

Решение задачи молекулярного докинга модифицированным методом роя частиц

04, апрель 2014

DOI: 10.7463/0414.0707258

профессор, д.ф.-м.н. Карпенко А. П., Матвеева К. О., Буланов В. А.

УДК 519.6

Россия, МГТУ им. Баумана

karpenko@rk6.bmstu.ru

kirandia1@gmail.com

vb001@rambler.ru

Введение

Википедия определяет молекулярный докинг (молекулярную стыковку) как метод молекулярного моделирования, позволяющий предсказать ориентацию и положение одной молекулы по отношению к другой, которые являются наиболее выгодными с точки зрения образования устойчивого молекулярного комплекса. Докинг молекул играет важную роль в процессе проектирования лекарственных препаратов, поскольку позволяет предсказать энергию связывания (аффинность) небольшой молекулы лекарства (лиганда) по отношению к белку-рецептору и, тем самым, оценить эффективность лекарства на основе белка-лиганда – чем сильнее молекула-лиганд связывается с белком-рецептором, тем эффективнее лекарство. Дорогостоящей альтернативой молекулярному докингу является метод экспериментального перебора потенциальных лигандов

Молекулярный докинг является вычислительно сложной процедурой, так как существует очень большое число вариантов расположения двух молекул относительно друг друга. С ростом чисел атомов во взаимодействующих молекулах число этих вариантов растет экспоненциально, так что исчерпывающий анализ всех вариантов взаимного расположения молекул невозможен.

Основной класс методов молекулярного докинга основан на отыскании такого расположения лиганда относительно молекулы-рецептора, которое минимизирует суммарную потенциальную энергию взаимодействующих молекул. Другими словами, методы докинга, как правило, основаны на решении задачи глобальной условной минимизации целевой функции (скоринг-функции), которая формализует указанную потенциальную

энергию. Для решения этой задачи широко используют различные эвристические методы оптимизации.

Цель работы состоит в исследовании эффективности известного эвристического метода оптимизации роением частиц и его двух авторских модификаций для решения задачи молекулярного докинга.

В зависимости от разрешенных положений и конфигураций взаимодействующих молекул в процессе решения задачи оптимизации различают следующие основные методы докинга.

Жесткий докинг (rigid protein docking) – конформация молекулы-рецептора (пространственное расположение атомов) в процессе процедуры докинга остаётся неизменной.

Гибкий докинг (flexible protein docking) – активный центр белка-рецептора в процессе докинга может изменять свою конформацию.

Полностью гибкий докинг (full protein flexibility) – все степени свободы белка-рецептора и лиганда используются в процессе докинга. Такой вид докинга в настоящее время не используют из-за неприемлемо высоких вычислительных затрат.

В работе применяем метод жесткого докинга, получивший в настоящее время наибольшее распространение.

Одной из основных задач, которую требуется решить для успешного проведения процедуры докинга, является задача выбора эффективного алгоритма глобальной минимизации скоринг-функции. Особенности этой задачи состоят в высокой размерности вектора варьируемых параметров, сложном ландшафте целевой функции, сложной топологии области допустимых значений вектора варьируемых параметров. В качестве указанного алгоритма могут быть использованы многие алгоритмы глобальной оптимизации – алгоритмы перебора, метод Монте-Карло, генетический алгоритм и др.

Наибольшее распространение среди алгоритмов, применяемых в программах докинга, получил генетический алгоритм оптимизации. Генетический подход используется, например, в программах *AutoDock* [1], *GOLD* [2], *MolDock* [3]. Этот же алгоритм используется в программе *SOL* [4]. Известны также программные комплексы молекулярного докинга, основанные на применении метода роя частиц. Так программа *SODOCK* [5] использует метод роя частиц с локальной оптимизацией, а программа *pso@autodock* [6] – метод роя частиц с топологией «клика», изменением коэффициента инерции и локальной оптимизацией.

В работе [6] показано, что в равных условиях при использовании метода роя частиц в программе докинга *AutoDock* 4.0 можно добиться более близкого положения лиганда к нативному (природному) положению, чем при использовании традиционного для этой программы генетического алгоритма.

Даем математическую постановку задачи молекулярного докинга, приводим схемы канонического метода роя частиц и двух его модификаций, представляем результаты вычислительного эксперимента по исследованию эффективности указанных методов и мо-

дификаций при решении тестовых задач оптимизации и одной задачи молекулярного докинга.

1. Постановка задачи

С математической точки зрения задача докинга состоит из трех подзадач:

- определить, какими степенями свободы будут наделены белок-рецептор и лиганд;
- составить скоринг-функцию, определяющую энергию связывания лиганда с белком-рецептором;
- выбрать эффективный алгоритм минимизации скоринг-функции.

