



ПИСЬМА РЕДАКТОРУ

УДК 577.112.5.087

ГРАФИЧЕСКАЯ ИДЕНТИФИКАЦИЯ КОНСЕРВАТИВНЫХ И ВАРИАБЕЛЬНЫХ УЧАСТКОВ В АМИНОКИСЛОТНЫХ ПОСЛЕДОВАТЕЛЬНОСТЯХ ГОМОЛОГИЧНЫХ БЕЛКОВ

*Костецкий П. В., Владимирова Р. Р.**Институт биоорганической химии им. М. М. Шемькина Академии наук СССР, Москва*

Нами ранее [1] предложен метод оценки достоверности неравномерно-го распределения аминокислотных замен в гомологичных фосфолипазах. При этом вероятность неслучайного возникновения консервативных зон в семействе четырех фосфолипаз доказывали с помощью численного эксперимента на искусственных гомологичных семействах, которые получали произвольной перестановкой неизменных столбцов остатков аминокислот в выравненных последовательностях исходных фосфолипаз, расположенных друг под другом. Недостатком этого метода является трудность визуального обнаружения консервативных и переменных зон.

В настоящей работе представлен улучшенный метод графической идентификации участков с резкой неравномерностью распределения аминокислотных замен в семействах гомологичных белков. Наиболее подробно были изучены фосфолипазы А2 яда змей [2, 3], представленные двумя группами последовательностей: семью фосфолипазами змей рода кобр *Naja* и тремя фосфолипазами змей рода гадюк *Bitis* (рис. 1).

Очевидно, что возможно сравнение белковых последовательностей, принадлежащих или разным группам (межгрупповое сравнение), или одной и той же группе (внутригрупповое сравнение). При построении профиля межгруппового сравнения (изменчивости) гомологичных последовательностей (рис. 2) каждая точка на графике соответствует изменчивости сегмента из 10 аминокислотных позиций (колонок), в котором подсчитывают сумму замен во всех колонках при сравнении 21 пары аминокислотных остатков. Полученное значение делят на максимально возможное число замен (21·10), чтобы изменчивость сегмента находилась в пределах от 0 до 1.

Видно (рис. 2а), что профиль изменчивости фосфолипаз характеризуется наличием отчетливых пиков и впадин, соответствующих наиболее переменным и консервативным участкам. При этом разброс вокруг среднего значения изменчивости весьма велик. Площадь, соответствующая этому разбросу, заключена между кривой изменчивости и ее средним значением. В рассматриваемом случае наблюдаемая площадь (S_n) значительно превосходит среднюю расчетную площадь (S_p) для серии из 1000 искусственных гомологичных семейств фосфолипаз, получаемых перестановкой столбцов остатков аминокислот исходного семейства белков [1]. Наблюдаемое различие ($S_n - S_p$) — общая неравномерность аминокислотных замен — составляет 4,8 стандартных отклонений (σ) от среднего значения расчетной площади S_p и заметно превышает величину 2σ , которая отвечает 95%-ной достоверности отличия S_n от случайного значения.

Это превышение логично связать с экстремальными участками графика, которые можно выделить с помощью горизонтальных прямых, отсекаю-

