

Член-корреспондент АН СССР М. В. ВОЛЬКЕНШТЕЙН

КОЛИЧЕСТВО И ЦЕННОСТЬ ИНФОРМАЦИИ, СОДЕРЖАЩЕЙСЯ В КОДОНАХ

Количество информации (далее именуемое информацией) в любом из 61 нуклеотидных триплетов, кодирующих включение аминокислот в белке, определяемое безотносительно к кодированию аминокислотного остатка, равно $\log_2 61 = 5,93$ бита. Информация канонического аминокислотного остатка равна $\log_2 20 = 4,32$ бита. Информация кодона применительно к процессу кодирования или, что то же самое, информация остатка, вычисляемая с учетом вырожденности кода, различна для различных кодонов или для различных остатков. Она равна $\log_2(61/n)$, где n — степень вырождения кодона, принимающая значения 1, 2, 3, 4, 6. Соответственно получаем 5,93; 4,93; 4,35; 3,93; 3,35 бита. Кодоны АУГ (Мет) и УГГ (Три), будучи невырожденными, дают наибольшую информацию 5,93 бита. Наименьшая информация 3,35 бита отвечает шестикратно вырожденным триплетам, кодирующим Арг, Сер и Лей. Средняя информация, приходящаяся на кодон или на аминокислотный остаток, вычисляется по формуле Шеннона и составляет

$$-\sum_{i=1}^{20} \frac{n_i}{61} \log_2 \frac{n_i}{61} = 4,14 \text{ бита.}$$

Естественно, что эта величина меньше $\log_2 20$, так как кодовый словарь определяет неодинаковые вероятности появления различных остатков, благодаря различному их вырождению. Отметим попутно, что средняя информация, приходящаяся на букву русского текста, равна 4,35 бита, на букву английского текста 4,03 бита ⁽¹⁾, — значения, весьма близкие к найденным для «аминокислотного текста».

Эти оценки небезынтересны, но мало что дают биологии. Принято говорить об «антиэнтропийности» биологических систем. Истинный смысл «антиэнтропийности» определяется невысокой упорядоченностью биологических систем, начиная с молекул белков и кончая организмами. Упорядочены и кристаллы. Суть деда заключается в том, что биологические системы являются не статистическими, но неравновесными динамическими, «машинными» системами. Соответственно усредненная статистическая характеристика — энтропия или эквивалентная ей информация недостаточны для описания поведения биологических систем. Для их теоретического анализа, прежде всего для рассмотрения проблем развития, необходимо понятие ценности, качества или содержания информации. Одним из первых на это обратил внимание И. И. Шмальгаузен ⁽²⁾. Очевидно, что рассмотрение ценности информации возможно лишь при наличии в системе некоторого количества информации. Информация может иметь ценность, если система обладает «информационной структурой», т. е. является определенным образом упорядоченной в пространстве и (или) во времени. В биологии информационная структура неоднородна и аperiodична. Простейший пример информационной структуры — первичная структура биополимера.

В отличие от количества информации, ее ценность может быть определена лишь применительно к конкретному физическому процессу рецессии и переработки информации. Поэтому универсальное определение ценности информации невозможно. Тем не менее конкретные количественные выражения ценности информации оказываются полезными для теоретической биологии.

Строя модельную теорию самоорганизации, селекции и эволюции информационных макромолекул, Эйген ввел величину «селективной ценности», выражаемую через кинетические параметры ⁽³⁾. Селективная ценность, однако, не эквивалентна ценности информации, содержащейся в цепи нуклеиновой кислоты или белка, так как она не учитывает первичную структуру всей цепи или достаточно большого ее отрезка. Несмотря на справедливость конечных выводов Эйгена, его теория требует частичной модификации ⁽⁴⁾. Более поздняя работа Куна ⁽⁵⁾ оперирует в неявной форме именно понятием ценности информации биологической макромолекулы.

Ценность информации, содержащейся в нуклеотидном триplete, определяется применительно к кодированию аминокислотного остатка, к программированию синтеза белка. Ценность связана с биологической значимостью мутации, состоящей в замене данного нуклеотида другими. Ценность нуклеотида зависит от его положения в кодоне *xuz*. При замещениях *z* в 61 значимом кодоне только 28% приводят к изменению аминокислотного остатка, при замещениях *x* остаток меняется в 95% и при замещениях в *y* — в 100% случаев. Следовательно, информационная ценность нуклеотида меньше всего для *z* и больше всего для *y*.

Мутационные замены аминокислотных остатков неравнозначны в смысле влияния на пространственную структуру белка и на его биологическую функцию. Наиболее опасны мутации, изменяющие класс остатка, в частности, — заменяющие гидрофобный остаток на гидрофильный и наоборот. Ранее было показано, что кодовый словарь обеспечивает вдвое меньшую вероятность таких опасных мутаций, чем вероятность мутаций, не сопровождающихся изменением класса остатка ⁽⁶⁻⁸⁾. В действительности разделение остатков на гидрофобные и гидрофильные условно, необходимо характеризовать остаток количественной мерой гидрофобности. В качестве таковой Генфорд предложил величину изменения свободной энергии, приходящуюся на боковую группу свободной аминокислоты, при переносе аминокислоты из этанола в воду ^(9, 10). Эти значения (в ккал/моль) указаны в табл. 1.

Будем считать, что опасность мутации для биологической функции белка тем больше, чем больше разность гидрофобностей исходного остатка и остатка, получаемого в результате замещения. Средняя разность гидрофобностей при произвольном замещении любого остатка на любой, без учета кодирования, равна 1,28 ккал/моль. Средние разности гидрофобностей при однократных замещениях нуклеотидов в кодоне *xuz* равны для *x* 1,00, для *y* 1,28 и для *z* 0,34 ккал/моль. В среднем по всем трем нуклеотидам 0,87, что существенно меньше 1,28 ккал/моль. Для 70 мутантов человеческого гемоглобина имеем 0,83, для шести цитохромов с разных видов 0,90 и т. д. Код уменьшает разность гидрофобностей при мутациях.

