



Стохастическая модель трансляции

***Владимиров Н.В.,
аспирант ИЦиГ СО РАН***

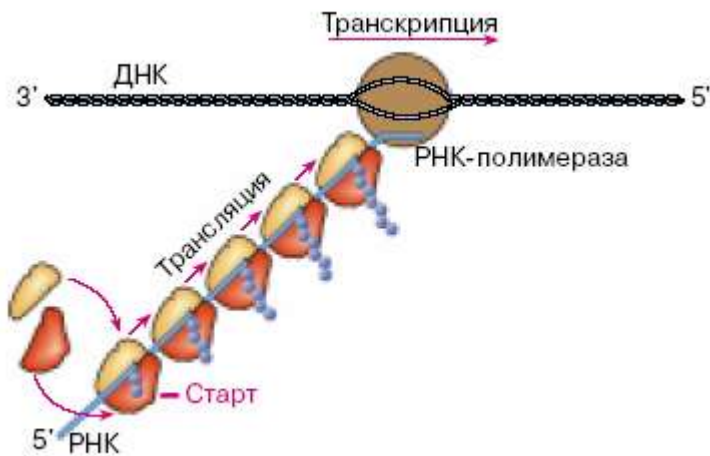
Кафедра информационной биологии ФЕННГУ



Механизм трансляции



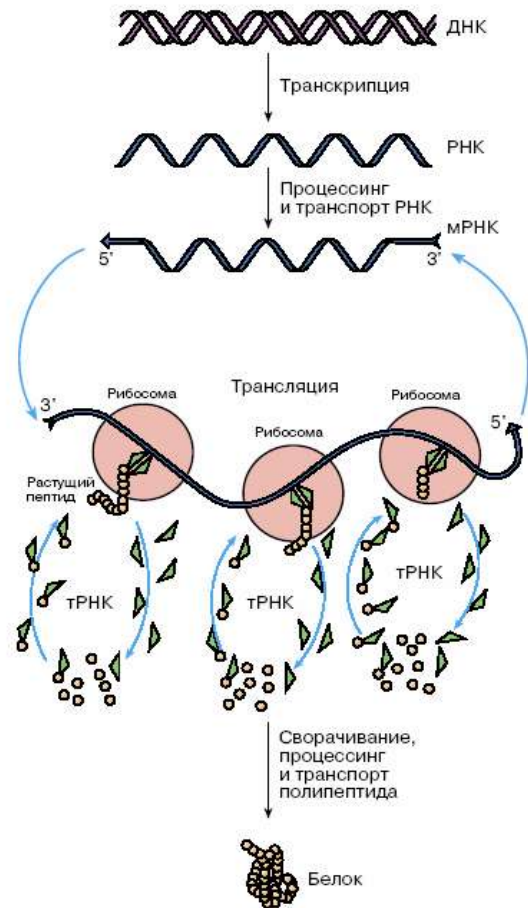
Трансляция – это процесс синтеза полипептидной цепи по матричной РНК.



Трансляция у *прокариот* начинается еще на стадии транскрипции (сопряженная транскрипция-трансляция). При этом скорость транскрипции у бактерий строго координирована со скоростью трансляции (10-15 триплетов в секунду).



Механизм трансляции



У эукариот сопряжение транскрипции и трансляция невозможно - мРНК сначала полностью синтезируется в ядре, затем происходит ее сплайсинг, и лишь после этого мРНК выходит из ядра в цитоплазму, где и встречается с рибосомами.



Механизм трансляции



Трансляция в обоих случаях включает три этапа:

- **инициацию**
- **элонгацию**
- **терминацию**



Механизм трансляции



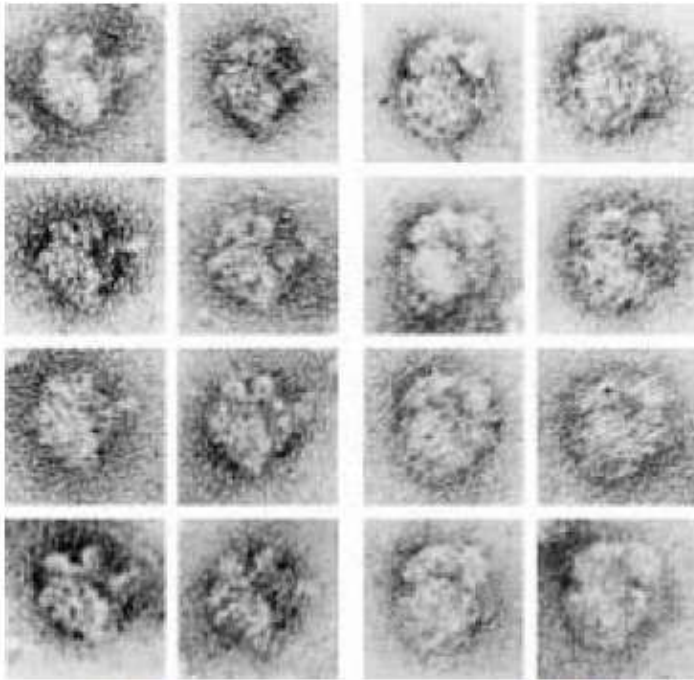
Инициация – это серия молекулярных событий, которые приводят к взаимодействию рибосомы с началом кодирующей последовательности мРНК, и последующему считыванию этой последовательности.

