The Author(s) *BMC Systems Biology* 2017, **11**(Suppl 4):93
DOI 10.1186/s12918-017-0459-4

BMC Systems Biology

RESEARCH Open Access

Hybrid method to solve HP model on 3D lattice and to probe protein stability upon amino acid mutations

Yuzhen Guo[1*†], Fengying Tao[1†], Zikai Wu[4,5] and Yong Wang[2,3]

From The 10th International Conference on Systems Biology (ISB 2016)
Weihai, China. 19-22 August 2016

Abstract

Background: Predicting protein structure from amino acid sequence is a prominent problem in computational biology. The long range interactions (or non-local interactions) are known as the main source of complexity for protein folding and dynamics and play the dominant role in the compact architecture. Some simple but exact model, such as HP model, captures the pain point for this difficult problem and has important implications to understand the mapping between protein sequence and structure.

Results: In this paper, we formulate the biological problem into optimization model to study the hydrophobic-hydrophilic model on 3D square lattice. This is a combinatorial optimization problem and known as NP-hard. Particle swarm optimization is utilized as the heuristic framework to solve the hard problem. To avoid premature in computation, we incorporated the Tabu search strategy. In addition, a pulling strategy was designed to accelerate the convergence of algorithm based on the characteristic of native protein structure. Together a novel hybrid method combining particle swarm optimization, Tabu strategy, and pulling strategy can fold the amino acid sequences on 3D square lattice efficiently. Promising results are reported in several examples by comparing with existing methods. This allows us to use this tool to study the protein stability upon amino acid mutation on 3D lattice. In particular, we evaluate the effect of single amino acid mutation and double amino acids mutation via 3D HP lattice model and some useful insights are derived.

Conclusion: We propose a novel hybrid method to combine several heuristic strategies to study HP model on 3D lattice. The results indicate that our hybrid method can predict protein structure more accurately and efficiently. Furthermore, it serves as a useful tools to probe the protein stability on 3D lattice and provides some biological insights.

Keywords: Protein structure prediction, HP model, 3D lattice, Particle swarm optimization, Protein stability

Background

Protein is the substantial basis of biological activity. The function of protein is determined by its structure which is believed to be decided by the amino acid sequence according to Anfinsen's experiments. So the research on protein structure prediction (also called protein folding problem) is very significant and fundamental in exploring the fundamental principle to map sequence, structure, and function.

To capture the backbone of protein structure prediction, Dill and his collaborators introduced HP lattice model to simplify real world complexity in 1995 [1]. HP lattice model is an abstracted scaffold, and eventually convert the protein structure prediction problem to an optimization problem on lattice. The aim is to find the optimal structure with the lowest energy. Computationally, solving this problem is NP-hard. For this reason many researchers have been attracted to study this problem by proposing

*Correspondence: guoyuzhen@nuaa.edu.cn
†Equal contributors
[1]Department of Mathematics, Nanjing University of Aeronautics and Astronautics, 210000 Nanjing, People's Republic of China
Full list of author information is available at the end of the article

many heuristic algorithms. In recent years, for 2D HP protein folding problem, many methods have been proposed, e.g. , PSO (Particle Swarm Optimization) [2], ACO (Ant Colony Algorithm) [3], ABO (Artificial Bee Colony) [4] and SOM (Self-Organizing Mapping) [5] etc.

One issue for 2D lattice model is that it's too simplified to constrain the amino acid sequence on a 2D plane. One step forward is to fold the sequence on 3D lattice and make it a better and native approximation. So far, several algorithms have been applied for 3D HP protein structure prediction problem, such as UEGO (Universal Evolutionary Global Optimization) [6], GA (Genetic Algorithms) [7], TS (Tabu Search) [8], EA (Evolutionary Algorithm) [9] and so on. Each method has its advantage to capture some special structure in the problem. In this paper, we aim to propose a hybrid method and improve the efficiency to solve the 3D HP protein structure prediction problem.

PSO was introduced by Kennedy and Eberhart [10]. It is a swarm intelligence optimization algorithm which imitates the foraging behaviors of birds and fish. As a simple meta-heuristic, it has been used to solve optimization problem with nonlinear, non-differentiable, and multi-modal function. Originally, this algorithm was designed for solving continuous optimization problem. Here, we started from the basic PSO framework and firstly extend the algorithm to the combinatorial optimization, into which we formally formulate the HP model on 3D lattice. In addition, we improved PSO as follows: a) redefined velocity for discrete model; b) employed modified Tabu search strategy to avoid premature convergence; c) designed pulling strategy to speed up convergence.

We showed that our hybrid algorithm can predict structures of amino acid sequences with different length efficiently. With this useful tool, we simulated the effects after single amino acid mutation and double amino acids mutation, respectively. Some biological insights are obtained.

The remainder of this paper is organized as follows. Firstly, a mathematical model was established for 3D HP problem. Secondly, we explained the PSO algorithm and proposed modified Tabu search method and pulling strategy. Thirdly, the performance of our algorithm was validated. Fourthly, the amino acid mutation result was obtained and analyzed. Finally, conclusions were presented.

Methods

Combinatorial optimization formulation for 3D HP lattice model

In HP model, every amino acid sequence is abstracted as an alphabetic string with H (hydrophobic amino acid) and P (hydrophilic amino acid). The protein conformation is a self-avoiding path on a 2D lattice. It is assumed that the main driving forces of the formation of the tertiary structure are the interactions among hydrophobic amino acids which are adjacent on lattice but not adjacent in the sequence, denoted as H-H interactions. The free energy of a protein conformation (X) is expressed by the number of H-H interactions. Based on Anfinsen's assumption [11], the configuration tends to form a core in the spatial structure shield from the surrounding solvent by hydrophilic amino acids with the minimal free energy. So the more H-H interactions, the lower the free energy. We assumed that the free energy equals to the minus number of H-H interactions. HP lattice model has been used for solving protein structure prediction problem on 2D and 3D lattices widely. In this paper, we focused on the 3D HP square lattice model.

At present, relative coordinates and space coordinates have been used to denote the protein conformation. For a sequence S with L amino acids, X is a string of length $L-1$ over the symbols $\{r(ight), l(eft), f(orward), d(own), u(p)\}$ in relative coordinates, these five symbols reflect the relative location of contiguous amino acids on lattice. In space coordinates, X records the 3D coordinates of L amino acids, namely, $X = (X(1), X(2) \cdots X(L))$ and $X(l) \in N^3$ $(l = 1, 2 \cdots L)$ is the coordinate of the l^{th} amino acid. In this paper, we chose the space coordinates. For example, Fig. 1 showed a conformation with 7 H-H interactions on 3D square lattice. Its conformation was denoted as $X = ((2,3,2),(3,3,2),(3,4,2),(3,4,3),(3,3,3),(2,3,3),(2,2,3),(3,2,3),(3,2,2),(3,1,2),(2,1,2),(2,2,2))$.