	10	20	30	40	50	60	65
а	NLYQFKNMIQCTVPNRSWDFADYGCYCGRGSGT	PVDDLDRCCQVHDNCYDEAEKISRCWPYFK					
бS.....N.....G.....					
вK...S...L...N.....I.....N..G...G.....					
гH.....P..H..N.....K.....I...K.....G...I..					
дH.....H..N.....I.....C.....G...I..					
еH....S.P..H.....K..A.....G...L-G...LT					
жH....S.P..H.....K.....EK.G.M-G....T					
з	..I..G...SAMTGK-.SLAY.S.....W..K.Q.K..T...F...C..CK.D.C.PKM---I						
и	D.T..G...NKMGQ--VF.YIY.....W..K.K.I.AT...F...C..GKMCTYDTK.---T						
к	D.T..G...NKMGQ--VF.YIY.....W..Q.K.R.AT...F...C..GKMCTYDTK.---T						
	70	80	90	100	110	120	
а	TYSYECSQGTLTCKNGNNAACAAVCDCLDLAAICFAGAFYNNN-NYNIDLKARCQ						
бG...CA.....D.-D...N.....E					
вGDD.N...S.....Y.....D-...N.....						
г	..T..SC.....D.G-K...S.....V..N...R.T..DK-.....FN....						
д	..T.DSC.....SCGAA.N...S.....V..N...R...IDK-.....FN....						
е	L.K.....K...SG...K.E...N..LV..N.....IDA-...VN..E...						
ж	L.K.K....K...SG..SK.G...N...W...N.....R.IDA-...NF.K...						
з	L...KFHN.NIV.-GDK...KKK..E...V.....ASKHSY.K.LWRYPSSK.TTGTAEKC						
и	S.N..IQN.GID.--DEDPQKKEL.E...V.....NNRNTY.S..FGHSSSK.TGTEQC						
к	S.N..FQD.DII.-GDKDPQKKEL.E...V.....NSRNTY.SK.FGYSSSK.TETEQCC						

Рис. 1. Аминокислотные последовательности фосфолипаз родов *Naja* и *Balis*: *N. n. kaouthia* (а), *N. n. atra* (б), *N. n. oxiana* (в), *N. melanoleuca* I (г), *N. melanoleuca* III (д), *N. m. mossambica* (е), *N. m. pallida* (ж), *B. caudalis* (з), *B. gabonica* (и), *B. nasicornis* (к). Для достижения максимума сходства в гомологичные последовательности введены дефисы, обозначенные дефисом. Одинаковые с последовательностью (а) остатки в последовательностях (б)–(к) обозначены точками

щих «излишек» площади, равный $S_n - (S_p + 2\sigma)$. Идентифицируемые таким образом консервативные и вариабельные участки гомологичных последовательностей можно с вероятностью 0,95 считать ответственными за наблюдаемую общую неравномерность аминокислотных замен при сравнении последовательностей фосфолипаз А2 змей *Naja* и *Bitis*.

Гораздо менее резким является профиль внутригрупповой изменчивости семи последовательностей фосфолипаз *Naja*, получаемый в результате 21 парного сравнения (рис. 2б). Значение общей неравномерности здесь составляет только 2,3σ, а число пиков и впадин, выходящих за уровень 95%-ного случайного возникновения, заметно меньше. Интересно, однако, что наиболее интенсивный пик изменчивости для группы последовательностей *Naja* совпадает по местоположению с одним из пиков межгруппового профиля изменчивости. В то же время очень интенсивный участок на графике межгруппового сравнения, отвечающий консервативному участку последовательностей 87–105, мало заметен на графике внутригруппового сравнения.

Важно отметить также, что при межгрупповом сравнении разница в количественном соотношении групп имеет ограниченное влияние. Так, профиль межгрупповой изменчивости сохраняет свои основные свойства