Числа рассматриваемых степеней свободы лиганда и белка-рецептора определяют размерность пространства поиска. Очевидно, что чем больше число этих степеней свободы допустимо, тем выше размерность пространства поиска.

Энергию связывания лиганда с белком-рецептором можно определить на основе «точных» и приближенных законов взаимодействия между молекулами. Если использовать «точные» законы, то для расчета внутримолекулярных взаимодействий комплекса из нескольких десятков атомов (белки, которые учувствуют в процессе докинга, состоят обычно из тысяч атомов) потребуются сотни часов процессорного времени. Для приближенного описания взаимодействия молекул используют модель *силового поля* (*force field*) – набор потенциалов, описывающих взаимодействие атомов и молекул, как классических частиц. Расчет взаимодействий в комплексе из тысяч атомов занимает в этом случае десятки минут. В настоящее время известны такие модели силовых полей, как *MMX*, *MM3*, *Amber*, *MMFF94* [7].

Как правило, модели силовых полей используют следующие параметры силового поля, которые определяют на основе экспериментальных данных и расчетов методами квантовой механики:

- длины связей и величины валентных углов;
- константы деформации длин связей и величин валентных углов;
- параметры торсионных потенциалов внутреннего вращения;
- параметры Ван-дер-Ваальсовых взаимодействий пар атомов;
- заряды атомов.

Точность описания экспериментальных данных силовыми полями иллюстрирует следующий пример: для модели силового поля *MMFF94* [7] ошибка воспроизведения длины связей моделью составляет около 0,014 Å, а ошибка воспроизведения валентных углов - около 1,2°.

В работе используем модель силового поля *MMFF94* (*Merck Molecular Force Field 94*). Энергия взаимодействия молекул в рамках этой модели имеет вид

$$E(X) = \sum_{i,j} EB_{i,j}(X) + \sum_{i,j,k} EA_{i,j,k}(X) + \sum_{i,j,l} EBA_{i,j,l}(X) + \sum_{i,j,k,l} EOO_{i,j,k,l}(X) + \sum_{i,j,k,l} ET_{i,j,k,l}(X) + \sum_{i,j} E_{vd}W_{i,j}(X) + \sum_{i,j} EQ_{i,j}(X), \quad (1)$$

где суммирование производится по группам атомов, и слагаемые имеют следующие смысл.

Энергия растяжения связи

$$EB_{i,j} = 143,9325 \frac{kb_{i,j}}{2} \Delta r_{i,j}^2 \left(1 + cs \Delta r_{i,j} + \frac{7}{12} cs^2 \Delta r_{i,j}^2 \right),$$

где $\Delta r_{i,j} = r_{i,j} - r_{i,j}^{(0)}$ – рассчитанная и равновесная длины связи соответственно;

$kb_{i,j}, A^{-2}$ ккал/моль – силовая постоянная растяжения связи; $cs = -2, A^{-1}$.

Энергия деформации валентного угла

$$EA_{i,j,k} = 0,043844 \frac{ka_{i,j,k}}{2} \Delta \vartheta_{i,j,k}^2 \left(1 + cb \Delta \vartheta_{i,j,k} \right),$$

где $\Delta \vartheta_{i,j,k} = \vartheta_{i,j,k} - \vartheta_{i,j,k}^{(0)}$ – рассчитанный и равновесный валентные углы в градусах соответственно, $ka_{i,j,k}$ – силовая постоянная деформации валентного угла; $cb = -0,0070, \text{град}^{-1}$.

Энергия, обусловленная одновременным растяжением связи и деформацией валентного угла,

$$EBA_{i,j,k} = 2,51210 \Delta \vartheta_{i,j,k} \left(kba_{i,j,k} \Delta r_{i,j} + kba_{k,j,i} \Delta r_{k,j} \right),$$

где $kba_{i,j,k}, kba_{k,j,i}, \text{град}^{-1} A^{-1}$ ккал/моль – силовые постоянные для одновременного растяжения связи и деформации валентного угла.

Энергия взаимодействия, обусловленная выходом атома из плоскости

$$EOO_{i,j,k,l} = 0,043844 \frac{koop_{i,j,k,l}}{2} \chi_{i,j,k,l}^2,$$

где χ – угол между валентной связью атомов j, l и плоскостью j, k, l ; $koop_{i,j,k,l}, \text{град}^{-2}$ ккал/моль – силовая постоянная деформации угла между валентной связью и плоскостью.

Энергия вращения вокруг связи

$$ET_{i,j,k,l} = 0,5 (V_1 (1 + \cos \phi) + V_2 (1 + \cos 2\phi) + V_3 (1 + \cos 3\phi)),$$

где V_1, V_2, V_3 , ккал/моль – силовые постоянные вращения вокруг связи, ϕ – угол вращения (торсионный угол).

