Folding simulation of small proteins by dissipative particle dynamics (DPD) with non-empirical interaction parameters based on fragment molecular orbital calculations

Koji Okuwaki1, Hideo Doi2, Kaori Fukuzawa3,4, and Yuji Mochizuki1,4

1Department of Chemistry and Research Center for Smart Molecules, Faculty of Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan
2CD-Fmat, National Institute of Advanced Industrial Science and Technology, 1-1-1 Umetsuno, Tsukuba, Ibaraki 305-8568, Japan
3School of Pharmacy and Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-ku, Tokyo, 142-8501, Japan
4Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan

E-mail: fullmoon@rikkyo.ac.jp

Received November 23, 2019; accepted December 2, 2019; published online December 11, 2019

Recently, we have developed a new simulation scheme with dissipative particle dynamics (DPD) based on non-empirical interaction parameters derived from a series of fragment molecular orbital (FMO) calculations. This approach (FMO–DPD) was applied to folding simulations of Chignolin and Supercichognin mini-proteins. Their characteristic hairpin structure was obtained from an elongated form within short computation time. Essential residue-residue interactions such as hydrogen bonding and CH/π were observed at the final form. FMO–DPD should have a potential applicability to nano-bio systems involving proteins. © 2019

Supplementary material for this article is available online

Proteins are fundamental substances widely involving in biological life phenomena, and their functionalities depend on the folded 3-dimensional structures. Nano-bio technology devices with sensor or receptor proteins have attracted practical interests in the field of applied physics as well. In molecular simulations, the reproduction of folded protein structures has long been one of the most demanding problem, because of large-scale configurational changes should take place. Amounts of studies with all-atom molecular dynamics (MD) have thus been made. In particular, Shaw’s group1 provided a revolutionary approach with customized computer chip (ANTON) to accelerate MD simulation and succeeded in folding simulations as long as ns or longer. The folding simulations with atomistic MD are still costly, unless using supercomputers or special accelerators. Therefore, the so-called coarse-grained (CG) approaches have attracted interests as alternative routes. For example, Trp-Cage2 (consisting of 20 amino acid residues) was successfully folded by CG–MD,3 and the method of dissipative particle dynamics (DPD)4 was applied to handle the folding of neurotoxic α-synuclein5 (140 residues). DPD could have a high efficiency in surveying wide configuration spaces of protein structures (not limited to the folding). As a brand new method, the folding of model proteins was solved through quantum annealing.6,7 Although the CG simulations for proteins were certainly cost-effective, the sets of parameters describing various interactions (among residue-residue and residue-water) were of empirical type. In other words, the applicability and usability could be limited potentially.

Recently, we have developed a non-empirical simulation scheme of DPD in which the set of effective interaction parameters $a_{ij}$ associated with $\chi$-parameters are evaluated by the fragment molecular orbital (FMO) calculations,8 abbreviated to FMO–DPD. The corresponding workflow to obtain $\chi$-parameters by FMO was available as a generic system named FCEWS (FMO-based Chi-parameter Evaluation Workflow System).9,10 Additionally, we have made an original DPD program, CAMUS11,12 by which specific interactions such as I–3 hydrogen bonds5 are easily incorporated. In this work, the FMO–DPD simulation was applied to the folding of Chignolin13 composed of 10 amino acid residues. Very recently, detailed dynamics of Chignolin has been reported.14 The speed-up of such extensive MD simulations still remains an important issue, however. Our FMO–DPD may be an efficient CG approach to the folding problem.

We performed the FMO–DPD simulation of Chignolin as follows. First, the amino acid residues of Chignolin (Gly-Tyr-Asp-Pro-Glu-Thr-Gly-Thr-Trp-Gly) were segmented to the respective particles in DPD (Fig. 1). According to the recipe of Ref. 5, the main-chain backbone was described in Gly unit, and the side chain moieties as well as the N- and C-terminals were individually treated with the respective chemical structures. The non-bonding interaction parameters among the residue particles and the environmental water particle were determined through the FMO-based $\chi$-parameter evaluations.9,10,15 The segment structures were pre-optimized with the GAUSSIAN 09 program16 at the B97D/6-31G(d’) level. The combination of FCEWS9,10 and ABINIT-MP (our FMO program)17 was employed in a series of evaluations of $\chi$-parameters, where the interaction energies among segmented particles were computed at the second-order perturbation [or MP2/6-31G(d’)] level. The hydration effect was incorporated by the Poisson–Boltzmann model (FMO-PB).18,19 In the DPD simulations with the CAMUS code,11 the force of $i$th particle $f_i$ contains four parts as

