

И. Е. Щечкин, Т. О. Гуца

## Метод конформационного поиска для длинных полипептидных цепей, основанный на анализе энергетической функции

*(Представлено академиком НАН Украины В. П. Кухарем)*

*The algorithm is proposed to search a conformational space of long flexible polypeptide chains. It consists in performing a series of geometry optimization runs by a gradient-based method using the atom-atom pairwise potential sum as a target function. Only a portion of pairwise terms is taken into account in each run. The algorithm was employed for the conformational analysis of the loop-like 125–167 fragment of human serum albumin treating the rest of the molecule as a rigid structure. Several stable conformers of the fragment were obtained that considerably differ by their 3D-structure from the original one. Some of the newly found conformers have no major attractive nonbonded contacts with the protein globule, and some of them form hydrogen bonds with other domains of the molecule. The latter enables us to suppose that the conformational transition of the 125–167 fragment can affect the mutual mobility of domains.*

Изменчивость конформационной структуры белков играет существенную роль в таких биохимических процессах, как иммунная защита, ферментативный катализ, клеточный транспорт. Известно, например, что активный центр белка перестраивается при связывании субстрата [1, 2], а поверхностные петли в белковой структуре могут служить воротами, обеспечивающими доступ к центру связывания [3, 4]. В ходе структурных превращений молекула белка проходит через ряд стабильных конформаций, определение которых уже давно стало неотъемлемой частью компьютерного моделирования биохимических процессов. Стабильные конформации представляют собой состояния молекулярной системы, соответствующие локальным минимумам функции ее потенциальной энергии. Вполне очевидно, что лишь некоторые из них могут оказаться продуктивными с точки зрения протекания изучаемого процесса.

В конформационном анализе белков используется множество различных численных методов поиска минимумов энергетической функции. Это, прежде всего, методы случайного поиска с использованием модельного отжига [5], метод геометрии межатомных расстояний [6], метод следования направлениям наименьших главных кривизн поверхности потенциальной энергии (низкочастотным модам) [7], методы сглаживания поверхности потенциальной энергии [8]. Все они используют энергетическую функцию для получения значений энергии при заданных значениях аргументов, не предполагая анализа роли отдельных составляющих функции в формировании рельефа поверхности потенциальной энергии.

В данной работе предложен метод конформационного поиска, основанный на анализе свойств отдельных составляющих потенциальной функции. В качестве примера для применения предложенного подхода проведен конформационный анализ участка пептидной цепи сывороточного альбумина человека.

**Подход к задаче минимизации функции.** Рассмотрим функцию как сумму нескольких компонент, каждая из которых имеет свою систему минимумов. Разумно предпо-

ложить, что области аргументов, в которых функция принимает малые значения, соответствуют удачному совпадению областей аргументов с низкими значениями ее компонент. Таким образом, области низких значений отдельных компонент могут служить опорными точками при проведении поиска наиболее низких локальных минимумов полной функции.

Наиболее быстрый метод минимизации функции основан на вычислении ее локальных производных. Однако таким способом можно найти только один минимум и именно тот, в области притяжения которого находится начальная точка. Это застревание градиентных методов в ближайшем к исходной точке локальном минимуме порождает известную “проблему локальных минимумов”. Если в точке локального минимума модифицировать функцию отключением некоторых из ее компонент, то эта точка уже не будет локальным минимумом модифицированной функции. Последующий процесс локальной минимизации приведет к ближайшему локальному минимуму, определяемому оставшимися компонентами. Таким образом, при проведении серии минимизаций с различными частичными наборами компонент полной функции мы будем передвигаться между локальными минимумами различных групп компонент. При этом передвижение, вероятно, будет происходить и по областям наиболее низких значений полной целевой функции, так как последние соответствуют значениям аргументов, при которых одновременно реализуются низкие значения многих компонент. Точки этого пути мы будем использовать в качестве начальных для последующей минимизации полной целевой функции.

Предложенный метод *альтернирования составляющих целевой функции* можно рассматривать как разновидность классического метода *мультистарта*, в котором начальные точки выбираются не на заданной заранее сетке, а находятся проведением *поисковой серии минимизаций частичных функций*.