Based on the abstraction and minimum energy principle, we established the optimization model (OM) for protein structure prediction problem on 3D square lattice as following:

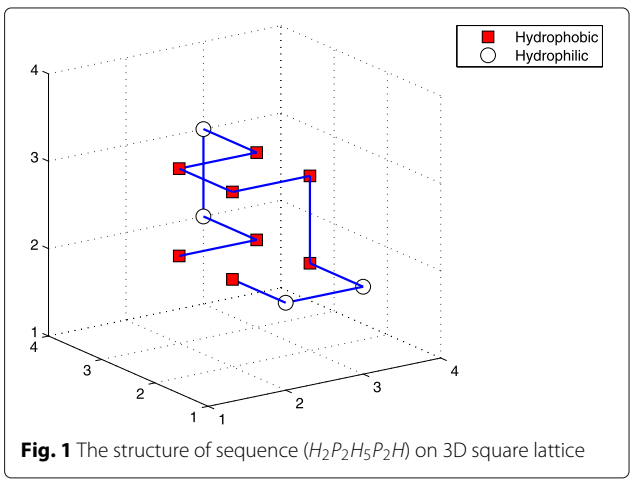

Fig. 1 The structure of sequence ($H_2P_2H_5P_2H$) on 3D square lattice

$$min \quad E(X) \tag{1}$$

$$s.t. \quad \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K} x_{i,j,k}(l) = 1 \qquad l = 1, 2 \cdots L \tag{2}$$

$$0 \le \sum_{l=1}^{L} x_{i,j,k}(l) \le 1 \qquad l = 1, 2 \cdots L \tag{3}$$

$$\sum_{d=1}^{3} |X(l+1)_d - X(l)_d| \cdot \|X(l+1) - X(l)\| = 1 \qquad l = 1, 2 \cdots L - 1 \tag{4}$$

Here,

$$E(X) = -M(X) \tag{5}$$

$$M = \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\sum_{l=1}^{L} x_{i,j,k}(l) f(l) \sum_{r=1}^{L} f(r)[x_{i,j,k+1}(r) + x_{i,j+1,k}(r) + x_{i+1,j,k}(r)] - h \tag{6}$$

$$h = \sum_{l=1}^{L-1} f(l) f(l+1) \tag{7}$$

$$x_{i,j,k}(l) = \begin{cases} 1 & \text{if the } X(l) = (i,j,k) \\ 0 & \text{else} \end{cases} \tag{8}$$

$$f(l) = \begin{cases} 1 & \text{if the } l^{th} \text{ amino acid is H} \\ 0 & \text{if the } l^{th} \text{ amino acid is P} \end{cases} \tag{9}$$

Where, $E(X)$ is the free energy of protein conformation X, $X(l)_d$ is the d^{th} component of $X(l)$, $M(X)$ is the number of H-H interactions in conformation X, r expresses the number of adjacent hydrophobic pairs in amino acid sequence and $\|\cdot\|$ is Hamming distance. Equations (2), (3) and (4) constrain that every amino acid occupies only one lattice point, each lattice point cannot be used more than once and adjacent amino acids in the chain occupy the adjacent points on the lattice. Equation (8) presents whether the l^{th} amino acid occupies point (i, j, k). In Eq. (9), $f(l)$ translates the l^{th} H (or P) of the amino acid sequence into 1 (or 0).

Solving the simplified HP model is NP-complete even on two dimensional lattice. Then we have to seek help from heuristic algorithms. Particle swarm optimization, one of the stochastic algorithm, serves as a powerful approximation method.

Hybrid algorithm

The basic PSO algorithm

Particle swarm optimization (PSO) is a heuristic framework that optimizes an objective function by iteratively improve a candidate solution. The motivation is to have a population of candidate particles, and move these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best known position, but is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. Finally it is expected to move the swarm toward the best solution. The advantage of PSO is that it makes no assumptions about the problem and can search very large spaces of candidate solutions.

In basic PSO algorithm (See Table 1), m particles search the optimal position simultaneously with dynamic velocity. Particle velocity is affected by iteration, own cognition, and social cognition of particle. Particularly, each particle can remember not only its own flight experience, but also the trajectories of all particles. In n dimensional search space, the position and velocity of the i^{th} particle are represented as $X_i \in R^n$ and $V_i \in R^n$, respectively. They are updated by the following two equations:

$$V_i^{t+1} = \omega V_i^t + c_1 r_1 \left(P_{ib}^t - X_i^t\right) + c_2 r_2 \left(P_{gb}^t - X_i^t\right) \tag{10}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{11}$$

Where P_{ib}^t and P_{gb}^t are the best position of the i^{th} particle and the best position of all particles in the t^{th} iteration, respectively. Inertia weight (ω), self confidence (c_1) and swarm confidence (c_2) are input parameters, r_1, r_2 are two separately generated uniformly distributed random numbers in the range [0,1].

The modified PSO algorithm

Definitions To solve the optimization model, we redefined position and velocity of PSO on 3D lattice. Particle position was orderly expressed by protein conformation (X). Velocity of particle was defined as a series of shift (j_1, j_2), which means that the j_1^{th} component of particle position becomes the j_2^{th} component, then the j_1^{th} component and the j_2^{th} component (including the j_2^{th} component) were changed subsequently. In addition, position X_1 was obtained by the sum of position X_2 and a series of shift, namely

Table 1 The process of basic PSO algorithm

Step 1	To **initialize** $\{X_i^0 \mid i = 1, 2 \cdots m\}$ and $\{V_i^0 \mid i = 1, 2 \cdots m\}$;
Step 2	To **calculate** $E(X_i^t)$, find P_{ib}^t and P_{gb}^t ;
Step 3	To **update** X_i^t and V_i^t;
Step 4	To **output** P_{gb}.

$X_1 = X_2 + \{(j_p, j_q)\}$. For example, $V = \{(2,4),(3,1)\}$ and $X = (X(1), X(2), X(3), X(4))$, then

$$\begin{aligned} X + V &= (X(1), X(2), X(3), X(4)) + \{(2,4),(3,1)\} \\ &= (X(1), X(3), X(4), X(2)) + \{(3,1)\} \\ &= (X(4), X(1), X(3), X(2)). \end{aligned}$$

Clearly, $X + V$ is a new position. Nevertheless, the new position may not satisfy the constraints in the OM model. An adjustment strategy is needed to ensure the new position was valid.

Modified Tabu search strategy Premature convergence is one of the major difficulty to solve OM model by PSO algorithm. To further improve the modified PSO, we adopted the idea of Tabu search which was proposed by Glover [12]. This method was briefly described as follows.

Tabu search is a meta-heuristic method that maintains only one solution in the iteratively searching process. Given an initial solution X, the idea is to calculate and compare its neighboring solutions $N(X)$. The best solution is chosen as candidate solution X_c. If X_c is satisfied with the aspiration rule, it will replace the current solution X and be added to tabu list T_{list}; Otherwise, the current solution X will be replaced by the best one X' ($E(X') = min\{E(X)|X \in N(X), X \notin T_{list}\}$) and X' will be added to T_{list}. Generally, T_{list} is a first-in first-out (fifo) memory with limited length. So particles would not search the solutions which have been found for a while, simultaneously, the better solutions would not always be taboo.

Neighbourhood of solution and aspiration rule are the key components of Tabu search. In our 3D HP problem, feasible solution is a 3D self-avoiding path. It was not easy to figure out its neighboring solutions from a given solution. According to Eqs. (10) and (11), we got similar solutions by changing r_1, r_2 at the same iteration for the same particle in PSO, then these solutions constituted a neighbourhood. When candidate solution was better than the current solution, we would ignore whether the candidate solution was taboo or not.

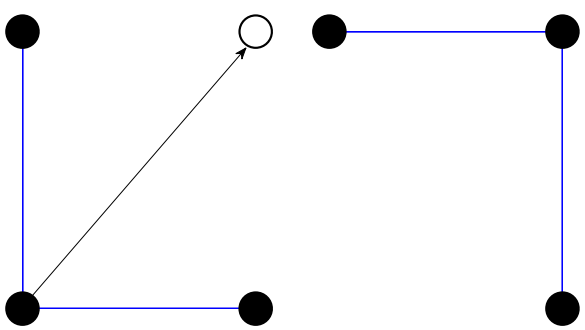

Fig. 2 The flow of pulling strategy. Left figure is for structure before pulled. Right figure is for structure after pulled

Table 2 The algorithm outline of TPPSO^2

Step 1	To **initialize** $\{X_i^0 \| i = 1, 2 \cdots m\}$, $\{V_i^0 \| i = 1, 2 \cdots m\}$ and $T_{list} = \emptyset$;
Step 2	To **calculate** $E(X_i^t)$, find P_{ib}^t and P_{gb}^t;
Step 3	To **update** $\{V_{ij}^t \| j = 1, 2 \cdots s\}$ and $\{X_{ij}^t \| j = 1, 2 \cdots s\}$;
Step 4	To **adjust** and pull $\{X_{ij}^t \| j = 1, 2 \cdots s\}$;
Step 5	To **calculate** $E(X_{ic}^t) = min\{E(X_{ij}^t) \| j = 1, 2 \cdots s\}$;
Step 6	**If** $E(X_{ic}^t) \le E(X_i^t)$ then $X_i^t = X_{ic}^t$;
Step 7	To **calculate** $E(P_{gbc}^t) = min\{E(X_i^t) \| i = 1, 2 \cdots m\}$;
Step 8	**If** $E(P_{gbc}^t) < E(X_{gb}^t)$ then $X_i^t = X_{ic}^t, T_{list} = \emptyset$;
Step 9	**If** $E(P_{gbc}^t) = E(X_{gb}^t)$ and $P_{gbc}^t \notin T_{list}$ then $T_{list} = T_{list} + X_{gb}^t, X_{gb}^t = X_{gbc}^t$;
Step 10	To **output** P_{gb}.

