

УДК 577.2

ХАОТИЧЕСКИЙ АЛГОРИТМ ИСКУССТВЕННАЯ “ПЧЕЛИНАЯ СЕМЬЯ”: НОВЫЙ ПОДХОД К ПРОБЛЕМЕ МИНИМИЗАЦИИ ЭНЕРГИИ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ БЕЛКА[#]

© 2013 г. Y. Wang^{1,2}, G. D. Guo^{1,2,*}, L. F. Chen^{1,2}

¹School of Mathematics and Computer Science, Fujian Normal University, Fuzhou, China

²Key Laboratory of Network Security and Cryptography, Fujian Normal University, Fuzhou, China

Поступила в редакцию 03.12.2012 г.

Принята к печати 19.04.2013 г.

Предсказание трехмерной структуры белка по его аминокислотной последовательности можно рассматривать как глобальную проблему оптимизации. В данной работе к предсказанию пространственной структуры белка применен хаотический алгоритм искусственная пчелиная семья. Основанный на вне-решеточной 3D-модели алгоритм сочетает в себе глобальный и локальный поиск искусственной пчелиной семьи с хаотическим алгоритмом поиска, чтобы избежать проблемы преждевременной сходимости и легкого попадания в локальное оптимальное решение. Эксперименты, проведенные с популярными последовательностями Фибоначчи, показывают, что предложенный алгоритм обеспечивает эффективный и высокопроизводительный метод предсказания структуры белков.

Ключевые слова: пространственная структура белков, вне-решеточные АВ-модели, искусственная пчелиная семья, хаотический поиск.

CHAOTIC ARTIFICIAL BEE COLONY ALGORITHM: A NEW APPROACH TO THE PROBLEM OF MINIMIZATION OF ENERGY OF THE 3D PROTEIN STRUCTURE, by Y. Wang^{1,2}, G. D. Guo^{1,2,*}, L. F. Chen^{1,2} (¹School of Mathematics and Computer Science, Fujian Normal University, Fuzhou, China; ²Key Laboratory of Network Security and Cryptography, Fujian Normal University, Fuzhou, China, *e-mail: ggd@fjnu.edu.cn). Prediction of the three-dimensional structure of a protein from its amino acid sequence can be considered as a global optimization problem. In this paper, the Chaotic Artificial Bee Colony (CABC) algorithm was introduced and applied to 3D protein structure prediction. Based on the 3D off-lattice AB model, the CABC algorithm combines global search and local search of the Artificial Bee Colony (ABC) algorithm with the Chaotic search algorithm to avoid the problem of premature convergence and easily trapping the local optimum solution. The experiments carried out with the popular Fibonacci sequences demonstrate that the proposed algorithm provides an effective and high-performance method for protein structure prediction.

Keywords: 3D protein structure, off-lattice AB model, artificial bee colony, chaotic search.

DOI: 10.7868/S0026898413060177

ВВЕДЕНИЕ

Задача предсказания структуры белка известна как проблема его фолдинга. Форма свернутого белка в значительной степени определяет его биологическую функцию. Быстрое получение информации о структуре имеет большое значение для изучения функции белка [1]. Таким образом, предсказание трехмерной структуры белка стало одной из самых важных задач исследования в области биоинформатики. В настоящее время основные методы исследования структуры белка – рентгеноструктурный анализ и ядерный магнит-

ный резонанс. Хотя в определении структуры белка наблюдается значительный прогресс, экспериментальные методы исследования его структуры все еще очень сложны и дороги. Поэтому предложено много методов с использованием компьютерного моделирования, в том числе гомологического выравнивания, узнавания укладки и *ab initio* [2].

В 1973 г. Анфинсеном (Anfinsen) предложена термодинамическая гипотеза, которая утверждает, что нативная структура белка соответствует глобальному минимуму свободной энергии его

[#] Текст представлен авторами на английском языке.

* Эл. почта: ggd@fjnu.edu.cn

поверхности [3], т.е. можно предсказать нативную структуру белка по его аминокислотной последовательности. Изучение предсказания фолдинга белков связано с двумя ключевыми проблемами. Первая – выдвигание математической модели, отражающей взаимодействие аминокислотных остатков и их окружения. Вторая проблема – разработка на основе термодинамической гипотезы мощного метода поиска, который может найти глобальный минимум среди множества локальных минимумов на энергетической поверхности белка [4]. Поскольку при увеличении длины последовательности белка расчет пространства поиска растет экспоненциально, нам нужна крупнозернистая модель, которая отличает нативную структуру от ненативной.

В настоящее время внерешеточную АВ-модель широко применяют для предсказания структуры белка. Считают, что главная сила формирования структуры белка – гидрофобные взаимодействия. Во внерешеточной модели АВ рассматривают два типа мономеров. Гидрофобные и гидрофильные мономеры обозначены как А и В соответственно. После принятия АВ-модели ключевой вопрос в предсказания 3D-структуры белков – это разработка алгоритма оптимизации для поиска лучшей нативной конформации белка. Однако при предсказании структуры белка также применяют правило NP-руки для упрощенных моделей [5].

