Predict Two-Dimensional Protein Folding Based on Hydrophobic-Polar Lattice Model and Chaotic Clonal Genetic Algorithm

Shuihua Wang[1], Lenan Wu[2], Yuankai Huo[1], Xueyan Wu[1], Hainan Wang[1], and Yudong Zhang[1(✉)]

[1] School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, Jiangsu, China
zhangyudong@njnu.edu.cn
[2] School of Information Science and Engineering, Southeast University, Nanjing, China

Abstract. In order to improve the performance of prediction of protein folding problem, we introduce a relatively new chaotic clonal genetic algorithm (abbreviated as CCGA) to solve the 2D hydrophobic-polar lattice model. Our algorithm combines three successful components—(i) standard genetic algorithm (SGA), (ii) clonal selection algorithm (CSA), and (iii) chaotic operator. We compared this proposed CCGA with SGA, artificial immune system (AIS), and immune genetic algorithm (IGA) for various chain lengths. It demonstrated that CCGA had better performance than other methods over large-sized protein chains.

Keywords: Protein folding · Chaotic clonal genetic algorithm · Clonal selection algorithm · Hydrophobic-polar model · Artificial immune system

1 Introduction

Protein folding (PF) is a physical process for a protein chain acquires its 3-dimensional structure [1, 2]. It imposes a challenge to biologists since the problem has an extremely large search space [3]. A successful solution to PF requires to solve two important problems [4]: (i) a series of free residues for the protein chain and (ii) an efficient optimization procedure. Now since the PF data are easily available, the latter is the most difficult thing.

Traditional optimizers will not work for the PF problem, because the model (See Sect. 2) is multimodal and non-differential. Besides, the problem is NP-hard [5]. Hence, advanced global optimizers are introduced to solve it.

The genetic algorithm (GA) and particle swarm optimization [6] are one of the most popular global optimizations. Moreover, swarm intelligence approaches [7]: biogeography-based optimization [8], firefly algorithm [9], artificial bee colony [10], and bacterial chemotaxis optimization [11], have attracted interest from scholars in many fields. Nevertheless, massive researchers had investigated the PF problems merely with GA, since the encoding in GA is more suitable for PF problems.

H. Yin et al. (Eds.): IDEAL 2016, LNCS 9937, pp. 10–17, 2016.
DOI: 10.1007/978-3-319-46257-8_2

Lin and Hsieh [12] presented a Taguchi-genetic algorithm (TGA). Huang, Yang and He [13] employed GA and optimal second structure. Narayanan, Krishnakumar and Judy [14] proposed an enhanced MapReduce framework using parallel genetic algorithm (PGA). However, GA may still encounter problems in converging into global optimal. To improve its performance, we introduced a novel chaotic clonal genetic algorithm (abbreviated as CCGA) in this study. Below we will show how the mechanism of CCGA and how it can be applied to solve PF problem.

2 Two Dimensional Hydrophobic-Polar Model

The hydrophobic-polar (HP) protein folding model [15] is a simplified version for exampling structures of protein folds in space. It stems from the fact that hydrophobic interactions between amino acids residues drive proteins to fold into native structure [16]. In this model, protein chains are composed of two types of residues, viz., polar (P) and hydrophobic (H) [17]. Figure 1 offers an example of a 10-residue chain with energy of −4.

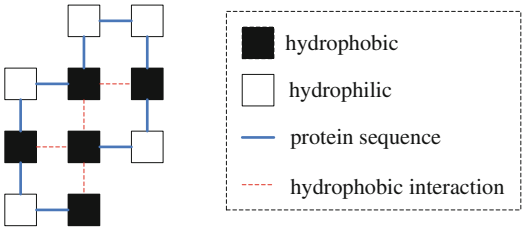


Fig. 1. A 10-residue chain of HPHPPHPHPH

Figure 2 indicates that protein chains can turn at each residue position 90° to either left (L) or right (R) or continue (C) ahead. The first interaction is always set as 'C'. Hence, the protein chain can be presented as 'CLCLLRCCL'. Note that clashes (i.e., residues overlap at the same position) is not allowed. In all, our object is to minimize the following expression

$$E = n(-1) \tag{1}$$

where n represents the number of hydrophobic interactions, E the energy function.