Энергия взаимодействия Ван-дер-Вальса

$$E_{vdW_{i,j}} = \varepsilon_{i,j} \left(\frac{1,07 R_{i,j}^*}{R_{i,j} + 0,07 R_{i,j}^*} \right) \left(\frac{1,12 (R_{i,j}^2)^7}{R_{i,j}^7 + 0,12 (R_{i,j}^*)^7} - 2 \right),$$

где $R_{i,j}$, Å – расстояние между i -м и j -м атомами; $\varepsilon_{i,j}$, ккал/моль, $R_{i,j}^*$, Å – параметры взаимодействия.

Энергия электростатического взаимодействия между атомами с парциальными зарядами q_1, q_2

$$EQ_{i,j} = 332,0716 \frac{q_i q_j}{R_{i,j}}.$$

Компонентами вектора варьируемых параметров X в модели силового поля *MMFF94* являются длины связей, валентные и торсионные углы Вильсона всех атомов, положения которых могут меняться в процессе решения задачи докинга.

Таким образом, ставим задачу докинга в виде задачи нелинейного программирования

$$\min_{X \in D} E(X) = E(X^*),$$

где $D \subset \mathbb{R}^{|X|}$ – множество допустимых значений вектора варьируемых параметров; X^* – искомый оптимальный вектор X , имеющий размерность $|X|$.

2. Метод роя частиц и его модификации

Канонический метод роя частиц. В основу алгоритма оптимизации роем частиц (*Particle Swarm Optimization, PSO*) положена социально-психологическая поведенческая модель толпы. Существует значительное число алгоритмов роя частиц. В каноническом алгоритме, предложенном в 1995 году Кеннеди (*J. Kennedy*) и Эберхартом (*R. Eberhart*), координаты частицы s_i на итерации t определяет вектор $X_i = (x_{i,1}, x_{i,2}, \dots, x_{i,|X|})^T$, а на итерации $(t + 1)$ – вектор $X'_i = (x'_{i,1}, x'_{i,2}, \dots, x'_{i,|X|})^T$. Начальные координаты частицы s_i равны $X_i(0)$, $i \in [1:|S|]$, где $|S|$ – число частиц в рое. Итерации в алгоритме *PSO* выполняются по схеме

$$X'_i = X_i + V_i, \quad (2)$$

$$V_i = b_I V_i^- + U_{|X|}(0; b_C) \otimes (X_i^* - X_i) + U_{|X|}(0; b_S) \otimes (X_i^{**} - X_i). \quad (3)$$

Здесь использованы следующие обозначения: $U_{|X|}(\mathbf{0}; \mathbf{b})$ - случайный $(|X| \times 1)$ -вектор, компоненты которого равномерно распределены в интервале $[0; b]$; \otimes - символ прямого произведения векторов; b_I, b_C, b_S - свободные параметры алгоритма; $V_i = V_i(t) = (v_{i,1}, v_{i,2}, \dots, v_{i,|X|})^T$ - $(|X| \times 1)$ -вектор приращения координат частицы; $V_i^- = V_i(t-1)$; X_i^* - вектор координат частицы s_i , соответствующий ее наилучшему значению фитнес-функции $\varphi(X)$ за время поиска $[0; t]$, то есть

$$\min_{\tau \in [0; t]} \varphi(X_i(\tau)) = \varphi(X_i^*); \quad (4)$$

X_i^{**} - вектор координат соседней с данной частицы с наилучшим за то же время значением фитнес-функции, то есть

$$\min_{j \in N_i} \varphi(X_j^*) = \varphi(X_i^{**}), \quad (5)$$

где N_i - множество номеров частиц, являющихся «соседями» данной частицы s_i [8].

Первое слагаемое в формуле (3) получило название *инерционной* компоненты, второе слагаемое - *когнитивной*, а третье слагаемое - *социальной* компоненты. Свободный параметр b_I определяет вес «инерционных» свойств частицы. Рекомендуемое значение параметра равно $b_I = 0,7298$. Значения параметров b_C, b_S определяют относительные веса когнитивной и социальной компонентов. Рекомендуемые значения параметров равны $b_C = b_S = 1,49618$.

Известно, что эффективность алгоритма *PSO* в значительной мере зависит от топологии соседства частиц в рое, определяемой неориентированным графом, вершины которого соответствуют частицам роя, а ребра связывают непосредственных соседей. Чаще всего используются топологии соседства частиц *клика* (глобально оптимальная топология), *кольцо* (локально оптимальная топология), *двумерный тор* (топология фон Неймана), *кластерная топология*. Известно, что от топологии соседства частиц в значительной мере зависит эффективность метода *PSO*, особенно при оптимизации многоэкстремальных функций, имеющих сложный ландшафт [8].

Схема алгоритма *PSO* имеет следующий вид.

1) Задаем значения свободных параметров алгоритма и инициализируем популяцию. Полагаем счетчик числа итераций $t = 1$.