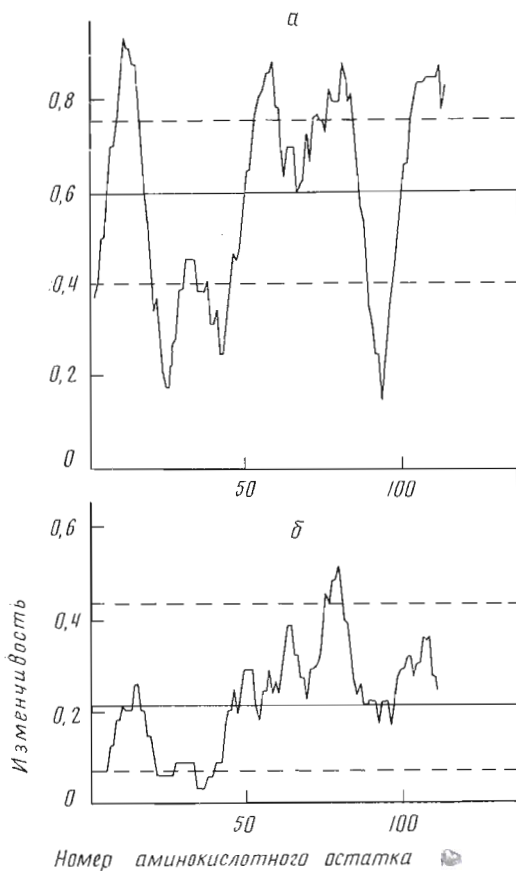


Рис. 2. Профили изменчивости аминокислотных последовательностей фосфолипаз A2 яда змей родов *Naja* и *Bitis*. Сплошная горизонтальная линия отвечает среднему значению изменчивости по всей длине сравниваемых последовательностей. Верхняя и нижняя прерывистые линии идентифицируют пики и впадины, отвечающие достоверным переменным и консервативным участкам гомологичных белковых последовательностей: а — профиль межгрупповой изменчивости фосфолипаз A2 при сравнении родов *Naja* и *Bitis*; б — профиль внутригрупповой изменчивости фосфолипаз рода *Naja*

даже при сравнении 7 фосфолипаз *Naja* с одной фосфолипазой *B. gabonica* и при сравнении 3 фосфолипаз *Bitis* с одной фосфолипазой *N. n. oxiana*. Напротив, профиль изменчивости последовательностей видов *Naja* достаточно чувствителен к количественному и видовому составу группы. Исключение из группы последовательностей двух фосфолипаз вида *N. melanoleuca* понижает значение общей неравномерности до 1,5σ и делает профиль изменчивости еще менее рельефным.

Наличие рельефных межгрупповых профилей изменчивости с высоким значением общей неравномерности оказалось характерным и для пяти других исследованных семейств гомологичных белков: бактериальных и митохондриальных аминотрансфераз [2], родопсинов человека [4, 5], α-субъединиц Na⁺, K⁺-АТФ-азы рыбы [6] и млекопитающих [7–9], L- и M-субъединиц фотореакционного центра пяти фотосинтезирующих бактерий [10–16]. Например, общая неравномерность профиля межгрупповой изменчивости митохондриальных аспаратаминотрансфераз трех млекопитающих (первая группа) и аспаратаминотрансферазы из *E. coli* (вторая группа) составляет 2,5σ. В профиле присутствуют 6 пиков и 4 впадины большой интенсивности, отвечающие переменным и консервативным участкам белковых последовательностей. Более подробные результаты изучения профилей изменчивости остальных семейств гомологичных белков будут сообщены в отдельной статье.

Авторы выражают признательность И. Б. Флоровой (биофак МГУ) за обсуждение некоторых результатов настоящей работы.

СПИСОК ЛИТЕРАТУРЫ

1. Мецеракова Е. А., Айанян А. Е., Костецкий П. В., Мирошников А. И. // Биоорганическая химия. 1982. Т. 8. № 3. С. 349–363.
2. Sidman K. E., George D. C., Barker W. C., Hunt L. T. // Nucl. Acids Res. 1988. V. 16. № 5. P. 1869–1871.

