

ПРЕДПОЧТИТЕЛЬНОСТЬ ResA-ФИЛАМЕНТА
К ПОСЛЕДОВАТЕЛЬНОСТЯМ ДНК КОРРЕЛИРУЕТ
С ГЕНЕТИЧЕСКИМ КОДОМ

© 1998 г. М. П. Пономаренко, Ю. В. Пономаренко, И. И. Титов,
Н. А. Колчанов, А. В. Мазин, С. Ч. Ковальчиковски

Представлено академиком В.К. Шумным 15. 12. 97 г.

Поступило 30. 12. 97 г.

Белок ResA является ключевым для гомологичной рекомбинации *E. coli* [1]. Связывание ResA с односторонней ДНК формирует ResA-филамент, виток спирали которого составляют 6 мономеров ResA и 18 нуклеотидов ДНК. Он осуществляет обмен нитей ДНК в гомологичной рекомбинации [2]. Обнаружение предпочтительности ResA-филамента к контексту ДНК [3] потребовало выявления особенностей ДНК, определяющих эту предпочтительность.

В настоящей работе с помощью компьютерной системы ACTIVITY [4] исследованы данные о сродстве ResA-филамента к ДНК [3]. Установлено, что это сродство убывает с ростом концентрации тринуклеотидов DRV = {AAA, AAC, AGA, AGC, GAA, GAC, GGA, GGC, TAA, TGA, AAG, AGG, TAG, TGG, GAG, GGG, TAC, TGC}. Показано, что тринуклеотиды DRV кодируют аминокислоты, формирующие поверхности белковых глобул с их функциональными сайтами, и не кодируют аминокислоты из глобулярных ядер белков. Это согласуется с общеизвестными консервативностью функциональных сайтов белков и блочными перестройками ядер белковых глобул.

Исследованные данные о сродстве ResA-филамента к ДНК [3] приведены в табл. 1. Они содержат 16 фрагментов ДНК длиной 32 нуклеотида $S = s_1 \dots s_{32}$ с величинами сродства ResA-филамента к ним $F(S)$ от -5.01 до 1.20 логарифмических единиц. Поскольку в ResA-филаменте на один ResA приводится 3 нуклеотида [3], то изучались взвешенные концентрации тринуклеотидов $Z = z_1 z_2 z_3$:

$$X_{Zw}(S) = \sum_{i=1}^{L-2} w(i) \times \Delta(s_i \in z_1) \times \Delta(s_{i+1} \in z_2) \times \Delta(s_{i+2} \in z_3), \quad (1)$$

здесь: $z \in \{A, T, G, C, W = A/T, R = A/G, M = A/C, K = T/G, Y = T/C, S = G/C, B = T/G/C, V = A/G/C, H = A/T/C, D = A/T/G, N = A/T/G/C\}$; $\Delta(\text{истина}) = 1$, $\Delta(\text{ложь}) = 0$; $w(i)$ – весовая функция, моделирующая вклад тринуклеотида Z в позиции i последовательности S в сродство ResA-филамента: “больше $w(i)$ ” \Leftrightarrow “больше вклад”.

На рис. 1 даны функции $w(i)$, моделирующие наибольший вклад в сродство ResA-филамента к ДНК для тринуклеотидов на 5'-конце этой ДНК (непрерывная линия) и в позиции 19 (штриховая линия), удаленной от 5'-конца ДНК на 18 нуклеотидов, составляющих виток спирали ResA-филамента [3]. Всего анализировалось 180 функций $w(i)$. Их комбинирование со всеми тринуклеотидами $Z = z_1 z_2 z_3$ дает $15^3 \times 180 \approx 10^6$ концентраций X_{Zw} . Чтобы концентрация $X_{Zw}(S_n)$ определяла

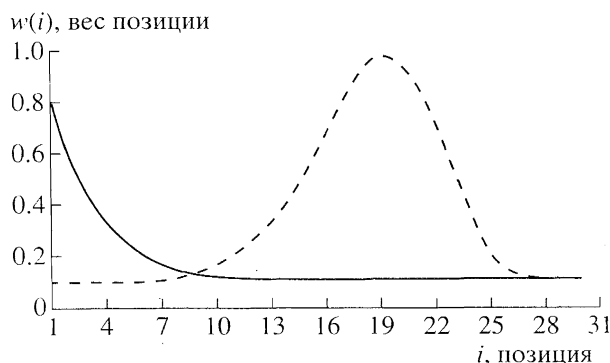


Рис. 1. Примеры весовых функций $w(i)$, моделирующих наибольший вклад в сродство ResA-филамента к ДНК длиной 32 нуклеотида для тринуклеотидов на 5'-конце (непрерывная линия) и в позиции 19 (штриховая линия), удаленной от 5'-конца ДНК на 18 нуклеотидов, составляющих виток спирали ResA-филамента [3].

