

Лекция 12-ая Часть 1 «Трансляция: компьютерный анализ и моделирование»

Часть 2 «Теоретико-компьютерные исследования процесса трансляции»

Часть 1

Лектор К.б.н., с.н.с. А.В.Кочетов

План лекции:

Лекция посвящена моделированию процесса трансляции и будет состоять из двух частей. В первой части я расскажу об основных характеристиках процесса трансляции. Вторая часть будет посвящена собственно моделированию, ее прочтет Юрий Георгиевич Матушкин.

СЛАЙД 1.

Трансляция - один из основных фундаментальных процессов живой клетки. Функцию хранителя генетической информации выполняет ДНК, РНК играет роль посредника между геномом и белок-синтезирующим аппаратом, белки выполняют подавляющее большинство жизненных функций клетки. Трансляция - это процесс перехода генетической информации от мРНК к белку, и в этом качестве изучение ее законов безусловно необходимо для создания сколь-нибудь внятного представления о структуре клетки.

СЛАЙД 2.

Основное содержание процесса трансляции заключается в следующем: в структуре мРНК расположена белок-кодирующая область, несущая информацию о структуре белка. Эта информация представляет собой непрерывную последовательность тринуклеотидных сочетаний - кодонов, генетический код - то есть соответствие кодонов аминокислотам - приведено здесь в виде таблицы. Напомню Вам, что триплетных комбинаций из 4 нуклеотидов может быть 64, три из них являются терминаторами, а каждой из 20 аминокислот соответствует от одного до 6 кодонов. В чем заключается это соответствие? Основные элементы трансляционной машины, это рибосома - сложный РНК-белковый комплекс и тРНК - посредники между пулом свободных аминокислот и трансляционным аппаратом. В структуре тРНК выделяют так называемый антикодон - триплет, комплементарный кодону и способный взаимодействовать с ним в силу этой своей комплементарности. Специальные ферменты - аминоацил тРНК-синтетазы - распознают "свои" тРНК, соответствующие каждой аминокислоте, и осуществляют связывание аминокислот и тРНК, в силу чего такие аминоацилированные тРНК несут аминокислоты. В этом, собственно, уже заключается связь между генетическим кодом - представленным антикодоном тРНК, и аминокислотой, с этой тРНК сцепленной.

СЛАЙД 3.

Далее, роль рибосомы заключается в обеспечении контакта между кодонами на мРНК и антикодонами тРНК, а также снятии аминокислот с тРНК и линковке их в аминокислотную последовательность. Для этого в структуре рибосомы есть два центра - аминоацильный (А) и пептидильный (Р). тРНК с синтезируемой аминокислотной цепью расположена в Р-центре, в А-центр поступает аминоацилированная тРНК, комплементарная следующему кодону. После этого аминокислотной отсоединяется от тРНК в Р-центре и переносится на новопришедшую тРНК, удлиняясь при этом на одну аминокислоту. тРНК из А-центра уходит, рибосома сдвигается вдоль мРНК на один кодон, и ситуация повторяется. В результате получается аминокислотная последовательность, соответствующая последовательности кодонов в мРНК и, соответственно, гене, которая укладывается в рабочую конформацию и начинает выполнять свою функцию.

Это - в самых общих чертах - описание сути процесса трансляции.

СЛАЙД 4. Теперь давайте подробнее остановимся на связи между структурой мРНК и процессом трансляции. На рисунке представлены про- и эукариотические мРНК. Как Вам, наверное, известно, зрелая эукариотическая мРНК состоит из трех частей: 5'-нетранслируемого района или лидера, кодирующей части (которая начинается со стартового кодона трансляции и заканчивается кодоном-терминатором), и 3'-нетранслируемого района (иногда называемого трейлером). Соответственно, процесс трансляции также делится на три этапа: инициация, элонгация и терминация. Каждый из районов мРНК выполняет в соответствующем процессе трансляции определенную роль, о чем я вкратце и расскажу.

Инициация трансляции.

Этот процесс весьма существенен, так как в составе мРНК нет специальных кодонов для начала процесса трансляции, эту роль играет метиониновый ATG кодон, а эта комбинация нуклеотидов присутствует и в некодирующих районах, и в кодирующей части в разных рамках считывания. Поэтому проблема выбора того самого особенного ATG из множества имеющихся, не тривиальна. Первое, о чем нужно сказать: инициация трансляции в про- и эукариотах качественно различна, и это различие обеспечивает существенную разницу в структуре их геномов в целом.

Прокариотические мРНК содержат специальный сигнал, с которым взаимодействуют рибосомы в поиске сайта инициации трансляции: так называемый, сайт Шайна-Дальгарно. Это короткая последовательность (5-6 нуклеотидов), похожая на GGAGGA и расположенная на расстоянии 6-10 нуклеотидов выше стартового кодона трансляции белок-кодирующей части. Как ее распознает рибосома? - опять же, используя комплементарные взаимодействия: в составе консервативного 3'-конца 16S рибосомной РНК, входящей в состав 30S субъединицы рибосомы, есть комплементарный участок, который и выполняет роль детектора сигнала. Именно комбинация AUG кодона и сайта Шайна-Дальгарно обозначает для рибосомы начало трансляции. Нужно отметить, что 30S субъединица рибосомы распознает AUG кодон с помощью инициаторной метиониновой тРНК - именно антикодон этой тРНК является детектором.