**Определение модели энергетической функции молекулы.** Пространственная структура молекулы вполне определена значениями ее внутренних координат — длин связей, валентных и двугранных (торсионных) углов. Значительные вариации последних, в отличие от вариаций длин связей и валентных углов, сопровождаются существенным изменением пространственной структуры молекулы без значительного изменения ее энергии, тем самым обеспечивая энергетически плавный путь “конформационного функционирования” за счет флуктуаций теплового движения. Поэтому поиск продуктивных конформаций биомолекулы целесообразно проводить в пространстве ее торсионных углов, приняв длины связей и валентные углы неизменными и равными их равновесным значениям.

Исходя из сказанного выше, в качестве энергетической функции будем использовать зависящую от торсионных углов сумму энергии невалентных взаимодействий, торсионный потенциал и энергию сольватации. Энергию невалентных взаимодействий будем вычислять в рамках метода атом-атомных потенциалов, а сольватационную составляющую — в рамках модели неявного растворителя как сумму поляризационной и гидрофобной энергий [9]. В таком приближении энергетическая функция представляет собой сумму вкладов отдельных пар атомов и торсионного потенциала.

Будем исходить из того, что низкая энергия всей системы обеспечивается в тех ее состояниях, в которых многие пары атомов испытывают значительное притяжение, а количество пар, испытывающих значительное отталкивание, мало. Поэтому конформации, оптимальные с точки зрения энергии отдельных пар атомов, могут служить опорными точками при поиске в конформационном пространстве.

В рамках описанного подхода к минимизации энергетической функции в качестве компонент, из которых сформируются частичные целевые функции, будем использовать пар-

ные взаимодействия, имеющие большую амплитуду изменения на больших протяженностях области конформационного пространства (низкочастотные с большой амплитудой). К ним относятся наиболее сильные электростатическое притяжение и отталкивание, а также поляризационная составляющая энергии сольватации. Указанные компоненты вместе с вандерваальсовым отталкиванием определяют *основной архитектурный мотив* поверхности потенциальной энергии. Будем называть их *главной частью* энергетической функции. Вандерваальсов парный потенциал, в отличие от электростатического и поляризационного, носит короткодействующий характер, быстро изменяется на протяжении малых областей конформационного пространства (высокочастотный с большой амплитудой). Поэтому при поиске в широком диапазоне значений независимых переменных эту составляющую учитывать нецелесообразно. Вандерваальсовы взаимодействия будут учтены при минимизации энергии на заключительном этапе. Слабо изменяющиеся парные вклады и компоненты торсионного потенциала на этапе поисковой минимизации в расчет не принимаются.

Таким образом, при проведении поисковой серии минимизаций в качестве частичных целевых функций предлагается использовать различные группы парных взаимодействий. Используемую группу будем называть текущей *фракцией* полного множества пар атомов, взаимодействующих в данной модели. Такие параметры, как количество пар во фракции, возможное неполное выключение взаимодействия внефракционных пар и степень его ослабления, оптимальное количество циклов перебора всех возможных комбинаций пар, а также количество и момент выбора начальных точек для минимизации полной энергии могут варьироваться. Оптимальные их значения требуют определения опытным путем.

На этапе поисковой минимизации вместо реальной энергии парного взаимодействия будем использовать более простую функцию, которая достигает малых значений в той же области конформационного пространства, что и реальная энергия. В отличие от полной энергии, для отдельных парных вкладов можно заранее сформулировать, при каком пространственном расположении атомов энергия будет достигать наименьших значений. Так, энергия притягивающейся пары достигает минимума при наибольшем пространственном сближении атомов, а отталкивающейся — при наибольшем удалении. Это справедливо как для атом-атомных потенциалов, так и для поляризационной составляющей сольватации в методе [9]. Кроме того, экранирование неполярными атомами полярных приводит к ухудшению сольватации и повышению энергии системы. Поэтому состояние удаленности слабо заряженных атомов от сильно заряженных соответствует минимуму энергии.

В качестве составляющей парного взаимодействия будем использовать простую параболическую функцию, значение и градиент которой вычисляются быстрее, чем те же величины для реальной энергетической функции:

для притягивающихся пар

$$e = a(r^2 - r_0^2) \quad \text{при } r \geq r_0 \quad \text{и} \quad e = 0 \quad \text{при } r < r_0,$$

для отталкивающихся пар

$$e = -a(r^2 - r_0^2) \quad \text{при } r \leq r_0 \quad \text{и} \quad e = 0 \quad \text{при } r > r_0,$$

где  $a > 0$  — постоянный коэффициент;  $r$  — расстояние между центрами атомов;  $r_0$  — расстояние, на котором атомы притягивающейся пары достаточно близки, чтобы обеспечить большую энергию притяжения, а взаимодействие отталкивающейся пары можно полагать малым. Величина  $r_0$  подбирается опытным путем.