2) Для каждой из частиц популяции s_i , $i \in [1: |S|]$ по формуле (4) находим лучшую локальную позицию X_i^* , а по формуле (5) - глобально лучшую позицию X_i^{**} .

3) По формулам (2), (3) находим новые позиции всех частиц популяции X_i' .

4) Проверяем выполнение условия окончания итераций. Если это условие выполнено, завершаем итерации, в противном случае полагаем $t = t + 1$ и переходим к шагу 2.

Инициализация популяции заключается в задании начальных позиций $X_i(0)$ и скоростей $V_i(0)$ частиц. Обычно точки $X_i(0)$ принимают случайными, равномерно распределенными в некотором гиперпараллелепипеде. Начальные скорости частиц также инициализируем случайными значениями.

В качестве условий окончания итераций обычно используют достижение максимально допустимого числа итераций \hat{t} или стагнацию итерационного процесса, то есть отсутствие улучшений фитнес-функции на протяжении δ_t итераций, где δ_t - свободный параметр алгоритма.

Модификация PSO-MI. Для каждой из частиц s_i , $i \in [1:|S|]$ введем в рассмотрение наборы частиц-соседей N_i^1, N_i^2 . Частицы роя S включаем в эти наборы случайным образом, допуская вхождение одной частицы в оба набора. Вероятность попадания частицы роя в набор N_i^j вычисляем по формуле

$$p_{i,j} = \frac{\exp\left(\ln|S| \frac{j}{3}\right)}{|S|}, \quad j = 1, 2.$$

Схему формирования наборов частиц-соседей N_i^1, N_i^2 иллюстрирует рисунок 1.

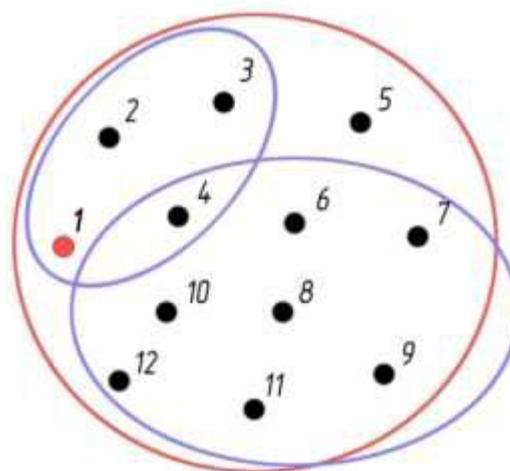


Рисунок 1 – Наборы соседей-частиц N_1^1, N_1^2 для частицы с номером 1 (синие овалы): $|S| = 12$

Формула (3) в алгоритме PSO-MI имеет вид

$$V_i = b_I V_i^- + U_{|X|}(0; b_C) \otimes (X_i^* - X_i) + U_{|X|}(0; b_S) \otimes (X_i^{**} - X_i) + U_{|X|}(0; b_1) \otimes (X_i^1 - X_i) + U_{|X|}(0; b_2) \otimes (X_i^2 - X_i), \quad i \in [1:|S|]. \quad (6)$$

Здесь X_i^1, X_i^2 - координаты лучших, в смысле формул (4), (5), частиц наборов N_i^1, N_i^2 , соответственно; $b_1, b_2 \in [0; 1]$ - свободные параметры алгоритма.

Модификация *PSO-M1* имеет целью диверсифицировать поиск и тем самым предупредить его преждевременную сходимость.

Модификация *PSO-M2*. Введем в рассмотрение среднеквадратическое отклонение координат частиц

$$\sigma = \frac{1}{|X||S|-1} \sqrt{\sum_{i=1}^{|S|} \sum_{j=1}^{|X|} (x_{i,j} - \bar{x}_j)^2},$$

где \bar{x}_j - j -я компонента вектора текущего центра роя:

$$\bar{x}_j = \frac{1}{|S|} \sum_{i=1}^{|S|} x_{i,j}, \quad j \in [1:|X|].$$

Модификация *PSO-M2* метода роя частиц заключается в варьировании в формуле (6) значений «инерционного» параметра b_I по правилу

$$b_I = \begin{cases} \alpha b_I, & \frac{\sigma(t)}{\sigma(t-\Delta t)} > \gamma, \\ \beta b_I, & \frac{\sigma(t-\Delta t)}{\sigma(t)} < \gamma. \end{cases}$$

Здесь $\alpha > 1$, $\beta, \gamma \in (0; 1)$, $\Delta t > 0$ - свободные параметры метода. Смысл данного правила заключается в следующем: если относительное значение величины σ в течение итераций с $t - \Delta t$ по t уменьшилась на величину, превышающую γ , то параметр b_I увеличиваем (с целью предотвратить преждевременную сходимость метода); если то же значение величины σ увеличилось на указанную величину, то параметр b_I уменьшаем (с целью интенсифицировать поиск). Рекомендованные значения величин $\alpha, \beta, \gamma, \Delta t$ равны 1,1; 0,9; 0,7 соответственно.

