

# Применение параллельных эволюционных алгоритмов оптимизации в задачах структурной биоинформатики

С.В. Полуян

ГБОУ ВО МО «Университет «Дубна»,  
институт системного анализа и управления

Настоящая работа посвящена исследованию области применения стохастических эволюционных алгоритмов оптимизации в задачах структурной биоинформатики: предсказание трехмерной структуры пептида по аминокислотной последовательности и предсказание структуры комплекса пептид-белок. Продемонстрированы основные положения, сводящие рассматриваемые задачи к задаче непрерывной глобальной оптимизации, и предложен способ использования эволюционных алгоритмов. Описана схема применения параллельных вычислений и проведена параллельная реализация алгоритмов. Представлены результаты разнообразных численных экспериментов. Показаны перспективы использования эволюционных алгоритмов в актуальных задачах структурной биоинформатики.

*Ключевые слова:* глобальная оптимизация, эволюционные вычисления, роевая оптимизация, структурная биоинформатика.

## 1. Введение

Белки являются макромолекулами состоящими из  $\alpha$ -аминокислот, соединенных в цепочку пептидной связью, тем самым образуя полипептидную цепь. Предсказание структуры белка – предсказание по аминокислотной последовательности трехмерной структуры белка, которая определяет нативное, т.е. функционально активное, состояние (выделяют вторичную, третичную и четвертичную). Короткие белки называют пептидами.

Пептиды играют ключевую роль во многих внутриклеточных процессах. Причем, как индивидуально, так и при взаимодействии с белками. Например, при иммунном ответе или в различных сигнальных системах клеток. В настоящее время изучение этих взаимодействий различными вычислительными методами приобретает все большую популярность в силу меньшей дороговизны относительно традиционных экспериментальных методов [1].

Наиболее широко принимаемая гипотеза, объясняющая процесс самоорганизации белковых молекул была сформулирована Анфинсеном [2]. Основные идеи предложенной им «термодинамической гипотезы» следующие: нативное состояние белка уникально; нативное состояние белка находится в глобальном минимуме свободной энергии. Таким образом, процесс сворачивания полипептидной цепи можно представить как процесс минимизации свободной энергии белка, при этом задача предсказания структуры сводится к задаче поиска глобального минимума свободной энергии, т.е. к задаче глобальной оптимизации.

Если поставить задачу классификации методов предсказания структуры белка, то можно выделить два основных подхода. Первый состоит в использовании информации из известных белковых структур. Такие методы предсказания называют моделированием по гомологии. Второй подход называют *ab initio*, где процесс сворачивания цепи рассматривается без привлечения каких-либо дополнительных эмпирических предположений. При таком подходе для предсказания структуры используются различные вычислительные методы, например, метод молекулярной динамики или методы стохастической оптимизации. Необходимо отметить, что в случае пептид-белок докинга применение методов молекулярной динамики является чрезвычайно ресурсоемкой задачей и может занимать несколько месяцев [3]. При этом применение методов оптимизации привлекательно не только в вы-

числительном отношении, но также позволяет гибко использовать в процессе поиска оптимальной структуры как целевую функцию, описывающее силовое поле, так и различные статистически известные данные и различные упрощения.

## 2. Постановка задачи

В настоящей работе рассматриваются две задачи структурной биоинформатики: задача поиска двух основных регулярных вторичных структур встречающихся у пептидов ( $\alpha$ -спирали и  $\beta$ -листа) и задача предсказания структуры комплекса пептид-белок.

В задаче предсказания структуры комплекса пептид-белок (или пептид-белок докинг) необходимо найти оптимальное место связывания белка и пептида при взаимодействии друг с другом, а также соответствующую этой связи конформацию комплекса. При этом представляется структура белка определяющая его нативное состояние и аминокислотная последовательность пептида. Оптимальное состояние комплекса пептид-белок находится в глобальном минимуме свободной энергии, т.е. данная задача также может быть сведена к задаче глобальной оптимизации.

С точки зрения наличия информации об месте связывания пептида и белка можно выделить два типа докинга: слепой и прямой докинг. В настоящей работе рассматривается докинг в окрестности места связывания. Важно отметить, что в при взаимодействии с пептидом в структуре белка и пептида могут происходить сильные взаимосвязанные неспецифичные для свободного состояния конформационные изменения, как правило, слабо поддающиеся статистическому анализу и моделированию по гомологии.

С момента появления термодинамической гипотезы исследователями были предложены различные способы применения методов оптимизации, в которых, на различных этапах, задача поиска оптимальной структуры представима как задача нелинейной непрерывной глобальной оптимизации. При этом используют различные стохастические алгоритмы оптимизации, в том числе эволюционные, а точность предсказания напрямую зависит от используемого алгоритма оптимизации. Актуальность применения и совершенствования современных методов оптимизации показана в различных работах [4]. Настоящая работа посвящена исследованию области применения стохастических эволюционных алгоритмов оптимизации в поставленных задачах.

Важно отметить, что основная цель исследования – сравнение различных эволюционных алгоритмов и выявление наиболее эффективных стратегий в рамках силового поля и рассматриваемых задач. В общем случае такого рода задачи решаются комбинированными методами, где может учитываться различная информация о представленных структурах, однако такого рода предсказания выходят за рамки текущего исследования.

### 2.1. Предсказание структуры

Задача поиска структуры пептида ставится в непрерывном пространстве степеней свободы пептида, которые описаны ниже.