**Приложения.** Алгоритм, подобный изложенному, уже применялся нами ранее при поиске конформации связывания дополнительного центра тромбина с комплементарным ему сайтом фибриногена [10]. В этом случае в качестве альтернативных компонент целевой функции выступали потенциал притяжения в парах лигандных групп, с одной стороны, и потенциал вандерваальсового отталкивания между всеми остальными парами атомов — с другой. Далее дано приложение сформулированного подхода для конформационного анализа участка пептидной цепи сывороточного альбумина человека (САЧ).

Визуальные исследования пространственной структуры молекулы САЧ, полученной в работе [12], показывают, что участки 125–167 и 317–359, лежащие в началах доменов Ib и IIb соответственно и представляющие собой петлеобразные структуры, особым образом пространственно расположены относительно глобулы белка. Корневые части каждого из них, “стреноженные” цистеиновым мостиком, лежат возле поверхности глобулы, а вершины глубоко внедрены внутрь молекулы. Такое расположение участков вызывает интерес к их возможным конформациям, в которых петлеобразные структуры выведены из глобулы белка. При реализации таких конформаций характер взаимодействий в молекуле, очевидно, существенно изменится и может нарушить баланс сил взаимодействия между доменами. Это, в свою очередь, может повлиять на характер доменной подвижности. Для проверки способности участка 125–167 принимать такого рода конформации нами был проведен конформационный поиск в пространстве его торсионных углов. При этом вся остальная часть белка рассматривалась в неизменной конформации, соответствующей кристаллической структуре.

В расчетах использованы атомные координаты САЧ, молекула А, с идентификатором 1A06 из банка данных белковых структур *Brookhaven Protein Data Bank*. Для проведения конформационного поиска мы использовали описанный выше алгоритм, реализованный в нашей программе *FlexiChain*. Все расчеты проводились при помощи компьютерного кластера, созданного при содействии компании *Intel* в вычислительном центре Киевского национального университета.

Были использованы два типа фракций парных взаимодействий. Каждая фракция первого типа состояла из одной пары отталкивания или притяжения, соответствующей сильному электростатическому или поляризационному взаимодействию, а каждая фракция второго типа — из всех пар отталкивания между одним сильно полярным и всеми слабо полярными атомами. При этом рассматривались только разделенные не менее чем шестью варьируемыми связями пары атомов внутри участка 125–167. Сильно полярными считались атомы с абсолютной величиной заряда более 0,75 а.е., а в качестве слабо полярных рассматривались атомы с абсолютной величиной заряда менее 0,15 а.е.

Нами выполнено 180 циклов поисковых минимизаций. Каждый цикл состоял из сеансов минимизации с целевой функцией, основанной на некоторой фракции. В каждом из циклов каждая фракция участвовала по одному разу. Порядок использования фракций в каждом цикле определялся случайным образом. Конформация, полученная после выполнения каждого из циклов, сохранялась. Сохраненные конформации были подвергнуты кластерному анализу по среднеквадратичному отклонению координат атомов. В результате были отобраны конформации (41 шт.), которые использовались в качестве начальных точек для минимизации полной энергии.

Минимизация полной энергии была проведена при помощи программы *Tinker* [<http://dasher.wustl.edu/tinker>]. В качестве независимых переменных использовались декар-

товы координаты атомов участка 125–167, а остальная часть молекулы была заморожена. Полная энергия внутримолекулярного взаимодействия вычислялась в рамках метода атом-атомных потенциалов с использованием силового поля *AMBER* с файлом параметров *amber96.prm*. Для вычисления энергии взаимодействия с водным окружением использовался неявный учет растворителя по методу, изложенному в работах [9, 11].

Из совокупности оптимизированных конформаций были выбраны девять, имеющие наиболее низкую энергию (табл. 1). Для них были определены наиболее эффективные *стягивающие контакты*, т. е. соответствующие наиболее сильным парным электростатическим взаимодействиям. Оказалось, что эти контакты представляют собой водородные связи между боковыми цепями остатков.