3. Исследование эффективности метода роя частиц и его модификаций

Программная реализация алгоритма выполнена на языке программирования высокого уровня C++, разработка выполнена в среде *GCC*.

Принимаем число части в рое (размер популяции) равным 50. Используем метод мультистарта с числом запуском алгоритма $m = 30$. Эффективность алгоритма оцениваем средним числом вычислений целевой функции \bar{n}_φ , а также средней и минимальной ошибками $\bar{\varepsilon}$, ε_{\min} локализации минимума целевой функции.

Сферическая функция. Рассматриваем сферическую целевую функцию

$$\varphi(X) = \sum_{i=1}^{|X|} x_i^2$$

в области генерации популяции

$$D = \{X \mid x_i^- \leq x_i \leq x_i^+, i \in [1:|X|]\}, \quad (6)$$

где $|X| = 16$.

Результаты вычислительного эксперимента представлены в таблице 1. Результаты показывают работоспособность предложенных модификаций канонического метода *PSO* и соответствующего программного обеспечения. Отметим преимущества модификации *PSO-M2* по сравнению с двумя другими методами по всем трем критериям эффективности.

Таблица 2 – Результаты исследования эффективности метода *PSO* и его модификаций: сферическая целевая функция

Метод <i>PSO</i>	Критерий эффективности		
	\bar{n}_φ	$\bar{\varepsilon}$	ε_{\min}
Канонический метод	6971	$7,0 \cdot 10^{-5}$	$8,1 \cdot 10^{-6}$
Модификация <i>PSO-M1</i>	6858	$8,2 \cdot 10^{-5}$	$3,5 \cdot 10^{-6}$
Модификация <i>PSO-M2</i>	4798	$6,3 \cdot 10^{-5}$	$3,3 \cdot 10^{-6}$

Функция Растригина. Исследование эффективности предложенных модификаций метода *PSO* выполнено для многоэкстремальной функции Растригина (*Rastrigin function*) [9]

$$f(X) = \sum_{i=1}^{|X|} (10 + x_i^2 - 10 \cos(2\pi x_i)),$$

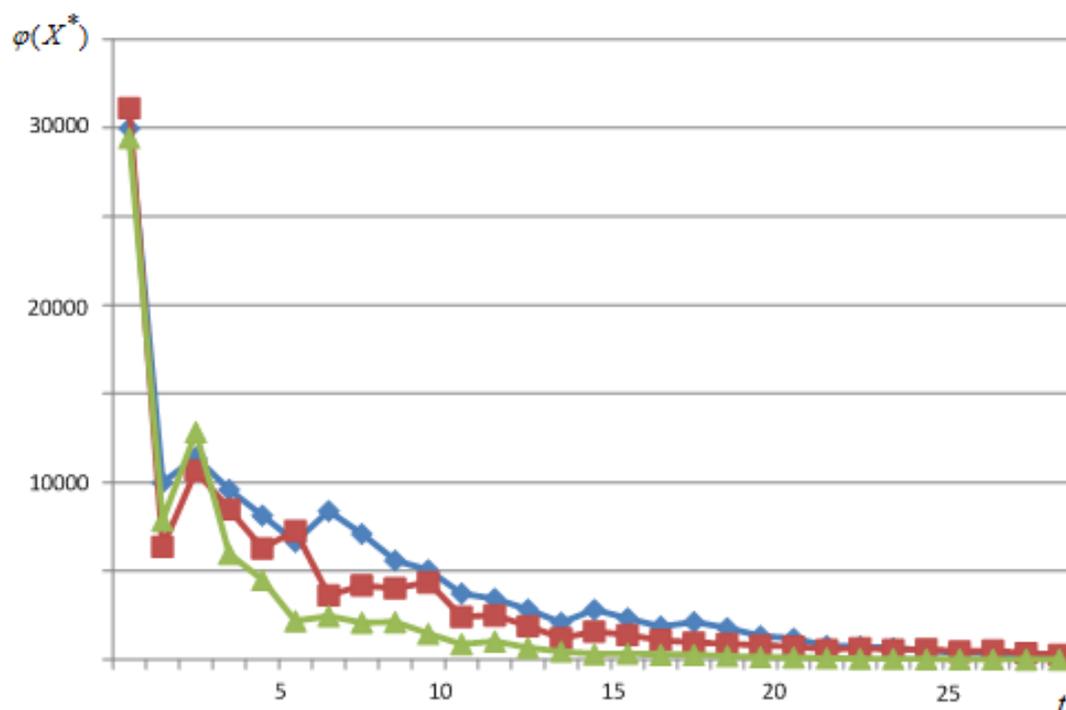
области (6) и $|X| = 16$. Минимум этой функции достигается в точке $(0, 0, \dots, 0)$ и равен нулю.