Институт цитологии и генетики
Сибирского отделения Российской Академии наук,
Новосибирск
Калифорнийский университет, Дэвис, США

Таблица 1. Сродство ResA-филамента к одонитевым ДНК [3]

№ п. п.	Вариант	Последовательность ДНК, S_n	Сродство F_n
1	A > T	CCTCCCGCTTTTTGTCCCTCTTTCTTTTGGT	1.20
2	dC	CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	0.54
3	#40	ACCACCACACACGCGCACACCACCACACACGC	0.48
4	htr#3	TTCACAAACGAATGGATCCTCATTAAGCCAG	0.34
5	#39	GCGTGTGTGGTGGTGTGCGCGTGTGGTGGT	0.33
6	dT	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	0.09
7	G > C	CCATCCCCAAAAATCACCTCTTATCAAAACCA	0.03
8	IDENT	CCATCCGCAAAAATGACCTCTTATCAAAAGGA	0.00
9	htr#4	CATGGAGCAGGTCGCGGATTCGACACAATTT	-0.02
10	G > T	CCATCCTCAAAAATTACCTCTTATCAAAATTA	-0.40
11	C > G	GGATGGGGAAAAATGAGGTGTTATGAAAAGGA	-1.00
12	C > T	TTATTTGTAAAAATGATTTTTTATTAAGGA	-1.20
13	#7	GGCGGGCGGCGCGGCCGGGCGGCGGGCGCGCG	-1.99
14	htr#2	AATTCTTCGAAGCTAGCCCTCAGGCCTAGGCA	-2.42
15	C > A	AAATAAGAAAAAATGAAATATTATAAAAAGGA	-3.40
16	dA	AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	-5.01

сродство F_n ResA-филамента к ДНК S_n , все пары $\{X_{Zw}(S_n), F_n\}$ должны отвечать требованиям регрессии:

$$F_{Zw}(S_n) = a + b \times X_{Zw}(S_n); \quad (2)$$

здесь a, b – коэффициенты регрессии [5] для $\{X_{Zw}(S_n), F_n\}$.

Формула (2) предсказывает сродство ResA-филамента к ДНК F_{Zw} по концентрации X_{Zw} . Для ее применения требуется соответствие предсказанного сродства F_{Zw} известному F . Всего проверяется 22 таких соответствия, в том числе линейная корреляция F_{Zw} и F , нормальность отклонений $\{F_{Zw} - F\}$. Для m -го соответствия [6] оценивается уровень его значимости α_m ($1 \leq m \leq 22$). Значимые соответствия ($\alpha_m < 0.05$) между F_{Zw} и F дают концентрации X_{Zw} , в терминах нечеткой логики Заде [7], положительную оценку $U_m(X_{Zw}, F) > 0$; незначимые – отрицательную. Всего получается 22 оценки $\{U_m(X_{Zw}, F)\}$, усреднение которых, в терминах теории принятия решений [8], дает полезность концентрации X_{Zw} для предсказания сродства ResA-филамента к ДНК, F :

$$U(X_{Zw}, F) = \sum_{m=1}^{22} U_m(X_{Zw}, F)/22. \quad (3)$$

Согласно формуле (3) полезность $U(X_{Zw}, F)$ концентрации X_{Zw} растет с числом значимых соответствий между предсказанным F_{Zw} и известным F сродством. Отрицательная полезность $U(X_{Zw}, F) < 0$ означает, что с помощью концентрации X_{Zw} нельзя предсказывать сродство ResA-фи-

ламента к ДНК: большинство соответствий между F_{Zw} и F не являются достоверными. Если полезность $U(X_{Zw}, F) > 0$ положительна, то по концентрации X_{Zw} можно предсказывать это сродство: большинство соответствий между F_{Zw} и F достоверны. Согласно биномиальному критерию [6] вероятность случайного появления среди 10^6 концентраций X_{Zw} одной X с $U(X, F) > 0$:

$$p[U(X, F) > 0] = 10^6 \times \sum_{m=12}^{22} C_{22}^m \times 0.05^m \times 0.95^{22-m} < 10^{-4}.$$

При этом наибольшая $U(X_{Zw}, F) > 0$ указывает концентрацию X_{Zw} , которая определяет предпочтительность ResA-филамента к ДНК.