С термодинамической точки зрения представленные в табл. 1 конформации слишком существенно разнятся по энергии, чтобы в одинаковой степени присутствовать в статистическом ансамбле. Такая разница между энергиями связана с тем, что в связи с ограниченностью вычислительных ресурсов мы рассматривали большую часть молекулы жесткой, и значительная часть степеней свободы системы была заморожена. Размораживание дополнительных степеней свободы и продолжение минимизации энергии приведет к дальнейшему снижению энергий. При этом, как показывает опыт подобных расчетов, основной архитектурный мотив пространственной структуры почти не изменится, а энергии различных конформаций, возможно, выровняются.

Следует обратить внимание на структурные особенности приведенных в табл. 1 конформаций. Если в исходной конформации рассматриваемый участок цепи имел стягивающие контакты лишь со своим субдоменом Ib, то в некоторых из альтернативных конформаций он имеет такие контакты также и с другими субдоменами и, следовательно, служит связкой между ними. Наибольшее число контактов за пределами своего субдомена завязывают конформация 9 — с субдоменами Ia и IIIb, конформация 4 — с Ia и IIa и конформация 2, которая завязывает три водородных связи с субдоменом Ia. Существование таких связей дает основание полагать, что данные конформационные переходы участка могут приводить к изменению динамики движения субдоменов САЧ друг относительно друга, уменьшая взаимную их подвижность.

Таблица 1. Полная энергия и ее составляющие, а также стягивающие контакты между участком 125–167 и глобулой САЧ для исходной и альтернативных конформаций

Конформация	Полная энергия	Ван-дер-Ваальс	Электростат	Сольватация	Контакты между петлей и глобулой
Исходная	–1075,4	98,0	–1152,2	–460,7	GLU188(Ib)
1	–1038,3	19,8	–1208,7	–531,6	Her
2	–1621,2	53,4	–3645,0	876,1	ARG117(Ia), ARG98(Ia), ASP72(Ia), ASP187(Ib)
3	–1610,6	72,9	–4277,9	1512,6	ASP259(IIa), ASP13(Ia)
4	–2087,7	64,0	–5960,0	2736,8	ARG10(Ia), ASP259(IIa), ASP255(IIa)
5	–1217,4	66,7	–2959,5	472,5	ARG117(Ia)
6	–1860,3	–114,2	–5542,8	2880,7	ASP187(Ib), ARG117(Ia)
7	–1049,5	268,4	–2376,0	159,1	ASP121(Ib), ASP563(IIIb)
8	–1659,7	–80,1	–4459,1	1748,3	Her
9	–1337,1	225,0	–5002,0	1861,9	ASP183(Ib), ARG186(Ib), ARG117(Ia), GLU520(IIIb)

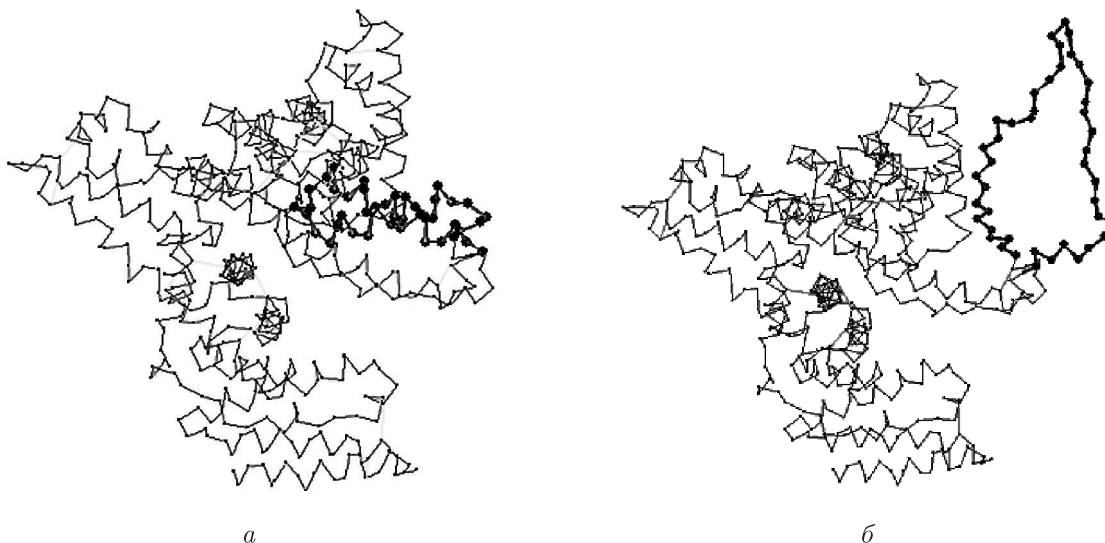


Рис. 1. Пространственная структура участка 125–167 сывороточного альбумина человека в двух конформациях, представленных в табл. 1:  
*a* — исходная конформация; *б* — конформация 1, найденная в результате конформационного поиска. Показан  $C_{\alpha}$  скелет молекулы, участок 125–167 выделен жирной линией