Результаты вычислительного эксперимента, представленные в таблице 2, показывают близость модификаций *PSO-M1*, *PSO-M2* по критериям $\bar{\varepsilon}$, ε_{\min} и значительное превосходство модификации *PSO-M2* над каноническим методом *PSO* и модификацией *PSO-M1* по критерию \bar{n}_φ .

Таблица 2 – Результаты исследования эффективности метода *PSO* и его модификаций: функция Растригина

Метод <i>PSO</i>	Критерий эффективности		
	\bar{n}_φ	$\bar{\varepsilon}$	ε_{\min}
Канонический метод	16371	27,9	8,95
Модификация <i>PSO-M1</i>	9634	20,8	8,95
Модификация <i>PSO-M2</i>	6688	24,6	8,95

Сходимость рассматриваемых методов иллюстрирует рисунок 2.



◇ - *PSO*; □ - *PSO-M1*; Δ - *PSO-M2*

Рисунок 2 – К сходимости методов *PSO*, *PSO-M1*, *PSO-M2*: функция Растригина; $|X| = 16$; $|S| = 50$

Решение задачи докинга. Рассматриваем задачу докинга для молекул протеазы *HIV1* и дарунавиром *3U7S* в качестве молекулы-рецептора и лиганда соответственно [10]. Учитываем взаимодействие указанных молекул с одной молекулой воды. Компоненты вектора варьируемых параметров X , полагаем, имеют следующий смысл:

- углы поворота лиганда вокруг осей Ox , Oy , Oz ;
- смещения центра лиганда по осям Ox , Oy , Oz ;
- валентные углы;
- торсионные углы.

Таким образом, размерность вектора варьируемых параметров задачи равна $|X| = 28$.

Используем, как более эффективный, метод *PSO-M2*, в котором рой состоит из 32 частиц ($|S| = 32$) и свободные параметрами имеют следующие значения: $b_C = 0,1; b_S = 0,3; b_1 = 0,1; b_2 = 0,1$.

Основные результаты вычислительного эксперимента представлены ниже:

- значение энергии лучшей частицы 111,81 ккал/моль;
- среднеквадратичное отклонение лиганда от нативного положения 1,86, Å;
- среднеквадратическое отклонение координат лиганда 0,000903;
- максимальное число итераций равно 472.

Заметим, что указанное значение среднеквадратического отклонения лиганда от его нативного положения следует считать хорошим результатом.

Полученный в результате вычислительного эксперимента молекулярный комплекс иллюстрирует рисунок 3.

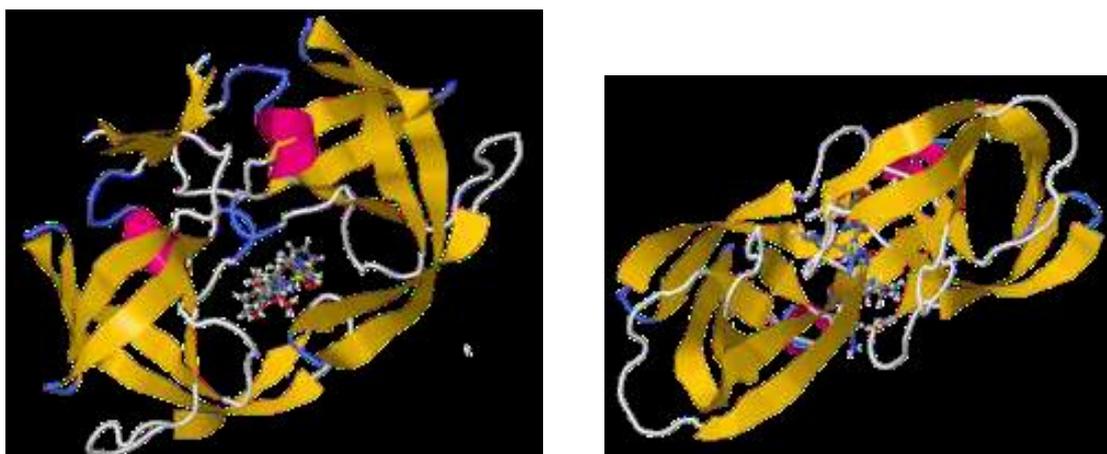


Рисунок 3 – Комплекс, полученный в результате вычислительного эксперимента: $|X| = 28; |S| = 32$; значение энергии лучшей частицы равно 111,81 ккал/моль

Результаты эксперимента показывают, что модификация метода роя частиц *PSO-M2* обеспечивает локализацию глобального минимума скоринг-функции с высокой точностью, как по критерию минимума полной потенциальной энергии полученного комплекса, так и по критерию минимума среднеквадратического отклонения лучшей частицы от нативного положения лиганда.