В течение последних десятилетий исследователи разработали множество эвристических алгоритмов для решения глобальной проблемы оптимизации 3D-предсказания структуры белков [6–21]. Например, алгоритм, разработанный в работах [17–18], является гибридной схемой объединения генетического алгоритма (GA) с имитацией отжига и обладает значительно более высокой эффективностью по сравнению с другими методами. В работе [19] использовали улучшенный алгоритм поиска с запретами, что позволило избежать “зацикливания” поиска и повысило его производительность и надежность. Для генетического алгоритма поиска с запретами в работах [20–21] применили алгоритм запретов, который быстро и точно предсказывает 3D-структуру белка. Однако эти высокосложные методы имеют временное ограничение, так как время поиска очень велико.

Алгоритм искусственной пчелиной семьи (АВС) введен для улучшения двумерного моделирования фолдинга белков в работе [22]. Показано, что АВС-алгоритм при решении задачи оптимизации более производителен по сравнению с генетическим алгоритмом и другими эволюционными алгоритмами. Мы представили алгоритм хаотической искусственной пчелиной семьи (САВС), который является усовершенствованием алгоритма АВС с учетом предсказания трехмерной структуры белка. Она сочетает в себе алгоритм искусственной пчели-

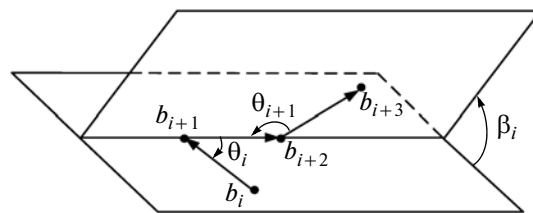


Рис. 1. Схема 3D-структуры белка.

ной семьи и хаотический алгоритм поиска, основанный на внерешеточной модели АВ. В процессе поиска, если алгоритм попадает в локальный оптимум, мы используем хаотические переменные, чтобы игнорировать локальное оптимальное решение. Такой метод не только способен вести глобальный и локальный поиск, но и позволяет избежать преждевременной сходимости и локального оптимума хаотического алгоритма поиска, который может осуществить глобальную оптимизацию.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Пространственная внерешеточная модель АВ. 3D-внерешеточная АВ-модель, предложенная Стиллинджером в работах [23, 24], разработана на основе 2D-решеточной модели АВ. В ней предполагается, что гидрофобные взаимодействия являются основной причиной сворачивания белка, аминокислотные остатки делятся на два типа – гидрофобные (А) и гидрофильные (В). Таким образом, первичную структуру белка можно представить как последовательность, состоящую из А и В. 3D-внерешеточная АВ-модель имеет следующие три преимущества [25]. Во-первых, неполнота баз данных белков не влияет на предсказания структуры белков. Мы можем определить структуру белка исключительно на полипептидной цепи. Во-вторых, модель реалистично имитирует основные особенности структуры белка. В-третьих, описание модели использует концепцию нейронных сетей, которая служит дальнейшим изучением сворачивания белка.

В трехмерном пространстве мономеры из аминокислотной последовательности соединены жесткими связями единичной длины с образованием линейных неориентированных полимеров и углом связи, образованным двумя связями, соединяющими три смежных мономера, которые могут быть заменены. При применении внерешеточной АВ-модели в качестве трехмерной модели предсказания структуры мы должны рассмотреть торсионные углы между двумя плоскостями, составленные из трех смежных со связью векторов. Как показано на рис. 1, форма белка из n аминокислотных остатков определяется $n-2$ валентными углами $\theta_1, \theta_2, \dots, \theta_{n-2}$ и $n-3$ торсионными угла-

ми $\beta_1, \beta_2, \dots, \beta_{n-3}$, где b_i ($i = 1, 2, \dots, n$) является i -тым мономером.

Таким образом, функционал энергии для n -мономеров белковой цепи можно описать следующим образом [24]:

$$E = \sum_{i=1}^{n-2} \frac{1}{4} (1 - \cos \theta_i) + \sum_{i=1}^{n-2} \sum_{j=i+2}^n 4[r_{ij}^{-12} - C(\xi_i, \xi_j)r_{ij}^{-6}], \quad (1)$$

где первый член обозначает потенциалы поворота основной цепи, которые не имеют отношения к последовательности, а второй член – потенциалы Леннарда-Джона несмежных мономеров. θ_i, β_i ($-\pi \leq \theta_i, \beta_i \leq \pi$) – валентные и торсионные углы соответственно. r_{ij} – расстояние между двумя несмежными мономерами, которое зависит как от валентного θ , так и от торсионного угла β . Постоянную $C(\xi_i, \xi_j)$ можно определить как:

$$C(\xi_i, \xi_j) = \begin{cases} +1, & \xi_i = \xi_j = 1 \\ +0.5, & \xi_i = \xi_j = -1, \\ -0.5, & \xi_i \neq \xi_j \end{cases} \quad (2)$$