$$f_i = \sum_{j=i} (f_{ij}^C + f_{ij}^D + f_{ij}^S + f_{ij}^R).$$

The first three forces of the right-hand-side are considered within a certain cutoff radius $r_c$. The conservative force $f_{ij}^C$ is a soft repulsion action as follows
describes the self-interaction of particle \( r_i \) and \( r_j \), and related definitions are \( r_i = r_j - r_i \) and \( r_j = |r_j| \), and \( n_{ij} = r_j / |r_j| \). The repulsion parameters \( a_{ij} \) between particles of different types are related to the Flory–Huggins \( \chi \)-parameter as:

\[
a_{ij} = \left( a_{ii} + a_{jj} \right) / 2 + 3.27 \chi_{ij}
\]

where \( a_{ii} \) or \( a_{jj} \) describes the self-interaction of particle \( i \) or \( j \). In Eq. (1), the dissipative force \( F_{ij}^D \) and the random force \( F_{ij}^R \) represent hydrodynamic drags and thermal noises of the Gaussian statistics, respectively. The fourth term in the right-hand side of Eq. (1) provides an additional spring force \( F_{ij}^S \) for directly bonded particles in polymers. For a certain connected particle pair \( i \) and \( j \), the corresponding Harmonic force \( F_{ij}^H \) and Morse force \( F_{ij}^M \) are given, respectively, as

\[
F_{ij}^H = C(r_i - r_j)n_{ij}, \tag{4}
\]

\[
F_{ij}^M = 2K_M\alpha e^{-\alpha(r_i - r_j)} (e^{-\alpha(r_i - r_j)} - 1)n_{ij}, \tag{5}
\]

where \( C \), \( r_i \), \( K_M \) and \( \alpha \) are the force constant, equilibrium distance, the well depth of Morse potential and the well width, respectively. \( F_{ij}^H \) in Eq. (1) thus becomes the summation of Harmonic and Morse contributions as \( F_{ij} = F_{ij}^D + F_{ij}^M \).

In our simulation, the \( a_{ij} \) parameters obtained from FMO calculations were used as essential interactions between non-bonded particles (refer to Table I). Note here that Trp (W) side chain was represented with a couple of particles because of its relative largeness (see Fig. 1). The 1–2 bonds and 1–3 bonds were of Harmonic type with \( K_M = 160 \) kT, \( r_i = 0.6 \) DPD-length unit \( (R_c) \) and \( C = 80 \) kT, \( r_i = 1.2 \) \( R_c \), respectively.\(^5\) To maintain the relative positions between main- and side-chains, a Morse potential was introduced for the 1–3 bonds (between main-chain Gly (G) particle and neighboring side chain particles) with setting of \( K_M = 12 \), \( \alpha = 8 \), and \( r_i = 0.85 \ R_c \) (our original setting). Similarly, another Morse potential \( (K_M = 12 \), \( \alpha = 8 \), and \( r_i = 1.34 \ R_c \) \) was used for the 1–4 bonds. The DPD time step was set as \( \Delta t = 0.005 \) DPD-time unit (at least 0.7 ps in real unit), and the number of particles in a cell \( [6.9 \ R_c \ (4.9 \text{ nm}) \text{ size box cell}] \) was 1000. The initial structure of mini-protein was an elongated line structure.

The numerical results in the \( a_{ij} \) form are listed for the combinations of particles in Table I, where two values for Gly (N-terminal) is set to zero because the calculated value was negative (or too attractive). Taking the C-terminal end of Gly (G) as an example, it is shown that the interaction value with itself \( (a_{ii}) \) shows a large value of repulsion with 145 kT and that the similar repulsions in \( a_{ij} \) are found for the negatively charged Asp (E) and Glu (D). Interestingly, the values involving water took various values, reflecting the respective chemical characteristics of the interacting particles.\(^5\) The bond potentials of 1–2 (direct bond), 1–3 (angle), and 1–4 (dihedral) were appropriately set to maintain the rigidity of protein skeleton in CAMUS.\(^11,12\)