Анализировались 10 ДНК из табл. 1 (№ 2, 3, 4, 5, 6, 8, 9, 13, 14 и 16), остальные 6 ДНК были контрольными. Для 10 проанализированных ДНК S_n вычислялись (1) все 10^6 возможных концентраций $X_{Zw}(S_n)$. По известному сродству ResA-филамента к ДНК F_n были построены регрессии (2) для предсказания $F_{Zw}(S_n)$ по $X_{Zw}(S_n)$. Сравнение $F_{Zw}(S_n)$ с F_n (формула (3)) дало полезности $U(X_{Zw}, F)$ концентрации X_{Zw} . Всего получилось 10^6 полезностей $U(X_{Zw}, F)$, лишь пять из них оказались положительными (табл. 2).

Наибольшей (X_{DRVw}, F) = 0.270 была полезность тринуклеотидов $DRV = \{AAA, AGA, TAA, TGA, GAA, GGA, AAG, AGG, TAG, TGG, GAG, GGG, AAC, AGC, TAC, TGC, GAC, GGC\}$, взвешенных функцией $wl(i)$ с максимумом на 5'-конце ДНК (рис. 1, непрерывная линия). Это означает,

Таблица 2. Выявленные концентрации тринуклеотидов X_{Zw} с положительной полезностью $U(X_{Zw}, F)$ для предсказания сродства RecA-филамента к одонитовой ДНК

№ п. п.	Выявленная концентрация тринуклеотидов X_{Zw}			Взаимосвязь с лучшим тринуклеотидом
	тринуклеотид Z	весовая функция $w(i)$ (см. рис. 1)	полезность $U(X_{Zw}, F) \pm$ ст. откл.	
1	DRV	Непрерывная линия	0.270 ± 0.011	Лучший
2	RVD	То же	0.229 ± 0.021	Циклическая перестановка
3	VDR	»	0.170 ± 0.033	
4	RRV	Пунктир	0.191 ± 0.033	
5	RRM	То же	0.180 ± 0.012	Частный случай

Таблица 3. Проекция тринуклеотида DRV на генетический код

Аминокислотный остаток		Генетический код (DRV)	Количество	
название	код		кодоны	DRV
Аланин	A	GCG GCA GCT GCC	4	
Аргинин	R	AGG AGA CGG CGA CGT CGC	6	2
Аспарагин	N	AAT AAC	2	1
Аспарагиновая кислота	D	GAT GAC	2	1
Цистеин	C	TGT TGC	2	1
Глютамин	Q	CAG CAA	2	
Глютаминовая кислота	E	GAG GAA	2	2
Глицин	G	GGG GGA GGT GGC	4	3
Гистидин	H	CAT CAC	2	
Изолейцин	I	ATA ATT ATC	3	
Лейцин	L	TTG TTA CTG CTA CTT CTC	6	
Лизин	K	AAG AAA	2	2
Метионин	M	ATG	1	
Фенилаланин	F	TTT TTC	2	
Пролин	P	CCG CCA CCT CCC	4	
Серин	S	AGT AGC TCG TCA TCT TCC	6	1
Треонин	T	ACG ACA ACT ACC	4	
Валин	V	GTG GTA GTT GTC	4	
Триптофан	W	TGG	1	1
Тирозин	Y	TAT TAC	2	1

что сродство RecA-филамента к ДНК определяется концентрацией тринуклеотидов DRV вблизи ее 5'-конца. Из табл. 2 видно, что два других тринуклеотида с положительными полезностями являются циклическими перестановками нуклеотидов (RVD и VDR) лучшего тринуклеотида DRV, взвешенными той же функцией $w1(i)$, а два оставшихся тринуклеотида RRV и RRM являются частными случаями DRV, взвешенными функцией $w2(i)$ с максимумом в позиции 19 (рис. 1, штриховая линия), удаленной от 5'-конца ДНК на 18 нуклеотидов, составляющих один виток спирали RecA-филамента [3]. Все это указывает на важ-

ность тринуклеотидов DRV для сродства RecA-филамента к ДНК. Простой регрессией для предсказания этого сродства была формула

$$F_{DRVw1}(S) = 0.54 - 1.03 \times X_{DRVw1}(S). \quad (4)$$

Предсказанное по формуле (4) сродство RecA-филамента к ДНК показано на рис. 2: светлые кружки для 10 проанализированных ДНК, темные для 6 контрольных ДНК. На контрольных данных коэффициент линейной корреляции между предсказанным F_{DRVw1} и известным F сродством был $r = 0.812$ ($\alpha < 0.05$). Таким образом, формула (4) достоверно предсказывает сродство RecA-фила-

Таблица 4. Соответствие между известными свойствами аминокислот [9, 10] и тринуклеотидами DRV

Свойства аминокислот	Аминокислоты		Число кодонов				Значимость*
	тринуклеотид DRV		DRV		другие		
	есть	нет	есть	нет	есть	нет	
DRV кодирует							
случайный клубок [9]	YDNEKG	AV	10	5	12	34	0.010
поверхность [10]	DEKNR	HQ	8	7	10	36	0.025
заряженные [10]	DEKR	H	7	8	7	39	0.025
DRV не кодирует							
ядро глобулы [9]	—	LIMFV	0	15	16	30	0.005

* Значимость α оценивалась по точному критерию Фишера [6].