где ξ_i – тип аминокислотных остатков, если аминокислота i гидрофобная (А), то $\xi_i = 1$, иначе $\xi_i = -1$. Как видно из уравнения (2), остатки АА сильно притягиваются, тогда, как остатки ВВ притягиваются слабо, а остатки АВ слабо отталкиваются. В некоторой степени уравнение отражает свойство реального белка, т.е. сворачивающийся белок образует гидрофобное ядро с сильным притяжением, а гидрофильные остатки находятся снаружи. Во внешней модели А предсказание 3D-структуры белка для n -мономеров белковой цепи эквивалентно нахождению оптимальных $n - 2$ валентных углов и $n - 3$ торсионных углов, которые минимизируют функционал энергии E . Следовательно, проблема предсказания 3D-структуры белка превращается в решение задачи условного экстремума целевой функции E [26]:

$$\min_{\theta_i, \beta_i \in (-\pi, \pi]} E(\theta_1, \theta_2, \dots, \theta_{n-2}, \beta_1, \beta_2, \dots, \beta_{n-3}). \quad (3)$$

Алгоритм искусственной пчелиной семьи (АВС) появился в 2005 г. [27]. Семья искусственных пчел состоит из трех групп: рабочие пчелы, пчелы-наблюдатели и разведчики. У каждой рабочей пчелы есть определенная еда, которая использована в итерации. После того как все рабочие пчелы завершат процесс поиска, они совместно с пчелами-наблюдателями используют информацию о

пищевых источниках. Пчела-наблюдатель выбирает источник пищи для дальнейшего использования. Если источник пищи не удалось использовать как лучший несколько раз, то он в дальнейшем не будет задействован, и пчелы-разведчики станут искать новые возможности для его замены случайным образом. Благодаря каждой рабочей пчеле, связанной только с одним источником питания, число рабочих пчел или пчел-наблюдателей равно числу источников пищи. В алгоритме АВС положение источника пищи представляет возможное решение проблемы оптимизации, и количество нектара в источнике пищи соответствует пригодности соответствующего решения.

Хаотический алгоритм поиска. Хаотичность – широко распространенная нелинейная система в природе, которая обладает характеристиками случайности, эргодичности и регулярности в соответствии со своими правилами, чтобы обойти все состояния в определенном диапазоне [28]. Одна из типичных хаотических систем – логистическая карта. Для N -мерного вектора размерности $Z_0, Z_1, Z_2, \dots, Z_n$, компоненты которых находятся в диапазоне $[0, 1]$, его логистическая итерационная формула:

$$Z_{n+1} = \mu Z_n (1 - Z_n), \quad n = 0, 1, 2, \dots, \mu \in (0, 4], \quad (4)$$

при $\mu > 3.5699456$ последовательность, итерированная с начальным значением Z_0 , хаотична [29], поэтому в этой статье принято $\mu = 4$.

Предсказания 3D-структуры белков с использованием алгоритма САВС. В этой работе алгоритм хаотической искусственной пчелиной семьи (САВС), который сочетает АВС-алгоритм и алгоритм хаотического поиска, применен к предсказанию 3D-структуры белка. Чтобы избежать попадания алгоритма в локальное оптимальное решение и ускорить сходимость алгоритма, хаотический алгоритм поиска вводится в АВС-алгоритм, который может эффективно предсказать структуру белка. 3D-предсказания структуры белка с применением САВС-алгоритма выглядят следующим образом.

Схема кодирования. В данной работе приняты декартовы координаты для представления источника пищи в САВС-алгоритме. Для n -остатков белковой цепи $h(\theta_1, \theta_2, \dots, \theta_{n-2}, \beta_1, \beta_2, \dots, \beta_{n-3})$ он может быть выражен с помощью $n - 2$ углов связи и $n - 3$ торсионных углов. Декартовы координаты остатка получены по уравнению (5) [19]:

$$pos(i) = \begin{cases} (0, 0, 0), & i = 1 \\ (0, 1, 0), & i = 2 \\ (\cos(\theta_1), \sin(\theta_2), 0), & i = 3 \\ \left(pos(i-1)_x + \cos(\theta_{i-2}) \cos(\beta_{i-3}), pos(i-1)_y + \sin(\theta_{i-2}) \cos(\beta_{i-3}), pos(i-1)_z + \sin(\beta_{i-3}) \right), & 4 \leq i \leq n \end{cases} \quad (5)$$

Координаты первых трех остатков $(0, 0, 0)$, $(0, 1, 0)$ и $(\cos(\theta_1), \sin(\theta_1), 0)$. Координаты последующих остатков вычисляются по предыдущим.