Элонгация – это следующий за инициацией процесс, в ходе которого происходит собственно трансляция мРНК рибосомой (синтез полипептидной цепи по последовательности кодонов мРНК).

Терминация – заключительный этап элонгации, когда сканирующая рибосома встречает т.н. незначащий триплет (не кодирующий никакой аминокислоты), или стоп-кодон. Всего таких кодонов 3.



Устройство рибосомы



Электронные микрофотографии
индивидуальных рибосом
Escherichia coli

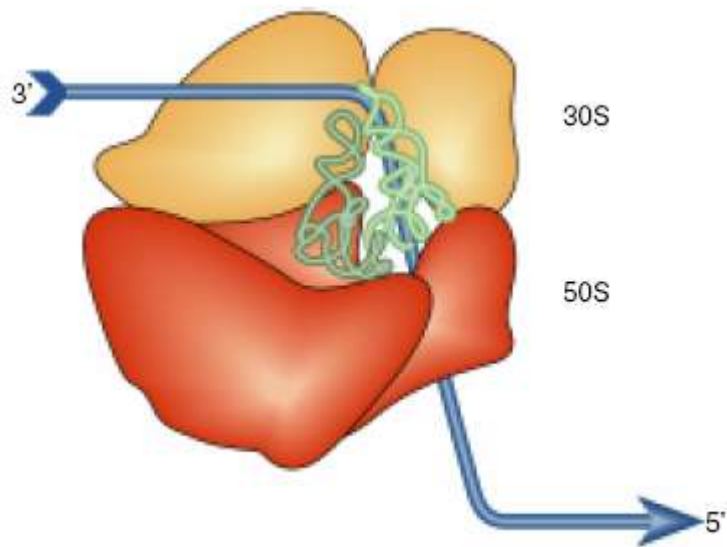
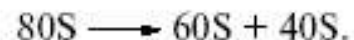
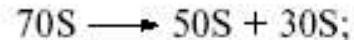


Модель рибосомы
в двух различных проекциях



Устройство рибосомы

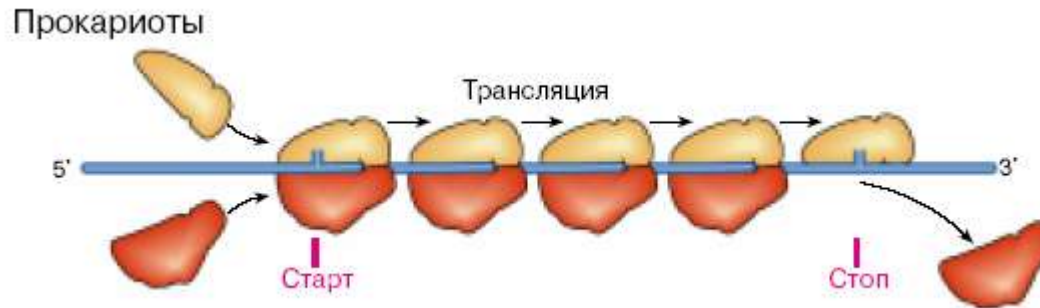
Рибосома состоит из большой и малой субчастиц. Их принято обозначать в соответствии с их коэффициентами седиментации в ультрацентрифуге (S). Бактериальная рибосома обозначается как 70S, а более крупная рибосома эукариотических организмов – 80S. Их диссоциация на субчастицы описывается следующим образом:



Размещение основных функциональных элементов – цепи мРНК и двух тРНК в рибосоме. Контуры рибосомы даны в соответствии с последними данными криоэлектронной микроскопии.



