

Молекулярная эволюция белков.

Д.А.Афонников,
лаборатория теоретической
генетики

Предмет молекулярной ЭВОЛЮЦИИ

Эволюция – развитие, постепенное изменение во времени.

Молекулярная эволюция – наука об эволюционных изменениях генетических макромолекул.

Основные задачи

- Выявление закономерностей эволюции генетических макромолекул
- Реконструкция эволюционной истории генов и организмов

Взаимосвязь с другими областями науки

- Палеонтология (датировка эволюционных событий)
- Генетика (принципы организации и передачи наследственной информации)
- Молекулярная биология, биофизика
(функционирование генетических молекул)
- Математика (построение моделей эволюции)
- Информатика (обработка и анализ данных)

Теория естественного отбора (Дарвин, 1859)



- Внутривидовые вариации (случайные)
- Приспособленность организмов (способность давать большее потомство)
- «Перевоспроизводство» потомства
- Естественный отбор наиболее приспособленных особей (по репродуктивности)

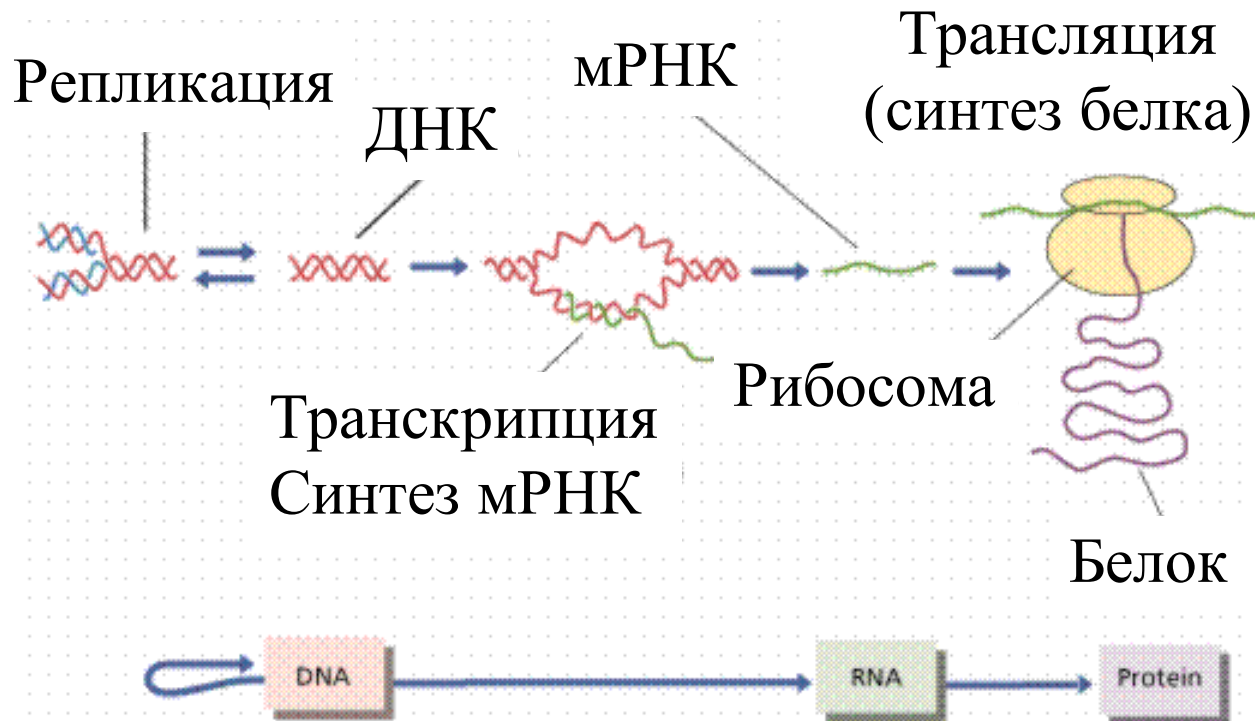
Резюме : **ЭВОЛЮЦИЯ как «хаос с обратной связью»**

Объект исследования молекулярной эволюции

- Последовательности ДНК (носитель генетической информации)
- Последовательности белков
- Структуры белков
- Геномы организмов

Реализация генетической информации

Центральная догма биологии



Организация геномной ДНК

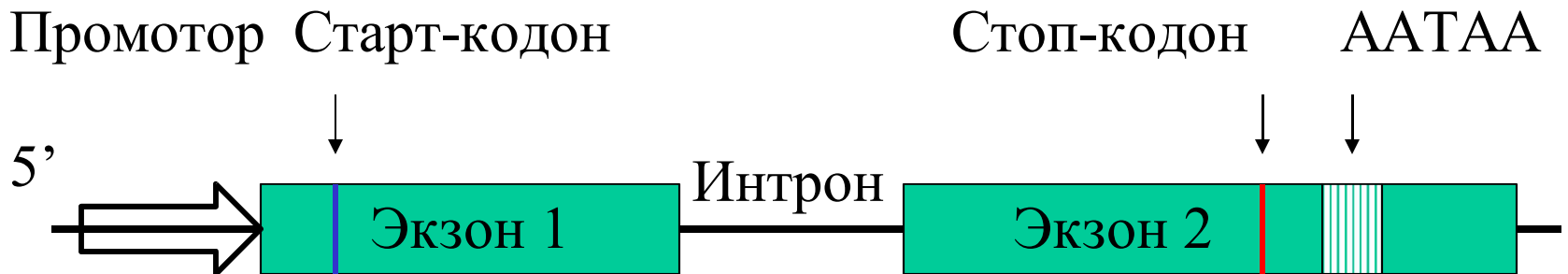
- Кодирующая часть (функция – кодирование РНК и белков). Доля генома, кодирующая белок у человека составляет $<2\%$
- Некодирующая часть (содержит регуляторные районы, повторы и районы, функция которых пока неизвестна). У человека составляет $\sim 50\%$.
- Соотношение кодирующей/некодирующей частей различно для разных организмов

Ген

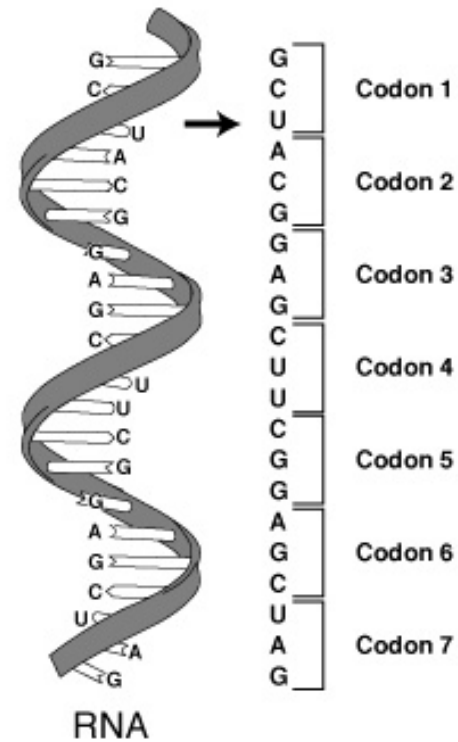
- Концепция гена изменялась на всем протяжении развития генетики и биологии
- «Ген - это сегмент ДНК, который обуславливает фенотип или функцию. В отсутствии данных о функции ген может быть охарактеризован последовательностью, транскриптом или по гомологии» (HUGO, J.A. White et. al. Guidelines for Human Gene Nomenclature HGNC Human Genome Nomenclature Committee, 1997) .
- Оценка числа генов ~30,000 (у человека)

Схема структуры гена

- Гены кодируют РНК или белки и могут содержать кодирующую часть (экзон), некодирующую часть (интрон) и регуляторные зоны (промотор)



Генетический код



Ribonucleic acid

Вторая позиция кодона

	U	C	A	G		
Первая позиция кодона	U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr STOP STOP	Cys Cys STOP Trp	Третья позиция кодона
	C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	
	A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	
	G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	

Рисунок из Молекулярная биология © 2000

Дымшиц Г.М.