Инициализация источников питания. Согласно вышеприведенной схеме кодирования, 3D-структуру белка можно абстрагировать точкой в $2n - 5$ -мерном пространстве. САВС-алгоритм генерирует начальную популяцию источников питания SN случайным образом. Каждый источник пищи $X_i(x_1, x_2, \dots, x_{2n-5})$, $i = 1, 2, \dots, SN$ есть $2n - 5$ -мерный вектор. Размер первоначальной популяции оказывает огромное влияние на предложенный алгоритм. Если размер популяции слишком мал, диапазон поиска ограничен, и алгоритм легко попадает в локальный оптимум. В случае слишком большой популяции количество вычислений увеличивается, и скорость сходимости будет медленной. Таким образом, в данной работе размер начальной численности населения SN устанавливается в три раза больше, чем количество белка остатков, то есть $SN = (2n - 5) \times 3$.

Фаза поиска рабочих пчел. На этом этапе каждая рабочая пчела X_i генерирует новый источник пищи Y_i в окрестности ее текущего положения с помощью следующего выражения:

$$y_{ij} = x_{ij} + r(x_{ij} - x_{kj}), \quad (6)$$

где $K \in \{1, 2, \dots, SN\}$ и $\in \{1, 2, \dots, 2n - 5\}$ – случайным образом выбранные индексы, но значение k не равно j ; r представляет собой случайное число в диапазоне $[-1, 1]$. Новый источник пищи Y_i обновляется путем изменения j -мерного значения X_i . По уравнению (6), чем меньше разница между x_{ij} и x_{kj} , тем меньше изменение в новом положении источника питания. Поэтому, когда все источники питания близки к оптимальным решениям, размер шага поиска постепенно уменьшается.

В предлагаемом САВС-алгоритме необходимо сделать выбор между новым источником Y_i и предыдущими X_i , приняв “жадный” алгоритм отбора. При условии, что значение пригодности новой выборки лучше, чем у предыдущей, Y_i заменит X_i и станет новым членом популяции, в противном случае X_i сохранится. Если источник питания успешно не обновляется, соответствующий счетчик поиска C_i растет на 1, иначе – сбрасывается.

Фаза поиска пчелы-наблюдателя. После завершения всеми рабочими пчелами поиска они делятся своей информацией об источниках пищи с пчелами-наблюдателями. Эти пчелы в свою очередь оценивают информацию, взятую от всех рабочих пчел, и выбирают по вероятности источник пищи. Вероятностный выбор зависит от пригодности значений источников пищи в популяции. В предлагаемом алгоритме САВС принята турнирная схема выбора для подбора источников пищи пчелами-наблюдателями [30].

Турнирная схема выбора основана на локальном конкурсном отборе. Она работает следующим образом: выберите k особей случайным образом из популяции, а затем выберите лучшую из k , где k – размер турнира. В этой работе все источники пищи случайно разбиваются на пары (т.е. $k = 2$). Тогда вычисляем их энергии в соответствии с уравнением (1) и выбираем минимальные для пчелы-наблюдателя. Как только пчела-наблюдатель получает источник пищи, она начинает искать новый источник пищи так же, как и рабочие пчелы.

По сравнению с другими схемами выбора отборочный турнир использует относительные значения пригодности как критерии отбора, а это не накладывает никакого требования на положительное или отрицательное значение пригодности, которая может избежать влияния отдельных отличных особей на итерации. В некоторой степени эта схема выбора может предохранять предложенный алгоритм от преждевременного решения и заикливания [31].

Хаотический поиск новых источников питания. В алгоритме САВС при условии, что источник питания не может быть улучшен в пределах заданного числа запросов, источником питания становится локальное оптимальное решение. Тогда пчела-разведчик будет искать новый источник пищи, чтобы заменить его. Значение заданного числа запросов – важный параметр, который называется “ограничением”. Здесь хаотические последовательности производятся на основании текущего локального оптимального решения с использованием хаотического поиска. Лучшее решение в хаотических последовательностях заменит текущее решение. По хаотическому алгоритму поиска “застойное” решение продолжает эволюционировать и скорость сходимости алгоритма улучшается. Предположим, что застойное решение – X_k , шаги хаотического поиска следующие:

– шаг 1: наносим на карту переменные решения $X_k(x_{k1}, x_{k2}, \dots, x_{kd})$ к хаотическим переменным $Z_k(z_{k1}, z_{k2}, \dots, z_{kd})$, локализованным во внутренней области $[0, 1]$, где

$$z_{ki}^0 = \frac{x_{ki} - a_i}{b_i - a_i}, \quad (a_i = \min(x_{ki}), b_i = \max(x_{ki}));$$

– шаг 2: создание хаотических переменных для следующей итерации с логистическим уравнением, где $C_{\max} - Z_k^m$ ($m = 1, 2, \dots, C_{\max}$) максимальное количество итераций;

– шаг 3: преобразование хаотических Z_k^m переменных в переменные решения $X_k^i(x'_{k1}, x'_{k2}, \dots, x'_{kd})$ по следующему уравнению:

$$x'_{ki} = a_i + (b_i - a_i)z_{ki}^m, \quad i = 1, 2, \dots, d; \quad (7)$$

– шаг 4: оценка новых решений с переменным решением, и лучшее решение заменяет $X_k(x_{k1}, x_{k2}, \dots, x_{kd})$.