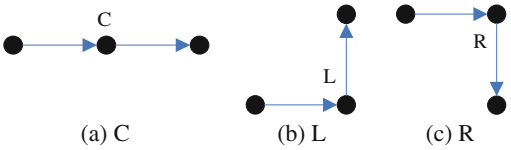


Fig. 2. Turning directions of a chain (C = Continue, L = Left, R = Right)

3 Materials

The materials consisted of five chains with residues sizes of 20, 36, 48, 64, and 85, respectively. Their information are listed in Table 1. The minimum energy was obtained by exhaustion method. The protein sequences can be found in Table 1 in reference [18].

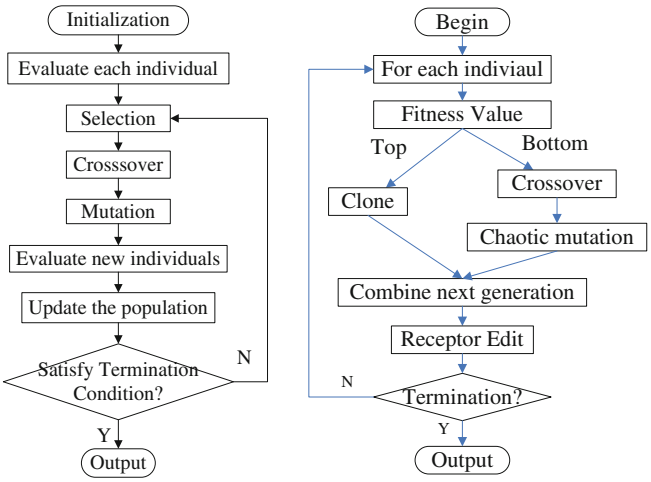
Table 1. Five protein chains

Index	1	2	3	4	5
Size	20	36	48	64	85
Energy	−9	−14	−23	−40	−52

4 Our Optimizor

4.1 Standard Genetic Algorithm

The individuals of standard genetic algorithm (SGA) are encoded as chromosomes [19, 20]. A set of those chromosomes is termed "population" [21]. A random population is created initially to represent solution candidates to PF problem. The energy function is associated with the objective function to measure each candidate [22]. At each step, selection, crossover, mutation, and evaluation are implemented in sequence as in Fig. 3a.



(a) SGA [23] (b) Proposed CCGA

Fig. 3. Diagram of algorithms

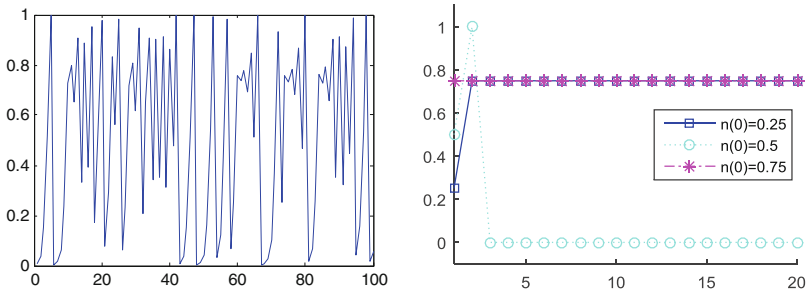
4.2 Two Improvements

The first improvement is we introduce the clonal selection mechanism from clonal selection algorithm (CSA), which is inspired by clonal selection theory of acquired immunity, which offers explanation how B and T lymphocytes enhances their response to antigens. CSAs are commonly applied to optimization fields, such as node detection [24], scheduling [25], feature selection [26], weight training [27], advanced intelligence turning [28], etc.

The next improvement is we introduce the chaos theory, because reproduction operator in clonal selection algorithm (CSA) and crossover operator in standard genetic algorithm (SGA) cannot generate any new variants to the current chromosome. To introduce mutations, we employed a chaotic number generator n_t on current chromosome:

$$n_{t+1} \leftarrow 4n_t(1 - n_t) \tag{2}$$

where $n_0 \in (0, 1)$ and $n_0 \notin \{0.25, 0.5, 0.75\}$. Figure 4 shows why the initial value of n_t cannot be the value of either 0.25, or 0.5, or 0.75.



(a) $n(0)$=0.01 (b) $n(0)$=(0.25,0.5,0.75)

Fig. 4. Time series of chaotic number

4.3 Our Chaotic Clonal Genetic Algorithm

The chaotic clonal genetic algorithm (CCGA) was proposed with above two improvements. On one hand, the chaotic operator of Eq. (2) was employed to add mutation to current chromosomes, with the aim of guaranteeing the dynamic ergodicity within solution space. On the other hand, the clonal selection and receptor edit mechanisms were included, so that a chromosome with larger affinity will have more chances to be reproduced meanwhile the population size keep unchanged.