Pulling strategy The convergence rate of modified PSO with Tabu search strategy is not fast enough and the conformations obtained by this modified PSO may be too loose. The following strategy was designed in order to improve the algorithm.

In native protein structure, hydrophobic amino acids concentrate inside of conformation and they were surrounded by hydrophilic amino acids. If hydrophilic amino acids were pulled out of the central of protein structure, the structure will be more compact and more stable. Without changing structure's legitimacy, this strategy was defined as pulling strategy. In order to make pulled structure to satisfy the self-avoiding constraints, only one amino acid could be pulled to its vacant diagonal position once. Figure 2 showed the move and result of one pulling.

Table 3 Sequences with 27 amino acids used in our study

Sequence ID	Amino acids sequence
A_1	$PHPHPH_3P_2HPHP_{11}H_2P$
A_2	$PH_2P_{10}H_2P_2H_2P_2HP_2HPH$
A_3	$H_4P_5HP_4H_3P_9H$
A_4	$H_3P_2H_4P_3HPHP_2H_2P_2HP_3H_2$
A_5	$H_4P_4HPH_2P_3H_2P_{10}$
A_6	$HP_6HPH_3P_2H_2P_3HP_4HPH$
A_7	$HP_2HPH_2P_3HP_5HPH_2PHPHPH_2$
A_8	$HP_{11}HPHP_8HPH_2$
A_9	$P_7H_3P_3HPH_2P_3HP_2HP_3$
A_{10}	$P_5H_2PHPHPHPHP_2H_2PH_2PHP_3$
A_{11}	$HP_4H_4P_2HPHPH_3PHP_2H_2P_2H$

Table 4 Comparing four algorithms in eleven sequences with 27 amino acids

Sequence ID	EN	hELP	TPPSO[1]	TPPSO[2]
A_1	-9	-9(18009)	-9(1983)	-9(177)
A_2	-10	-10(9447)	-10(1304)	-10(439)
A_3	-8	-8(1420)	-8(1249)	-8(44)
A_4	-15	-15(2125)	-15(795)	-15(19)
A_5	-8	-8(2877)	-8(104)	-8(61)
A_6	-11	-12(2610)	-11(940)	**-12**(812)
A_7	-13	-13(3967)	-12(721)	**-13**(805)
A_8	-4	-4(1070)	-4(6)	-4(3)
A_9	-7	-7(363)	-7(389)	-7(14)
A_{10}	-11	-11(416)	-11(2784)	-11(83)
A_{11}	-14	-16(285)	-14(957)	**-16**(2672)

The number in parentheses is the iteration number before the lowest free energy values are found. TPPSO[2] can find the optimal results of all sequences. TPPSO[1] can't obtain the minimal free energies for sequence A_6, A_7, and A_{11} (highlighted in bold)

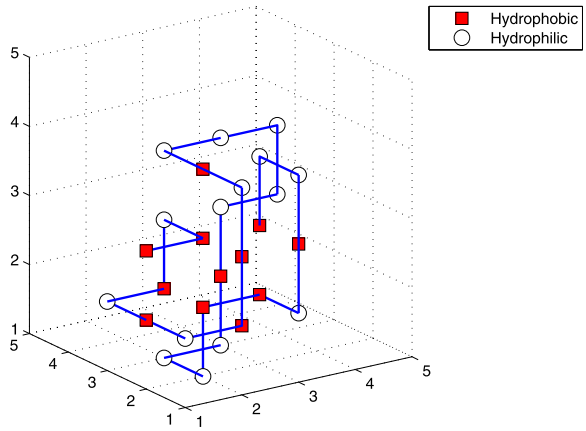

Fig. 4 This is one of structures for sequence A_7. This optimal conformation was simulated by TPPSO[2] with 13 H-H interactions. *Squares* are for hydrophobic amino acids, and *circles* are for hydrophilic amino acids. In this structure almost all hydrophobic amino acids are surrounded in center. It is stable with minimal free energy

Hybrid method A novel hybrid method was proposed by combining modified PSO with modified Tabu search strategy algorithm, denoted as TPPSO[1]. Another hybrid method was taken as TPPSO[2], which combined TPPSO[1] with pulling strategy. Both methods employed Tabu search strategy and were applied to solve protein structure prediction problem. In TPPSO[1] and TPPSO[2], when P_{ib} and P_{gb} were found, s alternative particles would be produced by Eqs. (10) and (11) for each particle.

We selected different r_1 and r_2 for finding alternative particles. These alternative particles might not satisfy the constraints, therefore they should be adjusted. Then the best alternative particle would replace the previous particle and P_{gb} would be taboo in a period of time. Differently, pulling strategy has been used in TPPSO[2], so each particle could be closer to optimal position. Table 2 showed the detailed procedures of TPPSO[2].

Results

Numerical simulations

In order to test the feasibility of the hybrid algorithms (TPPSO[1] and TPPSO[2]) and explore the properties of algorithms, we calculated two groups of amino acids sequences, respectively.

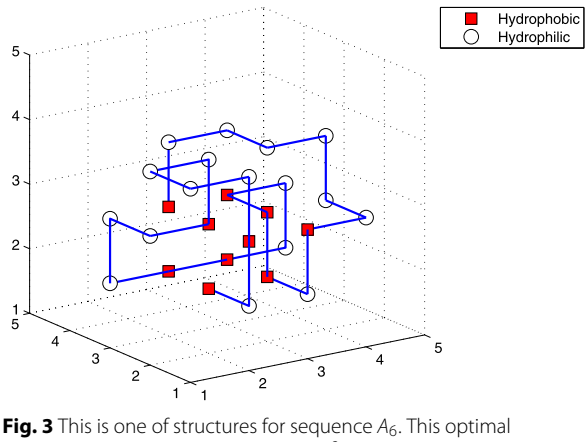

Fig. 3 This is one of structures for sequence A_6. This optimal conformation was simulated by TPPSO[2] with 12 H-H interactions. *Squares* are for hydrophobic amino acids, and *circles* are for hydrophilic amino acids. In this structure all hydrophobic amino acids are surrounded in center. It is stable with minimal free energy

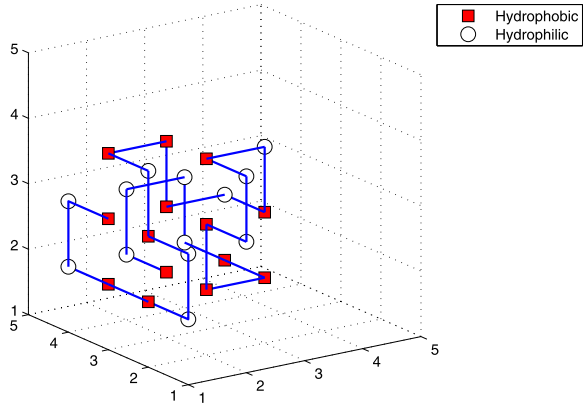

Fig. 5 This is one of structures for sequence A_{11}. This optimal conformation was simulated by TPPSO[2] with 16 H-H interactions. *Squares* are for hydrophobic amino acids, and *circles* are for hydrophilic amino acids. It is stable and compact with minimal free energy

