

КОНЦЕПТУАЛЬНЫЙ ПОДХОД К КЛАССИФИКАЦИИ МУТАЦИЙ В ГОМОЛОГИЧНЫХ БЕЛКАХ

Анализ первичных структур гомологичных белков дает представление об области макромолекулы, необходимой для осуществления ее основной функции, и содержит сведения о путях биологической эволюции. Введение понятия «концепт-белок», а также описание процесса точечных мутаций на языке алгебры логики позволили создать алгоритм классификации мутаций в гомологичных белках и выявить так называемые «дополнительные полипептиды», которые либо еще не обнаружены, либо уже не существуют, либо по каким-то причинам не могут существовать.

Введение. Как известно, гомологичные белки — это белки, выполняющие сходные неспецифические функции в различных организмах независимо от сложности внутреннего строения этих организмов и принадлежности к одному биологическому виду или даже классу.

Степень структурной близости гомологичных белков определяет их принадлежность к родственным видам, а степень различия в структуре и характере этих различий могут, по-видимому, свидетельствовать о пути, которым шла эволюция видов. Изменения в первичной структуре белков, в результате которых могут возникать новые виды, эволюционировать или исчезать старые, явились результатом спонтанных мутаций, порождающих вставки, делеции или замещения аминокислот. Изучение вопроса о том, какие из этих изменений были закреплены в процессе эволюции, что повлекли за собой в организме как целостной системе, чем они вызваны и как отражаются на матричном уровне в структуре нуклеиновых кислот, — все эти вопросы представляют большой интерес не только как способ познания истины, но и как средство, которое может быть использовано в последующем для целенаправленного изменения свойств организмов. Этому могут послужить средства генной инженерии, получившие широкое распространение в последнее время.

Хорошо известно также, что как при спонтанном мутационном процессе, так и при индуцированном мутагенезе генные или точечные мутации способны затрагивать область из одной или нескольких пар оснований. Эти мутации могут быть следствием замены, вставок или делеций пар оснований в структурной части гена, в результате чего образуется аномальный белок либо с заменой аминокислотных остатков, либо с нарушенной их последовательностью. Причиной точечных мутаций также могут быть транзиции, трансверсии, делеции или вставки на нуклеотидном уровне, происходящие спонтанно или вызванные химическими мутагенами, в частности, HNO_2 , ЛСД или коротковолновым излучением — ультрафиолетовым, рентгеновским, γ -лучами.

В настоящей работе рассматриваются точечные мутации, порождающие только транзиции и трансверсии, и предлагается способ (метод) классификации таких мутаций в зависимости от того, на каком месте в триплете произошла мутация.

Концептуальный подход. При построении моделей различных объектов молекулярной генетики нами используются концепция и подходы,

* Продолжение публикаций тематического выпуска. Начало см. № 6, 1990

развитые в работах по искусственному концептуальному интеллекту [1, 2], а также модели и гипотезы по использованию идей концептуализации, разработанные в целях построения так называемых «концепт-белков» [3—6].

Согласно этому подходу, понятие об объекте исследования может быть обобщено в отношении целостного восприятия одновременно нескольких объектов одного класса, и этот класс объектов, объединенных общими свойствами, рассмотрен как концепт-объект, как некий собирательный образ, который можно назвать «супергомобелок».

Для построения концепт-объекта гомологичных неспецифических белков у разных видов необходима формализация знаний. Так, для построения концепт-объектов белка и нуклеиновых кислот были использованы методы формальной логики и теории предикатов [7, 8]. В этих работах свойства нуклеотидов, входящих в состав ДНК, представляются следующим образом: \bar{P}_1 — свойства пуринов (А, G) быть длинными, т. е. иметь длину 0,7 нм; P_1 — свойство пиримидинов (С, Т(U)) — быть короткими, т. е. иметь длину 0,5 нм. Свойства нуклеотидов в двойной спирали образуют три водородные связи — \bar{P}_2 , таким «свойством» обладают G и C; свойство же \bar{P}_2 — присуще тем нуклеотидам, которые соединены в двойной спирали ДНК двумя водородными связями. Это А и Т. То же свойство \bar{P}_2 приписывается основанию U. Введение такого обобщения свойств нуклеотидов позволяет формализовать запись пяти оснований следующим образом:

$$\begin{aligned} G - P_1 \cdot P_2 = Q_1; \quad C - \bar{P}_1 \cdot P_2 = Q_3; \\ A - P_1 \cdot P_2 = Q_2; \quad T(U) - P_1 \cdot P_2 = Q_4. \end{aligned}$$