The diagram of CCGA was pictured in Fig. 3b. At each step, all the chromosomes are sorted with regards to the fitness values. Afterwards, the whole set was segmented into two parts: top part and bottom part as shown in Fig. 5.

Clone operations will perform over the top part, and crossover and chaotic mutation operations will perform over the bottom part.

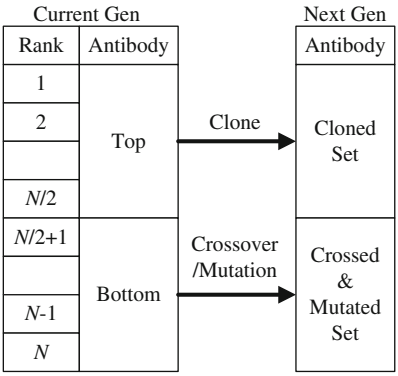


Fig. 5. Diagram of two parts

5 Experiments and Results

5.1 Parameter Setting

We compared the CCGA with standard genetic algorithm (SGA) [29], artificial immune system [30], and immune genetic algorithm [31]. Table 2 presented their parameters. Here S denotes the number of chromosomes, and MAX represents maximum iterative steps. Their values are equivalent for all algorithms for fair comparison. Besides, r denotes the receptor editing frequency, e denotes the chromosome elimination rate, C denotes for the crossover rate, and M denotes the mutation rate.

Table 2. Parameters setting

Approach	C	M	r	e	S	MAX
SGA [29]	0.3	0.08	–	–	1000	50000
AIS [30]	–	1	23	25 %	1000	50000
IGA [31]	0.3	0.15	–	10 %	1000	50000
CCGA (Proposed)	0.3	0.15	23	25 %	1000	50000

5.2 Algorithm Comparison

Table 3 lists the results of all four algorithms. We used "*success rate (SR)*" to measure all methods, which is defined as the ratio of success runs among all 100 runs.

Table 3. SR of different algorithms

Index	1	2	3	4	5
SGA [29]	94 %	2 %	0 %	0 %	0 %
AIS [30]	**100 %**	**99 %**	**81 %**	12 %	0 %
IGA [31]	**100 %**	78 %	22 %	6 %	1 %
CCGA (Proposed)	98 %	95 %	77 %	**14 %**	**3 %**

(Bold means the best)

Table 3 showed that CCGA achieves SR of 98 %, 98 %, 77 %, 14 %, and 3 % for all five protein-chains. It indicates that the CCGA has similar performance on small-size chains, but as the chain become longer (Index = 4 & 5), the proposed CCGA has better performance than SGA [29], IGA [31], and AIS [30]. The best structures found by CCGA are shown in Fig. 6.

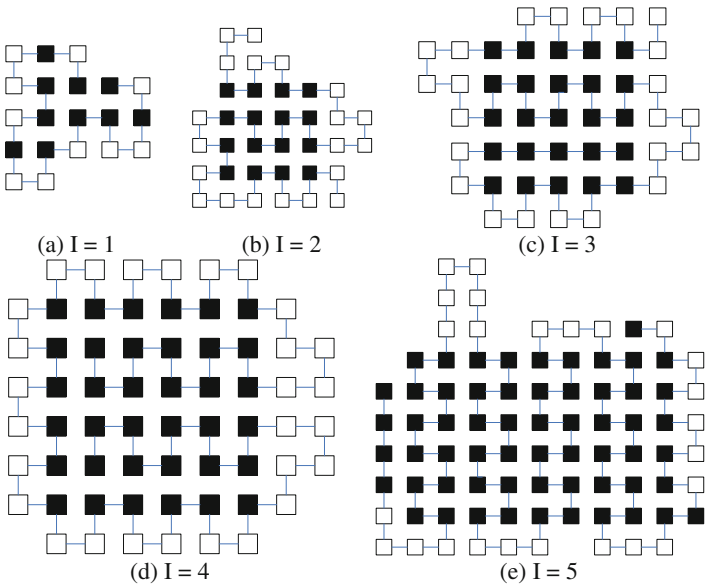


Fig. 6. The ideal solutions (I = Index)

6 Conclusions and Future Directions

We proposed a new global optimization method—chaotic clonal genetic algorithm (abbreviated as CCGA) based on GA, CSA, and chaotic operator. Experiment results show the superiority of CCGA to recent methods. Future work is composed of two folds. We shall try to increase the prediction performance for large-size chain, and we shall test other advanced global optimization algorithms, such as hybridization of swarm intelligence methods [32, 33].

