

A METHOD FOR SEARCH OF STRUCTURE MOTIFS IN AMINOACID SEQUENCES  
PROGRAM SITE OF THE GENBEE PACKAGE

*E. V. Koonin, K. M. Chumakov, A. E. Gorbalenya*

Institute of Microbiology, the USSR Academy of Sciences, Moscow;  
Institute of Poliomyelitis and Viral Encephalitis,  
Academy of Medical Sciences of the USSR, Moscow Region

Summary

A method is suggested to search for the structure motifs in amino acid sequences, based on their scanning by frequency profiles generated from aligned sequence segments. The program «SITE» implementing the proposed algorithm is discussed, exemplified by the search of an amino acid sequence database for the motif typical of a vast class of purine NTP-binding proteins. The superiorities of the proposed approach as compared to standard pattern-searching routines is demonstrated with respect to selectivity and completeness of extraction of relevant sequences.

СПИСОК ЛИТЕРАТУРЫ

1. *Staden R.* Methods to define and locate patterns of motifs and sequences // CABIOS.— 1988.— 4, N 1.— P. 53—60.
2. *Hodgman T. C.* The elucidation of protein function by sequence motif analysis // Ibid.— 1989.— 5, N 1.— P. 1—13.
3. *Кумура М.* Молекулярная эволюция: теория нейтральности.— М.: Мир, 1985.— 398 с.
4. *Distantly related sequences in the  $\alpha$ - and  $\beta$ -subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold / J. E. Walker, M. Saraste, M. J. Runswick, N. J. Gay // EMBO J.— 1982.— 1, N 8.— P. 945—951.*
5. *Gorbalenya A. E., Koonin E. V.* Viral proteins containing the NTP-binding sequence pattern // Nucl. Acids Res.— 1989.— 17, N 22.— P. 8413—8440.
6. *Dayhoff M. O., Barker W. C., Hunt L. T.* Establishing homologies in protein sequences // Meth. Enzymol.— 1983.— 91.— P. 524—549.
7. *HIV F/3 orf encodes a phosphorylated GTP-binding protein resembling an oncogene product / B. Guy, M. P. Kieny, Y. Riviere et al. // Nature.— 1987.— 330, N 6145.— P. 266—269.*
8. *Molecular analysis of the Escherichia coli recO gene / P. T. Morrison, S. T. Lovell, L. E. Gilson, R. Kolodner // J. Bacteriol.— 1989.— 171, N 7.— P. 3641—3649.*
9. *Fujisawa H., Yonesaki T., Minagawa T.* Sequence of the T4 recombination gene, *UvsX*, and its comparison with that of the *recA* gene of *Escherichia coli* // Nucl. Acids Res.— 1985.— 13, N 20.— P. 7443—7481.

Ин-т микробиологии АН СССР, Москва  
Ин-т полиомиелита и вирус. энцефалитов АМН СССР,  
Моск. обл.

Получено 28.05.90

УДК 577.32

**С. Г. Галактионов, В. М. Цейтин, И. А. Ваксер**

**КОМПЬЮТЕРНОЕ КОНСТРУИРОВАНИЕ  
БИОЛОГИЧЕСКИ АКТИВНЫХ ПЕПТИДОВ:  
НЕКОТОРЫЕ НОВЫЕ ВОЗМОЖНОСТИ**

*Развиты алгоритмы теоретического конформационного анализа для расчета стабильных конформаций молекул пептидов на поверхности раздела фаз «вода — липофильная среда». Поскольку взаимодействие многих пептидных биорегуляторов со специфическими рецепторами осуществляется главным образом за счет гидрофобных сил, развитая процедура может быть использована для компьютерного проектирования аналогов природных биорегуляторов, обладающих повышенным сродством к рецептору, измененным спектром проявления биологической активности, пролонгированным действием.*

**Введение.** Существующие стратегии конформационно направленного конструирования биологически активных пептидов основываются на поиске так называемых «биологически активных» конформаций.

© С. Г. ГАЛАКТИОНОВ, В. М. ЦЕЙТИН, И. А. ВАКСЕР, 1990

Элементы таких конформаций могут быть установлены на основании сопоставления конформационных возможностей активных и неактивных аналогов данного биорегулятора [1]. Этот подход, как известно, оказался вполне успешным при конструировании конформационно ограниченных аналогов многих природных пептидов.

Более целенаправленная модификация структуры пептидной молекулы для получения желаемых изменений в параметрах ее биологической активности стала бы возможной при известной ориентации молекулы пептида относительно рецептора. Однако установление топографии рецептора на основании данных об активности аналогов с объемными заместителями в различных положениях, как это сделано для низкомолекулярных биорегуляторов, практически невозможно из-за относительно больших размеров молекулы пептида.