Инициация

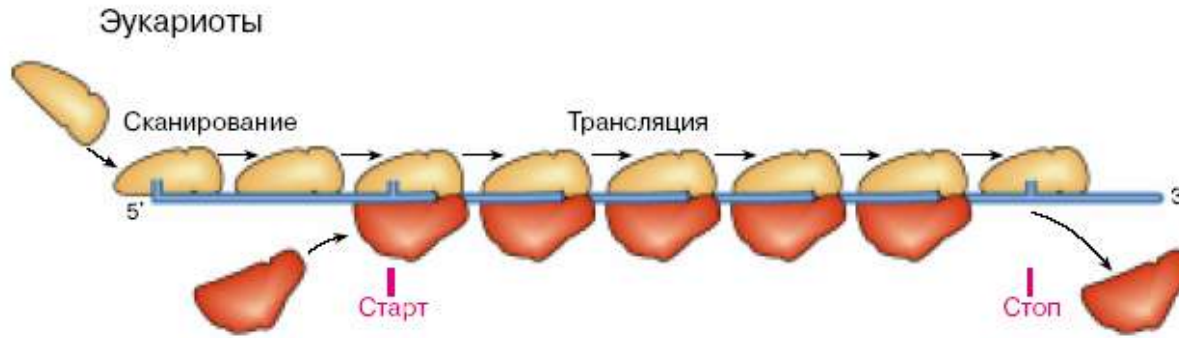


Инициация у прокариот – внутренняя: рибосомная частица связывается непосредственно с участком мРНК, содержащим иницирующий кодон (AUG).

Так как одна длинная цепь мРНК часто содержит несколько кодирующих последовательностей (полицистронные мРНК), этот способ обеспечивает инициацию трансляции сразу нескольких кодирующих последовательностей внутри мРНК независимо друг от друга.



Инициация



Инициация у эукариот – терминальная: рибосомная частица сначала присоединяется к 5'-концу мРНК, а затем движется вдоль цепи мРНК и сканирует ее, пока не встретит иницирующий кодон.



Элонгация

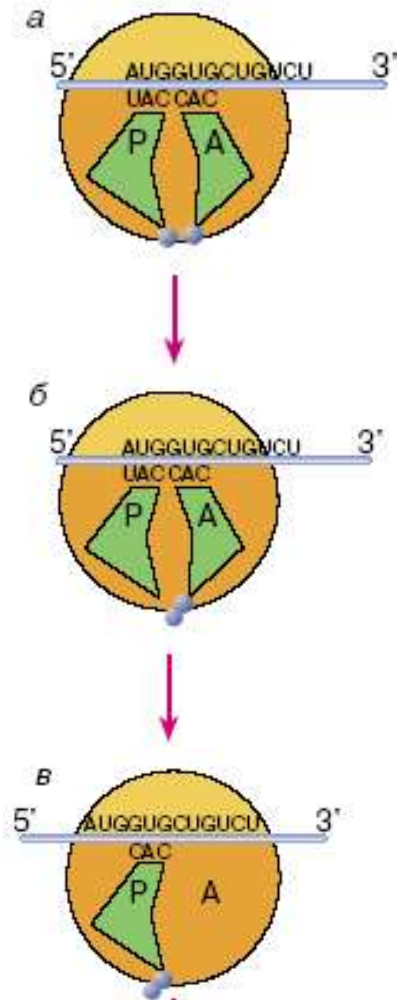


		Вторая буква								
		U		C		A		G		
Первая буква	U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
		UUC		UCC		UAC		UGC		C
		UUA	UCA	UAA		стоп	UGA	стоп	A	
		UUG	UCG	UAG		стоп	UGG	Trp	G	
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U	
	CUC		CCC		CAC		CGC		C	
	CUA		CCA		CAA	Gln	CGA		G	A
	CUG		CCG		CAG		CGG			G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U	
	AUC		ACC		AAC		AGC		C	
	AUA		ACA		AAA	Lys	AGA		Arg	A
	AUG		ACG		AAG		AGG			G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U	
	GUC		GCC		GAC		GGC		C	
	GUA		GCA		GAA	Glu	GGA		G	A
	GUG		GCG		GAG		GGG			G

Кодовый словарь (кодон-аминокислота).