<http://www.nsu.ru/education/biology/molbiol/>

Рисунок Darryl Leja

<http://www.accessexcellence.org/AB/GG/codon.html>

Генные мутации

- Изменения сегментов гена
 - Транслокации
 - Дупликации
 - Инверсии
 - Делеции
- Точечные мутации - замены одного нуклеотида

Типы точечных мутаций

Мутации, приводящие к кистозному фиброзу (пример)

Patient	Mutation	Result
A	482 C G C ↓ C A C	Arg-117 ↓ His-117
B	1609 C A G ↓ T A G	Gln-493 ↓ STOP
C	Insertion of 2 nucleotides (AT) at 2566	Frameshift
D	Deletion of one C at 3659	Frameshift
E	Deletion of 3 nucleotides at 1654-1656	Deletion of Phe-508

Миссенс-мутация
(замена аминокислоты)

Нонсенс-мутация
(замена на стоп-кодон)

Вставка 2 нуклеотидов
(сдвиг рамки)

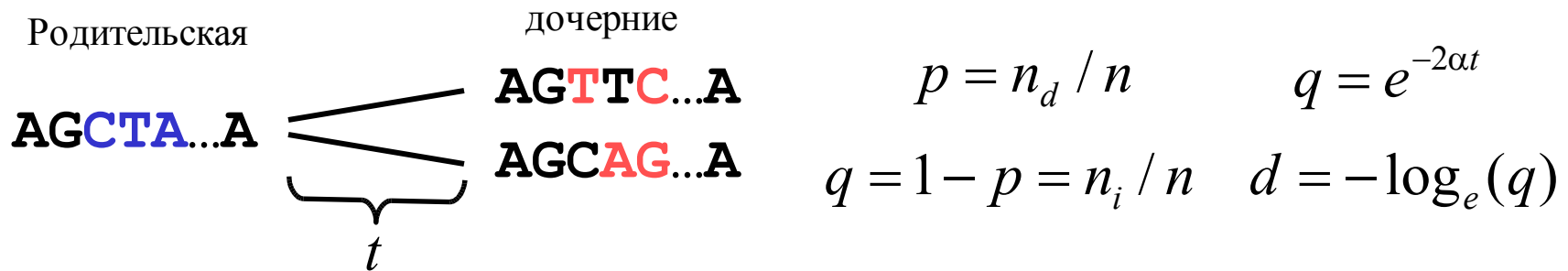
Делеция нуклеотида
(сдвиг рамки)

Делеция 3 нуклеотидов
(делеция аминокислоты)

Модели эволюции последовательностей

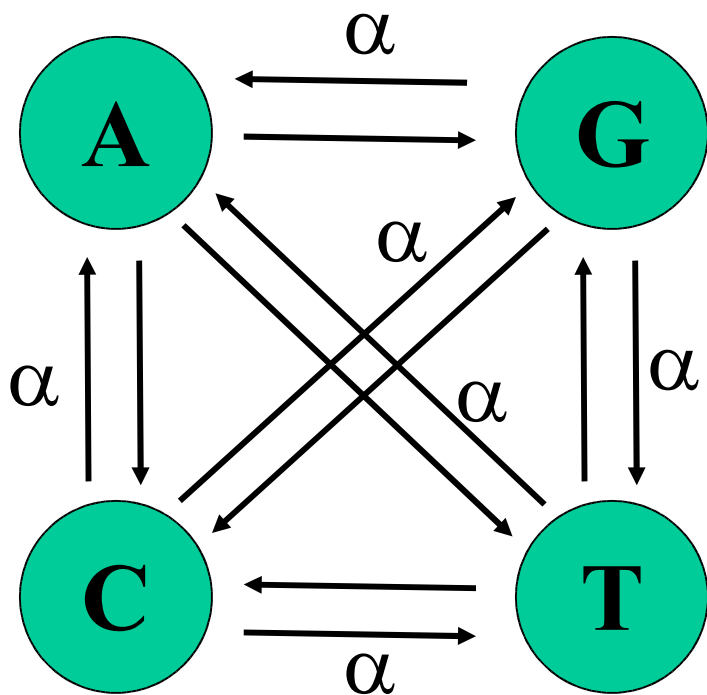
- Описывают временные изменения в последовательностях генетических макромолекул
- Для одиночных замен используются Марковские модели с конечным числом состояний
- Развитие от простого к сложному

Расстояние между двумя последовательностями ДНК



- n_d – число несовпадающих пар, p – их доля.
- n_i – число совпадающих пар, q – их доля. $p+q=1$.
- Направления \pm времени равноправны, поэтому при сравнении двух дочерних последовательностей расстояние d между ними эквивалентно времени эволюции $2t$.
- В простейшем приближении можно использовать расстояние, основанное на частотах совпадений аминокислот (модель Пуассона).
- Зная время и d можно установить скорость замен и наоборот, зная α и d можно установить t .
- Недостаток – не учитываются повторные и обратные замены, применимо на небольших временах дивергенции