Описание алгоритма САВС. Основную идею САВС-алгоритма можно описать следующим образом.

1. Инициация: численность популяции SN , максимальное количество циклов $Maxloop$; ограниченные числа предельных поисков; максимальное количество хаотических поисков C_{max} ; текущий H_{min} оптимального решения и его энергии E_{min} ; поисковый счетчик $c_i = 0$ ($i = 1, 2, \dots, SN$).

2. Создание случайным образом SN источников питания для формирования популяции P .

3. Расчет энергетической ценности каждой еды по уравнению (1) и обновление текущего оптимального решения H_{min} и его энергии E_{min} .

4. Для счетчика цикла $loopcount = 1$: $Maxloop$:

1) Создание новых решений для рабочих пчел с помощью уравнения (6). Если новое решение лучше, чем старое, то замена старых решений, в противном случае соответствующие $c_i = c_i + 1$;

2) Генерация новых решений для пчел-наблюдателей из выбранных по турнирной схеме решений и их оценка;

3) Для $i = 1$: SN .

Если ($c_i == limit$), тогда получают новые решения для замены соответствующего решения для пчел-разведчиков с помощью хаотического поиска, и затем $c_i = 0$.

End.

4) Расчет энергетической ценности всех решений по уравнению (1) и обновление H_{min} и E_{min} .

End.

На выходе глобальный оптимум H_{min} решения и его энергии E_{min} .

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Последовательности Фибоначчи для экспериментов

В данной работе мы выполняем наши эксперименты с использованием последовательности Фибоначчи для проверки эффективности алгоритма САВС. Последовательности Фибоначчи определены рекурсивно

$$S_0 = A, S_1 = B, S_{i+1} = S_{i-1} \oplus S_i,$$

где \oplus означает оператор конкатенации. Первые несколько последовательностей $S_2 = AB$, $S_3 = BAB$, $S_4 = ABBA$ и т.д. В этих последовательностях Фибоначчи A и B – гидрофобные и гидрофильные остатки соответственно. Гидрофобный остаток встречается только по одному вдоль цепи, в то время как B – либо в одиночку, либо парами.

Подбор экспериментальных параметров

Алгоритм САВС реализован в 2008 г. с помощью MATLAB Windows XP. Экспериментально установлены следующие параметры: размер популяции – $(2n - 5) \times 3$ (n – длина последовательности), максимальное количество циклов – $1000n$, ограниченное количество поисков – 200, а максимальное число хаотических поисков – 300. Каждую последовательность выполняли 50 раз случайным образом, а затем мы брали средний результат как экспериментальный.

Экспериментальные результаты

Минимальные значения энергии E_{min}^{CABC} , полученной по алгоритму САВС на трехмерной внерешеточной модели AB , перечислены в таблице. Для сравнения мы также представили перечень минимальных значений энергии, получаемой с помощью отжига конформационного пространства (CSA) [11], алгоритма поиска с запретами (TS) [19] и генетического поиска с запретами (GATS) [21] соответственно. Все эти методы зависели от одной и той же энергии функционала (уравнение (1)), так же и одинаковая конформация дает такое же значение энергии всеми методами.

В таблице показано, что минимальные значения энергии E_{min}^{CABC} , полученной с помощью САВС-алгоритма, меньше, чем у полученных другими методами CSA, TS, и GATS для последовательностей длиной 13, 21, 34. В случае длины в 55 остатков алгоритмы CSA и GATS лучше, чем алгоритм САВС. Несмотря на то, что минимальное значение энергии, полученное по алгоритму САВС, не так мало, как полученное по предыдущим алгоритмам для длины 55, он имеет более

Сравнение минимальных значений энергии для четырех алгоритмов

N	Последовательность	E_{min}^{CSA}	E_{min}^{TS}	E_{min}^{GATS}	E_{min}^{CABC}
13	ABBAABBAABBA	-4.9746	-6.5687	-6.9539	-7.0025
21	BABABBAABBAABBA	-12.3266	-13.4151	-14.7974	-14.9570
34	ABBAABBAABBAABBAABBAABBAABBA	-25.5113	-27.9903	-27.9897	-28.0055
55	BABABBAABBAABBAABBAABBAABBAABBAABBAABBAABBAABBA	-42.3418	-41.5098	-42.4746	-42.2769

