (1962)

17. Kumar, S., Bouzida, D., Swendsen, R.H., Kollman, P.A., Rosenberg, J.M.: The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. J. Comput. Chem. 13, 1011–1021 (1992)

18. Liwo, A., Baranowski, M., Czaplewski, C., Gołaś, E., He, Y., Jagieła, D., Krupa, P., Maciejczyk, M., Makowski, M., Mozolewska, M.A., Niadzvedtski, A., Ołdziej, S., Scheraga, H.A., Sieradzan, A.K., Ślusarz, R., Wirecki, T., Yin, Y., Zaborowski, B.: A unified coarse-grained model of biological macromolecules based on mean-field multipole-multipole interactions. J. Mol. Model. 20(8), 2306 (2014)

19. Liwo, A., Czaplewski, C., Ołdziej, S., Rojas, A.V., Kaźmierkiewicz, R., Makowski, M., Murarka, R.K., Scheraga, H.A.: 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, chap. 8, pp. 1391–1411. CRC Press (2008)

20. Liwo, A., Czaplewski, C., Pillardy, J., Scheraga, H.A.: Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field. J. Chem. Phys. 115, 2323–2347 (2001)

21. Liwo, A., Khalili, M., Czaplewski, C., Kalinowski, S., Ołdziej, S., Wachucik, K., Scheraga, H.: 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 (2007)

22. Liwo, A., Ołdziej, S., Czaplewski, C., Kleinerman, D.S., Blood, P., Scheraga, H.A.: Implementation of molecular dynamics and its extensions with the coarse-grained UNRES force field on massively parallel systems; towards millisecond-scale simulations of protein structure, dynamics, and thermodynamics. J. Chem. Theor. Comput. 6, 583–595 (2010)

23. Lopez, C.A., Rzepiela, A., de Vries, A.H., Dijkhuizen, L., Hunenberger, P.H., Marrink, S.J.: Martini coarse-grained force field: Extension to carbohydrates. J. Chem. Theor. Comput. 5, 3195–3210 (2009)

24. Lubecka, E.A., Liwo, A.: New UNRES force field package with FORTRAN 90. TASK Quart 20(4), 399–407 (2016)

25. Markutsya, S., Baluyut, J., Windus, T.L., Gordon, M.S., Lamm, M.H.: Evaluation of coarse-grained mapping schemes for polysaccharide chains in cellulose. J. Chem. Phys. 138, 214108 (2013)

26. Marrink, S.J., Tieleman, D.P.: Perspective on the martini model. Chem. Soc. Rev. 42, 6801–6822 (2013)

27. Molinero, V., Goddard III, W.A.: M3b: a coarse-grain force field for molecular simulations of malto-oligosaccharides and their water mixtures. J. Phys. Chem. B 108, 1414–1427 (2004)

28. Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., Marrink, S.J.: The martini coarse-grained force field: Extension to proteins. Journal of Chemical Theory and Computation 4(5), 819–834 (2008)

29. Mozolewska, M.A., Krupa, P., Scheraga, H.A., Liwo, A.: Molecular modeling of the binding modes of the iron-sulfur protein to the Jac1 co-chaperone from Saccharomyces cerevisiae by all-atom and coarse-grained approaches. Proteins: Struct., Funct., Bioinf. 83(8), 1414–1426 (2015)

30. Rakowski, F., Grochowski, P., Lesyng, B., Liwo, A., Scheraga, H.A.: Implementation of a symplectic multiple-time-step molecular dynamics algorithm, based on the united-residue mesoscopic potential energy function. J. Chem. Phys. 125, 204107 (2006)

31. Rhee, Y.M., Pande, V.S.: Multiplexed-replica exchange molecular dynamics method for protein folding simulation. Biophys. J. 84, 775–786 (2003)

32. Rojas, A., Liwo, A., Browne, D., Scheraga, H.A.: Mechanism of fiber assembly; treatment of aβ-peptide peptide aggregation with a coarse-grained united-residue force field. J. Mol. Biol. 404, 537–552 (2010)

33. Samsonov, S.A., Bichmann, L., Pisabarro, M.T.: Coarse-grained model of glycosoaminoglycans. J. Chem. Inf. Model, 55, 114–124 (2015)

34. Sieradzan, A.K., Krupa, P., Wales, D.J.: What makes telomeres unique? J. Phys. Chem. B 121, 2207–2219 (2016)

35. Sieradzan, A.K., Makowski, M., Augustynowicz, A., Liwo, A.: 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(12), 124106 (2017)

36. Takada, S., Kanada, R., Tan, C., Terakawa, T., Li, W., Kenzaki, H.: Modeling structural dynamics of biomolecular complexes by coarse-grained molecular simulations. Accounts of Chemical Research 48(12), 3026–3035 (2015)

37. Yong, D., Chun, W., Shibasish, C., C., L.M., Guoming, X., Wei, Z., Rong, Y., Piotr, C., Ray, L., Taisung, L., James, C., Junmei, W., Peter, K.: A pointcharge force field for molecular mechanics simulations of proteins based on condensedphase quantum mechanical calculations. Journal of Computational Chemistry 24(16), 1999–2012 (2003)

38. Zhou, R., Maisuradze, G.G., Suñol, D., Todorovski, T., Macias, M.J., Xiao, Y., Scheraga, H.A., Czaplewski, C., Liwo, A.: Folding kinetics of WW domains with the united residue force field for bridging microscopic motions and experimental measurements. Proc. Natl. Acad. Sci. U.S.A. 111, 18243–18248 (2014)