Модель Джукса-Кантора



Замены случайны,
независимы о других
позиций, равновероятны с
вероятностью α

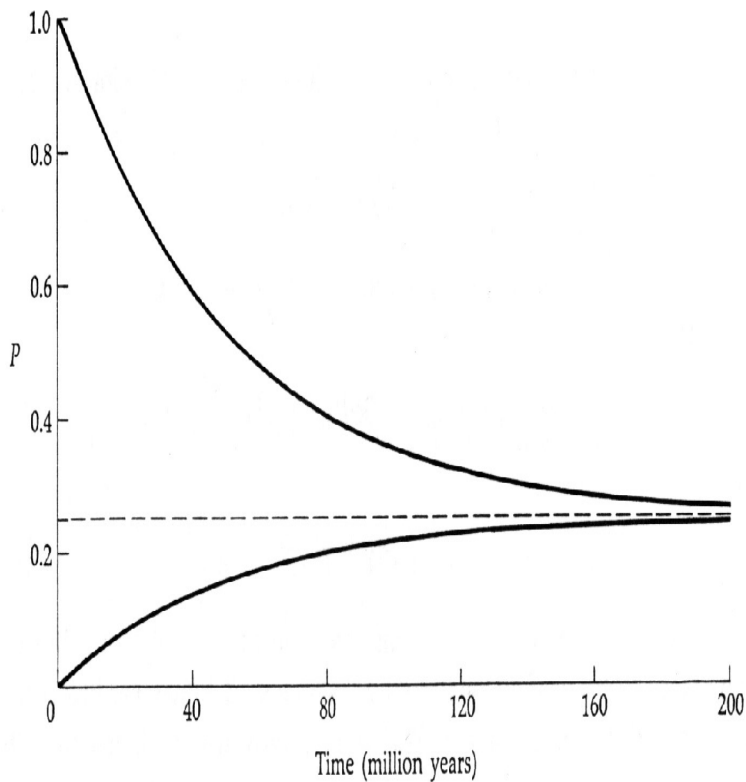
Дискретное приближение

$$P_A(t+1) = (1-3\alpha)P_A(t) + \alpha[1-P_A(t)] = -4\alpha P_A(t) + \alpha$$

Непрерывное приближение

$$\frac{dP_A(t)}{dt} = -4\alpha P_A(t) + \alpha$$

Зависимость частоты нуклеотида 'А' от времени



$$P_A(t) = 0.25 + (P_A(0) - 0.25)e^{-4\alpha t}$$

$$P_A(t) = 0.25 + 0.75e^{-4\alpha t} \quad (P_A(0)=1)$$

$$P_A(t) = 0.25 - 0.25e^{-4\alpha t} \quad (P_A(0)=0)$$

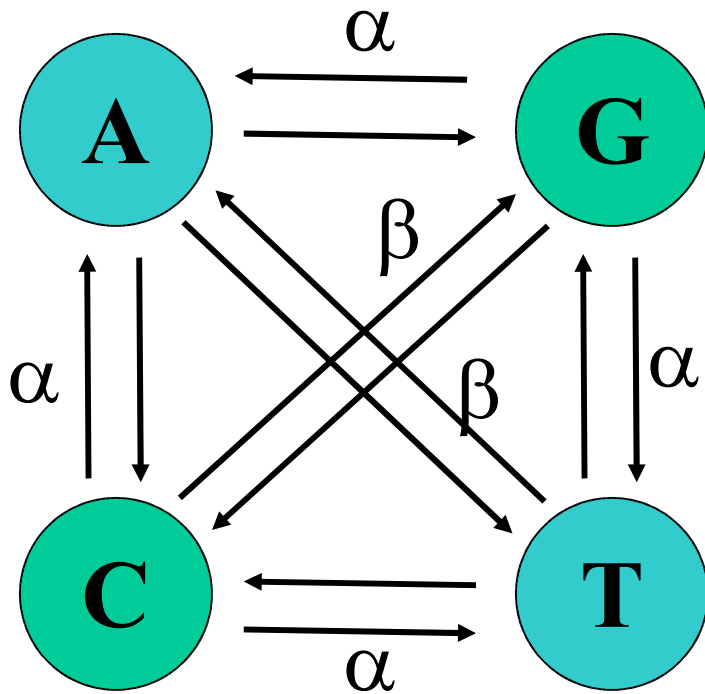
Особенности:

$$P_A(\infty) = P_T(\infty) = P_G(\infty) = P_C(\infty) = 1/4$$

$$P_{AA}(t) = P_{ii}(t) = 0.25 + 0.75e^{-4\alpha t}$$

$$\begin{aligned} P_{AG}(t) &= P_{AT}(t) = P_{AC}(t) = \\ &= P_{ij}(t) = 0.25 - 0.25e^{-4\alpha t} \end{aligned}$$

Модель Кимуры



Учет разных частот

Для транзиций - α

Для трансверсий (A-T, G-C) - β

$$P_{ii}(t) = 0.25 + 0.25e^{-4\beta t} + \\ + 0.5e^{-2(\alpha + \beta)t}$$

Однако P_{ij} различаются по парам замен (транзиции Y, трансверсии Z):

Матрица замен и ее свойства

$$p_A' = (1 - \lambda_1)p_A + \lambda_{21}p_T + \lambda_{31}p_G + \lambda_{41}p_C$$

$$p_T' = \lambda_{12}p_A + (1 - \lambda_2)p_T + \lambda_{32}p_G + \lambda_{42}p_C$$

$$p_G' = \lambda_{13}p_A + \lambda_{23}p_T + (1 - \lambda_3)p_G + \lambda_{43}p_C$$

$$p_C' = \lambda_{14}p_A + \lambda_{24}p_T + \lambda_{34}p_G + (1 - \lambda_4)p_C$$

$$\mathbf{p}'(t + \Delta t) = \mathbf{M}_0 \mathbf{p}(t)$$

$$\mathbf{M} = \begin{bmatrix} (1 - \lambda_1) & \lambda_{21} & \lambda_{31} & \lambda_{41} \\ \lambda_{12} & (1 - \lambda_2) & \lambda_{32} & \lambda_{42} \\ \lambda_{13} & \lambda_{23} & (1 - \lambda_3) & \lambda_{43} \\ \lambda_{14} & \lambda_{24} & \lambda_{34} & (1 - \lambda_4) \end{bmatrix}$$

Свойства:

- Сумма по строкам равна 1
- Эволюция за n шагов эквивалентна умножению на $\mathbf{M}_n = (\mathbf{M}_0)^n$.
- Если $\mathbf{M} = \text{const}$, то существуют равновесные частоты, которые находятся из уравнения

$$\mathbf{p}_0 = \mathbf{M}_0 \mathbf{p}_0$$

Матрица замен для модели JK

$$\mathbf{M}_0 = \begin{bmatrix} 1-3\alpha & \alpha & \alpha & \alpha \\ \alpha & 1-3\alpha & \alpha & \alpha \\ \alpha & \alpha & 1-3\alpha & \alpha \\ \alpha & \alpha & \alpha & 1-3\alpha \end{bmatrix} \quad \mathbf{p}_0 = [1/4, 1/4, 1/4, 1/4]$$

$$\mathbf{M}(t) = \begin{bmatrix} 0.25 + 0.75e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} \\ 0.25 - 0.25e^{-4\alpha t} & 0.25 + 0.75e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} \\ 0.25 - 0.25e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} & 0.25 + 0.75e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} \\ 0.25 - 0.25e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} & 0.25 - 0.25e^{-4\alpha t} & 0.25 + 0.75e^{-4\alpha t} \end{bmatrix}$$

Долю совпадающих пар q можно вычислить как $\sum \lambda_i p_i =$

$$q = 4(0.25 + 0.25e^{-4\alpha t}) \cdot 0.25 = (0.25 + 0.25e^{-4\alpha t})$$

Тогда время эволюции можно оценить как

$$\alpha t = -1/4 \log(4q/3 - 1/3) = -1/4 \log(1 - 4p/3)$$

$$t = \frac{-1}{4\alpha} \log\left(1 - \frac{4}{3} p\right)$$

Различные меры расстояний последовательностей ДНК

Эволюционное расстояние принято иногда выражать как среднее число замен за время t .