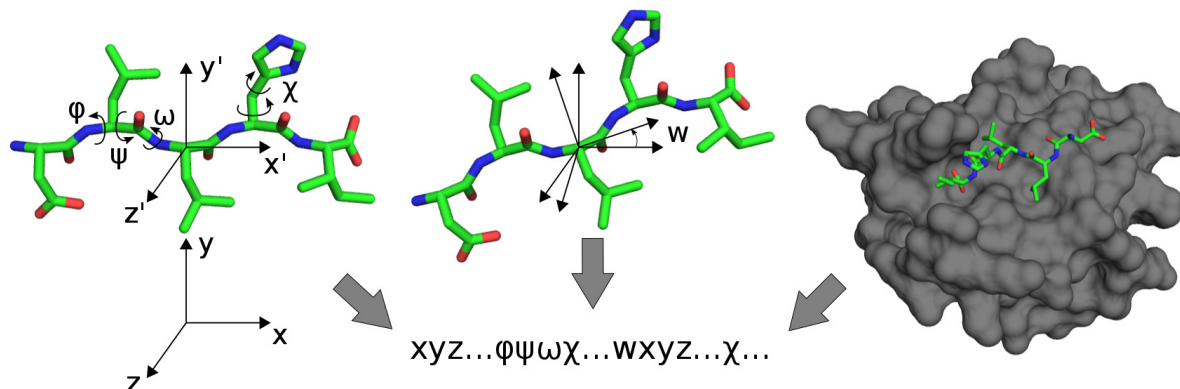
- Торсионные углы главной цепи пептида. Углы  $\phi$ ,  $\psi$  и  $\omega$  для каждого аминокислотного остатка пептида, за исключением  $\phi$  первого и  $\psi, \omega$  последнего остатка. Вводится ограничение  $[-\pi, \pi]$  на диапазон значений для углов  $\phi$  и  $\psi$ . Для угла  $\omega$  вводится ограничение  $[\pi - \delta, \pi + \delta]$ , где  $\delta = 0.2$  рад., поскольку в рассматриваемых задачах данный угол находится в *trans*-конформации и стремится быть планарным.
- Торсионные углы для каждой боковой цепи пептида. Углы  $\chi_{1-4}$  с ограничением  $[-\pi, \pi]$ , за исключением пролина, для которого вводятся аналогичные углу  $\omega$  отклонения от идеализированных значений используемых в силовом поле.

В задаче поиска структуры комплекса пептид-белок к параметрам пептида, описанным выше, добавляются следующие параметры с непрерывными значениями.

- Вектор для смещения пептида относительно белка. При этом вводятся ограничения в

несколько ангстрем относительно начального положения белка.

- Вектор и угол поворота пептида в трехмерном пространстве относительно начального положения. Ограничения  $[-1, 1]^3$  и  $[0, \pi]$  соответственно.
- Торсионные углы боковых цепей белка, которые находятся в непосредственной близости к пептиду.



**Рис. 1.** Описывающие смещение и поворот пептида в трехмерном пространстве, а также степени свободы пептида и белка параметры на примере пептида DLLHI и белка 1JWG (код PDB).

## 2.2. Силовое поле

В численных экспериментах использовалось силовое поле Rosetta 3.8 [5]. Отличительной особенностью данного силового поля является использование, при вычислении энергии макромолекулы, неявного растворителя, различных потенциалов и статистически полученных данных.

Целевая функция (именуемая также скоринг-функцией) представляет собой сумму так называемых термов, которые входят в состав суммы с определенным весом. Веса термов калибруются на определенной выборке белков. Термы описывают межатомные взаимодействия с использованием классической молекулярной механики (силы отталкивания и притяжения Леннарда-Джонса, электростатические взаимодействия), и эмпирически известных данных (планарность торсионного угла  $\omega$  главной цепи и водорода в гидроксильной группе). Водородные связи разбиты на четыре группы: взаимодействия между атомами основной цепи в зависимости от положения в первичной структуре (близкие и дальние); взаимодействия между атомами главной цепи и боковыми цепями; взаимодействия между боковыми цепями. В рассматриваемых скоринг-функциях использовалось приблизительно 15 термов. В связи с тем, что при вычислении целевой функции используются эмпирические термы и все веса термов откалиброваны, невозможно говорить о получаемой энергии пептида как о потенциальной энергии выражаемой в килокалориях на моль. Вместо этого рассматривается просто получаемое значение скоринг-функции.

В качестве целевых функций использованы две скоринг-функции – *score12* и *talaris2014*, соответствующие предыдущему и текущему стандарту скоринг-функции у силового поля Rosetta. Принципиальное различие *score12* и *talaris2014* заключается в способе вычисления электростатических взаимодействий. В первом случае используется терм, описывающий статистически полученные данные из PDB [5], во втором случае в явном виде вычисляется кулоновский потенциал.

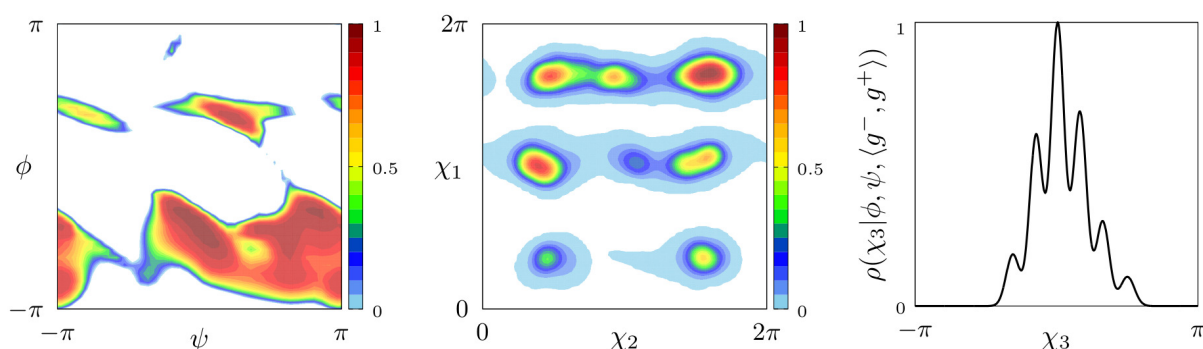
Выбор силового поля обусловлен широкой распространенностью, быстродействием и ориентированностью к проблеме предсказания пространственной структуры белков. Также из преимуществ стоит отметить удобный фреймворк для разработки с возможностью использования Python или C++ современного стандарта.