Элонгация



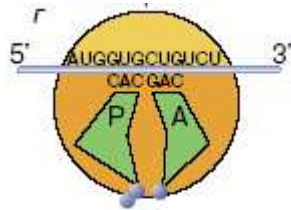
a – инициаторная аминоацил-тРНК находится в *P*-участке, и первая элонгаторная аминоацил-тРНК приходит в *A*-участок;

б– транспептидация приводит к переносу аминокислотного остатка от инициаторной тРНК на аминоацил-тРНК в *A*-участке;

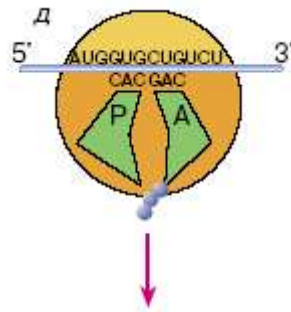
в– транслокация перемещает тРНК из *A*-участка в *P*-участок, и эта тРНК увлекает за собой связанный с ней (комплементарный) кодон мРНК.



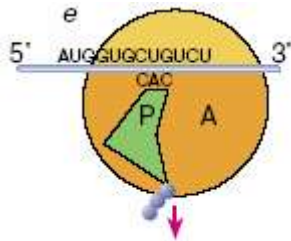
Элонгация



γ –аминоацил-тРНК, комплементарная этому кодону, связывается с *A*-участком;



δ –транспептидация переносит дипептид на аминоацил-тРНК в *A*-участке: образуется трипептидил-тРНК

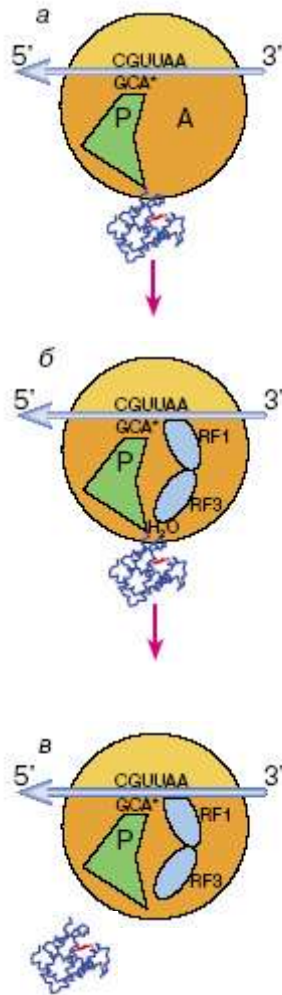


ϵ –транслокация перемещает тРНК из *A*-участка в *P*-участок, что приводит к сдвигу мРНК еще на один триплет. В *A*-участке устанавливается новый кодон.

Далее процесс продолжается по схеме (г-д-е).



Терминация



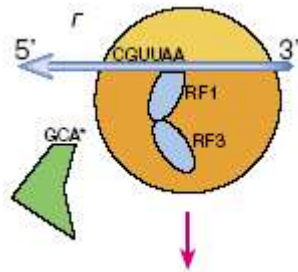
а – после добавления последнего аминокислотного остатка к растущему полипептиду и завершающей транслокации, в *A*-участке устанавливается кодон терминации (UAG, UAA или UGA).

б – *A*-участок с кодоном терминации воспринимает специальные белки – факторы терминации, непосредственно узнающие кодон терминации и взаимодействующие с рибосомой;

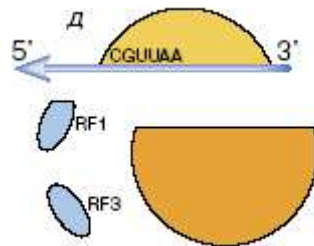
в – происходит гидролиз связи между тРНК и полипептидом и последний освобождается;