Acknowledgment. This paper was supported by Natural Science Foundation of Jiangsu Province (BK20150983), Program of Natural Science Research of Jiangsu Higher Education Institutions (15KJB470010), Nanjing Normal University Research Foundation for Talented Scholars (2013119XGQ0061, 2014119XGQ0080), Open Project Program of the State Key Lab of CAD&CG, Zhejiang University (A1616).

References

1. Cumberworth, A., Bui, J.M., Gsponer, J.: Free energies of solvation in the context of protein folding: implications for implicit and explicit solvent models. J. Comput. Chem. **37**, 629–640 (2016)
2. Dumitriu, A., Golji, J., Labadorf, A.T., Gao, B.B., Beach, T.G., Myers, R.H., Longo, K.A., Latourelle, J.C.: Integrative analyses of proteomics and RNA transcriptomics implicate mitochondrial processes, protein folding pathways and GWAS loci in Parkinson disease. BMC Med. Genomics **9**, Article ID: 5 (2016)
3. An, P., Winters, D., Walker, K.W.: Automated high-throughput dense matrix protein folding screen using a liquid handling robot combined with microfluidic capillary electrophoresis. Protein Expr. Purif. **120**, 138–147 (2016)
4. Rivas-Pardo, J.A., Eckels, E.C., Popa, I., Kosuri, P., Linke, W.A., Fernandez, J.M.: Work done by titin protein folding assists muscle contraction. Cell Rep. **14**, 1339–1347 (2016)
5. Santos, J., Villot, P., Dieguez, M.: Emergent protein folding modeled with evolved neural cellular automata using the 3D HP model. J. Comput. Biol. **21**, 823–845 (2014)
6. Ji, G.: A comprehensive survey on particle swarm optimization algorithm and its applications. Math. Probl. Eng. **2015**, Article ID: 931256 (2015)
7. Agarwal, P., Bhatnagar, V., Balochian, S.: Swarm Intelligence and Its Applications 2014. Sci. World J. **2014**, Article ID: 204294 (2014)
8. Wang, S., Yang, J., Liu, G., Du, S., Yan, J.: Multi-objective path finding in stochastic networks using a biogeography-based optimization method. Simulation **92**, 637–647 (2016)
9. Othman, M.M., El-Khattam, W., Hegazy, Y.G., Abdelaziz, A.Y.: Optimal placement and sizing of voltage controlled distributed generators in unbalanced distribution networks using supervised firefly algorithm. Int. J. Electr. Power Energy Syst. **82**, 105–113 (2016)
10. Wu, L.: Magnetic resonance brain image classification by an improved artificial bee colony algorithm. Prog. Electromagn. Res. **116**, 65–79 (2011)
11. Zhang, Y.: Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. Expert Syst. Appl. **36**, 8849–8854 (2009)
12. Lin, C.J., Hsieh, M.H.: An efficient hybrid Taguchi-genetic algorithm for protein folding simulation. Expert Syst. Appl. **36**, 12446–12453 (2009)
13. Huang, C.H., Yang, X.B., He, Z.H.: Protein folding simulations of 2D HP model by the genetic algorithm based on optimal secondary structures. Comput. Biol. Chem. **34**, 137–142 (2010)
14. Narayanan, A.G.H., Krishnakumar, U., Judy, M.V.: An enhanced mapreduce framework for solving protein folding problem using a parallel genetic algorithm. In: Satapathy, S.C., Avadhani, P.S., Udgata, S.K., Lakshminarayana, S. (eds.) ICT and Critical Infrastructure: Proceedings of the 48th Annual Convention of Computer Society of India - Vol I. AISC, vol. 248, pp. 241–250. Springer, Heidelberg (2014)
15. Garza-Fabre, M., Rodriguez-Tello, E., Toscano-Pulido, G.: Constraint-handling through multi-objective optimization: the hydrophobic-polar model for protein structure prediction. Comput. Oper. Res. **53**, 128–153 (2015)