Каждое основание представляется в виде двухместного предиката Q_i , $i = \bar{1}, 4$. Используя эти обозначения, можем представить первичную структуру нуклеиновых кислот в виде конъюнкции предикатов:

$$Q_1 \& Q_2 \& Q_3 \& Q_4 \dots,$$

образованной четырьмя структурными единицами. В аналогичной форме может быть представлена и первичная структура белка с той лишь разницей, что структурных единиц будет двадцать:

$$B^* \& H \& P \& M \& M \& F \& K \dots,$$

где каждой аминокислоте соответствует один или несколько вектор-столбцов $\begin{pmatrix} Q_i \\ Q_j \\ Q_k \end{pmatrix}$, $i, j, k = \bar{1}, 4$.

Вырожденность генетического кода будет изображаться дизъюнкцией вектор-матриц соответствующих триплетов, т. е. объединением триплетов, являющихся кодом для данной аминокислоты [3].

Учитывая свойство коммутативности дизъюнкций, результат объединения элементов матрицы можно записать так:

$$\begin{aligned} Q_1 \vee Q_2 = P_1 \quad Q_1 \vee Q_4 = R \quad R \vee \bar{R} = \check{R} \\ Q_3 \vee Q_4 = \bar{P}_1 \quad \underline{Q}_2 \vee Q_3 = \bar{R} \quad Q_1 \vee Q_2 \vee Q_3 \vee Q_4 = \check{Q} \\ Q_1 \vee Q_3 = P_2 \quad P_1 \vee \bar{P}_1 = \check{P}_1 \quad Q_m \vee Q_n \vee Q_p = \check{Q} - Q_i \\ Q_2 \vee Q_4 = \bar{P}_2 \quad P_2 \vee \bar{P}_2 = \check{P}_2 \quad m, n, p, i = \bar{1}, 4; \quad m \neq n \neq p \neq i. \end{aligned}$$

* Использовано общепринятое в настоящее время однобуквенное обозначение аминокислот.

Соответственно, триплеты оснований будут иметь вид:

$$\begin{array}{cccc} Q_i & P_1 & P_2 & R \\ Q_j & \bar{P}_1 & R & P_1 \dots \\ Q_h & \check{Q} & \check{Q} & \check{Q} \end{array}$$

Согласно введенному выше определению, концепт-белок представляет собой конъюнкцию «обобщенных аминокислот», описываемую «обобщенными триплетами». Обобщенным триплетом мы называем объединение (логическую сумму) тех триплетов, которые кодируют всю совокупность аминокислот, расположенных в данной позиции (локусе) в белке.

Очевидно, что на каждом уровне обобщенного триплета может находиться один или несколько нуклеотидов. Число нуклеотидов, входящих в состав обобщенного триплета, является логической суммой степени вырожденности триплетов, кодирующих всю совокупность аминокислот, входящих в состав обобщенной аминокислоты.

При вычислении обобщенного триплета, соответствующего обобщенной аминокислоте, для n гомологичных белков (n — мерный концепт-белок), руководствовались следующими правилами [3].

1. Элемент каждой строки обобщенного триплета является объединением (логической суммой) соответствующих элементов триплетов, кодирующих каждую из аминокислот, расположенных в данной позиции рассматриваемого семейства гомологичных белков.

2. Триплеты, входящие в состав обобщенного триплета, но не являющиеся кодом ни для одной из аминокислот, расположенных в данной позиции рассматриваемого семейства гомологичных белков, называются дополнительными триплетами. Обобщенный триплет может и не содержать дополнительных триплетов.

3. Аминокислоты, которые кодируются дополнительными триплетами, называются дополнительными аминокислотами данной обобщенной аминокислоты. Обобщенная аминокислота может и не содержать дополнительных аминокислот.