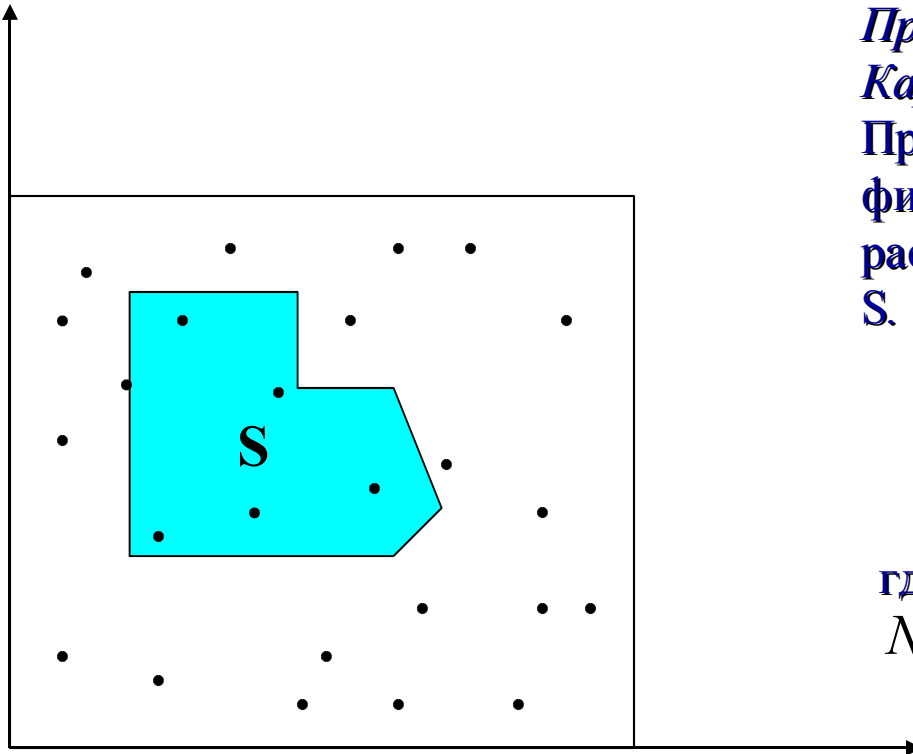
Терминация



г – тРНК освобождается из рибосомы;



д – “пустая” рибосома легко диссоциирует на субчастицы, Малая субчастица может некоторое время оставаться в лабильной ассоциации с мРНК и скользить вдоль нее, находя следующий кодон инициации (в полицистронных мРНК).



Пример простейшего метода Монте-Карло.

Приближенное вычисление площади фигуры S бросанием точек, равномерно распределенных в квадрате, содержащем S .

$$|S| \approx \frac{n_s}{N}$$

где n_s - число точек, попавших в S ,
 N - число всех точек в квадрате.



Задачи, решаемые ММК:

- 1) ММК позволяет моделировать любой процесс, на протекание которого влияют случайные факторы.
- 2) Для задач, не содержащих случайностей, можно придумать вероятностную модель, с помощью которой решается эта задача (см. пример).

Особенности ММК:

- простая структура вычислительного алгоритма (как правило, составляется программа для вычисления одного случайного события, которое затем многократно повторяется)
- ошибка вычислений обычно пропорциональна корню из (D/N) , где D – дисперсия построенной случайной величины, а N – число испытаний.



Случайная величина x называется *дискретной*, если она принимает некоторое конечное или счетное число значений $x=a_1, a_2, a_3, \dots$ с вероятностями p_1, p_2, p_3, \dots , соответственно.

Случайная величина x называется *непрерывной*, если она принимает значения из некоторого интервала (a, b) . Непрерывная с.в. определяется интервалом значений (a, b) и плотностью $p(t)$:

$$P(x \in (a_1, b_1)) = \int_{a_1}^{b_1} p(t)dt, \quad (a_1, b_1) \subset (a, b)$$



Для серии из N испытаний (x_1, \dots, x_N) математическое ожидание с.в. x приблизительно равно

$$Mx \approx \frac{1}{N} \sum_{i=1}^N x_i ,$$

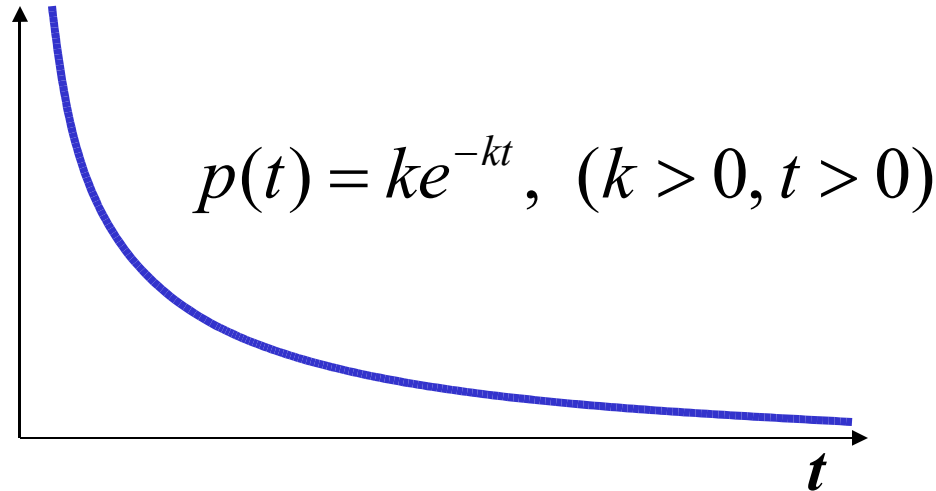
Дисперсия приблизительно равна

$$Dx \approx \frac{1}{N} \sum_{i=1}^N x_i^2 - \left(\frac{1}{N} \sum_{i=1}^N x_i \right)^2 ,$$

Сущность ММК заключается в том, чтобы построить случайную величину x с нужным распределением, и путем многократного генерирования на ЭВМ этой случайной величины приблизительно вычислить ее математическое ожидание Mx и дисперсию Dx .