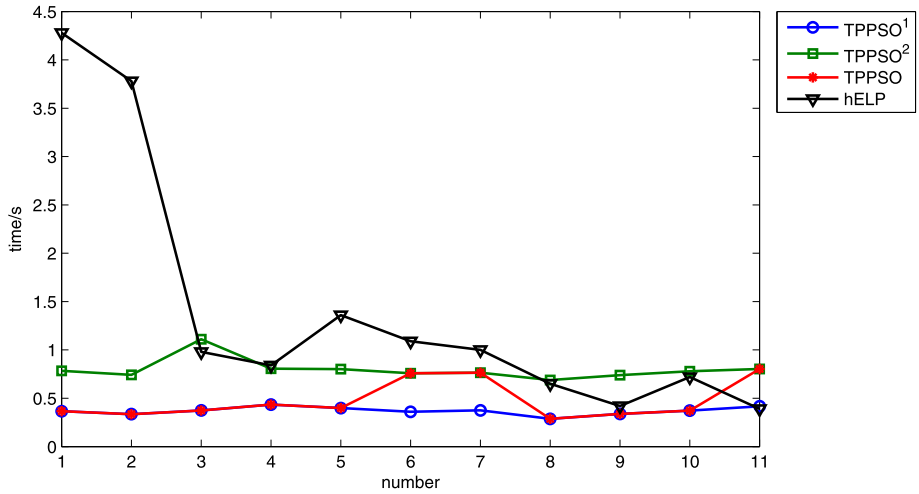

Fig. 6 The average CPU time of our methods and hELP. The abscissa is the number of sequence, and the ordinate is CPU time. Because TPPSO[1] can't obtain the minimal free energy of sequence A_6, A_7 and A_{11}, we chose smaller CUP time with minimal free energy for all sequences, denoted as TPPSO. In the figure, the CPU time of TPPSO[1] and TPPSO[2] are stable with respective optimal structure. CPU time of TPPSO also is stabler. But CPU time of hELP is fluctuant

Simulation of sequences with 27 amino acids

We selected 11 sequences with 27 amino acids (See Table 3) which were also computed by EN [13] and hELP [14]. These sequences were used to test the performances of TPPSO[1] (without pulling strategy) and TPPSO[2] (with pulling strategy), respectively. In TPPSO[1] and TPPSO[2], the inertia weight ω was updated by the following formula:

$$\omega = 0.1 - 0.05\frac{Time}{Maxtime} \tag{12}$$

The *Time* is the circular times and *Maxtime* is the maximum number of iterations which is 3000 in our implementation. For each particle, we chose $c_1 = c_2 = 1$, $r_{11} = rand(0.9, 1)$, $r_{12} = rand(0.82, 0.92)$, $r_{13} = rand(0.74, 0.84)$, $r_{21} = rand(0.9, 1)$, $r_{22} = rand(0.85, 0.95)$, $r_{23} = rand(0.8, 0.9)$ to produce three similar but not identical alternative particles. In this test, T_{list} only contained ten particles.

According to Table 4, we knows that all the sequences in Table 3 were simulated by EN, hELP, and our method TPPSO[1] and TPPSO[2]. hELP and TPPSO[2] can obtain the minimal free energy of every sequence, but EN and TPPSO[1] can't find the minimal free energies of sequence A_6, A_7, and A_{11} which are bigger. It illustrated our method can successfully predict the protein structure on 3D square lattice. The number in parentheses is the iteration number when the lowest free energy values are found. By comparing the results of hELP with TPPSO[1] and TPPSO[2], TPPSO[1] is superior to hELP, and TPPSO[2] can fold stable structures earlier than TPPSO[1].

Especially, TPPSO[2] found the lowest free energies of sequences A_6, A_7, and A_{11}, while TPPSO[1] or EN did not. So we enumerated the structures of these sequences (See Figs. 3, 4, and 5), which were computed by TPPSO[2]. It can be seen that these conformations were more native, furthermore, the hydrophobic amino acids were concentrated and surrounded by hydrophilic amino acids. Because pulling strategy of TPPSO[2] has not been employed in TPPSO[1], it is understandable that pulling strategy was able to accelerate the convergence of algorithm and optimize the protein structure.

The average CPU time of hELP was summarized in reference [14]. We also computed the average CPU time of TPPSO[1] and TPPSO[2]. The average CPU time of all methods were shown in Fig. 6. It is obvious that the average CPU time of every sequence of TPPSO[1] is the shortest, and that of TPPSO[2] is longer, because TPPSO[2] added the pulling strategy. For every sequence, the average CPU time

Table 5 IBE number of TPPSO[2]

Sequence ID	A_8	A_9	A_3	A_5	A_1	A_2	A_{10}	A_6	A_7	A_4	A_{11}
H-H[a]	4	7	8	8	9	10	11	12	13	15	16
IBE number[b]	3	14	44	61	177	439	83	812	805	19	2672

[a]H-H means the number of hydrophobic-hydrophobic amino acid interactions for optimal structure
[b]iteration numbers before the lowest free energy values are found

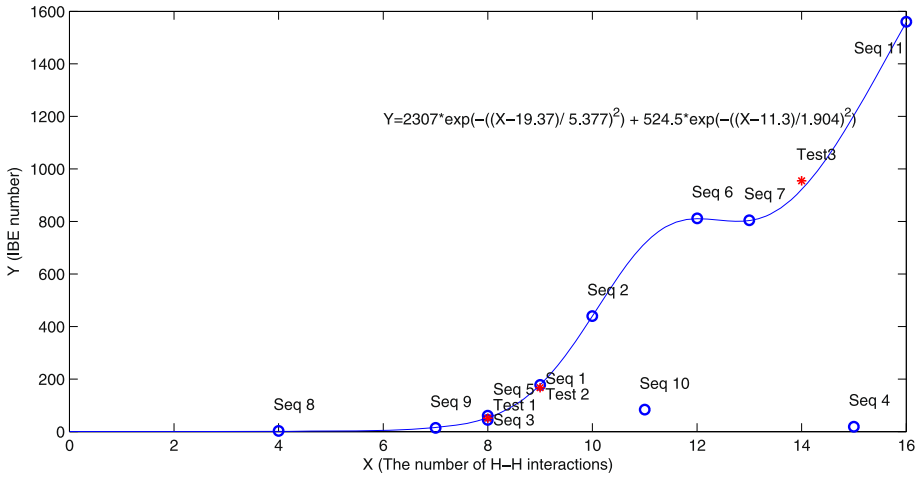

Fig. 7 The fitting figure of IBE number and H-H interaction for sequences in table 1 by TPPSO^2. The abscissa is H-H interaction of every sequence, and the ordinate is IBM number. Almost all sequences are satisfied with this fitting figure. IBE number will increase with H-H interaction for sequences with the same length. The three stars is the IBE number of test sequence with the same length 27. Their fitting function values are almost matched to computed IBE numbers

of TPPSO^1 and TPPSO^2 are stable and vary around 0.4 and 0.8 s respectively. However, the average CPU time of hELP is not stable. Since TPPSO^1 can't obtain the lowest free energy of sequence A_6, A_7, and A_{11}, we made TPPSO as the method which can fold the optimal structures of all sequences by PSO. The average CPU time of TPPSO was taken as less average CPU time of TPPSO^1 and TPPSO^2,which was also showed in Fig. 6. We know that the average CPU times of all sequences of TPPSO and hELP are 0.475 and 1.41 s respectively. It indicated that our method TPPSO is faster.