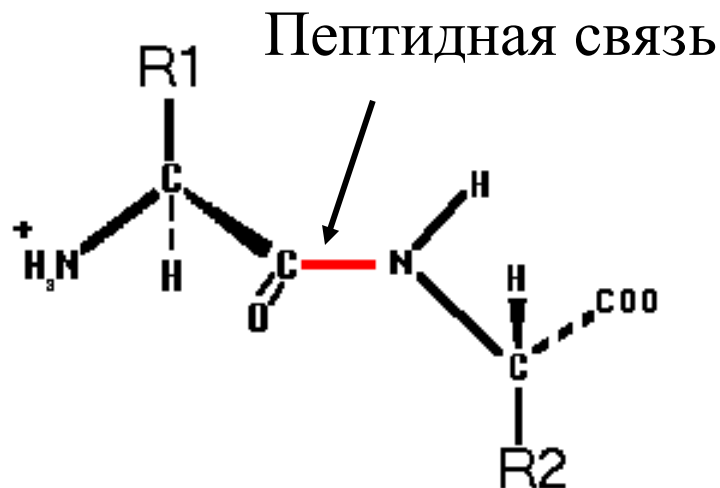
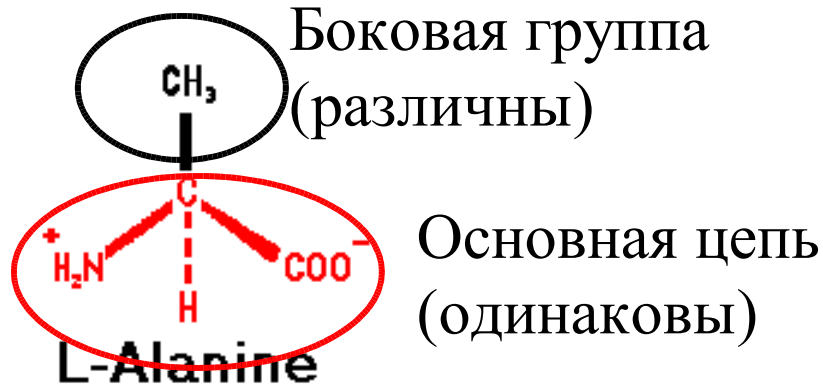
Определение расстояния зависит от типа модели. Примеры:
Модель Джукса-Кантора:

$$d = -\frac{3}{4} \log_e \left(1 - \frac{4}{3} p\right)$$

Модель Кимуры (P-частота транзиций, Q-частота трансверсий):

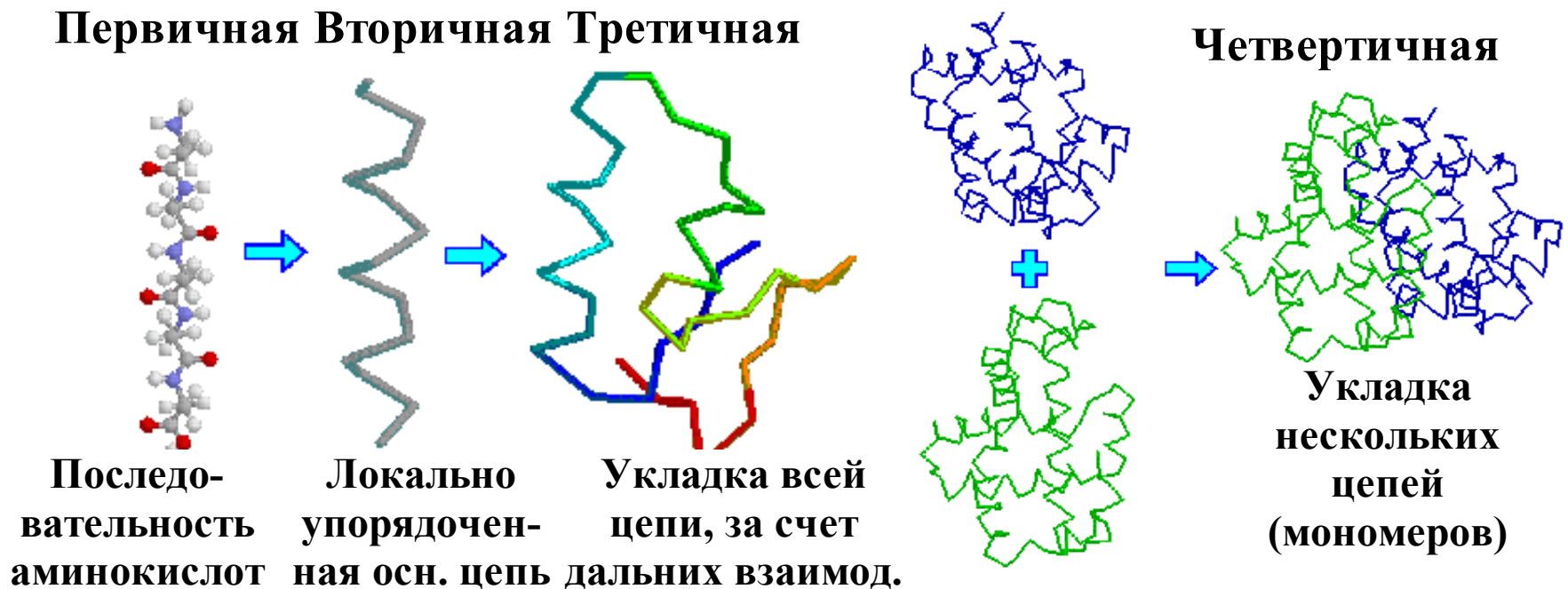
$$d = -\frac{1}{2} \log_e [(1 - 2P - Q) \sqrt{1 - 2Q}]$$