Ряд данных позволяет предположить, что многие биологически активные пептиды взаимодействуют со своими специфическими рецепторами главным образом благодаря гидрофобным силам. Молекулы этих пептидов способны принимать конформацию, обеспечивающую пространственное разделение гидрофобных и гидрофильных функциональных групп; таким образом, эти соединения обладают выраженной поверхностной активностью [2]. Следовательно, мы можем получить «подсказку» об ориентации молекулы биорегулятора относительно рецептора, рассматривая взаимодействие этой молекулы с границей фаз «вода — липофильная среда» в качестве модели поведения на поверхности рецептора. В самом деле, корреляция между сродством фрагментов в АКТГ к липофильной поверхности и их биологической активностью показана Гисин и Швицером [3].

Целью работы является демонстрация полезности этой концепции для конформационно направленного конструирования биологически активных пептидов с помощью методов теоретического конформационного анализа.

**Материалы и методы.** Для расчета стабильных конформаций молекул в вакууме и в водном окружении использовали описанные ранее алгоритмы и параметры [4].

Энергию гидратации рассчитывали методом Хопфингера — Шераги [5]. Границу фаз рассматривали как плоскость, разделяющую две среды со свойствами воды и со свойствами октанола. Для расчета стабильных конформаций пептидной молекулы на такой границе развит алгоритм, выполняющий минимизацию суммы конформационной энергии и энергии сольватации по торсионным углам, углам  $\Theta$  и  $\Phi$ , определяющим ориентацию молекулы относительно границы фаз и глубины погружения молекулы в липофильную фазу  $l$  (рис. 1).

**Результаты и обсуждение.** Рассчитаны спектры стабильных конформаций в водном растворе и на границе фаз двух выражено амфифильных пептидных биорегуляторов — ангиотензина и брадикинина, а также их шести фрагментов. Энергию переноса каждой молекулы из водного окружения на границу раздела фаз вычисляли в виде

$$U^{hb} = \min_i U^b(\bar{\Phi}_i) - \min_j U^h(\bar{\Phi}_j),$$

где  $U^b(\Phi_k)$ ,  $U^h(\Phi_k)$  — энергии  $k$ -й стабильной конформации,  $\bar{\Phi}_k = \{G_1, G_2, \dots, G_n\}$  на границе фаз и в водном окружении соответственно.

Сравнение рассчитанных величин и энергии переноса  $\Delta U^{hb}$  с величинами  $pD_2$  для соответствующих соединений [6] обнаружило плотную корреляцию между этими двумя параметрами (рис. 2) в полном согласии с уже упомянутыми результатами Гисин и Швицера.

Иерархия стабильности основных структур ангиотензина в основном одинакова в вакууме, воде и на границе фаз (табл. 1). В случае брадикинина эта иерархия претерпела коренные изменения (табл. 2).

Анализ данных о биологической активности ряда аналогов ангиотензина, предпринятый в [1], указал на две структуры как вероятные

«биологически активные» конформации VBBRRBVB и VRBRRBVB. Последняя, имеющая довольно высокую энергию в вакууме, еще более нестабильна на границе фаз, поэтому, вероятно, рецептор-связываемой структурой является VBBRRBVB (рис. 3), являющаяся наиболее стабильной как в вакууме, так и на границе раздела фаз.

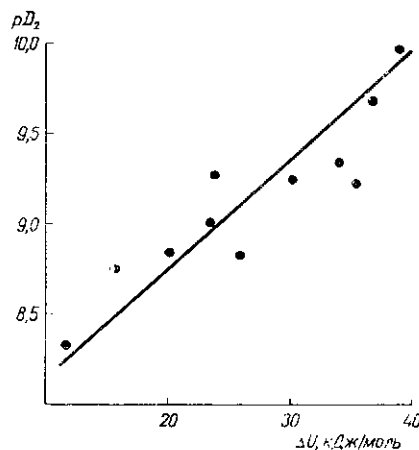
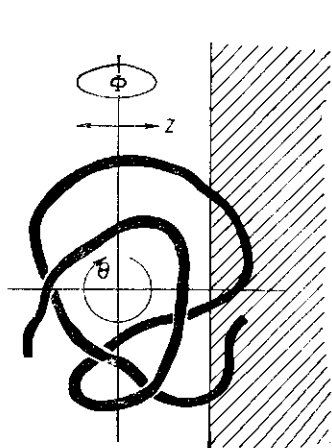


Рис. 1. Молекула условного пептида на поверхности раздела фаз «вода — липофильная среда»

Fig. 1. Symbolic peptide molecule at the interphase «water — lipophilic medium»

Рис. 2. Сопоставление значений  $pD_2$  с оценками энергий переноса некоторых биологически активных пептидов

Fig. 2. Correlation between  $pD_2$  values and transfer energy  $\Delta U$  for some biologically active peptides.