3. Овчинников Ю. А., Мирошников А. И., Назимов И. В., Ансалон У. Р., Солдатов Л. Н. // Биоорг. химия. 1979. Т. 5. № 6. С. 805–813.
4. Nathans J., Hogness D. S. // Proc. Nat. Acad. Sci. USA. 1984. V. 81. № 15. P. 4851–4855.
5. Nathans J., Hogness D. S. // Science. 1985. V. 232. № 4747. P. 203–210.
6. Kawakami K., Noguchi S., Noda M., Takahashi H., Ohta T., Kawamura M., Nojima H., Nagano K., Hirose T., Inayama S., Nayashida H., Miyata T., Numa S. // Nature. 1985. V. 316. № 6030. P. 733–736.
7. Shull G. E., Schwartz A., Lingrel J. B. // Nature. 1985. V. 316. № 6030. P. 891–895.
8. Овчинников Ю. А., Модянов Н. Н., Бродье Н. Е., Петрушкин К. Е., Гришин А. В., Арзамасова Н. М., Алданова Н. А., Монастырская Г. С., Свердлов Е. Д. // FEBS Lett. 1986. V. 201. № 2. P. 237–245.
9. Kawakami K., Ohta T., Nojima H., Nagano K. // J. Biochem. Tokyo. 1986. V. 100. № 2. P. 389–398.
10. Овчинников Ю. А., Абдулаев Н. Г., Золотарев А. С., Шмуклер Б. Е., Заргаров А. А., Кутузов М. А., Тележинская И. Н., Левина Н. Б. // FEBS Lett. 1988. V. 231. № 1. P. 237–242.
11. Овчинников Ю. А., Абдулаев Н. Г., Шмуклер Б. Е., Заргаров А. А., Кутузов М. А., Тележинская И. Н., Левина Н. Б., Золотарев А. С. // FEBS Lett. 1988. V. 232. № 2. P. 364–368.
12. Belanger G., Berard J., Corriveau P., Gingras G. // J. Biol. Chem. 1988. V. 263. № 16. P. 7632–7638.
13. Youvan D. C., Bylina E. J., Alberti M., Begush H., Hearst J. E. // Cell. 1984. V. 37. № 6. P. 949–957.
14. Williams J. C., Steiner L. A., Odgen R. C., Simon M. I., Feher G. // Proc. Nat. Acad. Sci. USA. 1983. V. 80. № 21. P. 6505–6509.
15. Williams J. C., Steiner L. A., Feher G., Simon M. I. // Proc. Nat. Acad. Sci. USA. 1984. V. 81. № 23. P. 7303–7307.
16. Michel H., Weyer K. A., Gruenberg H., Dunger I., Oesterhelt D., Lottspeich F. // EMBO J. 1986. V. 5. № 6. P. 1149–1158.

Поступило в редакцию
8.VI.1989

A METHOD OF A GRAPHIC IDENTIFICATION OF VARIABLE AND CONSTANT REGIONS IN HOMOLOGOUS PROTEINS SEQUENCES

KOSTETSKY P. V. VLADIMIROVA R. R.

*M. M. Shemyakin Institute of Bioorganic Chemistry, Academy of Sciences
of the USSR, Moscow*

An algorithm is presented for localizing variable and constant regions in homologous protein sequences. A set of aligned protein sequences is divided into two groups consisting of m and n sequences. Each group contains sequences of most related species. Value of the position dissimilarity of proteins from different groups of m and n sequences is defined as a number of failures to coincide in comparison with all possible $m \times n$ pairs of amino acid residues in the position (each from different group) divided by $m \times n$. The position dissimilarity value of m protein sequences within a group is defined as the number of failures to coincide in comparison with all possible $m \times (m-1)/2$ pairs of amino acid residues divided by $m \times (m-1)/2$. Ten position average of dissimilarity values is plotted vs. the first position number. Area of the figure included between the profile of dissimilarity values and its mean value line characterizes the overall irregularity of amino acid substitutions along the protein sequences. If the area value is greater than the average area for 1000 random profiles by more than two standard deviation units, the profile extrema containing the «surplus» of area are cut off. The cut off stretches are likely to be variable and constant regions.

In case of «between groups» comparisons it is found that the overall irregularity of amino acid substitutions is very high for all considered families of proteins; phospholipases A₂, aspartat – a minotransferases, alpha – subunits of Na⁺, K⁺-ATPase, L- and M-subunits of photosynthetic bacteria photoreaction centre, human rhodopsins.