Структура белков

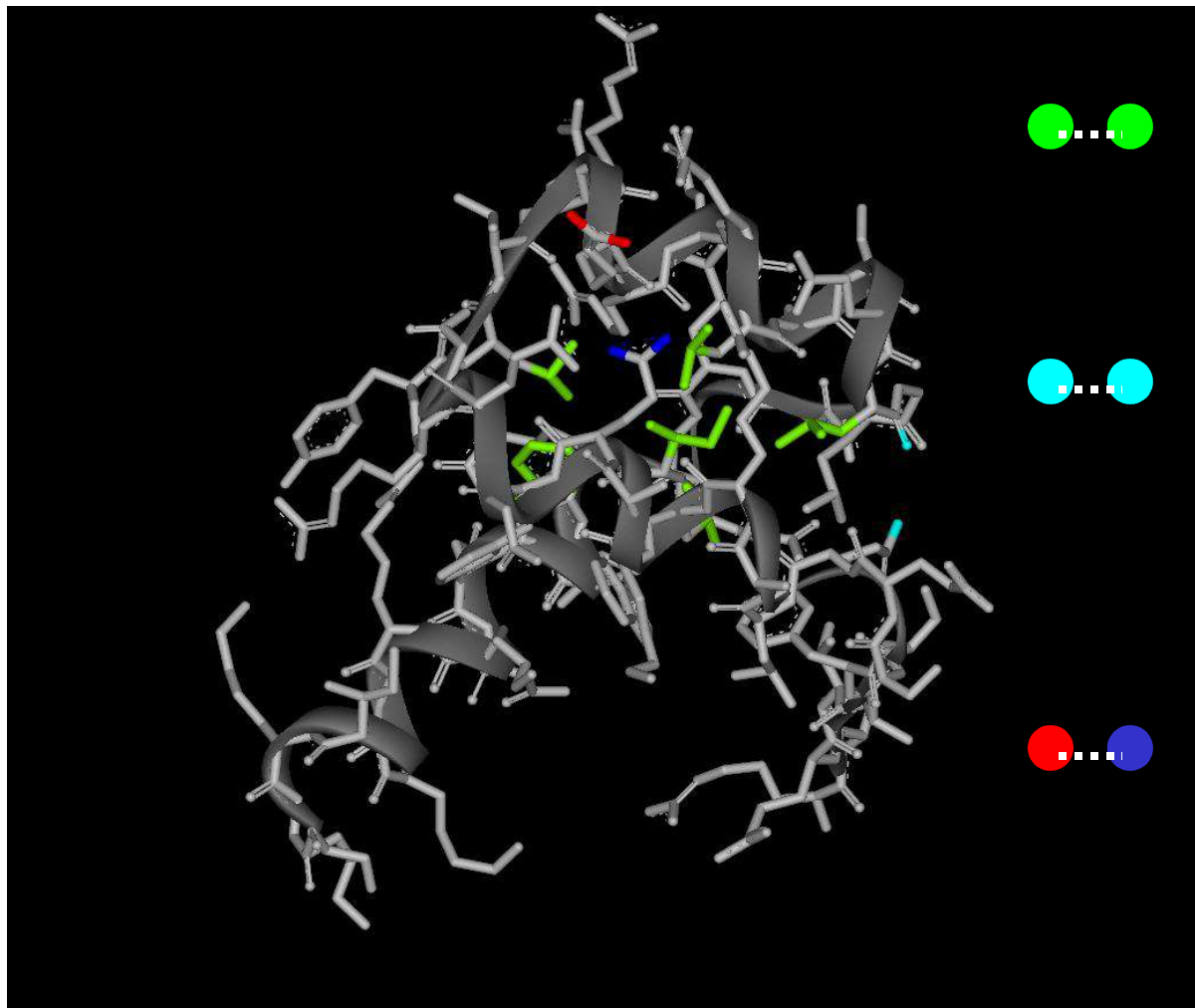


- Относительно короткие цепи (~100 – 1000 а.к.)
- 20 мономеров (различаются боковыми группами)
- Пространственная структура нелинейна
- Вторичная структура: α, β, τ
- Третичная структура (глобула) : дальние контакты
- Гетерогенность пространственной структуры

Уровни структурной организации белков



Типы взаимодействий в белках и их вклад в стабильность глобулы



Гидрофобные,
~1.2 ккал/моль на
одну группу $-\text{CH}_2-$

Водородные связи,
 1.6 ± 1 ккал/моль на
одну водородную
связь

Солевые мостики,
1 ккал/моль на
поверхности, 3-5
ккал/моль внутри

В среднем 1.7 ккал/моль на остаток. Для сравнения: $kT=0.6$ ккал/моль

Влияние аминокислотных замен на стабильность и функцию белка.

По фенотипическому проявлению мутации можно разделить на следующие классы:

- 1) нейтральные;
- 2) снижающие эффективность функционирования;
- 3) повышающие эффективность функционирования;
- 4) приводящие к возникновению новой функции.

По молекулярному механизму действия замены остатков можно разделить на три основных группы:

- 1) изменяющие конформационную стабильность белков;
- 2) влияющие на процесс укладки (самоорганизации) полипептидной цепи в нативную глобулу;
- 3) изменяющие функцию белков за счет нарушения локальной конформации функциональных центров;

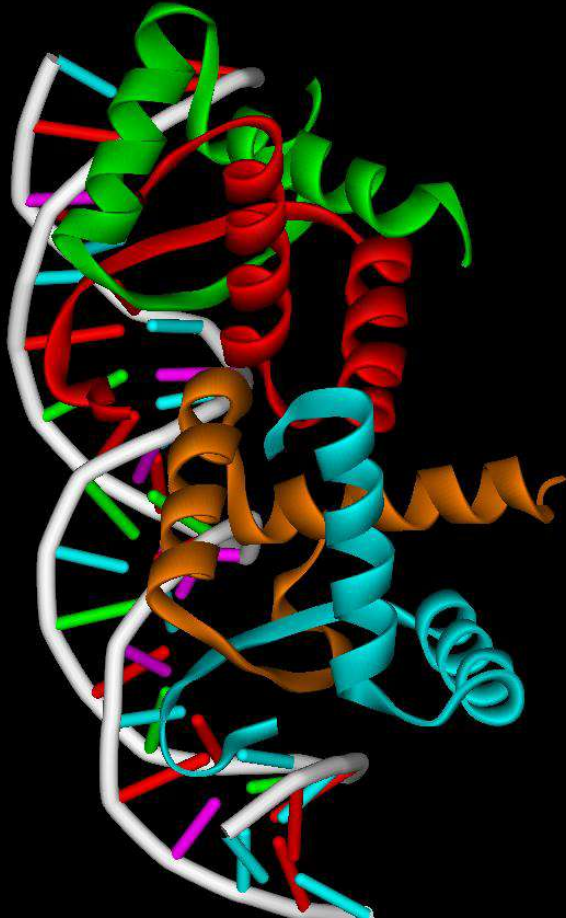
Влияние аминокислотных замен на стабильность и функцию белка.

Эффект замены существенно зависит от структурного контекста. Менее всего влияют на стабильность замены на поверхности белка и замены в неструктурных участках белка (высокое значение температурного фактора).

Дестабилизирующие замены – такие, что увеличение свободной энергии в результате замены происходит более чем на 1 ккал/моль.

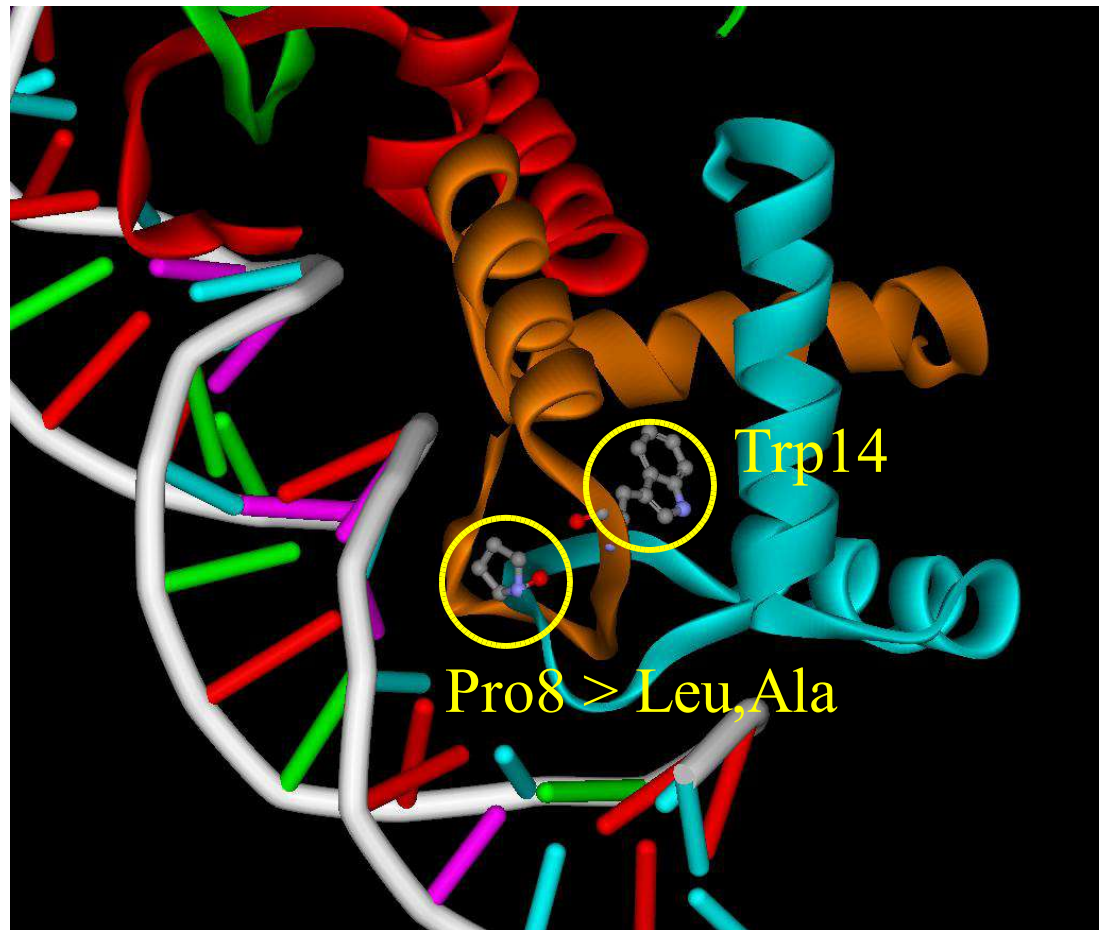
Дестабилизирующие замены наблюдаются в гидрофобном ядре, затрагивают формирование солевых мостиков и водородных связей, влияют на конформацию основной цепи (замены Pro, Gly и наоборот).