Table 5 summarized the number of H-H interactions and iteration number before the lowest free energy values are found by TPPSO^2 (denoted as IBE number) for eleven sequences with 27 amino acids in the Table 3. It is obvious that the more H-H interactions, the more IBE numbers. The fitting function of the number of H-H interactions and IBE number was given as follows.

$$y = 2307 * e^{-\frac{(x-19.37)^2}{5.377^2}} + 524.5 * e^{-\frac{(x-3)^2}{1.904}} \quad (13)$$

where x is the number of H-H pairs, and y is the IBE number.

Table 6 Test sequences

Sequence ID	Amino acids sequence	H-H	IBE number	Relative error
Test 1	$H_4P_5HP_5H_3P_8H$	8	51 (52.3829)	0.0271
Test 2	$H_4P_5HP_5H_3P_4HP_3H$	9	167 (177.8417)	0.0649
Test 3	$(HP_2HP)_5HP$	14	956 (921.1219)	0.0365

IBE number is the iteration number of every sequence by TPPSO^2 before the minimal free energy was found. The number in parentheses is IBE number calculated by fitting function

The figure of fitting function was exhibited in Fig. 7. Except for sequences A_4 and A_{10}, the IBE number of others are all close to the fitting function. The IBE number of sequence A_4 and A_{10} are not satisfied with the fitting function, because in these sequences H amino acids and P amino acids are very dispersive, but in other sequences H segments or P segments are longer. We believed that IBE number of TPPSO^2 is mainly affected by the number of H-H interactions for sequences with the same length. It tends to be larger with more H-H interactions. Moreover, the length of H or P segments will affect the IBE number.

In order to further verify the above conclusion, we simulated three test sequences with the same length (See Table 6). It is obvious that H segments and P segments of these test sequences are longer. IBE numbers for test sequences are close to the fitting curve (See Fig. 7) and all relative errors were showed in Table 6. It means that our inference about IBE number of TPPSO^2 is reasonable.

Table 7 Sequences with different lengths

Sequence ID	Amino acids sequence	Length	H-H	IBE number
B	$H_4P_2H_7P_3H$	17	9	2
C	$HPHP_2H_2PHP_2HPH_2P_2HPH$	20	11	11
D	$P_2HP_2H_2P_4H_2P_4H_2P_4H_2$	25	9	139
E	$P_3H^2P_2H_2P_5H_7P_2H_2P_4H(HP_2)_2$	36	17	432
F	$P_2H(P_2H_2)_2P_5H_{10}P_6(H_2P_2)_2HP_2H_5$	48	29	976

H-H interactions were the same by TPPSO^1 and TPPSO^2. IBE numbers were computed by TPPSO^2. These sequences were simulated by different methods, and every method only folds a part of sequences. TPPSO^2 found the minimal free energy of all sequences

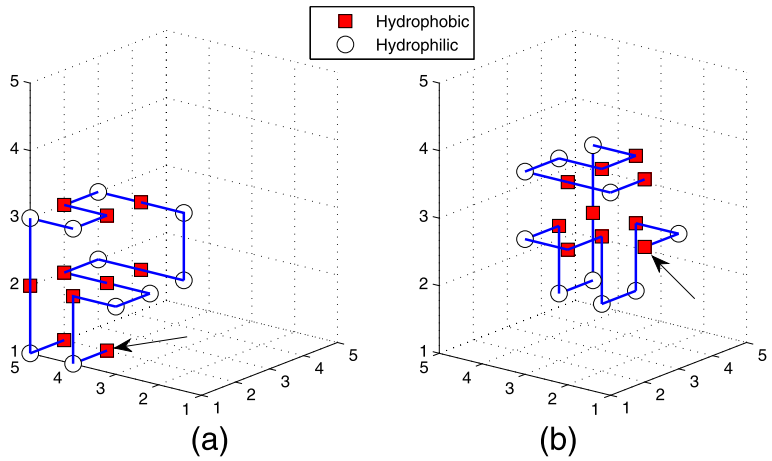

Fig. 8 These figures are the structures of sequence B with 20 amino acids. **a** is one of structures by TPPSO[1] with 11 H-H interactions. **b** is one of structures by TPPSO[2] with 11 H-H interactions. By comparing, *left* structure is more compact, *right* structure is looser

Simulation of sequences with different length

We also computed several sequences with different length which have not been solved by EN and hELP. Moreover, Table 7 recorded the H-H interactions and IBE number of TPPSO[2]. These sequences were simulated by TPPSO[1] and TPPSO[2] respectively. The results of two methods are the same (See Table 7). They have the same H-H interactions. But we knows that the CPU time of TPPSO[2] is shorter than one of TPPSO[1], because TPPSO[2] includes pulling strategy. For this reason, the structure obtained by TPPSO[2] is more compact. It is illustrated in Fig. 8.

These results shows that: a) TPPSO[2] is able to solve sequences with different length and the obtained characteristic of protein structure is significant. b) pulling strategy improved the performance. c) Tabu search strategy

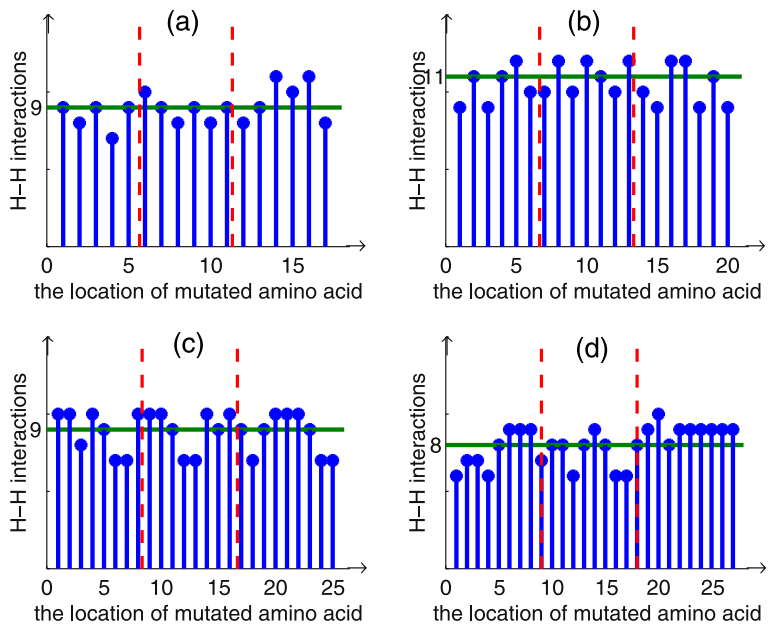

Fig. 9 These figures are H-H interactions of sequences after single amino acid mutation. The abscissa is the location of mutated amino acid, and the ordinate is the number of H-H interaction for mutated sequence. The *horizonal line* is the number of H-H interaction of original sequence. Two vertical lines spit sequence into three equal parts. **a** is mutational results of sequence B. In this figure, 100% pivotal amino acids locate at beginning or ending of sequence B. **b** is mutational results of sequence C. In this figure, 100% pivotal amino acids locate at beginning or ending of sequence C. **c** is mutational results of sequence D. In this figure, 71.4% pivotal amino acids locate at beginning or ending of sequence D. **d** is mutational results of sequence A_8. In this figure, 50% pivotal amino acids locate at beginning or ending of sequence A_8

Table 8 The single amino acid mutation results for sequence B

D-value	-3	-2	-1	0	1	2	3
Q-value	0	1	5	7	2	2	0
R-value	0%	**6%**	2%	41%	12%	**12%**	0%

There are 9 H-H interactions by TPPSO^2 for original sequence B. Every amino acid would be mutational, namely H (P) was changed into P (H). D-value is the deviation of H-H interactions between new sequence and original sequence when single amino acid was mutated. Q-value is the number of amino acids caused the deviation. R-value is the ratio of amino acids. The ratio of amino acids which caused the maximal deviation is 18% (summarization of the numbers highlighted in bold)

avoided prematurity effectively. d) For TPPSO^2, the longer the sequence, the more the IBE number.