мента к ДНК на независимых контрольных данных, что однозначно указывает на важность тринуклеотидов DRV для этого средства.

Поскольку большая часть генома *E. coli* кодирует белки, то было весьма интересно сравнить тринуклеотид DRV, определяющий средство ResA-филамента к ДНК, с генетическим кодом (табл. 3). Тринуклеотиды DRV совпадают с кодонами аргинина, аспарагина, лизина, глицина, цистеина, серина, тирозина аспарагиновой и глютаминовой кислот, триптофана. Между свойствами аминокислот [9, 10] и тринуклеотидами DRV имеется четыре достоверных соответствия (табл. 4): тринуклеотиды DRV являются кодонами аминокис-

лот, формирующих случайный клубок ($\alpha < 0.01$), поверхность ($\alpha < 0.025$) и заряд ($\alpha < 0.025$) белка, и не являются кодонами аминокислот из ядра глобулы ($\alpha < 0.005$).

Таким образом, между контекстной предпочтительностью ResA-филамента к ДНК и генетическим кодом имеется корреляция: ResA-филамент имеет повышенное средство к DRV-бедным районам кодирования в генах белков ядер их глобул и пониженное средство к DRV-богатым районам кодирования в генах белков их поверхностей и функциональных сайтов. Этот результат согласуется с общеизвестным фактом, что в ходе эволюции белков их глобулярные ядра подвергаются блочным перестройкам, тогда как их функциональные сайты остаются консервативными.

Настоящую работу поддерживали Российский фонд фундаментальных исследований и ГНТП "Геном человека".

СПИСОК ЛИТЕРАТУРЫ

1. Cox M.M. // BioEssay. 1993. V. 15. P. 617–623.
2. West S.C. // Cell. 1994. V. 76. P. 9–15.
3. Mazin A.V., Kowalczykowski S.C. // PNAS USA. 1996. V. 93. P. 10673–10678.
4. Ponomarenko M.P. et al. // J. Comput. Biol., 1997. V. 4. P. 83–90.
5. Forster E., Ronr B. Methoden der Korrelations- und Regressionsanalyse B.: Verlag Die Wirtschaft, 1979.
6. Lehman E.L. Testing Statistical Hypotheses. N.Y.: Wiley, 1959.
7. Zadeh L.A. // Information and Control. 1965. V. 8. P. 338–353.
8. Fishburn P.C. Utility Theory for Decision Making. N.Y.: Wiley, 1970.
9. Karlin S. et al. Mathematical Methods for DNA Sequences. Boca Raton: CRC Press, 1989. P. 133.
10. Cohen B.I. et al. // Meth. in Enzimol. 1991. V. 202. P. 252–278.

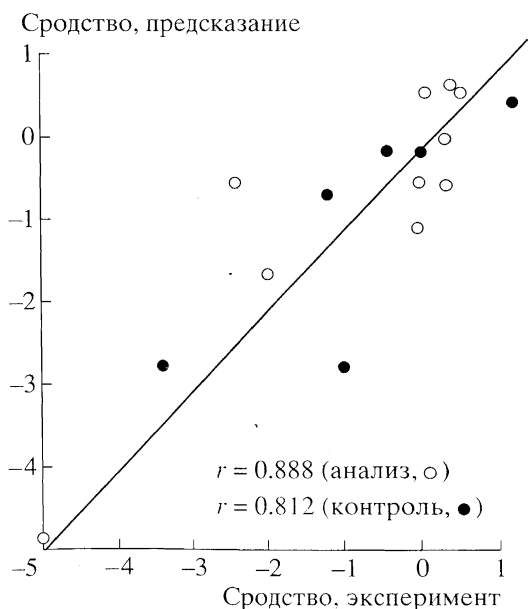


Рис. 2. Сравнение предсказанных (формула (4)) величин средства ResA-филамента к ДНК с его экспериментальными значениями [3]: светлые кружки для 10 проанализированных ДНК, темные кружки для 6 контрольных ДНК.