Таким образом, концепт-белок не только содержит каждый из составляющих его гомологичных белков в качестве «траектории» или «реализации» в пространстве концепта, но и так называемые «дополнительные белки», которые либо еще не найдены, либо уже не существуют, либо по каким-то причинам не могут существовать.

Результаты и обсуждение. Приведенные нами алгоритмы и формы представления знаний позволили получить интересный биологический результат — осуществить классификацию точечных, т. е. изменяющих только одно основание в триплете, мутаций в гомологичных белках (табл. 1): а) молчащая мутация либо не вызывает изменения смысла кодирования, либо приводит к замещению одной аминокислоты другой, отличающейся основанием, расположенным на III месте (уровне) в триплете.

В этом случае обобщенный триплет имеет вид: Q_i , где $Z = \check{P}_1, \check{P}_2,$
 Q_j
 Z

$\check{R}, (\check{Q} - Q_h)$ или \check{Q} ; $k = \overline{1, 4}$; б) простая точечная мутация, приводящая к замещению аминокислоты и возникающая в результате замещения нуклеотидов, расположенных на первом или втором месте в триплете. Обобщенный триплет в этом случае имеет вид: \check{Q}_j или y , где $x = \check{P}_1, \check{P}_2, \check{R}, (\check{Q} -$
 Q Q_i
 Q

$Q_j)$ или \check{Q} , $y = \check{P}_1, \check{P}_2, \check{R}, (\check{Q} - Q_i)$ или \check{Q} ; $i, j = \overline{1, 4}$; в) сложная мутация, т. е. мутация, которая также приводит к замене аминокислот в бел-

ке, однако причиной сложной мутации является одновременное изменение нуклеотидов, расположенных на первом и втором уровнях в триплете. Обобщенный триплет в сложной мутации может иметь вид: x .

y
 Q

Изучаемые нами объекты молекулярной биологии являются результатом исследований, проводимых традиционно специалистами различных направлений физики, химии, биологии. Нами создана программа, позволяющая классифицировать основные причины изменений в первичной структуре белка, т. е. определить точечные мутации как «мол-

Таблица 1
Классификация мутаций гомологичных белков

Белок	Число гомологичных белков, n	Длина полипептидной цепи	Пробелы	Молчащие мутации			Точечные мутации		Сложные мутации
				Изменяющие аминокислоту	Не изменяющие аминокислоты	Транзиции	Трансверсии	Транзиции и трансверсии	
Аномальные гемоглобины человека	4	148	—	2	56	19 (15)	18 (13)	7 (7)	46
Цитохром C	26	112	—	2	37	5 (4)	4 (1)	5 (5)	59
Цитохром C последовательности общих предков	17	119	—	2	70	5 (5)	5 (4)	5 (4)	32
Цитохром C ₂ из <i>Rhodospirillum rubrum</i>	9	123	16	2	25	9 (6)	11 (6)	7 (7)	53

Цифры в скобках соответствуют изменениям, расположенным на 1-м месте в триплете.

Таблица 2
Дополнительные аминокислоты локусов 11—16 цитохрома C₂ из *Rhodospirillum rubrum*

Локус	Аминокислоты	Обобщенные триплеты	Дополнительные триплеты	Дополнительные аминокислоты	Обобщенные аминокислоты	Классификация мутаций
11	VIA	$\frac{P_1}{P_1}$ \checkmark Q	Q_2 Q_2 Q_4 Q_3 Q_1 Q	MT	Ψ_{VIAMT}	Сложная
12	EK	$\frac{P_1}{P_1}$ Q_2 P_1	—	—	Ψ_{EK}	Точечная (транзиция)
13	G	Q_1 Q_1 \checkmark Q	—	—	Ψ_G	Молчащая
14	KA	$\frac{P_1}{R}$ Q	Q_2 Q_1 Q_2 Q_3 Q_2 Q_2 Q Q P	TEDN	Ψ_{TEDN}	Сложная
15	KT	Q_2 R \checkmark Q	Q_2 Q_2 P	N	Ψ_{KTN}	Точечная (трансверсия)
16	IL	\checkmark — Q_1 Q_4 \checkmark Q	Q_2 Q_4 Q_4 Q_1 Q_1 P_1	MF	Ψ_{ILMF}	Точечная (транзиция и трансверсия)

чащие», «простые» и «сложные». Банк данных содержит сведения о 10 гомологичных белках: инсулине, цитохромах, гемоглобине (α -, β -, γ -цепи) и др.