Заключение

В работе представлены две модификации канонического метода роя части, имеющие целью повышение эффективности этого метода при решении многоэкстремальных задач оптимизации высокой размерности.

Суть модификации *PSO-M1* заключается в формировании двух новых точек притяжения для частиц роя (наряду с точками, отвечающими за инерциальную, когнитивную и социальную компоненты канонического метода). Указанные новые точки представляют собой лучшие точки наборов частиц-соседей данной точки. Модификация имеет целью диверсифицировать поиск.

Если все свободные параметры метода *PSO-M1* (как и канонического метода) являются статическими, то один из таких параметров модификации *PSO-M2* является динамическим, так что данная модификация представляет собой пример самоадаптивного метода оптимизации [11]. Модификация имеет целью интенсифицировать поиск.

Выполнена программная реализация канонического методов роя части и двух его указанных модификаций. Вычислительный эксперимент по исследованию эффективности метода и модификаций при решении тестовых задач оптимизации показал преимущества предложенных модификаций по сравнению с каноническим методом, выявил превосходства модификации *PSO-M2* как над каноническим методом, так и над модификацией *PSO-M1*.

С помощью метода *PSO-M2* решена 28-мерная задача докинга для молекул протеазы *HIV1* и дарунавиром *3U7S* в качестве молекул-рецептора и лиганда соответственно. Результаты вычислительного эксперимента показали, что метод *PSO-M2* успешно находит положение лиганда, близкое к нативному и может быть рекомендован для решения задач молекулярного докинга как альтернатива генетическому алгоритму.

В развитие работы авторы планируют решение с помощью разработанного алгоритмического и программного обеспечения ряда других задач молекулярного докинга.

Список литературы

1. Ewing T.J., Makino S., Skillman A.G., Kuntz I.D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases // Journal of Computer-Aided Molecular Design. 2001. Vol. 15. P. 411-428.
2. Verdonk M.L., Cole J. C., Hartshorn M.J., Murray C. W., Taylor R.D. Improved protein-ligand docking using GOLD // Proteins: Structure, Function and Genetics. 2003. Vol. 52. P. 609-623.
3. Thomsen R., Christensen M.H. MolDock: A New Technique for High-Accuracy Molecular Docking // Journal of Medicinal Chemistry. 2006. Vol. 49, no. 11. P. 3315-3321.
4. Романов А.Н., Кондакова О.А., Григорьев Ф.В., Сулимов А.В., Луцкеина С.В., Мартынов Я.Б., Сулимов В.Б. Компьютерный дизайн лекарственных средств: программа докинга SOL // Вычислительные методы и программирование. 2008. Т. 9. С. 213-233.

5. Chen H.-M., Liu B.-F., Huang H.-L., Hwang S.-F., Ho S.-Y. SODOCK: Swarm optimization for highly flexible protein-ligand docking // *Journal of Computational Chemistry*. 2007. Vol. 28, no. 2. P. 612-623.
6. Namasivayam V., Günther R. pso@autodock: a fast flexible molecular docking program based on Swarm intelligence // *Chemical Biology & Drug Design*. 2007. Vol. 70, no. 6. P. 475-484. DOI: [10.1111/j.1747-0285.2007.00588.x](https://doi.org/10.1111/j.1747-0285.2007.00588.x)
7. Halgren T. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94 // *Journal of Computational Chemistry*. 1996. Vol. 17. P. 490-519.
8. Карпенко А.П., Селиверстов Е.Ю. Глобальная оптимизация методом роя частиц. Обзор // *Информационные технологии*. 2010. № 2. С. 25-34.
9. Tang K., Yao X., Suganthan P.N., MacNish C., Chen Y.P., Chen C.M., Yang Z. Benchmark Functions for the CEC'2008 Special Session and Competition on Large Scale Global Optimization: Technical Report // *Nature Inspired Computation and Applications Laboratory, USTC, China, 2007*. Available at: http://sci2s.ugr.es/eamhco/cec2010_functions.pdf, accessed 01.03.2014.
10. HIV-1 Протеаза // База знаний по биологии человека: сайт. Режим доступа: <http://humbio.ru/humbio/peptides/00060494.htm> (дата обращения 01.03.2014).
11. Eiben A.E., Michalewicz Z., Schoenauer M., Smith J.E. Parameter Control in Evolutionary Algorithms // *Parameter Setting in Evolutionary Algorithms*. Springer Berlin Heidelberg, 2007. P. 19-46. DOI: [10.1007/978-3-540-69432-8_2](https://doi.org/10.1007/978-3-540-69432-8_2)