16. Giaquinta, E., Pozzi, L.: An effective exact algorithm and a new upper bound for the number of contacts in the hydrophobic-polar two-dimensional lattice model. J. Comput. Biol. **20**, 593–609 (2013)
17. Zhang, Y., Wu, L., Wang, S.: Solving two-dimensional hp model by firefly algorithm and simplified energy function. Math. Probl. Eng. **2013**, Article ID: 398141 (2013)
18. Huo, Y., Zhu, Q., Wang, S., Wu, L.: Polymorphic BCO for protein folding model. J. Comput. Inf. Syst. **6**, 1787–1794 (2010)
19. Zhang, Y., Wang, S., Ji, G., Dong, Z.: Genetic pattern search and its application to brain image classification. Math. Probl. Eng. **2913**, Article ID: 580876 (2013)
20. Al Adwan, F., Al Shraideh, M., Al Saidat, M.R.S.: A genetic algorithm approach for breaking of simplified data encryption standard. Int. J. Secur. Appl. **9**, 295–303 (2015)
21. Gibbs, M.S., Maier, H.R., Dandy, G.C.: Using characteristics of the optimisation problem to determine the genetic algorithm population size when the number of evaluations is limited. Environ. Modell. Softw. **69**, 226–239 (2015)
22. Javaid, W., Tariq, A., Hussain, I.: A comparison of a standard genetic algorithm with a hybrid genetic algorithm applied to cell formation problem. Adv. Mech. Eng. **11**, Article ID: 301751 (2014)
23. Wang, S., Lu, Z., Wei, L., Ji, G., Yang, J.: Fitness-scaling adaptive genetic algorithm with local search for solving the multiple depot vehicle routing problem. Simulation **92**, 601–616 (2016)
24. Sindhuja, L.S., Padmavathi, G.: Replica node detection using enhanced single hop detection with clonal selection algorithm in mobile wireless sensor networks. J. Comput. Netw. Commun. **13**, Article ID: 1620343 (2016)
25. Shui, X.G., Zuo, X.Q., Chen, C., Smith, A.E.: A clonal selection algorithm for urban bus vehicle scheduling. Appl. Soft Comput. **36**, 36–44 (2015)
26. Marinaki, M., Marinakis, Y.: A hybridization of clonal selection algorithm with iterated local search and variable neighborhood search for the feature selection problem. Memetic Comput. **7**, 181–201 (2015)
27. Cozma, P., Dragoi, E.N., Mamaliga, I., Curteanu, S., Wukovits, W., Friedl, A., Gavrilescu, M.: Modelling and optimization of CO2 absorption in pneumatic contactors using artificial neural networks developed with clonal selection-based algorithm. Int. J. Nonlinear Sci. Numer. Simul. **16**, 97–110 (2015)
28. Kim, D.H., Cho, J.H.: Advanced intelligence tuning using hybrid of clonal selection and genetic algorithm, GM and PM. Int. J. Comput. Intell. Appl. **14**, 19 (2015)
29. Lin, C.D., Anderson-Cook, C.M., Hamada, M.S., Moore, L.M., Sitter, R.R.: Using genetic algorithms to design experiments: a review. Qual. Reliab. Eng. Int. **31**, 155–167 (2015)
30. Zhang, Y., Huo, Y.: Artificial immune system for protein folding model. J. Converg. Inf. Technol. **6**, 55–61 (2011)
31. Zhang, T., Ding, Y., Shao, S.: Protein subcellular location prediction based on pseudo amino acid composition and immune genetic algorithm. In: Huang, D.-S., Li, K., Irwin, G.W. (eds.) ICIC 2006. LNCS (LNBI), vol. 4115, pp. 534–542. Springer, Heidelberg (2006)
32. Feng, C., Du, S., Yan, J., Wang, Q., Phillips, P.: Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. Int. J. Imaging Syst. Technol. **25**, 153–164 (2015)
33. Phillips, P., Dong, Z., Yang, J.: Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization. Progr. Electromagn. Res. **152**, 41–58 (2015)

http://www.springer.com/978-3-319-46256-1

Intelligent Data Engineering and Automated Learning – IDEAL 2016
17th International Conference, Yangzhou, China, October 12–14, 2016, Proceedings
Yin, H.; Gao, Y.; Li, B.; Zhang, D.; Yang, M.; Li, Y.; Klawonn, F.; Tallón-Ballesteros, A.J. (Eds.)
2016, XVI, 647 p. 209 illus., Softcover
ISBN: 978-3-319-46256-1