### 2.3. Использование статистической информации

В силу большого числа экспериментально полученных структур белков достаточно высокого разрешения известны наиболее часто встречающиеся значения торсионных углов у аминокислотных остатков. Существует ряд библиотек [5–7] предоставляющих эти эмпирические данные в различной форме. Например, с помощью термина *rama2b* средствами Rosetta можно получить плотность распределения наиболее часто встречающихся углов  $\phi, \psi$  главной цепи и, тем самым, исключить из рассмотрения физически невозможные конформации. На рис. 2. показана нормализованная плотность распределения углов главной цепи аспарагина в окружении тирозина и пролина.

Торсионные углы боковых цепей центрируются за счет  $sp^3-sp^3$  гибридизации атомных орбиталей и образуют три приблизительно симметричных пика. Например, на рис. 2 показана плотность распределения углов гистидина, где  $\chi_1$  находится в состоянии  $gauche^+ \{g^+\}$ ,  $trans \{t\}$  и  $gauche^- \{g^-\}$ , в позициях  $60^\circ$ ,  $180^\circ$  и  $300^\circ$  соответственно. Такие углы называют ротамерными (rotameric). Неротамерные (non-rotameric) углы боковых цепей центрируются за счет  $sp^3-sp^2$  гибридизации и демонстрируют более широкие и ассиметричные плотности, например, угол  $\chi_2$  гистидина. На рис. 2 показано независимое от углов главной цепи распределение углов боковой цепи гистидина.

Гораздо больший интерес представляют библиотеки [6], в которых можно получить вероятность нахождения угла боковой цепи в зависимости от значений углов главной цепи и от позиции предшествующих торсионных углов аминокислотного остатка. Например, на рис. 2 показаны возможные состояния для неротамерного угла  $\chi_3$  глутамина, в зависимости от углов главной цепи  $\phi = -40^\circ$ ,  $\psi = 10^\circ$  и положения предшествующих ротамеров: первый в состоянии  $\{g^-\}$ , второй в состоянии  $\{g^+\}$ .



**Рис. 2.** Нормализованная плотность распределения углов главной цепи аспарагина. Распределение углов боковой цепи гистидина. Предпочтительные значения угла  $\chi_3$  глутамина.

С помощью представленной выше информации о торсионных углах можно значительно сократить пространство поиска, чтобы в процессе предсказания структуры стохастическим алгоритмом оптимизации не рассматривать невозможные конформации. Для этого достаточно построить соответствующие функции распределения и, перед вычислением целевой функции, сделать прерасчет значений. Такой подход был реализован в рамках исследования. Для случая углов главной цепи с использованием термина *rama2b* выполняется построение двумерных функций распределения с зависимостью от соседних остатков в главной цепи. Для случая углов боковой цепи: одномерные функции распределения, так как библиотекой Dunbrack [6] предоставляются значения ротамеров в виде среднего значения и отклонения. Для случая библиотеки Top500 [7] размерность функции распределения составляет от одного до четырех.

Поскольку предоставляемые библиотеками значения дискретны, для получения однозначного отображения и непрерывности пространства поиска использовалась линейная интерполяция. Так как значения в функции распределения отсортированы по возрастанию,

для получения нового значения достаточно использовать бинарный поиск.

Сложность распределения одного параметра в среднем и худшем случае будет составлять  $O(\log_2 n)$  операций, где  $n$  – число шагов сетки, что несущественно, так как максимальное  $n = 720$ . Увеличение числа шагов не будет играть никакой роли из-за дискретности предоставляемых библиотеками значений. Недостатком такого метода является значительное потребление оперативной памяти. Однако, для рассматриваемых задач, оно несущественно.

Важно отметить, что реализованные функции распределения позволяют получать обратные значения, что позволяет на любом этапе работы алгоритмов добавлять измененные вне процесса работы решения в популяцию.

### 3. Эволюционные алгоритмы оптимизации

Для поиска оптимальной структуры использовались следующие эволюционные алгоритмы оптимизации: адаптивная дифференциальная эволюция JADE [8], эволюционная стратегия ESCH [9], метод роя частиц PSO [10] с локальным поиском SW [11], алгоритм бактериального поиска с адаптивным изменением шага SABFO [12], алгоритм роевой оптимизации со стратегией соревнования особей CSO [13], неоднородный клеточный генетический алгоритм NCGA [14], эволюционная стратегия с адаптацией матрицы ковариаций CMAES [15], гибрид дифференциальной эволюции с CMAES для локальной оптимизации JDE-CMAES [16]. Выбор рассматриваемых алгоритмов обусловлен хорошими результатами при решении различных практических и тестовых задач оптимизации [13].