Реализованы алгоритмы, позволяющие получать обобщенный белок, обобщенную аминокислоту, дополнительные аминокислоты (табл. 2).

Создана программа, дающая возможность оценивать частоту встреч тех или иных аминокислот в данной позиции, т. е. получать данные о «наиболее» и «наименее» вероятных составах полипептидов для данного семейства гомологичных белков, относящихся к данному «супергомомеру» как к своему «концепт-белку». Что же касается проблемы интерпретации обобщенного набора свойств конкретного «супергомомера», то это предмет экспериментальных и теоретических исследований будущего. Граница различия и отождествления тех элементов, которые организуют и обеспечивают неспецифические гомологичные белковые структуры, останется непостижимой, пока не будет решена задача идентификации, распознавания и классификации материальных белковых структур лабильно-устойчивой природы. Индивидуальность и коллективность проявляются уже на уровне белков и особенно гомологичных.

Резюме

Аналіз первинних структур гомологічних білків дає уяву про область макромолекули, необхідну для здійснення її основної функції, і вміщує відомості про шляхи біологічної еволюції. Введення поняття «концепт-білка», а також описання процесу точкових мутацій на мові алгебри логіки дало змогу створити алгоритм класифікації мутацій в гомологічних білках і виявити так звані «додаткові поліпептиди», які або ще не знайдені, або вже не існують, або з якихось причин не можуть існувати.

Summary

As based on a notion of concept-object a computable notion «concept-biopolymer», a hypothetical generalized biopolymer, is introduced. It may be a nuclear acid (DNA, RNA), proteins, etc.

This approach makes it possible: 1) to reproduce changes, taking place in the primary structure of m-RNA in the process of evolution; 2) to classify point mutations in nucleotid matrix stipulating changes in the primary protein structure; 3) to supplement the family of homologous proteins, i. e. to reproduce new proteins which, evidently, have taken place in the process of evolution, or now exist in the period of morphogenesis and differentiation.

СПИСОК ЛИТЕРАТУРЫ

1. Чавчанидзе В. В. К теории естественного и искусственного концептуального интеллекта // Материалы IV МОКИИ.— Тбилиси, 1976.— С. 7—17.
2. Chavchanidze V. V. Towards the general theory of conceptual systems // Kybernetes.— 1972.— 3, N 1.— P. 17—25.
3. Квинихидзе К. С. Использование концептуального подхода для выявления дополнительных аминокислотных последовательностей // Сообщ. АН ГССР.— 1985.— 118, № 1.— С.— 57—60.
4. Квинихидзе К. С., Чавчанидзе В. В. Концептуальный подход к описанию эволюции белковых структур // Там же.— 1976.— 84, № 2.— С. 337—340.
5. Kvinikhidze K. S., Chavchanidze V. V. Application of conceptual approach to describe the evolution of protein structure // Rep. at 8-th Int. Congr. on Cybernetics.— Namur, 1976.— P. 839—847.
6. Chavchanidze V. V., Kvinikhidze K. S. Biological structures computer modelling // Междунар. симпоз. «Физико-химия ДНК»: Тез. докл.— Тбилиси, 1987.— С. 223.
7. Carnap R. The logical foundation of probability.— Chicago: Univ. Chicago press, 1950.— P. 122—126.
8. Чавчанидзе В. В. К вопросу о расшифровке кода ДНК(РНК) и биологические соединения включения // Сообщ. АН ГССР.— 1966.— 44, № 1.— С. 19—26.
9. Чавчанидзе В. В., Квинихидзе К. С. О создании вычислимого понятия «концепт-объект» // Теория и устройства систем автомат. управления.— Тбилиси: Мецниереба, 1989.— С. 156—162.
10. Dayhoff M. O. Atlas on protein sequences and structure.— New York: Silver Spring Md., 1969.— P. D8—D96.

Ин-т систем. управления АН ГССР, Тбилиси

Получено 06.06.90