Определим относительную ценность информации кодона как сумму средних разностей гидрофобностей, возникающих при всех однократных замещениях *xuz*. Наиболее ценен кодон, мутации которого сильнее всего меняют гидрофобности аминокислотных остатков. Таковым оказывается УГГ, кодирующий Met. В табл. 2 приведены количества информации в битах (первое число) и ценности информации в ккал/моль (второе число) осмысленных кодонов. Ясно, что ценность информации, определенная таким образом, не коррелирует с ее количеством. Триплеты, кодирующие одну и ту же аминокислоту, могут иметь различную ценность. Можно думать, что в ходе эволюции кода ценность информации кодона возростала.

Таблица генетического кода *

Таблица 1

x	y				z
	А	Ц	Г	У	
А	Лиз 1,50	Тре 0,44	Арг 0,73	Иле 2,97	А Ц Г У
	Асп -0,01		Сер 0,04	Иле 2,97	
	Лиз 1,50		Арг 0,73	Мет 1,30	
	Асп -0,01		Сер 0,04	Иле 2,97	
Ц	Гли -0,10	Про 2,60	Арг 0,73	Лей 2,42	А Ц Г У
	Гис 1,40				
	Гли -0,10				
	Гис 1,40				
Г	Глу 0,55	Ала 0,73	Гли 0,00	Вал 1,69	А Ц Г У
	Асп 0,54				
	Глу 0,55				
	Асп 0,54				
У	Тир —	Сер 0,04	—	Лей 2,42	А Ц Г У
	Тир 2,87		Цис 0,65	Фен 2,65	
	Тир —		Три 3,00	Лей 2,42	
	Тир 2,87		Цис 0,65	Фен 2,65	

* Числа — гидрофобности аминокислот по Тенфорду ккал/моль.

Таблица 2

Количества информации в битах (первая цифра) и ценности информации в ккал/моль (вторая цифра) кодонов

x	y				z
	А	Ц	Г	У	
А	4,9 3,5	3,9 2,6	3,4 1,9	4,4 3,4	А Ц Г У
	4,9 3,9	3,9 2,1	3,4 2,0	4,4 4,1	
	4,9 2,9	3,9 1,7	3,4 2,0	5,9 3,1	
	4,9 3,9	3,9 2,1	3,4 2,0	4,4 4,1	
Ц	4,9 4,2	3,9 3,8	3,4 1,8	3,4 1,8	А Ц Г У
	4,9 3,2	3,9 3,3	3,4 1,9	3,4 1,5	
	4,9 4,2	3,9 3,8	3,4 2,5	3,4 2,1	
	4,9 3,2	3,9 3,3	3,4 1,9	3,4 1,5	
Г	4,9 1,4	3,9 1,6	3,9 1,7	3,9 2,2	А Ц Г У
	4,9 1,6	3,9 1,6	3,9 1,4	3,9 2,3	
	4,9 1,4	3,9 1,6	3,9 2,5	3,9 1,9	
	4,9 1,6	3,9 1,6	3,9 1,4	3,9 2,3	
У	— —	3,4 3,6	— —	3,4 3,0	А Ц Г У
	4,9 4,0	3,4 3,2	4,9 3,2	4,9 2,3	
	— —	3,4 3,9	5,9 6,6	3,4 2,3	
	4,9 4,0	3,4 3,2	4,9 3,2	4,9 2,3	

Согласно модельной теории Куна, первоначальное кодирование в добиологической эволюции определялось только нуклеотидом у⁽⁵⁾.

Конечно, приведенное определение ценности информации кодона условно. В ряде случаев любая замена некоторых аминокислотных остатков в белке оказывается летальной в том смысле, что полностью нарушает его биологическую функцию. В этих случаях понятие ценности информации подлежит дальнейшей конкретизации и уточнению. Сказанное относится

и к мутациям части кодонов, которые приводят к кодонам УАА, УАГ и УГА, определяющим терминацию цепи. В сущности, подход, основанный на определении ценности информации, должен применяться к процессу биосинтеза белка в целом, с учетом роли всех участвующих в этом процессе факторов. Так или иначе, есть основания думать, что определение ценности информации и анализ, исходящий из такого определения, полезны для рассмотрения мутагенеза и молекулярной биологической эволюции.

Институт молекулярной биологии
Академии наук СССР
Москва

Поступило
15 III 1974

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- ¹ А. М. Яглом, И. М. Яглом, Вероятность и информация, «Наука», 1973. ² И. И. Шмальгаузен, Кибернетические вопросы биологии, «Наука», 1968. ³ М. Эйген, Самоорганизации материи и эволюция биологических макромолекул, М., 1973. ⁴ М. В. Волькенштейн, УФН, т. 109, 499 (1973). ⁵ Н. Кuhn, Angew. Chem., v. 84, 838 (1972). ⁶ М. В. Волькенштейн, Генетика, № 2, 54 (1965); № 4, 119 (1966); Biochim. et biophys. acta, v. 119, 421 (1966). ⁷ М. В. Волькенштейн, Ю. Б. Румер, Биофизика, т. 12, 10 (1967). ⁸ М. В. Волькенштейн, Физика ферментов, «Наука», 1967. ⁹ С. Tanford, J. Am. Chem. Soc., v. 84, 4240 (1962); v. 86, 2050 (1964). ¹⁰ P. Dunnill, Biophys. J., v. 8, 865 (1968).