Необходимо отметить, что, несмотря на разнообразие стратегий поиска, приведенные эвристики имеют схожую структуру, напоминающую структуру классического генетического алгоритма, в котором популяция решений несколько раз проходит три этапа – мутацию, скрещивание и отбор. Практически во всех рассматриваемых алгоритмах в явном виде присутствует оператор отбора решений, а в модификациях дифференциальной эволюции присутствует оператор скрещивания. При этом шаги, определяющие движение членов популяции в роевых алгоритмах, выполняют роль мутации и осуществляют локальный поиск. Во всех приведенных алгоритмах операторы мутации разные: в случае с JADE используется стратегия *current-to-best*; в JDE – классическая для дифференциальной эволюции стратегия *rand*; в NCGA – классическая стратегия для генетического алгоритма. Отдельно следует отметить алгоритм CMAES, который показывает в начале оптимизации самую высокую скорость сходимости среди всех алгоритмов, однако даёт один из худших результатов, показывая, тем самым, только локальную эффективность.

Следует подчеркнуть, что сходимость у эволюционных алгоритмов в значительной степени зависит от используемых параметров, причем их число варьируется от двух (CSO) до 12 (SABFO). В проводимых экспериментах часть параметров подбирались с учетом размерности, границ рассматриваемой задачи, рекомендаций авторов и собственных исследований. Для максимальной объективности сравнения эксперименты проводились на одинаковом размере популяции решений. Чувствительность алгоритмов к размеру популяции, в данном случае, нивелируется достаточно большим числом итераций.

### 4. Параллельная реализация

Поскольку у эволюционных алгоритмов операция вычисления целевой функции для каждого члена популяции обычно выносится из основных операторов в общий для популяции шаг вычисления целевой функции, практически все из них довольно просто распараллеливаются. В результате выполненной работы произведена параллельная реализация данного этапа у большинства рассматриваемых алгоритмов.

Используемая схема распараллеливания следующая. Популяция решений сначала раз-

бывается на части пропорционально количеству используемых узлов с помощью технологии MPI по принципу: один узел – один MPI процесс. Внутри каждого процесса, с применением технологии OpenMP, вычисляется значение целевой функции для полученной из главного процесса части популяции.

Все вычисления выполнены на кластере ОИЯИ HybriLIT [17], где каждый вычислительный узел имеет два 12-ядерных процессора Intel Xeon. Результаты приведены в табл. 1 для алгоритма JADE. Важно отметить, что некоторые операторы алгоритмов также поддаются параллелизации, в связи с этим можно добиться более заметных результатов.

**Таблица 1.** Производительность применяемой схемы параллельных вычислений.

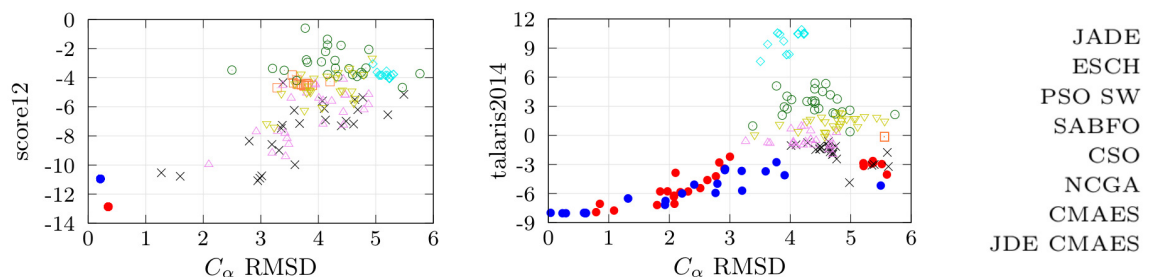
| Количество потоков / узлов | 4 / 1 | 8 / 1 | 12 / 1 | 16 / 1 | 24 / 1 | 24 / 2 | 48 / 2 |
|----------------------------|-------|-------|--------|--------|--------|--------|--------|
| Ускорение                  | 3.31  | 6.28  | 9.03   | 11.84  | 17.1   | 15.85  | 31.1   |
| Эффективность              | 0.82  | 0.78  | 0.75   | 0.74   | 0.71   | 0.66   | 0.64   |

## 5. Результаты численных экспериментов

Как указывалось выше, основным мотивом исследования является проверка работы различных эволюционных алгоритмов в рамках довольно сложного энергетического ландшафта предоставляемого используемым силовым полем. Обычно алгоритмы оптимизации сравниваются на широком ряде тестовых функций [12], которые носят отчасти синтетический характер в силу присутствия однотипных базовых функций. Даже при сравнении с использованием сложных по структуре функций или на ряде практических задач различные алгоритмы дают очень неоднозначные результаты [13]. Таким образом становится актуальным выявление эффективных эволюционных операторов в рамках решения поставленных задач.

### 5.1. Предсказание пространственной структуры пептидов

На первом этапе вычислительных экспериментов ставилась задача нахождения оптимальной структуры модельного декапептида A<sub>10</sub> [18], с искомой структурой –  $\alpha$ -спираль. Размерность задачи составила 27 параметров. Количество вызовов целевой функции ограничено одним миллионом. Для каждого алгоритма выполнено 25 независимых запусков. Результаты показаны на рис. 3.



**Рис. 3.** Результаты 25 независимых запусков.

Оптимальная структура для рассматриваемого пептида получена с помощью сервера PEP-FOLD [19], предсказывающим структуру пептида с использованием статистической информации (низкоэнергетических фрагментов) для получения пула крупнозернистых моделей и последующей полноатомной минимизацией.