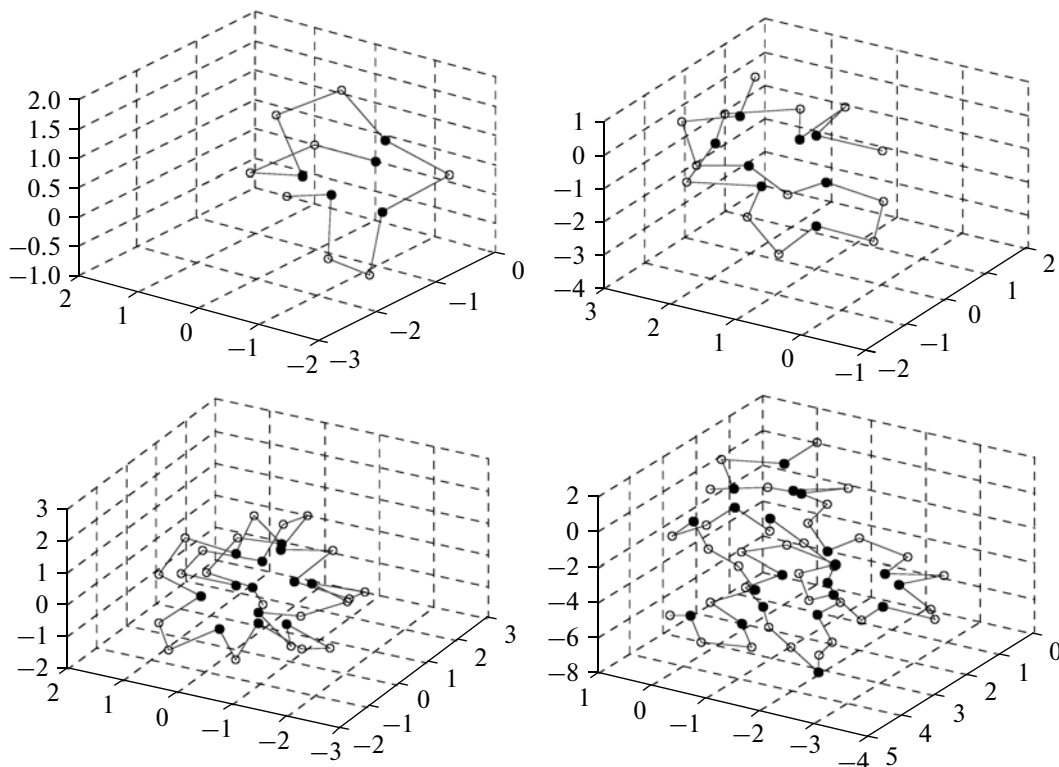


Рис. 2. Минимальные энергетические конформации для четырех последовательностей Фибоначчи.

высокую производительность для первых трех последовательностей, которые показывают, что САВС – эффективный алгоритм для предсказания 3D-структуры белков.

На рис. 2 показаны минимальные конформационные энергии в 3D-внерешеточной модели АВ, полученные с помощью алгоритма САВС, темные точки соответствуют гидрофобным мономерам А, в то время как белые точки представляют собой гидрофильные мономеры В. Для конформаций четырех последовательностей Фибоначчи мы видим, что гидрофобные остатки складываются в гидрофобное ядро, окруженное гидрофильными остатками, что наблюдается и в реальных белках. Результаты показывают, что от 3D-внерешеточную модель АВ целесообразно применять для имитации реального белка.

ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

В целях дальнейшего сравнения производительности алгоритма мы берем мультипопуляционный генетический алгоритм для сравнительных экспериментов. Как АВС-алгоритм, так и алгоритм GA – умные алгоритмы поиска оптимального решения. Существенно улучшенный генетический алгоритм применили для предсказания 3D-структуры белка. Мультипопуляционный генетический алгоритм обрабатывает

несколько популяций, которые имеют различный независимый и эволюционный механизм для одновременного развития. Гибридный алгоритм может поддерживать разнообразие популяции и избегать попадания в локальный оптимум

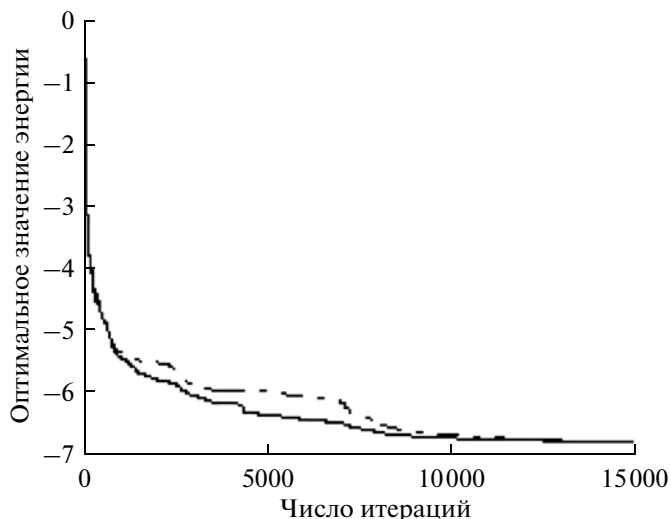


Рис. 3. Оптимальная энергия процесса эволюции кривых двух алгоритмов. Сплошная линия – предложенный нами алгоритм САВС, прерывистая – мультипопуляционный генетический алгоритм.