Solving a molecular docking problem by the modified PSO method

04, April 2014

DOI: 10.7463/0414.0707258**A.P. Karpenko, K. O. Matveeva, V.A. Bulanov**

Bauman Moscow State Technical University, 105005, Moscow, Russian Federation

karpenko@rk6.bmstu.rukirandia1@gmail.comvb001@rambler.ru

The paper presents an canonical method of the swarm particles in two modifications to raise this method efficiency in solving multi-extreme problems of high dimension optimization. The essence of PSO-M1 modification is to form two new points to attract swarm particles (along with the points which are responsible for inertial, cognitive, and social components of canonical method). These new points represent the best points of sets of particles-neighbours of a given point. The modification aims to diversify search. All free parameters of the PSO-M1 method (as well as an canonical method) are static. In contrast, one of such parameters of PSO-M2 modification is dynamic. So this modification represents an example of a self-adaptive method of optimization. The modification aims to intensify search. A computing experiment to study the method efficiency and its abovementioned modifications at solving the test problems of optimization showed advantages of offered modifications in comparison with canonical method, revealed a superiority of PSO-M2 modification both over canonical method, and over PSO-M1 modification. Using the PSO-M2 method allows us to solve the 28-dimensional molecular docking problem of HIV1 protease and darunavir 3U7S as the molecules of receptor and a ligand, respectively. Results of computing experiment have shown that the PSO-M2 method successfully finds the position of ligand close to native and can be recommended for solving the molecular docking problems as an alternative to genetic algorithm.

Publications with keywords: [global optimization](#), [Method of a plenty of particles](#), [molecular docking](#), [proteases HIV1](#), [darunavir 3U7S](#)**Publications with words:** [global optimization](#), [Method of a plenty of particles](#), [molecular docking](#), [proteases HIV1](#), [darunavir 3U7S](#)

References

1. Ewing T.J., Makino S., Skillman A.G., Kuntz I.D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *Journal of Computer-Aided Molecular Design*, 2001, vol. 15, pp. 411-428.
2. Verdonk M.L., Cole J. C., Hartshorn M.J., Murray C. W., Taylor R.D. Improved protein-ligand docking using GOLD. *Proteins: Structure, Function and Genetics*, 2003, vol. 52, pp. 609-623.
3. Thomsen R., Christensen M.H. MolDock: A New Technique for High-Accuracy Molecular Docking. *Journal of Medicinal Chemistry*, 2006, vol. 49, no. 11, pp. 3315-3321.
4. Romanov A.N., Kondakova O.A., Grigor'ev F.V., Sulimov A.V., Lushchekina S.V., Martynov Ya.B., Sulimov V.B. [The SOL docking package for computer-aided drug design]. *Vychislitel'nye metody i programmirovaniye - Numerical methods and programming*, 2008, vol. 9, pp. 213-233. (in Russian).
5. Chen H.-M., Liu B.-F., Huang H.-L., Hwang S.-F., Ho S.-Y. SODOCK: Swarm optimization for highly flexible protein-ligand docking. *Journal of Computational Chemistry*, 2007, vol. 28, no. 2, pp. 612-623.
6. Namasivayam V., Günther R. pso@autodock: a fast flexible molecular docking program based on Swarm intelligence. *Chemical Biology & Drug Design*, 2007, vol. 70, no. 6, pp. 475-484. DOI: [10.1111/j.1747-0285.2007.00588.x](https://doi.org/10.1111/j.1747-0285.2007.00588.x)
7. Halgren T. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *Journal of Computational Chemistry*, 1996, vol. 17, pp. 490-519.
8. Karpenko A.P., Seliverstov E.Yu. [Global Optimization by the Particle Swarm Optimization Method. Review]. *Informatsionnye tekhnologii*, 2010, no. 2, pp. 25-34. (in Russian).
9. Tang K., Yao X., Suganthan P.N., MacNish C., Chen Y.P., Chen C.M., Yang Z. Benchmark Functions for the CEC'2008 Special Session and Competition on Large Scale Global Optimization: Technical Report. *Nature Inspired Computation and Applications Laboratory, USTC, China*, 2007. Available at: http://sci2s.ugr.es/eamhco/cec2010_functions.pdf , accessed 01.03.2014.
10. *HIV-1 Proteaza* [HIV-1 Protease]. Baza znaniy po biologii cheloveka [Knowledge Base on Human Biology]: website. Available at: <http://humbio.ru/humbio/peptides/00060494.htm> , accessed 01.03.2014. (in Russian).
11. Eiben A.E., Michalewicz Z., Schoenauer M., Smith J.E. Parameter Control in Evolutionary Algorithms. In: *Parameter Setting in Evolutionary Algorithms*. Springer Berlin Heidelberg, 2007, pp. 19-46. DOI: [10.1007/978-3-540-69432-8_2](https://doi.org/10.1007/978-3-540-69432-8_2)