На рис. 3 показано среднеквадратичное отклонение координат атомов  $\alpha$ -углерода главной цепи получаемое после оптимизации пептидов относительно найденной с помощью PEP-FOLD структуры. Так как вторичную структуру определяет конформационное расположение атомов главной цепи, такой способ сравнения наиболее объективен. Силовое поле, используемое в методе PEP-FOLD, отличается от Rosetta, поэтому перед сравнением здесь и далее получаемая PEP-FOLD структура проходит процедуру локальной оптимизации в Rosetta стандартными средствами пакета. Следует отметить, что первичная цепочка для оптимизации алгоритмами порождалась средствами Rosetta с идеализированными значениями валентных углов и длин ковалентных связей. В рассматриваемом случае при оптимизации эти значения не изменялись и конечные структуры несколько отличаются от получаемой с помощью PEP-FOLD.

Можно заметить, что меньшее значение целевой функции соответствуют большему отклонению атомов главной цепи. Здесь необходимо учесть локальный характер оптимизации для структуры полученной с помощью PEP-FOLD.

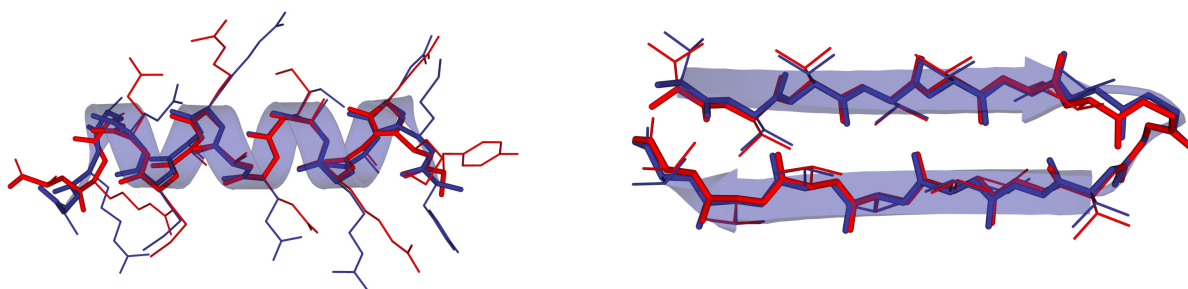


Рис. 4. Найденные модельные  $\alpha$ -спираль Y(MEARA)<sub>3</sub> и  $\beta$ -лист V<sub>7</sub>G<sub>2</sub>V<sub>7</sub>.

На рис. 4 представлены получаемые с помощью алгоритма JADE структуры для модельных пептидов [18] длиной 16 аминокислотных остатков. Использовалось силовое поле *talaris2014*. Размерность задач составила 78 и 59 параметров, количество вызовов целевой функции  $10^7$  и  $5 \cdot 10^6$  для  $\alpha$ -спирали и  $\beta$ -листа соответственно. Время запуска на одном вычислительном узле составило  $\sim 40$  мин. для  $\alpha$ -спирали. Среднеквадратичное отклонение атомов главной цепи относительно структуры, найденной с помощью PEP-FOLD, для которой показана вторичная структура и выделен синий цвет, меньше одного ангстрема. Поиск осуществлялся с использованием библиотеки Dunbrack и с сокращением пространства поиска для углов главной цепи.

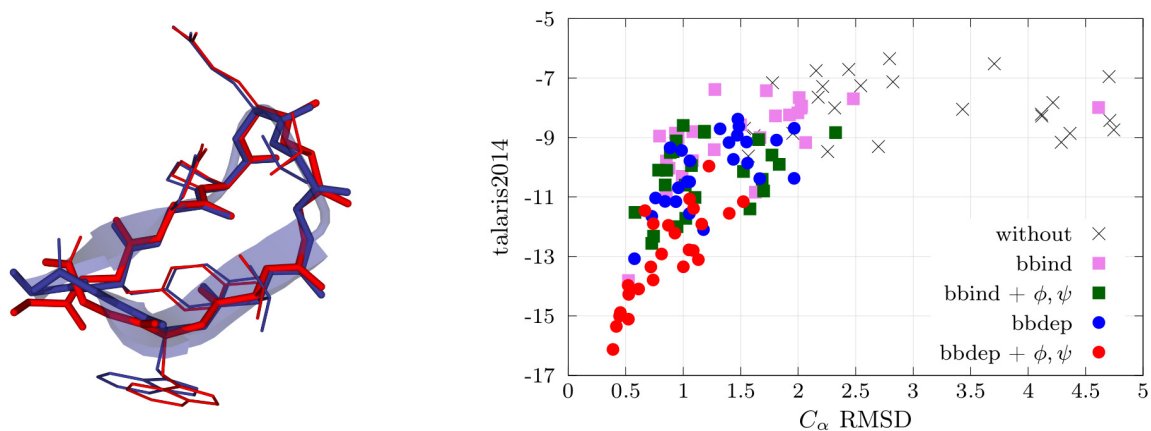


Рис. 5. Результаты для пептида 1UAO.

На рис. 5 представлены результаты для пептида 1UAO (код PDB) длиной десять аминокислотных остатков.