Экспоненциальное распределение



Э. р. очень часто встречается в природе. По этому закону распределена средняя продолжительность жизни живого организма, длительность исправной работы механизмов, длина временного интервала между последовательными химическими реакциями и т.д. (большинство случайных процессов во времени). Коэффициент k имеет смысл частоты.



Случайным процессом назовем такую последовательность событий, что временной интервал между двумя последовательными событиями является случайной величиной с заданным распределением (здесь и далее – с экспоненциальным).

Пример 1. Разреженный газ

Молекулы газа хаотически движутся, сталкиваясь между собой. Состояние газа описывается средней скоростью и плотностью большого ансамбля частиц, из которых состоит газ. Для каждой частицы время между двумя последовательными столкновениями (t_i, t_{i+1}) является случайной величиной $\tau = (t_{i+1} - t_i)$. Эта случайная величина распределена по экспоненциальному закону с плотностью

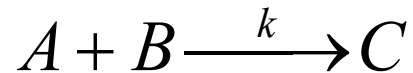
$$p(t) = ke^{-kt}$$

где частота k зависит от некоторых физических параметров частицы (скорость, площадь сечения и т.д.).



Пример 2. Химические реакции.

Рассмотрим простейшую реакцию



где k - константа скорости реакции. Вероятность того, что данная молекула вещества A прореагирует с данной молекулой в-ва B за короткий промежуток времени dt есть

$$kdt + o(dt)$$



Если в растворе имеется $[A]$ молекул А и $[B]$ молекул В, то вероятность реакции между веществами А и В за время dt есть

$$k[A][B]dt + o(dt)$$

Если обозначить момент n -й реакции через t_n , то момент следующей, $(n+1)$ -й реакции есть случайная величина t_{n+1} с плотностью

$$p(t) = k_n e^{-k_n(t-t_n)}$$

$$\text{где } k_n = k[A]_n[B]_n$$

$[A]_n$ и $[B]_n$ – количества А и В сразу после n -й реакции

Таким образом, в примерах 1 и 2, несмотря на различную природу физических процессов, они описываются на языке случайных величин одинаковым образом, образуя случайные процессы во времени с одинаковым (экспоненциальным) распределением. Это позволяет создать общую структуру стохастической модели для моделирования всех процессов такого рода.



Генерация случайных чисел на ЭВМ

Существует ряд алгоритмов получения псевдослучайных чисел, которые близки по своим распределениям к «настоящим» случайным числам и удовлетворяют принятым тестам. Все распределения получаются из равномерного путем алгебраических преобразований.

Так, если есть равномерно распределенная на отрезке $[0,1)$ случайная величина y , то с. в. x с экспоненциальной функцией распределения

$$p(t) = ke^{-kt}$$

получается по простой формуле

$$x = -\frac{1}{k} \ln(1 - y)$$

Это позволяет с минимальными затратами генерировать на ЭВМ большие выборки случайных величин с заданным экспоненциальным распределением, для стохастического моделирования случайных процессов.



Метод МК с непрерывным временем

Моменты химических реакций в растворе образуют случайный процесс с экспоненциальным распределением. Научившись моделировать такой процесс и зная константы реакций плюс начальные концентрации субстратов, можно моделировать на ЭВМ практически любую систему химических реакций, включая сложные биохимические системы.



Любая стохастическая модель такого процесса содержит следующие этапы:

- 1) Инициализация (задание констант реакций, начальных концентраций и др. условий)*
- 2) Выбор реакции*
- 3) Выполнение реакции*
- 4) Обновление данных*
- 5) Окончание или повтор шагов 2)-4)*

Алгоритмы различаются в основном по реализации шагов 2) и 4).



Стохастическая модель элонгации

Основные переменные:

t – момент последнего случившегося события

$t < t_1, t_2, \dots, t_n$ - упорядоченный по возрастанию список моментов будущих событий
(n – число всех типов событий)

t_{max} – время счета



Общий алгоритм модели (на основе алгоритма Гибсона):

- 3) Инициализация переменных, $t=0$; Вычисление списка моментов t_1, t_2, \dots, t_n
- 4) Выбор события с минимальным временем (t_1). Если событие допустимо, идем на шаг 3). Если нет, ищем первое допустимое (t_i). Полагаем $t=t_i$.
- 5) Реализация выбранного события.
- 6) Обновление изменившихся переменных. Назначение событию нового значения (t_i), которое встраивается в список t_1, t_2, \dots, t_n согласно его величине.
- 7) Остановка счета, если $t > t_{max}$, иначе на шаг 2).