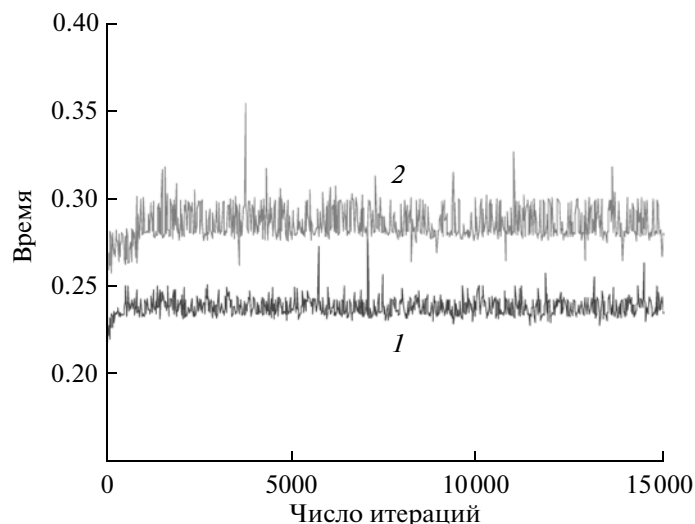


Рис. 4. Время работы двух алгоритмов в итерационном процессе.

решения с помощью обмена информацией между лучшими особями в различных популяциях.

На рис. 3 показана оптимальная энергия процесса эволюции кривой САВС-алгоритма и мультипопуляционного генетического алгоритма при $N = 13$ последовательности Фибоначчи. На начальной стадии развития скорость сходимости обоих алгоритмов велика, и затем они постепенно сливаются, но сходимость САВС-алгоритма лучше, чем мультипопуляционного генетического. Из-за особенностей целевой функции у алгоритма есть недостаток — легкое скатывание в локальный оптимум. В работе вводится хаотическая стратегия поиска, чтобы пройти локальное оптимальное решение как можно скорее, тем самым повышая скорость сходимости и точность. На рис. 4 представлено время работы двух алгоритмов в итерационном процессе. Эффективность алгоритма САВС выше, чем мультипопуляционного генетического алгоритма. Между тем, алгоритм САВС обладает небольшими отклонениями и хорошей стабильностью.

ВЫВОДЫ

В данной работе для борьбы с несколькими экстремумами и решения многопараметрической задачи оптимизации 3D-предсказания структуры белка использован хаотический алгоритм искусственной семьи пчел. На основе внерешеточной модели АВ принят хаотический алгоритм поиска, чтобы заставить задержанное решение выпрыгнуть из локального оптимума, который может хранить многообразие популяции и избегать преждевременной сходимости алгоритма. По сравнению с известными в литературе алгоритма-

ми, алгоритм САВС, кажется, обладает сильным глобальным потенциалом поиска и высокой производительностью. Это показывает, что это исследование предсказания структуры белка является возможным и эффективным. Кроме того, данный алгоритм показывает характеристику реального белка в некоторой степени. В дальнейшем мы будем для реализации алгоритма САВС использовать многоядерные вычислительные платформы для прогнозирования более сложных последовательностей. Кроме того, для ограничения внерешеточной модели АВ мы только разделили остатки на два типа и игнорировали взаимосвязь потенциалов изгиба основной цепи и последовательности белка. Поэтому получение лучших моделей, которые учтут разницу между гидрофобными взаимодействиями различных остатков, также составляет предмет дальнейших исследований. Кроме того, так как значения энергии зависят от функционала энергии, мы должны попытаться сравнить энергию и доступную поверхность гидрофобного растворителя при предсказании структуры белка.

Работа выполнена при поддержке Национального фонда естественных наук Китая (61070062) и ключевого проекта по сотрудничеству промышленности и Университета провинции Фуцзянь (2010H6007).

СПИСОК ЛИТЕРАТУРЫ

1. Koehl P. 2010. Protein structure prediction. *Biomed. Appl. Biophysics*. **3**, 1–34.
2. Lopes H.S. 2008. Evolutionary algorithms for the protein folding problem: A review and current trends. *Comput. Intelligence Biomedicine Bioinform.* **151**, 297–315.