Probing protein stability upon amino acid mutation

Protein stability determines whether a protein will be in its native folded conformation or a denatured state. The folded, biologically active conformation of a protein is believed more stable than the unfolded, inactive conformations [15]. Thus, making proteins more stable is important in medicine and basic research. Amino acid mutations are widely used in protein design and analysis techniques to increase or decrease stability. These mutations are carried out experimentally using site-directed mutagenesis and similar techniques. This is time-consuming and often requires the use of computational prediction methods to select the best possible combinations [16–19]. With the efficient hybrid method at hand, we aim to probe the protein stability on 3D lattice. Particularly, we will simulate how single-site or double amino acid mutation affects protein stability. i.e., predicting the protein stability changes upon amino acid mutations with TPPSO^2.

Single amino acid mutation

The hybrid method TPPSO^2 has been tested to solve protein structure prediction problem. Now, we focused on single amino acid mutation, whether and which amino acid affects the stability of protein structure. The experiments is designed as follows. We firstly calculate the optimal H-H interactions of original sequence by TPPSO^2. Then we choose one amino acid to mutate, i.e., we change it from H (P) into P (H). Then we calculate the optimal H-H interactions of mutated sequence by TPPSO^2. Finally the deviation of H-H interactions between mutated sequence and original sequence was recorded.

Table 9 The single amino acid mutation results for sequence C

D-value	-3	-2	-1	0	1	2	3
Q-value	0	5	5	4	6	0	0
R-value	0%	**25%**	25%	20%	30%	**0%**	0%

There are 11 H-H interactions by TPPSO^2 for original sequence C. Every amino acid would be mutational, namely H (P) was changed into P (H). The ratio of amino acids which caused the maximal deviation is 25% (summarization of the numbers highlighted in bold)

Table 10 The single amino acid mutation results for sequence D

D-value	-3	-2	-1	0	1	2	3
Q-value	0	7	1	6	11	0	0
R-value	0%	**28%**	4%	24%	44%	0%	0%

There are 9 H-H interactions by TPPSO^2 for original sequence D. Every amino acid would be mutational, namely H (P) was changed into P (H). The ratio of amino acids which caused the maximal deviation is 28% (summarization of the numbers highlighted in bold)

Sequences with different length In order to probe the stability of amino acid mutation, we chose four sequences with different lengths. These sequences were mentioned in the above section. They are sequence *B*, *C*, *D* and A_8.

Figure 9 recorded the H-H interactions of every single amino acid mutational sequence. From the results, we found that some mutational amino acids will result in a bigger deviation. We call those pivotal amino acids. The ratios of pivotal amino acids are 100%, 100%, 71.4% and 50% respectively for the four sequences. Also we deduced that those pivotal amino acids tend to locate at the beginning or end of sequence.

Tables 8, 9, 10 and 11 recorded the characters of mutated sequences including the deviation between mutated sequence and original sequence, the quantity of amino acid which mutated to cause the deviation and the ratio of every deviation. According to the results, we found that single amino acid mutation has the maximal and minimal deviation 2 and -2. The results also indicated that the ratio of maximal deviation is around 22%.

Table 12 summarized the effect of hydrophobic (hydrophilic) amino acid mutation on H-H interactions. We know that hydrophobic amino acid mutation would not make the H-H interactions increase and hydrophilic amino acid mutation would not lead the H-H interactions decrease. It means that hydrophilic amino acid mutation will result in more compact structure, while hydrophobic amino acid mutation will result in the looser structure. By comparing the results of H^2 and P^2 in the Table 12, we suppose that hydrophobic amino acid is more impressible than hydrophilic amino acid to reflect stability of protein structure for sequence with different lengths.

The structures of sequence B before and after mutation are showed in Fig. 10. Since the 14th amino acid was changed from P to H, the number of H-H interaction increases and the deviation is 2, which is the maximal

Table 11 The single amino acid mutation results for sequence A_8

D-value	-3	-2	-1	0	1	2	3
Q-value	0	5	3	7	11	1	0
R-value	0%	**19%**	11%	26%	41%	**3%**	0%

There are 8 H-H interactions by TPPSO^2 for original sequence A_8. Every amino acid would be mutational, namely H (P) was changed into P (H). The ratio of amino acids which caused the maximal deviation is 22% (summarization of the numbers highlighted in bold)

Table 12 Summary of the single mutation results

Sequence ID	H^0	P^0	H^1	P^1	H^2	P^2
B	12	5	12	5	1	2
C	10	10	10	10	5	0
D	9	9	16	16	7	0
A_8	6	21	6	21	5	1

H^0 and P^0 are the number of hydrophobic and hydrophilic in the original sequence. H^1 is the number of mutational hydrophobic amino acid whose H-H interactions is not more than the original one. P^1 is the number of mutational hydrophilic amino acid whose H-H interactions is not less than the original one. H^2 is the number of hydrophobic amino acid which caused the maximal deviation with original H-H pairs. P^2 is the number of hydrophilic amino acid which caused the maximal deviation with original H-H pairs

deviation. It is obvious that optimal structures of 14th amino acid mutation is more compact.

The structures of sequence D before and after mutation are showed in Fig. 11. Since the 6th amino acid was changed from H to P, the number of H-H interaction decreases and the deviation is 2 which is the maximal deviation. It is obvious that optimal structure of original sequence is more compact.

Sequences with the same length We selected five sequences from Table 3 to test what kind of protein structures are more stable upon single amino acid mutation by TPPSO^2. We changed every amino acid of these sequences, then recalculated and recorded the H-H interactions of every mutated sequence.

Table 13 showed that: 1) For the sequences with the same length, UC number is likely larger with more mass number. The exception might be caused by the fact that UC number reflected the stability of protein structure. Namely, sequence with larger UC number was susceptible to single amino acid mutation. 2) According to H-H interactions and the number of hydrophobic amino acids, the sequence with more hydrophobic amino acid usually had more H-H interactions. 3) From P→H and H→P, most of mutational hydrophobic amino acids made the H-H interactions changed. Relatively, only a part of hydrophilic amino acids affect the number of H-H interactions. We conclude that hydrophobic amino acid is more impressible than hydrophilic amino acid to reflect stability of protein structure for sequence with the same length.

All these results illustrated that: a) the more hydrophobic amino acids, the more H-H interactions; b) sequence with more H-H interactions tends to be more stable when single amino acid is mutated; c) hydrophobic amino acid mutation tends to alter the protein structure largely.

According to the above observations, we summarize that the sequence with more hydrophobic amino acids will be less susceptible to single amino acid mutation.

Double neighbouring amino acids mutation

Amino acid does not work alone and multiple amino acids coordinate to maintain stability and perform function. Our in-silicon simulation allows us to go beyond single amino acid mutation and explore the combinatorial effect of amino acid mutation. In this section, we explore the effect of double neighbouring amino acids mutation (two adjacent amino acids are mutated) in protein folding. Double neighbouring amino acids mutations were classified as HH → PP, PP → HH, HP → PH, PH → HP.

We simulated three sequences (B, C, D) with different length when adjacent amino acids mutated. The maximal deviation and locations of pivotal amino acids were conserved (See Tables 14, 15, 16).