Брадикинин взаимодействует с двумя типами рецепторов  $V_1$  и  $V_2$ , имеющими различные требования к пространственной структуре. Наиболее вероятными «биологически активными» конформациями для

Таблица 1

Стабильные конформации ангиотензина в различном окружении

Stable backbone conformations of angiotensin

Конформация	Энергия, кДж/моль		
	В вакууме	В воде	На границе фаз
VBBRRBVB*	11,8	-154,6	-190,3
VBLRRBVB	21,0	-136,9	-163,0
VRBRRBVB*	15,5	-79,8	-111,7
VBLRRBVB	20,2	-73,1	-100,4
VRBRRBVB	0,0	-124,3	-148,7

\* Отмечены «продуктивные» конформации.

рецепторов  $V_1$  и  $V_2$  являются VBBHRBBLV и VBBHRBRRV соответственно [1]. Структура VBBHRBBLV на границе раздела вполне стабильна, а структура VBBHRBRRV имеет энергию

74 кДж/моль по сравнению с наиболее стабильной конформацией, поэтому маловероятно, что она является «биологически активной» в отношении рецептора типа  $V_2$ .

Таблица 2

Стабильные конформации брадикинина в различном окружении

Stable bradykinine backbone conformations

Конформация	Энергия, кДж/моль		
	В вакууме	В воде	На границе фаз
VVBHBBRBB	5,5	-123,5	-163,8
VBRRRBRBL	20,2	-74,8	-110,0
VBBHBBBRB	0,8	-96,2	-141,5
VBBHRBRRB	34,0	-117,6	-147,4
VBRRRBVVV	5,0	-104,2	-117,2
VBBHRBRRB*	42,0	-56,7	-92,8
VBBHBBRRL	15,5	-126,0	-166,7
VBBHRBRRBL	37,8	-60,9	-98,7
VBRRBLVVV	0,0	-95,3	-143,6
VBBHRBBLB*	10,1	-119,7	-155,0
VBRRBLRBB	8,4	-118,9	-154,1
VBRLBVVV	55,9	-102,5	-132,7

\* Отмечены «продуктивные» конформации.

Во всех описанных рецептор-связываемых конформациях обеих молекул в гидрофобной части молекулы есть функциональные группы, обладающие донорными, и группы, обладающие акцепторными свойствами, ориентированные по направлению к рецептору. Мы можем предположить существование на рецепторе соответствующих функциональных групп-партнеров, которые могут образовывать водородную связь, дополнительно стабилизирующую комплекс.

Развитый подход позволяет ввести некоторые новые элементы в идеологию конформационно направленного конструирования биологически активных пептидов. Можно указать функциональные группы, непосредственно вовлеченные во взаимодействие пептида с рецептором, и на этой основе предложить такую модификацию

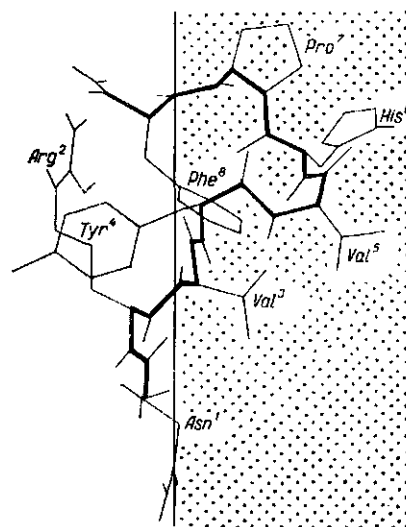
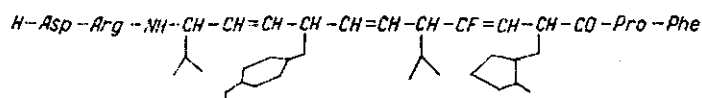


Рис. 3. Молекула ангиотензина на границе фаз  
Fig. 3. Angiotensin molecule at the interphase.

структуры молекулы, которая приведет к увеличению или уменьшению сродства пептидной молекулы к рецептору.

Для иллюстрации этого подхода мы можем воспользоваться описанными выше результатами, относящимися к ангиотензину. Для усиления гидрофобного взаимодействия некоторые полярные группы основной цепи молекулы, предпочтительно находящейся в контакте с поверхностью рецептора, могут быть заменены на метиленовые группы; более того, предполагая наличие водородной связи между карбонильной группой остатка Val<sup>5</sup> и функциональной группой — донором рецептора, мы можем попытаться усилить ее введением в соответствующее положение, например, атома фтора:



Можно ожидать увеличения стабильности комплекса этого аналога с рецептором примерно на 10 кДж/моль, а величины  $pD_2$  (с учетом данных рис. 2) — на 3,7.