Figure 2(a) illustrates the hairpin-shape structure of folded Chignolin,\(^13\) where the internal attractive interactions among residues such as hydrogen bond between Asp3 (E) and Gly7 (G) and CH/\( \pi \) interaction between Tyr2 (Y) and Trp9 (W) are crucial. Several DPD simulations with the FMO-derived effective parameters provided the folded structures, as shown in Figs. 2(b) and 2(c), starting from the elongated ones. It is notable that the computation time for this folding simulation was only 8 s (running 10000 steps) on 8 cores (8 OpenMP threads of parallelization\(^11,12\) ) of an in-house Xeon (E5-2690) server. The FMO–DPD simulation was applied also to folding of Superchignolin\(^20\) (Tyr-Tyr-Asp-Pro-Glu-Thr-Gly-Thr-Trp-Tyr), and the hairpin like structure was reproduced successfully, as illustrated in Fig. 3. Figure 4 plots the distances between N- and C-terminal particles against the time step of DPD for both Chignolin and Superchignolin, indicating the rapid decrease of distance from the elongated structure. Comparison between Chignolin and Superchignolin suggests that the step-dependent fluctuation of Superchignolin is smaller than that of Chignolin, and this situation is consistent with the fact that the internal interactions in the former is stronger (by the intended design\(^20\)) than that in the latter.
Table I. List of interaction parameters among segments \((a_c, a_q)\) for Chignolin.

|       | G       | G (N-term) | G (C-term) | Y       | D       | P       | E       | T       | W       | Water |
|-------|---------|------------|------------|---------|---------|---------|---------|---------|---------|--------|
| G     | 110.6   | 111.4      | 141.0      | 152.0   | 116.8   | 105.2   | 125.6   | 113.2   | 150.0   | 170.8  |
| G (N-term) | 170.4 | 0.0        | 169.3      | 0.0     | 126.6   | 59.3    | 130.5   | 154.5   | 98.7    | 26.1   |
| G (C-term) | 145.2 | 114.3      | 139.5      | 103.4   | 152.0   | 70.0    | 98.7    | 26.1    |         |        |
| Y     | 138.1   | 81.0       | 131.2      | 160.3   | 132.8   | 154.5   | 98.7    | 26.1    |         |        |
| D     | 130.2   | 95.8       | 129.1      | 95.2    | 78.5    | 39.2    |         |         |         |        |
| P     | 66.2    | 103.6      | 82.0       | 111.5   | 132.8   |         |         |         |         |        |
| E     | 140.2   | 107.7      | 156.4      | 36.1    |         |         |         |         |         |        |
| T     | 58.6    | 122.0      | 117.0      |         |         |         |         |         |         |        |
| W     |         |            |            |         |         |         |         |         |         |        |
| Water |         |            |            |         |         |         |         |         |         | 78.5   |

Fig. 2. (Color online) Crucial residue-residue interactions (a) and folded structure [interactions indicated with circles in (b) and (c)].

Fig. 3. (Color online) Structure of Superchignolin (a) and DPD-folded structure (b).

Fig. 4. (Color online) Plots of distances between C- and N-terminal particles for Chignolin and Superchignolin.

Fig. 5. (Color online) RMSD results of 10 trajectories up to 100 000 steps (see text).
In this paragraph, verification of geometrical features obtained by our DPD simulation is described for Chignolin. In comparison with the NMR structure, Fig. 5 plots the root mean square deviation (RMSD) of Cα atoms for 10 trajectories (up to 100 000 steps), where the average value (bold blue color plot) reaches the range of 3–4 Å at about 5000 steps and the minimal value is evaluated as 1.6 Å (in black color plot). These values may be acceptable, because both interaction parameters and distances between particles were not set at atomistic level in DPD (unlike usual MD). Related verification with RMSD plots was done as well, and the results are given in the supporting information (Figs. S1 and S2), available online at stacks.iop.org/APEX/13/017002/mmedia. Another comparison is made for time evolutions of mutual positions (see the corresponding Fig. S3) of Asp3 - Gly7 pair (hydrogen bonding) and Tyr2 - Trp9 pair (weaker CH/π interaction). Rather stable behavior is found for the former, but several different conformations (including misfolded type as shown in Fig. S4) are observed for the latter. These characteristics are consistent with atomistic MD results of Ref. 21.

In this brief letter, we have reported an application of the FMO–DPD simulation to the folding of Chignolin and Superchignolin.9,13 where the non-bonding particle-particle interaction parameters were determined in non-empirical way.9,10 The folding was accomplished in short DPD runs, indicating an efficiency in handling proteins. FMO–DPD was already successfully applied by us to lipid membranes and vesicles,22–24 and the technique of reverse mapping25 (by which meso-scale atomic structures are to be reproduced from meso-scale CG structures) has been recently settled for lipids. Preliminary works for lipid membranes containing mini-proteins have been in progress. Although nonbonding interaction parameters have been determined with the FMO calculations in the present study, direct bonding parameters as well as structure-constraint parameters have still been of empirical type.5 Good balance among all these parameters is important in reproducing proper structures (refer to previous Figs. S1 and S2 and also Fig. S5 of α-synuclein). Non-empirical determination of bonding and constraint parameters should be our future issue. We would continue a series of developments associated with FMO–DPD simulations, considering that this protocol should have a potential to efficiently model the crucial parts (e.g. interfaces) of nano-bio device systems consisting of lipids, proteins and substances.