Элементарными событиями в модели являются:

- посадка i -й рибосомы на старт-кодон мРНК
- присоединение заряженной тРНК к кодону в А-сайте i -й рибосомы
- транслокация рибосомы (если рибосома перед ней не блокирует путь)
- терминация трансляции



Синонимичные кодоны

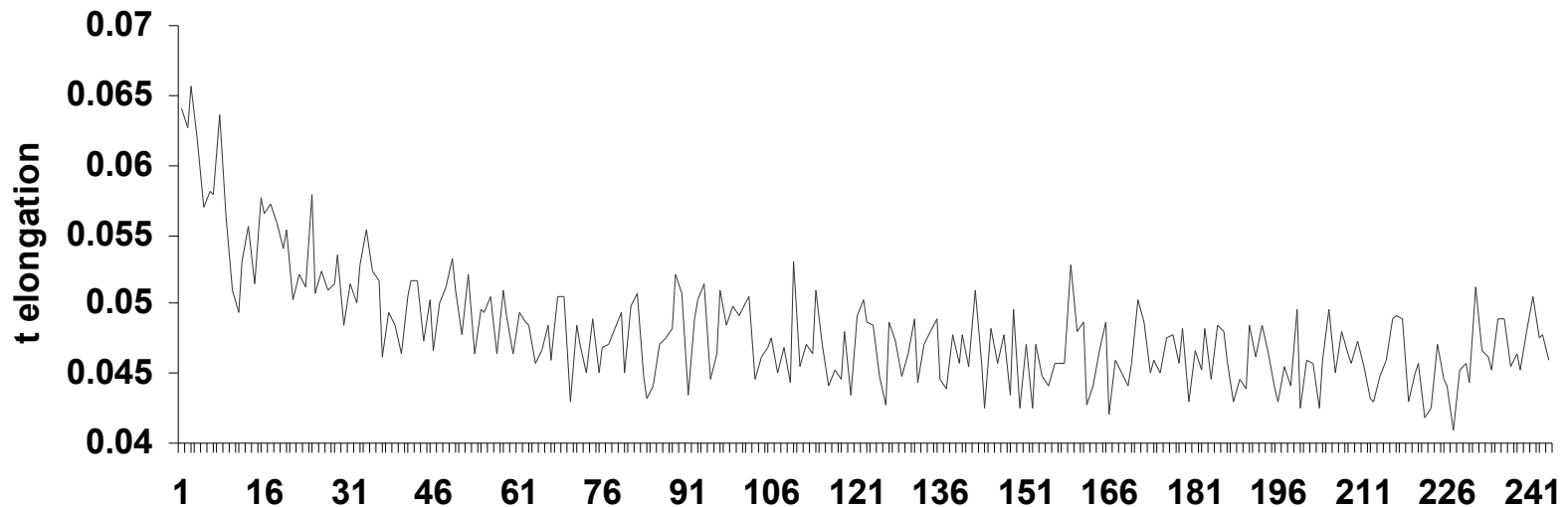
Использование синонимичных кодонов с разными скоростями элонгации является распространенным механизмом оптимизации трансляции. В мРНК с высоким уровнем экспрессии чаще используются «оптимальные» кодоны, имеющие наибольшую скорость считывания, которым соответствуют аминоацил-тРНК в наибольших концентрациях. Для ряда организмов (*E. coli*, *S. Cerevisiae* и др.) уровень экспрессии генов коррелирует с частотой использования в них оптимальных кодонов. Это позволяет оценивать уровень экспрессии генов по их кодонному составу.