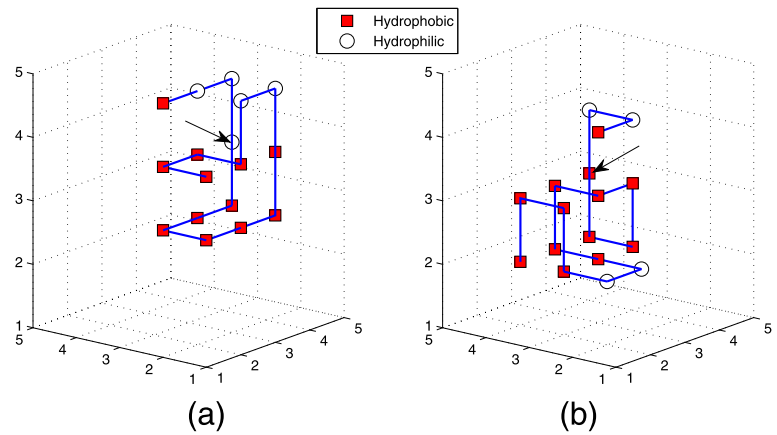

Fig. 10 These figures are the structures of sequence B before mutating and after mutating respectively. **a** is the optimal structure with 9 H-H interactions predicted for original sequence B by TPPSO^2. **b** is one of optimal structures predicted for mutated sequence B by TPPSO^2, in which the 14th amino acid was mutated. The mutated amino acid is denoted by *arrow* in figures. Since the 14th amino acid was changed from P to H, the number of H-H interaction increases and the deviation is 2. It is obvious that right structure is more compact

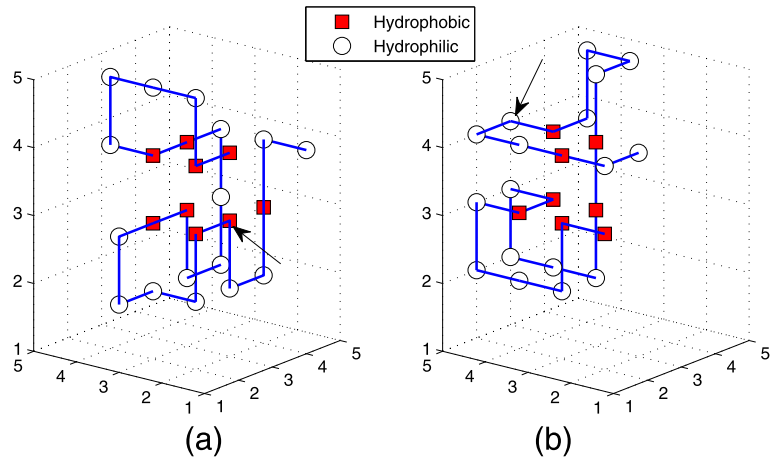

Fig. 11 These figures are the structures of sequence D before mutating and after mutating respectively. **a** is the optimal structure with 9 H-H interactions predicted for original sequence B by TPPSO[2]. **b** is one of optimal structures predicted for mutated sequence C by TPPSO[2], in which the 6th amino acid was mutated. The mutated amino acid is denoted by *arrow* in figures. Since the 6th amino acid was changed from H to P, the number of H-H interaction decreases and the deviation is 2. It is obvious that left structure is more compact

Tables 14, 15 and 16 recorded the variation of H-H interactions and the position of pivotal double amino acids . According to these tables, we concluded that: a) If double amino acids mutation was HH → PP or PP → HH, the H-H interactions must be changed. But PH→HP and HP→PH maybe have variation. b)HH → PP and PP → HH must make the H-H interactions decrease and increase, respectively. c) The effect of double adjacent amino acids mutation which belongs to HP → PH or PH → HP was finite. d) The position of pivotal double adjacent amino acids mutation tend to locate be at the head or tail of sequence.

Double arbitrary amino acids mutation

We continued to explore the combinatorial effect of amino acid mutation. In this section, we check the effect of double amino acids mutations with arbitrary distance in protein folding. The amino acid mutations were classified ed as HH → PP, PP → HH, HP → PH, PH → HP. We simulated the sequence B in Table 7 with 20 amino acids. There are 10 hydrophobic amino acids and 10 hydrophilic amino acids in sequence B. We folded the conformations of all of mutation sequences.

Form Table 17, we knew that combinations of H+H and P+P are more sensitive and easier to affect the stability of protein structure. We simulated every mutational sequence. The results showed that a) all HH→PP mutations will decrease HH interactions, namely HH interactions won't increase; b) 43 PP→HH mutations will increase HH interactions, 2 PP→HH mutations won't change HH interactions, in other words, HH interactions won't decrease; c) HP→PH or PH→HP mutations will not influence HH interactions.

The simulation results in Table 18 indicated that **a)** closer H and H (or P and P) can result in D-value; **b)** amino acid H_{18} respectively matches amino acid H_{20} and amino acid P_{10} to obtain maximal deviation of structure, so amino acid H_{18} is the most sensitive amino acid, which

Table 13 Single amino acid mutation of sequences with 27 amino acids

Sequence ID	H-H	mass number	UC number	H	P	P→H	H→P
A_8	4	9	4	6	21	19	4
A_3	8	7	2	9	18	17	8
A_5	8	8	7	9	18	12	8
A_{10}	11	17	7	11	16	9	11
A_4	15	13	10	14	13	3	14

H-H is the number of H-H pairs of original sequence. The same continuous amino acids were taken as a mass. Mass number is for original sequence. UC number is the number of mutational amino acid which does not change the H-H interactions. H (P) is the number of hydrophobic (hydrophilic) amino acids in the original sequence. P→H (H→P) is the number of mutational hydrophilic (hydrophobic) amino acid which affected the H-H interactions

Table 14 Double amino acids mutation results for sequence B

D-vale	HH → PP (9)[a]	PP → HH (3)[a]	HP → PH (2)[a]	PH → HP (2)[a]
-2	4(H,M,T)[b]	0	0	0
-1	5	0	1	0
0	0	0	0	0
+1	0	1	1	1
+2	0	2(T)	0	1(T)

[a]The number in parentheses is the number of adjacent amino acids in sequences. D-value is the deviation of minimal free energy caused by neighboring mutational amino acids

[b]The position of mutational double amino acids. H (M,T) means that the mutational double amino acids is at head (middle, tail) of the sequence

Table 15 Double amino acids mutation results for sequence C

D-vale	HH → PP (2)	PP → HH (3)	HP → PH (7)	PH → HP (7)
-3	1(H)	0	0	0
-2	1(T)	0	0	0
-1	0	0	3	3
0	0	0	4	4
+1	0	2	0	0
+2	0	1(H)	0	0

is at the tail of sequence; **c)** amino acid H_3 will cause D-value with hydrophilic (P), so amino acid H_3 is very sensitive to polar, it is at the head of sequence; **d)** matching amino acid H_7 with arbitrary amino acid P, HH interactions are invariable, so H_7 is obtuse, but by combining H_7 with arbitrary H, it is sensitive, this amino acid is in the middle of sequence; **e)** H_1 and H_{20} are impressible for other H, but they are stable for arbitrary P, the mutations of H_1 and H_{20} with all of P lead to decrease one more HH interaction, H_1 and H_{20} are at the head and tail of sequence, respectively.

According to the above observations, we summarized that a) double arbitrary amino acids mutation will be more sensitive to affect protein stability; b) double amino acids mutation with the same hydrophilic or hydrophobic property is more unstable than double amino acids mutation with different property; c) most of sensitive combinations are at the head or tail of sequence.

Discussion

As many research results indicate, HP model is very useful for modelling protein properties though it is simple and has many disadvantages. It captures the main difficulty of the real world problem. HP model has been applied in investigation of ligand binding to proteins [20]. The distinct influences of function, folding, and structure on the evolution of HP model are studied, by exhaustive enumeration of conformation and sequence space on a two dimensional lattice, which costs four week's computation [21]. These research all show that our effort to fold the HP chain by a hybrid method on 3D lattice is necessary and important.