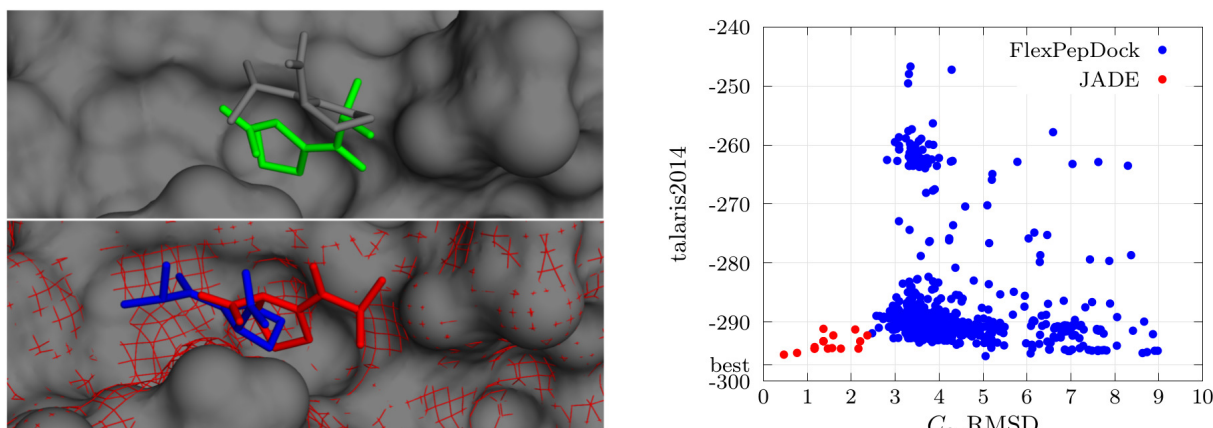


кислотных остатков. Параметры нативной структуры оптимизированы средствами Rosetta. Показаны результаты без сокращения пространства поиска (without), с распределением для углов главной цепи ( $\phi, \psi$ ), с использованием библиотеки Top500 (bbind) и Dunbrack (bbdep), т.е. данные независимые и зависимые от углов главной цепи соответственно. Представлены 25 независимых запусков алгоритма JADE для каждого случая. Количество вызовов целевой функции равно  $2 \cdot 10^6$  для каждого запуска.

## 5.2. Предсказание структуры комплекса пептид-белок

На втором этапе вычислительных экспериментов ставилась задача пептид-белок докинга. Необходимо отметить, что докинг даже коротких пептидов, длиной 2-5 аминокислотных остатка, представляет собой сложную задачу даже для специально разработанных пакетов [1, 4]. Сравнение эволюционного алгоритма JADE производилось с протоколом Rosetta FlexPepDock [20], выполняющим докинг в окрестности места связывания алгоритмом Монте-Карло с локальной оптимизацией. Выбор алгоритма JADE обусловлен результатами, приведенными на рис. 3.

Рассматривался докинг пептида AP в белок 2СУН [1] (код PDB) со стартовой позицией пептида в 3-х ангстремах от места связывания в случайной конформации. Поиск производился в 5-ти ангстремах от стартовой позиции с участием боковых цепей белка находящихся в радиусе 10 ангстрем от стартовой позиции. Размерность задачи составила 25 параметров. Время  $10^3$  запусков FlexPepDock соответствуют 20-ти запускам JADE с количеством вызовов целевой функции  $5 \cdot 10^6$  для каждого запуска. Преимущество использования JADE заключается в получении меньших значений скоринг-функции, что показано на рис. 6.



**Рис. 6.** Стартовая позиция (серый цвет) и нативное положение (зеленый цвет) пептида. Результат JADE (красная сетка и пептид) и FlexPepDock (синий пептид и серая поверхность).

## 6. Заключение

В результате выполненной работы проведено исследование применимости эволюционных алгоритмов оптимизации к поставленным задачам. Для некоторых рассматриваемых алгоритмов произведена параллельная реализация.

На основании проведенных исследований можно заключить, что часть рассматриваемых алгоритмов способна находить оптимальную структуру коротких пептидов длиной порядка 10 аминокислотных остатков, а при использовании сокращения пространства поиска длина может быть увеличена до 15 остатков. Также эволюционные алгоритмы оптимизации применимы к задаче пептид-белок докинга при сопоставимом количестве остатков, с пределом в 30-50 параметров. Важно отметить, что в актуальных задачах пептид-белок докинга необходимо рассматривать пептиды длиной 10-15 аминокислотных остатков, а с



учетом степеней свободы белка, размерность задачи возрастает как минимум в 2-3 раза, что приводит к невозможности их использования в полноатомном разрешении.

На основании представленных результатов и перечисленных выше аргументов можно сделать вывод, что при решении поставленных задач принципиальными оператором является оператор мутации дифференциальной эволюции, причём со стратегией *current-to-best*. Необходимо отметить, что в рассматриваемом алгоритме вероятность мутации и скрещивания являются адаптивными параметрами и, как показывают исследования, вероятность скрещивания практически все время работы алгоритма крайне мала в отличие от вероятности мутации. Однако, как показывают исследования [8, 14], именно оператор скрещивания осуществляет глобальный поиск и его применение напрямую влияет на эффективность работы алгоритма. Описанные выше результаты указывают на предположения [21], что эволюционные алгоритмы используют в своей работе неявную вероятностную модель распределения, образующуюся в процессе оптимизации.

Целью дальнейшей работы является проверка идей алгоритмов оценки распределений, позволяющих строить вероятностные модели с использованием, например, байесовских сетей. При таком подходе можно добиться эффективного использования статистических известных данных и привлечь к поиску локальную оптимизацию.

Автор выражает благодарность:

- научному руководителю Н.М. Ершову, к.ф.-м.н., с.н.с. ВМК МГУ им. Ломоносова, за полезные обсуждения и замечания;
- А.В. Головину, д.х.м., с.н.с. НИИ ФХБ им. А.Н. Белозерского, преподавателю ФББ МГУ им. Ломоносова за ценные консультации, указание актуальных задач в области докинга и подбор силового поля;
- коллективу гетерогенного кластера HybriLIT ОИЯИ, за поддержку и предоставление вычислительных ресурсов.