3. Anfinsen C.B. 1973. Principles that govern the folding of protein chains. *Science*. **181**, 223–227.
4. Liwo A., Lee J., Ripoll D.R., Pillardy J., Scheraga H.A. 1999. Protein structure prediction by global optimization of a potential energy function. *Proc. Nat. Acad. Sci. USA*. **96**, 5482–5485.
5. Gabriel P.H.R., Melo V.V., Delbem A.C.B. 2012. Evolutionary algorithms and HP Model for protein structure prediction. *Sba: Controle & Automação Sociedade Brasileira de Automatica*. **23**, 25–37.
6. Khimasia M.M., Coveney P.V. 1997. Protein structure prediction as a hard optimization problem: the genetic algorithm approach. *Mol. Simulation*. **19**, 205–226.
7. Chen B., Johnson M. 2009. Protein local 3D structure prediction by Super Granule Support Vector Machines (Super GSVM). *BMC Bioinformatics*. **10**(Suppl 1), S15.
8. Chen X., Lv M.W., Zhao L.H., Zhang X.D. 2011. An improved particle swarm optimization for protein folding prediction. *Int. J. Inform. Engineering Electronic Business (IJIEEB)*. **3**, 1–8.
9. Dorn M., Buriol L.S., Lamb L.C. 2011. A hybrid genetic algorithm for the 3-D protein structure prediction problem using a path-relinking strategy. *Evol. Comput. (CEC)*. 2709–2716.
10. Islam M., Chetty M., Murshed M. 2011. Novel local improvement techniques in clustered memetic algorithm for protein structure prediction. *Evol. Comput. (CEC)*. 1003–1011.
11. Kim S.Y., Lee S.B., Lee J. 2005. Structure optimization by conformational space annealing in an off-lattice protein model. *Physical. Rev. E*. **72**, 011916.
12. Liu J.F., Huang W.Q. 2007. Quasi-physical algorithm of an off-lattice model for protein folding problem. *J. Comp. Sci. Technol.* **22**, 569–574.
13. Marks D.S., Colwell L.J., Sheridan R., et al. 2011. 3D protein structure predicted from sequence. Arxiv preprint arXiv: 1110.5091.
14. Novosam T., Snamel V., Abraham A., Yang J.Y. 2010. Searching protein 3-D structures for optimal structure alignment using intelligent algorithms and data structures. *Inf. Technol. Biomed., IEEE Transactions on*. **14**, 1378–1386.
15. Perdomo-Ortiz A., Dickson N., Drew-Brook M., Rose G., Aspuru-Guzik A. 2012. Finding low-energy conformations of lattice protein models by quantum annealing. Arxiv preprint arXiv:1204. 5485.
16. Pérez-Hernández L.G., Rodríguez-Vázquez K., Garduño-Juárez R. 2009. Parallel particle swarm optimization applied to the protein folding problem. *GECCO'09 Proceedings of the 11th Annual Conference on Genetic and Evolutionary Computation*, ACM New York. N.Y., USA, 1791–1792.
17. Zhang X.L., Lin X.L. 2010. Effective 3D protein structure prediction with local adjustment genetic-annealing algorithm. *Interdisciplinary Sciences: Comput. Life Sci.* **2**, 256–262.
18. Zhang X.L., Lin X.L., Wan C.P., Li T.T. 2007. Genetic-annealing algorithm for 3D off-lattice protein folding model. *Emerging Technol. Knowledge Discovery Data Mining*. **4819**, 186–193.
19. Zhang X., Cheng W. 2008. Protein 3D structure prediction by improved tabu search in off-lattice AB model. *The 2nd Intern. Conference of IEEE ICBBE*, pp. 184–187.
20. Wang T., Zhang X. 2011. A case study of 3D protein structure prediction with genetic algorithm and Tabu search. *Wuhan University J. Nat. Sciences*. **16**, 125–129.
21. Zhang X.L., Wang T., Luo H.P., Yang J.Y., Deng Y.P., et al. 2010. 3D protein structure prediction with genetic tabu search algorithm. *BMC Systems Biology*. **4**(Suppl 1), S6.
22. Zhang Y.D., Wu L. 2012. Artificial bee colony for two dimensional protein folding. *Adv. Electrical Engineering Systems*. **1**, 19–23.
23. Stillinger F.H., Head-Gordon T., Hirshfeld C.L. 1993. Toy model for protein folding. *Physical. Rev. E*. **48**, 1469–1477.
24. Stillinger F.H., Head-Gordon T. 1995. Collective aspects of protein folding illustrated by a toy model. *Physical. Rev. E*. **52**, 2872–2877.
25. Huang W., Liu J. 2006. Structure optimization in a three-dimensional off-lattice protein model. *Biopolymers*. **82**, 93–98.
26. Zhang X.L., Cheng W. 2009. Protein 3D structure prediction based on improved tabu search. *Computer Engineering*. **35**, 31–34.
27. Karaboga D., Basturk B. 2007. Artificial bee colony (ABC) optimization algorithm for solving constrained optimization problems. *Foundations Fuzzy Logic Soft Comput.* **4529**, 789–798.
28. Fei C.G., Han Z.Z. 2006. Improved chaotic optimization algorithm. *Control Theory Appl.* **23**, 471–474.
29. Wu X., Guan Z.H. 2007. A novel digital watermark algorithm based on chaotic maps. *Physics Letters A*. **365**, 403–406.
30. Blicke T., Thiele L. 1995. A mathematical analysis of tournament selection. *Proceedings of the Sixth International Conference on Genetic Algorithms*. 9–16.
31. Kolendo P., Jaworski B., Śmierzchalski R. 2011. Comparison of selection schemes in evolutionary method of path planning. *Comput. Collective Intelligence. Technol. Applications*. **6923**, 241–250.