Also we propose to use HP model to probe the protein stability. HP model serves as a very efficient tool here. The

Table 16 Double amino acids mutation results for sequence D

D-vale	HH → PP (4)	PP → HH (11)	HP → PH (4)	PH → HP (5)
-2	4(H,M,T)	0	0	0
-1	0	0	0	3
0	0	0	2	1
+1	0	5	1	1
+2	0	6(M,T)	1(T)	0

Table 17 Double arbitrary amino acids mutation results for sequence B

Combination	O-num	V-num	V-rate
H+H	45	45(↓)	100%
P+P	45	43(↑)	96%
H+P	50	29(↑↓)	58%
P+H	50	18(↑↓)	38%

H+H(P+P) means that arbitrary double hydrophobic(hydrophilic) amino acids will be mutated. H+P(P+H) means that hydrophobic(hydrophilic) will match with hydrophilic(hydrophobic) behind of it to mutate. O-num is original combination number in sequence. V-num is the number of combinations with which minimal free energy were altered after mutating. V-rate is the rate of V-num. The arrows in parentheses indicate increase or decrease of the free energy

simplification of 20 amino acids to H, and P types dramatically reduce the possible mutation pattern. Especially we can easily perform the double mutation only considering four combinations. Those insights from the HP model can serve as novel hypothesis to guide experiments. We also need to point out that the protein stability results and conclusions are heavily depending on the optimal solution of 3D HP model. We demonstrate the results in some small scale problems. When we want to generalize the study, we need to further improve the hybrid algorithm.

In our study, the computational experiments show that the new hybrid algorithm is efficient for short sequences. When the input space is bigger, there will be some suboptimal solutions and more difficult to find the minimal energy configurations. It's really a challenge for large scale HP model. The conformation space grows rapidly as the chain length increases. A possible method is to introduce divide-and-conquer strategy. We can also consider to combine with other algorithms or start from a good initial point from biological view. It will be our future work in devising such an algorithm for large protein.

Conclusion

In this paper, we studied protein structure prediction problem on 3D square lattice. We summarize the findings of this work as follows. Firstly, we formulated the

Table 18 Combination D-value and pivotal amino acids results for sequence B

Combination	D-value and pivotal amino acids	
H+H	-4	$H_1H_3, H_{18}H_{20}$
P+P	+2	$P_4P_5, P_4P_{13}, P_5P_8, P_8P_{17}, P_{10}P_{13}, P_{11}P_{16}, P_{13}P_{16}, P_{16}P_{19}$
H+P	-2	$H_3P_{10}, H_3P_{16}, H_3P_{19}, H_6P_{19}$
P+H	-3	$P_{10}H_{18}$

H+H(P+P) means that arbitrary double hydrophobic(hydrophilic) amino acids will mutate. H+P(P+H) means that hydrophobic(hydrophilic) will match with hydrophilic(hydrophobic) behind of it to mutate. D-value is the maximal deviation of H-H interactions between new sequence and original sequence when double arbitrary amino acids mutation. H_iH_j means that the i^{th} animo acid matches the j^{th} amino acid to mutate, in which two mutational amino acids are hydrophobic

protein structure prediction problem on 3D lattice into a combinatorial optimization problem; secondly, basic PSO algorithm has been enhanced to deal with discrete optimization problem; thirdly, we proposed a novel hybrid method (TPPSO^2) and proved its feasibility by simulating; fourthly, we derived some interesting insights for protein stability via single and double amino acid mutation perturbation.

Acknowledgements
The open access fee was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB13000000). This work is also supported by National Natural Fund under grant number 11601288, 11422108, 61621003, and 61304178.

Availability of data and materials
Not applicable

About this supplement
This article has been published as part of *BMC Systems Biology* Volume 11 Supplement 4, 2017: Selected papers from the 10th International Conference on Systems Biology (ISB 2016). The full contents of the supplement are available online at https://bmcsystbiol.biomedcentral.com/articles/supplements/volume-11-supplement-4.

Authors' contributions
YG developed and implemented the methods. YW and FT participated in the development of the methods. YW conceived the protein stability experiment. All authors draft, read, and approved the final manuscript.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
The authors declare that they have no competing interests.

Author details
[1]Department of Mathematics, Nanjing University of Aeronautics and Astronautics, 210000 Nanjing, People's Republic of China. [2]National Center for Mathematics and Interdisciplinary Sciences, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, 100190 Beijing, People's Republic of China. [3]University of Chinese Academy of Sciences, 100049 Beijing, People's Republic of China. [4]University of Shanghai for Science and Technology, 200433 Shanghai, People's Republic of China. [5]Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, 200433 Shanghai, People's Republic of China.

Published: 21 September 2017

References

1. Dill KA, Bromberg S, Yue K, et al. Principles of protein folding a perspective from simple exact models. Protein Sci. 1995;4(4):561–602.
2. Guo YZ, Wu Z, Wang Y, et al. Extended particle swarm optimization method for folding protein on triangular lattice. IET Sys Bio. 2016;10(1):30–33.
3. Nardelli M, Tedesco L, Bechini A. Cross-lattice Behavior of General ACO Folding for Proteins in the HP Model. Proc of the 28th Annual ACM Symp on Appl Comput. 2013;18(22):1320–1327.
4. Zhang Y, Wu L. Artificial Bee Colony for Two Dimensional Protein Folding. Adv Electr Eng Syst. 2012;1(1):19–23.
5. Zhang XS, Wang Y, Zhan ZW, Wu LY, Chen LN. Exploring protein's optimal HP configurations by self-organizing mapping. J Bioinf Comput Biol. 2005;3(02):385–400.
6. García-Martínez JM, Garzón EM, Cecilia JM, et al. An efficient approach for solving the HP protein folding problem based on UEGO. J Math Chem. 2015;53(3):794–806.
7. Lin CJ, Su SC. Protein 3D HP model folding simulation using a hybrid of genetic algorithm and particle swarm optimization. Int J Fuzzy Syst. 2011;13:140–147.
8. Benítez CMV, Lopes HS. Protein structure prediction with the 3D-HP side-chain model using a master-slave parallel genetic algorithm. J Braz Comput Soc. 2010;16:69–78.
9. Tsay JJ, Su SC. An effective evolutionary algorithm for protein folding on 3D FCC HP model by lattice rotation and generalized move sets. Proteome Sci. 2013;11(1):1.
10. Eberhart RC, Kennedy J. A new optimizer using particle swarm theory. Proc of the sixth Int Symp on micro Mach Hum Sci. 1995;1:39–43.
11. Anfinsen CB. Principles that govern the folding of protein chains. Sci. 1973;181(4096):223–230.
12. Glover F. Tabu search-part I. J Comput. 1989;1(3):190–206.
13. Guo YZ, Feng EM. The simulation of the three-dimensional lattice hydrophobic-polar protein folding. J Chem Phys. 2006;125(23):234703.
14. Liu J, Li G, Yu J, et al. Heuristic energy landscape paving for protein folding problem in the three-dimensional HP lattice model. Comput Biol Chem. 2012;28:17–26.
15. Pascal L, Stefan G, Abdullah K, Valentina C, Paul JB, Christian M, Mering C, Paola P. Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science. 2017;355(6327):812.
16. Parthiban V, Michael MG, Schomburg D. CUPSAT: prediction of protein stability upon point mutation. Nucleic Acids Res. 2006;34(suppl 2):W239-W242.
17. Cheng J, Randall A, Baldi P. Prediction of Protein Stability Changes for Single Site Mutations Using Support Vector Machines, Proteins: Str. Func Bioi. 2006;62:1125–32.
18. Shortle D, Stites WE, Meeker AK. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. Biochem. 1990;29:8033–41.
19. Perl D, Mueller U, Heinemann U, Schmid FX. Two exposed amino acid residues confer thermostability on a cold shock protein. Nat Struct Bio. 2000;7(5):380–3.
20. Miller DW, Dill KA. Ligand binding to proteins: The binding landscape model. Prot Sci. 1997;6(10):2166–79.
21. Blackburne BP, Hirst JD. Evolution of functional model proteins. J Chem Phys. 2001;115(4):1935–42.