Исследование мутаций на примере Arc репрессора (Sauer *et al.*, 1996, Milla *et al.*, 1994)



Проведены одиночные замены остатков на аланин в 51-ой из 53-х позиций белка.
25 мутаций – нейтральные (<1 ккал/моль)
Повышают стабильность – единицы.
Около половины – дестабилизируют белок.

Пример мутации, повышающей стабильность белка (Arc репрессор).



Замена Pro8 на Leu или Ala увеличивает стабильность димера на 2.5 ккал/моль. Это происходит в результате образования водородной связи атомами основной цепи с Trp14 другого мономера. Такие мутации чрезвычайно редки. В стафилококковой нуклеазе из 290 замен только 33 привели к относительно небольшому увеличению стабильности.

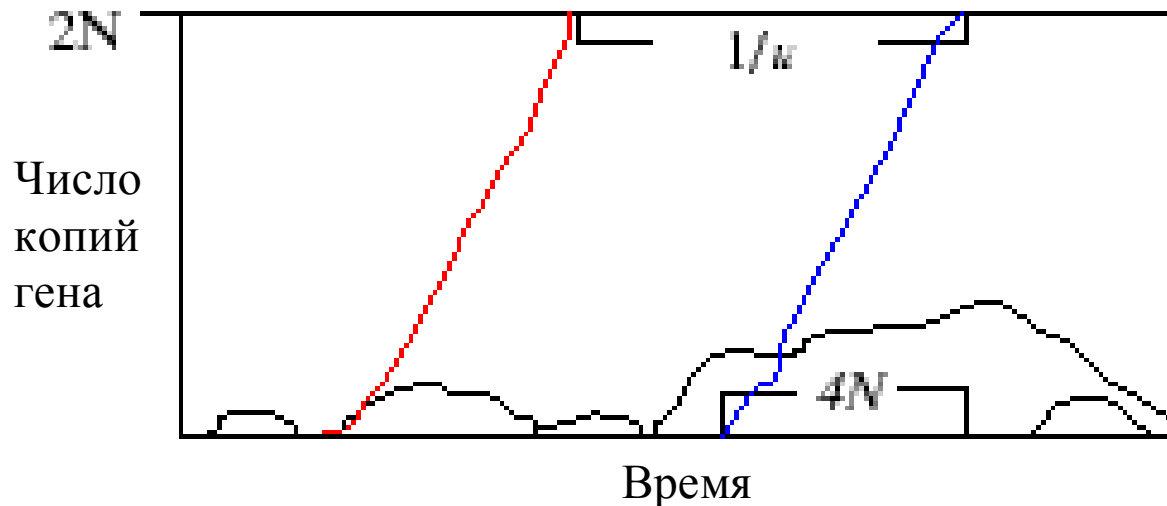
Влияние мутаций на активность белка (тирозил тРНК синтетаза, (Fersht, 1987).

Таблица 3. Энергия взаимодействия тирозил-тРНК синтетазы с различными реагентами (Fersht, 1987).

Остаток	Энергия взаимодействия боковых групп ^a в комплексе с				
	Tyr	АТФ	[Tyr-АТФ]	PP	Tyr-АМФ
	Сайт связывания тирозина				
Tyr34	+	0	+	0	+
Asp78	++++	++*	++++	++*	++++
Tyr169	++++	0	++++	0	++++
Gln173	++++	++*	++++	+	++++
	Сайты связывания нуклеотида и пирофосфата				
Cys35	0	0	++	0	+++
Thr40	0	0	++++	++++	0
His45	0	0	++++	++++	0
His48	0	0	+++	0	+++
Thr51	0	0	0	0	-
Lys82	0	++	++++	++++	0
Arg86	0	0	++++	++++	-
Asp194	0	0	++++	+	+++
Lys230	0	0	++++	++++	0
Lys233	0	++++	++++	++++	0
Относительная энергия стабилизации от взаимодействия с боковой группой (ккал/моль): 0 = -0.5...+0.5; + = 0.5...1.0; ++ = 1.0...1.5; +++ = 1.5...2.0; ++++ > 2.0; - = -0.5...-1.0; * - при мутации нарушается структура белка.					

Нейтральная теория Кимуры

Теория Кимуры объясняет высокую степень полиморфизма белков. Пример схематически показан на диаграмме:

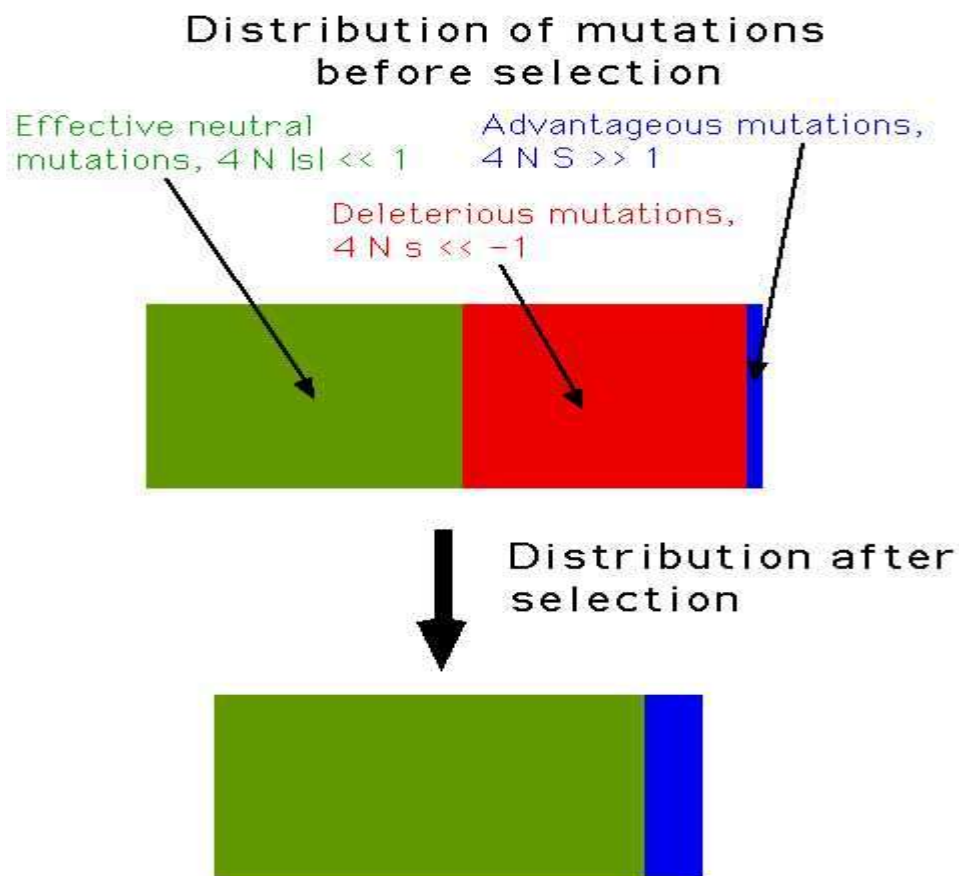


Время фиксации *нейтральных* мутаций $t_{\text{fix}} = 4N$ поколений

Скорость фиксации *нейтральных* замен (число фиксации за единицу времени) $K = u$, u - частота возникновения мутации

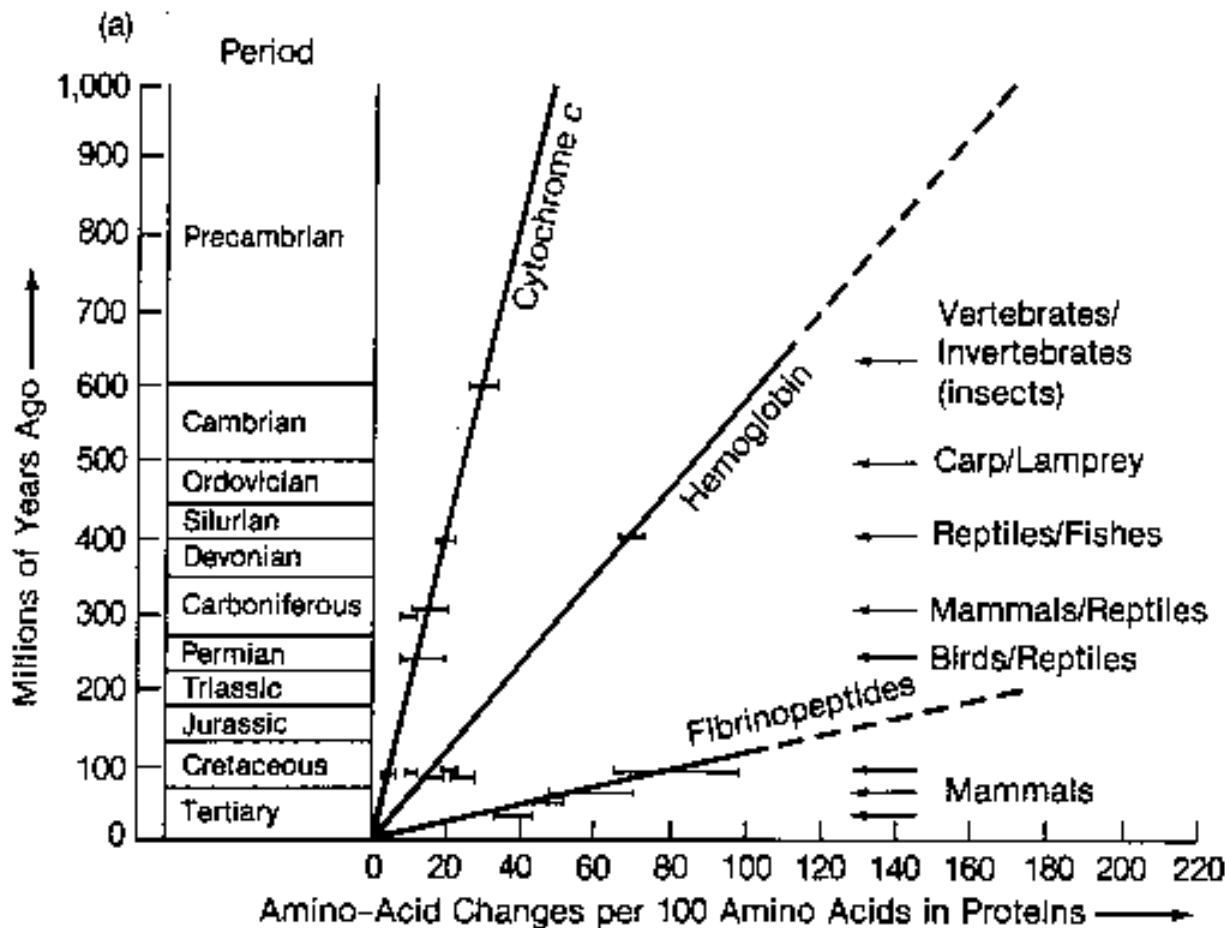
Для адаптивных замен $K = 4Nsu$

Постулаты нейтральной теории



- Адаптивные мутации очень редки
- Деструктивные мутации быстро элиминируются
- Подавляющее большинство фиксированных мутаций селективно нейтральны
- Скорость фиксации замен в генах является постоянной (гипотеза молекулярных часов, Zuckercandle & Poling)
- Функционально важные районы функционируют медленнее, менее важные - быстрее

Является ли скорость замен постоянной?



Для большого числа белков наблюдаются постоянные скорости замен, однако для разных белков они различны

Сравнение скоростей

СИНОНИМИЧЕСКИХ/НЕСИНОНИМИЧЕСКИХ

ЗАМЕН

Оценки скоростей нуклеотидных замен на 1 позицию на 1000 млн. Лет при сравнении некоторых белков человека и грызунов (время дивергенции ~80 млн.лет согласно ископаемым)

Ген	Число кодонов	К(несин.)	К(син.)
histone 3	135	0.00	6.38
actin beta	349	0.03	3.13
myoglobin	153	0.56	4.44
In-gamma	136	2.79	8.59