## Литература

1. Rentzsch R., Renard B.Y. Docking small peptides remains a great challenge: an assessment using AutoDock Vina // Briefings in Bioinformatics. – Vol. 16, No. 6, – 2015. – Pp. 1045-1056. DOI: 10.1093/bib/bbv008.
2. Anfinsen C. Principles that Govern the Folding of Protein Chains // Science. – Vol. 181(4096). – Jul. 1973. – Pp. 330-331. DOI: 10.1126/science.181.4096.223.
3. Huangz Z., Wong C.F. Docking Flexible Peptide to Flexible Protein by Molecular Dynamics Using Two Implicit-Solvent Models: An Evaluation in Protein Kinase and Phosphatase Systems // The Journal of Physical Chemistry B, No. 113(43), Pp. 14343–14354. DOI: 10.1021/jp907375b.
4. Lopez-Camacho E., Garcia Godoy M.J., et al. Solving molecular flexible docking problems with metaheuristics: A comparative study // Applied Soft Computing – Mar. 2015. DOI: 10.1016/j.asoc.2014.10.049
5. Alford R.F., Leaver-Fay A., Jeliaskov R., et al. The Rosetta all-atom energy function for macromolecular modeling and design. Feb. 2017. DOI: 10.1101/106054.
6. Shapovalov M., Dunbrack R.L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions // Structure, 19, – Jun. 2011 – Pp. 844-858. DOI: 10.1016/j.str.2011.03.019.
7. Hintze B.J., Lewis S.M., Richardson J.S., Richardson D.C. MolProbity's ultimate rotamer-library distributions for model validation // Proteins: Structure, Function, and Bioinformatics – Vol. 84(9). – June 2016 – Pp. 1177-1189. DOI: 10.1002/prot.25039.

8. Zhang J., Sanderson A. JADE: Adaptive differential evolution with optional external archive // IEEE Transactions on Evolutionary Computation. – Vol. 13(5). – 2009. – Pp. 945-958. DOI: 10.1109/TEVC.2009.2014613.
9. Silva-Santos C. H., Goncalves M. S., Hernandez-Figueroa H. E. Designing Novel Photonic Devices by Bio-Inspired Computing // IEEE Photonics Technology Letters. – Vol. 22(10). – 2010. – Pp. 1177-1179. DOI: 10.1109/IMOC.2009.5427618.
10. Kennedy J., Eberhart R. Particle swarm optimization // Proceedings of IEEE International Conference on Neural Networks. – Vol. 4. – 1995. – Pp. 1942-1948. DOI: 10.1109/ICNN.1995.488968.
11. Solis F.J., Wets R.J-B. Minimization by random search techniques // Mathematics of Operation Research. – Vol. 6(1) – 1981. – Pp. 19-30. DOI: 10.1287/moor.6.1.19.
12. Полуян С.В., Рейнгард Н.М., Ершов Н.М. Самоадаптация в алгоритмах роевой оптимизации // Вестник Российского университета дружбы народов: Серия Математика, информатика, физика. – № 2. – 2014. – с. 415-418.
13. Cheng R., Jin Y. A Competitive Swarm Optimizer for Large Scale Optimization // IEEE Transactions on Cybernetics – Vol. 45(2). – 2015. – Pp. 191-204. DOI: 10.1109/TCYB.2014.2322602.
14. Ершов Н.М. Неоднородные клеточные генетические алгоритмы // Компьютерные исследования и моделирование. – 2015. – Т. 7, № 3. – с. 775-780.
15. Hansen N., Ostermeier A. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation // Proceedings of the 1996 IEEE International Conference on Evolutionary Computation. – 1996. – Pp. 312-317. DOI: 10.1109/ICEC.1996.542381.
16. Brest J., Zamuda A., et al. Large scale global optimization using self-adaptive differential evolution algorithm // IEEE World Congress on Computational Intelligence. – 2010. – Pp. 1-8. DOI: 10.1109/CEC.2010.5585927.
17. Heterogeneous Computing Cluster HybriLIT. URL: <http://hybrilit.jinr.ru/en/> (дата обращения: 04.04.2017).
18. Galzitskaya O.V., Higo J., Finkelstein A.V.  $\alpha$ -Helix and  $\beta$ -Hairpin Folding from Experiment, Analytical Theory and Molecular Dynamics Simulations // Current Protein and Peptide Science. – Vol. 2(3). – Apr. 2002. – Pp. 191-200. DOI: 10.2174/1389203024605340.
19. Lamiable A., Pierre T., et al. PEP-FOLD3: faster denovo structure prediction for linear peptides in solution and in complex // Nucleic Acids Research, 2016, Vol. 44, Apr. 2016. DOI: 10.1093/nar/gkw329.
20. Raveh B., London N., et al. Rosetta FlexPepDock ab-initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors // PLoS ONE – Vol. 6(4) – Apr. 2011. DOI: 10.1371/journal.pone.0018934
21. Moradabadi B., Beigy H. A new real-coded Bayesian optimization algorithm based on a team of learning automata for continuous optimization // Genetic Programming and Evolvable Machines – June 2014. – Vol. 15(2). – Pp. 169-193. DOI: 10.1007/s10710-013-9206-9

# Parallel evolutionary algorithms for solving optimization problems in structural bioinformatics

S.V. Poluyan

Dubna State University,  
Institute of System Analysis and Control

The paper presents an exploration of the scope of application of the evolutionary optimization algorithms in the problems in structural bioinformatics: prediction of three-dimensional peptide structure from amino acid sequence and peptide-protein docking. This paper presents the main assumptions that reduce the above-described tasks to the continuous global optimization problem and provides a way of using evolutionary optimization algorithms. The used scheme of parallel computations was described. The parallel realization of the considered algorithms was carried out. The paper presents the results of various numerical experiments and emphasizes the most effective. The prospect of using evolutionary algorithms in contemporary issues of structural bioinformatics was shown.