Отметим также, что стягивающие контакты участка с субдоменами IIb и IIIa отсутствуют во всех найденных конформациях. Кроме того, обнаружены две “висящих”, т. е. не имеющих существенных стягивающих контактов с глобулой, конформации — 1 и 8 (см. табл. 1). Одна из последних показана на рис. 1 в сравнении с исходной. В “висящих” конформациях рассматриваемый участок подвижен и может служить “щупальцем” при взаимодействиях с лигандами.

Таким образом, предложенный метод конформационного поиска для длинных полипептидных цепей, основанный на анализе энергетической функции, показал свою способность находить значительно различающиеся конформеры длинной, состоящей из 43 остатков, полипептидной цепи. В результате применения этого метода к участку 125–167 цепи САЧ получены стабильные конформации участка, отличающиеся по своим структурным характеристикам от исходной. Среди этих конформаций имеются такие, в которых участок за счет завязывания водородных связей образует стяжки между субдоменами вида IIb-Ia, IIb-IIIb IIb-Ia-IIIb, и IIb-Ia-IIIa. Благодаря образованию такого рода стяжек характер взаимной подвижности доменов может изменяться. Кроме того, обнаружены несвязанные с глобулой конформации петли, обладающие вследствие этого значительной подвижностью.

1. Getzoff E. D., Geysen H. M., Rodda S. J. et al. Mechanisms of antibody binding to a protein // *Science*. – 1992. – **235**. – P. 1191–1196.
2. Rini J. M., Schulze-Gahmen U., Wilson I. A. Structural evidence for induced fit as a mechanism for antibody-antigen recognition // *Ibid.* – 1992. – **255**. – P. 959–965.
3. Nicholson L. K., Yamazaki T., Torchia D. A. Change C-H Flexibility and function in HIV – 1 protease // *Struct. Biol.* – 1995. – **2**. – P. 274–280.
4. Wagner G. The importance of being floppy // *Struct. Biol.* – 1995. – **2**. – P. 255–257.
5. Chang G., Guida W. C., Still W. C. An internal coordinate Monte Carlo method for searching conformational space // *J. Amer. Chem. Soc.* – 1989. – **111**. – P. 4379–4386.
6. Crippen G. M., Havel T. F. *Distance Geometry and Molecular Conformation*. – New York: Wiley, 1988.

7. *Kolossvary G., Guida W.* Low mode search-an efficient, automated computational method for conformational analysis: application to cyclic and acyclic alkanes and cyclic peptides // *J. Amer. Chem. Soc.* – 1996. – **118**. – P. 5011–5019.
8. *Kostrowicki J., Scheraga H. A.* Application of the Diffusion Equation Method for Global Optimization to Oligopeptides // *J. Phys. Chem.* – 1992. – **96**. – P. 7442–7449.
9. *Qiu D., Shenkin P. S., Hollinger F. P., Still W. C.* The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii // *J. Phys. Chem. A.* – 1997. – **101**. – P. 3005–3014.
10. *Shchekhin I. E., Karabut L. V., Chetyrkina S. N., Sereyskaya A. A.* Theoretical conformational investigation of thrombin-fibrinogen recognition complex // *Ukrain. Biochem. J.* – 1995. – **67**, No 5. – P. 110–114.
11. *Still W. C., Tempczyk A., Hawley R. C., Hendrickson T.* A Semiempirical Treatment of Solvation for Molecular Mechanics and Dynamics // *J. Amer. Chem. Soc.* – 1990. – **112**. – P. 6127–6129.
12. *Sugio S., Kashima A., Mochizuki S. et al.* Crystal structure of human serum albumin at 2.5 Å resolution // *Prot. Eng.* – 1999. – **12**, No 6. – P. 439–446.

*Институт биоорганической химии  
и нефтехимии НАН Украины, Киев*

*Поступило в редакцию 23.01.2007*