Этот пример является лишь иллюстративным; очевидно, что знание ориентации молекулы биорегулятора относительно рецептора может оказаться полезным и во многих других отношениях при конформационно направленном конструировании аналогов биологически активных пептидов.

#### COMPUTER-AIDED DESIGN OF BIOLOGICALLY ACTIVE PEPTIDES: SOME NEW POSSIBILITIES

S. G. Galaktionov, V. M. Tseytin, I. A. Vakser

All-Union Research Institute of Microbiological Industry, Minsk

#### Summary

Algorithms for the calculation of peptide molecules stable conformation on the «water-non-polar phase» boundary has been developed. Since many biologically active peptides are supposed to interact with specific receptors mainly due to hydrophobic forces, the model

of receptor binding site can be obtained as a mould of the hydrophobic part of the molecule, immersed into lipid phase, in the so-called «biologically active» conformation. Therefore, it is possible to make analogs with enhanced affinity to receptor; this possibility is exemplified by newly-constructed carba-analog of angiotensin.

#### СПИСОК ЛИТЕРАТУРЫ

1. Никифорович Г. В., Галактионов С. Г., Чипенс Г. И. Конформации пептидных биорегуляторов.— М.: Медицина, 1983.— 192 с.
2. Surface properties of an amphiphilic peptide hormone and its analog: corticotropin releasing factor and sauvagine / S. H. Lau, J. Rivier, W. Vale et al. // Proc. Nat. Acad. Sci. USA.— 1983.— 80, N 23.— P. 7070—7074.
3. Gysin B., Schwyzer R. Liposome-mediated labelling of adrenocorticotropin fragments parallels their biological activity // FEBS Lett.— 1983.— 158, N 1.— P. 12—16.
4. Theoretical conformational analysis of oxytocin molecule / G. V. Nikiforovich, V. I. Leonova, S. G. Galaktionov, G. I. Chipens // Int. J. Pept. Prot. Res.— 1979.— 13, N 3.— P. 363—373.
5. Hopfinger A. J., Battershell R. D. Application of SCAP to drug design. I. Prediction of octanol-water partition coefficients using solvent-dependent conformational analysis // J. Med. Chem.— 1976.— 19, N 5.— P. 569—573.
6. Структура и функции низкомолекулярных пептидов.— Рига: Зинатне, 1980.— 327 с.

Всесоюз. науч.-исследоват. и проектно-конструкт. ин-т микробиол. пр-в, Минск

Получено 28.03.90

УДК 577.323:681.3

Н. И. Луккина, Т. Р. Сойдла

#### ПОИСК ГИГАНТСКИХ ДНК-БЕЛКОВЫХ КОМПЛЕКСОВ ПРИ ПОМОЩИ ЭВМ

*Созданы две программы для выявления и визуализации пространственной асимметрии ДНК. С их помощью анализировали 2 мкм плазмиду дрожжей-сахаромицетов. Выявлен участок особой ориентации G—C-пар. Расположение этого участка совпадает с локализацией нуклеосомного ДНК-белкового комплекса, выявленного ранее в экспериментальной работе. Созданные программы могут быть использованы для поиска крупных ДНК-белковых комплексов.*

**Введение.** Начало секвенирования огромных геномов эукариот ставит перед исследователем заманчивый, но трудный вопрос: существуют ли правила, которые позволяют на основе одной лишь нуклеотидной последовательности выявить участки ДНК, взаимодействующие с регуляторными, а, возможно, и другими нуклеосомными белками. Известно, что белки узнают на ДНК самые разные последовательности, которые невозможно свести к одному или нескольким элементарным консенсус-последовательностям. Список белков, взаимодействующих с ДНК, уже долгие годы остается очень и очень неполным [1]. Поэтому основной вопрос состоит в том, можно ли создать алгоритм для поиска участков ДНК, узнаваемых белками, при условии, что нам не известно, какие нуклеотидные последовательности эти белки узнают на ДНК.

При попытке создания наших программ поиска ДНК-белковых комплексов мы исходили из двух положений. Во-первых, что взаимодействие ДНК с белком чаще всего ограничено только одной стороной ДНК. Во-вторых, что на узнаваемой белком стороне ДНК нуклеотидный состав отличается от усредненных данных по рассматриваемому району. Таким образом, мы искали пространственно-ориентированные аномалии нуклеотидного состава ДНК.

© Н. И. ЛУККИНА, Т. Р. СОЙДЛА, 1990