Acknowledgments The authors would thank to Dr. Yuto Komeiji (AIST Japan) for fruitful comments on the manuscript. The present work was supported by Ministry of Education, Culture, Sports, Science and Technology (MEXT) as a social and scientific priority issue #6 (Accelerated Development of Innovative Clean Energy Systems) to be tackled by using post-K computer (“Fugaku”). Additional supports were due to the Grant-in-Aid for Scientific Research (Kakenhi-16H04635) as well as the Rikkyo SFR.

ORCID iDs  Koji Okuwaki https://orcid.org/0000-0002-4510-5717 Hideo Doi https://orcid.org/0000-0002-5885-1967 Kaori Fukuzawa https://orcid.org/0000-0001-5357-8250 Yuji Mochizuki https://orcid.org/0000-0002-7310-5183

1) K. Lindorf-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, Science 334, 517 (2011).
2) J. W. Neidigh, R. M. Fesmirey, and N. H. Andersen, Nat. Struct. Biol. 9, 425 (2002).
3) W. Han and K. Schulten, J. Phys. Chem. B 117, 13367 (2013).
4) R. D. Groot and P. B. Warren, J. Chem. Phys. 107, 4423 (1997).
5) A. Vishnyakov, D. S. Talaga, and A. V. Neimark, J. Phys. Chem. Lett. 3, 3081 (2012).
6) A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik, Sci. Rep. 2, 1 (2012).
7) T. Babje, C. Ing, and M. Fingerhuth, (2018), arXiv:1811.00713.
8) D. G. Fedorov and K. Kitaura, The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems (CRC Press, Boca Raton, FL, 2009).
9) K. Okuwaki, H. Doi, and Y. Mochizuki, J. Comput. Chem. Jpn. 17, 1 (2018).
10) K. Okuwaki, Y. Mochizuki, H. Doi, and T. Ozawa, J. Phys. Chem. B 122, 2924 (2018).
11) H. Doi, S. Saitou, K. Okuwaki, T. Naito, and Y. Mochizuki, J. Comput. Chem. Jpn. 16, 126 (2018).
12) H. Doi, K. Okuwaki, T. Naito, S. Saitou, and Y. Mochizuki, Chem-Bio Inf. J. 18, 70 (2018).
13) S. Honda, K. Yamasaki, Y. Sawada, and H. Morii, Structure 12, 1507 (2004).
14) T. Sumi and K. Koga, Sci. Rep. 9, 5186 (2019).
15) K. Okuwaki, Y. Mochizuki, H. Doi, S. Kawada, T. Ozawa, and K. Yamasuka, RSC Adv. 8, 34582 (2018).
16) M. J. Frisch et al., Gaussian 09, Revision B.01 (Gaussian, Inc., Wallingford CT, 2016).
17) S. Tanaka, Y. Mochizuki, Y. Komeiji, Y. Okiyama, and K. Fukuzawa, Phys. Chem. Chem. Phys. 16, 10310 (2014).
18) Y. Okiyama, T. Nakano, C. Watanabe, K. Fukuzawa, Y. Mochizuki, and S. Tanaka, J. Phys. Chem. B 122, 4457 (2018).
19) Y. Okiyama, C. Watanabe, K. Fukuzawa, Y. Mochizuki, T. Nakano, and S. Tanaka, J. Phys. Chem. B 123, 957 (2019).
20) S. Honda, T. Akiba, Y. S. Kato, Y. Sawada, M. Sekijima, M. Ishimura, A. Oishi, H. Watanabe, T. Odahara, and K. Harata, J. Am. Chem. Soc. 130, 15327 (2008).
21) D. Satoh, K. Shimizu, S. Nakamura, and T. Terada, FEBS Lett. 580, 3422 (2006).
22) H. Doi et al., Chem. Phys. Lett. 684, 427 (2017).
23) H. Doi, K. Okuwaki, Y. Mochizuki, and T. Ozawa, J. Comput. Chem. Jpn. 16, 28 (2017).
24) E. Shinso, K. Okuwaki, H. Doi, Y. Mochizuki, T. Furushii, K. Fukuzawa, and E. Yonemochi, J. Comput. Chem. Jpn. 17, 172 (2018).
25) P. Doruker and W. L. Mattice, Macromolecules 30, 5520 (1997).