Результаты



Расчеты последовательностей мРНК *E.coli* демонстрируют известный эффект концентрации неоптимальных кодонов в начале мРНК у генов с высоким уровнем экспрессии. Это, по-видимому, предотвращает излишнее скопление рибосом на кодирующей части мРНК, которое снижало бы эффективность трансляции клетки в целом:



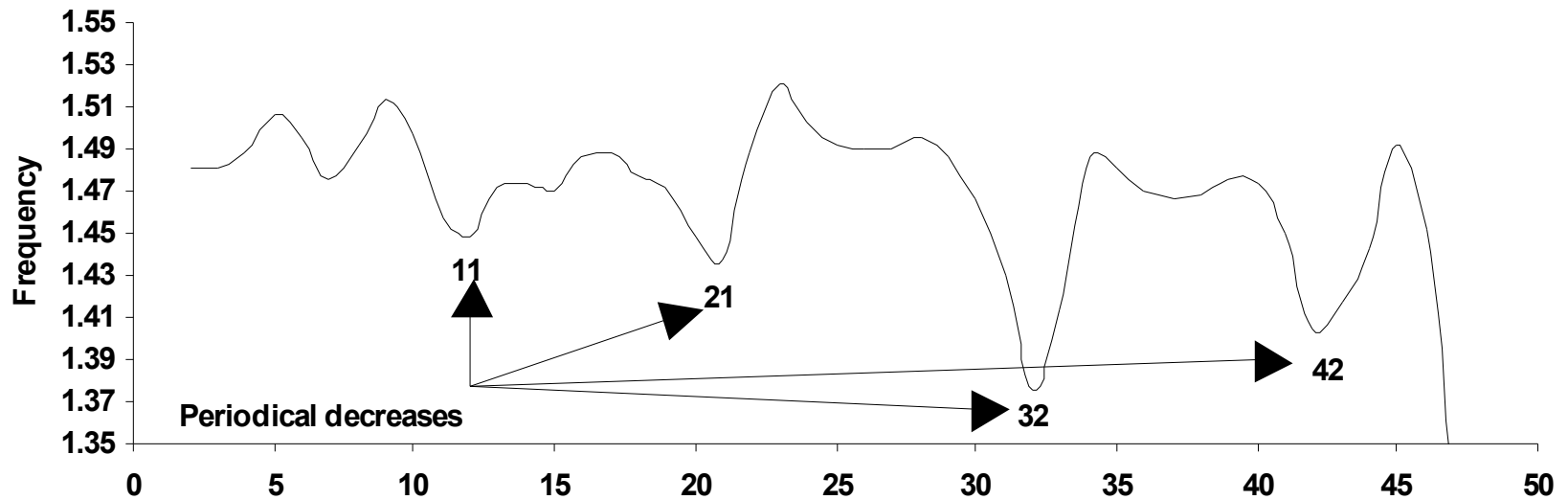
Усредненный профиль элонгации мРНК высокоэкспрессирующихся генов E. coli. Ось абсцисс – номер кодона в рамке трансляции: №1 соответствует кодоону, следующему за инициаторным кодоном, ось ординат – усредненное время, затрачиваемое рибосомой на присоединение одного аминокислотного остатка к растущему полипептиду



Результаты



Анализ частот кодонов в генах *E. coli* показал наличие периодичностей в распределении неоптимальных кодонов в мРНК клетки.



Средние частоты кодонов, лежащих перед редким кодоном на расстоянии D. Ось абсцисс – расстояние D (в кодонах), ось ординат – средняя частота кодона.



Результаты



Найденная длина периода равна 11 ± 1 кодонов. Понижения частот повторяются на расстояниях D , $2D$, $3D$, $4D$ перед медленным кодоном (см. рис.), что согласуется с предсказанным поведением частот на «идеальной» мРНК. Найденное значение $D=11 \pm 1$ кодонов (33 ± 3 нуклеотида) можно интерпретировать как стерический размер рибосомы. Это значение согласуется с размером рибосомы на мРНК, полученным в экспериментальных работах других авторов – около 30 нуклеотидов.



Разработанная модель может быть полезна при исследовании задач теории эволюции и решении задач биоинформатики, связанных с оптимизацией кодонного состава мРНК.

В лекции были использованы материалы публикаций проф. А.С. Спирина в Соросовском образовательном журнале.



Литература



- Льюин Б., Гены: Пер. С англ. – М.: Мир, 1987
- Албертс Б., и др., Молекулярная биология клетки, т. 1, М.: Мир, 1994.
- Соболь И.М., Метод Монте-Карло, М., Наука, 1968 («Популярные лекции по математике», вып. 6).
- Спирин А.С., Биосинтез белка: элонгация полипептида и терминация трансляции. // Соросовский Образовательный Журнал. 1999. №6. С 2-7.
- Спирин А.С., Принципы функционирования рибосом // Соросовский Образовательный Журнал. 1999. № 4. С. 2–9
- Спирин А.С., Принципы структуры рибосом // Соросовский Образовательный Журнал. 1998. №11. С 65-70



**Спасибо за
внимание**