*Keywords:* global optimization, evolutionary computation, swarm optimization, structural bioinformatics.

## References

1. Rentzsch R., Renard B.Y. Docking small peptides remains a great challenge: an assessment using AutoDock Vina // Briefings in Bioinformatics. – Vol. 16, No. 6, – 2015. – Pp. 1045-1056. DOI: 10.1093/bib/bbv008.
2. Anfinsen C. Principles that Govern the Folding of Protein Chains // Science. – Vol. 181(4096). – Jul. 1973. – Pp. 330-331. DOI: 10.1126/science.181.4096.223.
3. Huangz Z., Wong C.F. Docking Flexible Peptide to Flexible Protein by Molecular Dynamics Using Two Implicit-Solvent Models: An Evaluation in Protein Kinase and Phosphatase Systems // The Journal of Physical Chemistry B, No. 113(43), Pp. 14343–14354. DOI: 10.1021/jp907375b.
4. Lopez-Camacho E., Garcia Godoy M.J., et al. Solving molecular flexible docking problems with metaheuristics: A comparative study // Applied Soft Computing – Mar. 2015. DOI: 10.1016/j.asoc.2014.10.049
5. Alford R.F., Leaver-Fay A., Jeliazkov R., et al. The Rosetta all-atom energy function for macromolecular modeling and design. Feb. 2017. DOI: 10.1101/106054.
6. Shapovalov M., Dunbrack R.L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions // Structure, 19, – Jun. 2011 – Pp. 844-858. DOI: 10.1016/j.str.2011.03.019.
7. Hintze B.J., Lewis S.M., Richardson J.S., Richardson D.C. MolProbity's ultimate rotamer-library distributions for model validation // Proteins: Structure, Function, and Bioinformatics – Vol. 84(9). – June 2016 – Pp. 1177-1189. DOI: 10.1002/prot.25039.
8. Zhang J., Sanderson A. JADE: Adaptive differential evolution with optional external archive // IEEE Transactions on Evolutionary Computation. – Vol. 13(5). – 2009. – Pp. 945-958. DOI: 10.1109/TEVC.2009.2014613.

9. Silva-Santos C. H., Goncalves M. S., Hernandez-Figueroa H. E. Designing Novel Photonic Devices by Bio-Inspired Computing // IEEE Photonics Technology Letters. – Vol. 22(10). – 2010. – Pp. 1177-1179. DOI: 10.1109/IMOC.2009.5427618.
10. Kennedy J., Eberhart R. Particle swarm optimization // Proceedings of IEEE International Conference on Neural Networks. – Vol. 4. – 1995. – Pp. 1942-1948. DOI: 10.1109/ICNN.1995.488968.
11. Solis F.J., Wets R.J-B. Minimization by random search techniques // Mathematics of Operation Research. – Vol. 6(1) – 1981. – Pp. 19-30. DOI: 10.1287/moor.6.1.19.
12. Poluyan S.V., Reinhard N.M., Ershov N.M. Self-Adaptation in Swarm Optimization Algorithms // Vestnik RUDN. Seriya «Matematika. Informatika. Fizika.» – Vol. 2. – 2014. – с. 415-418.
13. Cheng R., Jin Y. A Competitive Swarm Optimizer for Large Scale Optimization // IEEE Transactions on Cybernetics – Vol. 45(2). – 2015. – Pp. 191-204. DOI: 10.1109/TCYB.2014.2322602.
14. Ershov N.M. Non-uniform cellular genetic algorithms // Computer Research and Modeling. – 2015. – Vol. 7(3) – с. 775-780.
15. Hansen N., Ostermeier A. Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation // Proceedings of the 1996 IEEE International Conference on Evolutionary Computation. – 1996. – Pp. 312-317. DOI: 10.1109/ICEC.1996.542381.
16. Brest J., Zamuda A., et al. Large scale global optimization using self-adaptive differential evolution algorithm // IEEE World Congress on Computational Intelligence. – 2010. – Pp. 1-8. DOI: 10.1109/CEC.2010.5585927.
17. Heterogeneous Computing Cluster HybriLIT. URL: <http://hybrilit.jinr.ru/en/> (дата обращения: 04.04.2017).
18. Galzitskaya O.V., Higo J., Finkelstein A.V.  $\alpha$ -Helix and  $\beta$ -Hairpin Folding from Experiment, Analytical Theory and Molecular Dynamics Simulations // Current Protein and Peptide Science. – Vol. 2(3). – Apr. 2002. – Pp. 191-200. DOI: 10.2174/1389203024605340.
19. Lamiable A., Pierre T., et al. PEP-FOLD3: faster denovo structure prediction for linear peptides in solution and in complex // Nucleic Acids Research, 2016, Vol. 44, Apr. 2016. DOI: 10.1093/nar/gkw329.
20. Raveh B., London N., et al. Rosetta FlexPepDock ab-initio: Simultaneous Folding, Docking and Refinement of Peptides onto Their Receptors // PLoS ONE – Vol. 6(4) – Apr. 2011. DOI: 10.1371/journal.pone.0018934
21. Moradabadi B., Beigy H. A new real-coded Bayesian optimization algorithm based on a team of learning automata for continuous optimization // Genetic Programming and Evolvable Machines – June 2014. – Vol. 15(2). – Pp. 169-193. DOI: 10.1007/s10710-013-9206-9