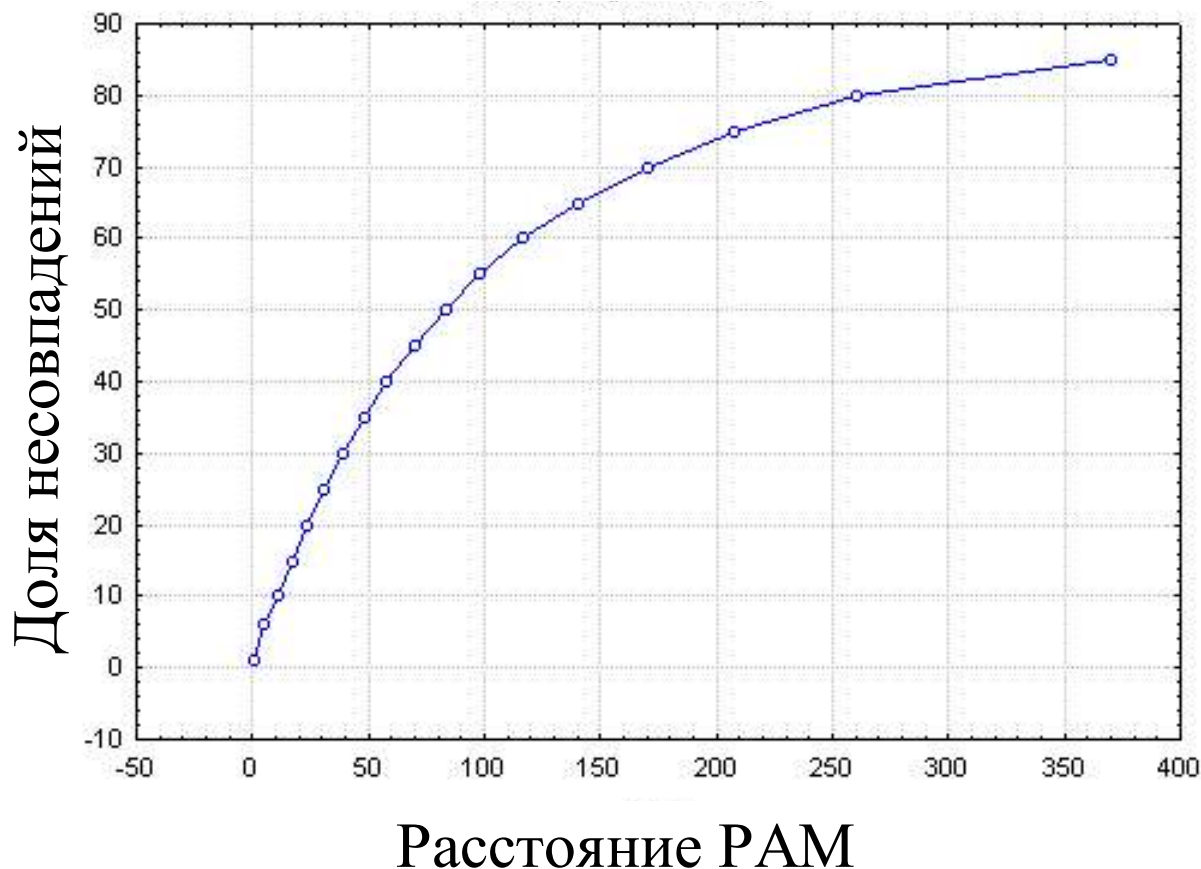
Модель замен в белках

20 аминокислот. Замены в позициях независимы и определяются матрицей одинаковой для всех белков и всех позиций $M(20 \times 20)$. Матрица замен M была определена эмпирически на основе анализа нескольких семейств гомологичных белков Дайхофф и сотр. (1978).

Свойства матрицы Дайхофф:

- Равновесные частоты равны частотам встречаемости аминокислот в последовательностях белков.
- Наиболее часты замены аминокислот на аминокислоты, сходные по физико-химическим свойствам.
- Исходная матрица нормирована на время, эквивалентное 1 замене на 100 позиций (1РАМ).
- Для оценки вероятности замен через время $t=n$ надо матрицу 1РАМ возвести в степень n .

Расстояния белковых последовательностей



- Матрицы серии РАМ-N отражают вероятность замены аминокислот $a_i \rightarrow a_j$ за время эволюции, эквивалентное N РАМ единиц.

Множественное выравнивание последовательностей

Множественное выравнивание позволяет:

- Выявлять более слабую гомологию белков.
- Обеспечивает информацию для разбиения последовательностей по паттернам дивергенции на подгруппы (подсемейства)
- Позволяет выделять вырожденные функциональные мотивы

Множественное выравнивание последовательностей

Пример: дегалогеназы и еноил-Со-А гидратазы.

Dehalogenases	GGG	LGISLAS	DMAICADS	AKFVCAWHTI	GI	IGNDTAT
	GGG	LGMSLAC	DLAVCTDR	ATFLPAWMSI	GI	ANDASS
Enoyl CoA Hydratases	GGG	CELAMMC	DIYAGEKA	QFGQPEILL	GT	IPGAGG
	GGG	NELAMMC	DIYAGEKAR	FGQPEINIG	GT	IPGAGG
	GGG	CELAMMC	DFIIASETAK	FGLPEITL	GV	IPGMGG
	GAG	CELALLC	DVVVAGENAR	FGLPEITL	GI	MPGAGG

Множественное выравнивание последовательностей

Консервативные блоки выравнивания соответствуют функционально важным участкам белков. Пример:

дегалогеназы	10	151	176
	*	*	*
Cu ⁺⁺ ATPase .Ec	LDTVVFDKTGTLTEG	VIAGVLPDGKAEAIKHL	AMVG D GINDAPAL
Cu ⁺⁺ ATPase .Hs	VKVVVFDKTGTITHG	VFAEVLPSHKVAKVKQL	AMVG D GINDSPAL
Ca ⁺⁺ ATPase .At	ATTICSDKTGTLTTN	VMARSSPMDKHTLVRL	AVTG D GTNDAPAL
Urf .Mj	KVAIVFDSAGTLVKI	E - - AHQELKRD LIRNL	IMVG D GANDVPAM
PhosSerPhos .Hs	ADAVCFD V D S T V I R E	TAE - SGGKG K V I K L L K E	IMIG D GATDMEAC
2-D0-6-PPhos .Sc	V D L C L F D L D G T I V S T	I T G F D V K N G K P D P E G Y S	V V F E D A P V G I K A G
DL-Gly-3-Phos .Sc	I N A A L F D V D G T I I I S	I T A N D V K Q G K P H P E P Y L	V V F E D A P A G I A A G
Phosphon .Pa	L Q A A I L D W A G T V V D F	A T D E V - P N G R P W P A Q A L	V K V D D T W P G I L E G
Phosphon .St	I H A V I L D W A G T T V D F	A T D D L A A G G R P G P W M A L	V K V D D A A P G I S E G
Phosphon .Bc	I E A V I F D W A G T T V D Y	T P D D V - P A G R P Y P W M S Y	I K V G D T V S D M K E G
PhosGlycolPhos .Rs	M P G V V F D L D G T L V H S	I G G E S L P Q R K P D P A P L A	L Y V G D S E V D A A T A
NtermDom .IGPD .Pp	V Q A L L L D M D G V M A E V	L E D C P P - - - K P S P E P I L	A M V G D T V D D I I A G
B-PhosGlucoMut .L1	F K A V L F D L D G V I T D T	A E V A A S - - - K P A P D I F I	I G L E D S Q A G I Q A I
HaloAcidDehal .PspYL	I K G I A F D L Y G T L F D V	L S V D P V Q V Y K P D N R V Y E	L F V S S N A W D A T G A
NtermDomEpoxyld .Hs	L R A A V F D L D G V L A L P	I E S C Q V G M V K P E P Q I Y K	V F L D D I G A N L K P A
EnolasePhos .Ko	I R A I V T D I E G T T S D I	F D - - T L V G A K R E A Q S Y R	L F L S D I H Q E L D A A

В дегалогеназах позиции 10, 151 и 176 формируют активный сайт.