Вторая характеристика прокариот, несвойственная эукариотам - оперонная структура мРНК, то есть в составе многих мРНК находятся две и более белок-кодирующие последовательности разных белков. Рибосома распознает сайт инициации трансляции в начале матрицы (комбинацию сайта Ш-Д и AUG кодона) и транслирует проксимальную белок-кодирующую последовательность, затем часть рибосом диссоциирует с матрицы, а часть может реиницировать трансляцию на следующей кодирующей последовательности. Альтернативно, в межцистронном промежутке может располагаться независимый сайт инициации трансляции. Тогда рибосомы будут садиться в межцистронном промежутке. В совокупности, около половины генов *E. coli* входят в состав оперонов. Это дает определенные возможности для контроля экспрессии генов - например, гены одной метаболической цепи транскрибируются с одного промотора, а соотношение белков, считываемых с одной полицистронной мРНК, может определяться структурой межцистронных стыков.

Что касается эукариотических генов, то там ситуация другая. Во-первых, в эук. мРНК нет никакого аналога сайта Ш.-Д. - соответствующая этому сайту область в эукариотической 18S рибосомной РНК отсутствует. Во-вторых, эукариотические мРНК моноцистронны и содержат только одну белок-кодирующую рамку. В экспериментах показано, что при трансляции модельной полицистронной матрицы в клетках эукариот дистальные цистроны почти не транслируются. Таким образом, структура мРНК про- и эукариот и особенности трансляционных аппаратов этих таксонов отражаются и на структуре генома.

Инициация трансляции в клетках эукариот - вопрос - как и большинство описанных - неоднозначный. Существенными характеристиками эук. матриц являются специфические модификации - на 5'конце у них расположен так называемый кэп - инвертированный гуанин, а на 3'конце зрелой молекулы мРНК расположен поли(А)-хвост из нескольких сотен

нуклеотидов.

СЛАЙД 5.

На этом рисунке представлена структура кэпа - видно, что вместо обычно 3'-5' концевой связи между сахарофосфатным остовом здесь сформирована перевернутая - 5'-5'.

СЛАЙД 6.

Кэп сайт является меткой, по которой рибосомы распознают большинство мРНК в качестве матрицы для трансляции: есть и другие механизмы, открытые в последнее время, но этот является одним из основных. На этом рисунке представлена схема взаимодействия трансляционного аппарата эукариотической клетки и мРНК: считается, что с кэпом связан фактор инициации трансляции 4F, состоящий из трех белков - кэа-связывающего фактора 4E, РНК-геликазы 4A и белка 4G. 40S субъединица рибосомы также несет факторы инициации трансляции и метиониновую тРНК, антикодон которой способен распознать AUG кодон. И малая субъединица рибосомы, и фактор инициации трансляции 4A обладают ограниченной геликазной активностью и способны расплетать вторичную структуру мРНК, которая может формироваться вследствие комплементарных взаимодействий.

СЛАЙД 7.

Считается, что большинство мРНК транслируются по механизму линейного сканирования: ее последняя модификация выглядит примерно так. мРНК представляет собой кольцо - ее 5' и 3' концы сближены за счет бело-белкового взаимодействия между поли(А)-связывающим белком и кэп-связывающим фактором 4F. Полагают, что это способствует рециркуляции рибосом - то есть после терминации трансляции субъединицы рибосом оказываются вблизи 5'конца и - после реактивации - могут снова участвовать в инициации.

Согласно модели линейного сканирования, 40S субъединица рибосомы движется от 5'конца вдоль лидерной последовательности до тех пор, пока не встретит подходящий AUG кодон. "Подходящий" в данном случае означает то, что вероятность его распознавания в качестве сайта инициации трансляции зависит от контекста - например, у млекопитающих особое значение имеют аденин в -3 положении перед AUG и гуанин в плюс четвертом положении. Если контекст другой, часто 40S субъединицы его не распознают и продолжают сканирование в поиске другого AUG.

СЛАЙД 8.

Не так давно стало известно, что возможны и другие механизмы инициации трансляции, отличные от линейного сканирования. На этом рисунке приведен механизм внутренней инициации трансляции, открытый у пикорнавирусов - у них мРНК содержит длинный - до 800 нуклеотидов, высокоструктурированный и насыщенный AUG кодонами лидер, поэтому эффективная инициация трансляции с помощью линейного сканирования невозможна. Показано, что 5'НТП таких мРНК формирует сложную структуру, функционально замещающую кэп и обеспечивающую посадку 40S субъединицы рибосомы в непосредственной близости от стартового кодона трансляции. При этом между местом посадки рибосомы и стартом трансляции нет ни стабильных шпилек, ни паразитных AUG кодонов.

СЛАЙД 9.

На этом рисунке изображена модель сайта внутренней инициации трансляции одного из пикорнавирусов. Видно, что эта структура очень сложна. Однако, в последнее время найдены другие варианты сайтов внутренней инициации трансляции со значительно менее сложной структурой, так что ни механизмы работы, ни структура этих сайтов до конца не известны.

СЛАЙД 10.

На этом слайде отображены различия между линейным сканированием и внутренней инициацией трансляции. Показаны уязвимые места линейного сканирования - стабильные шпильки и AUG кодоны в составе 5'НТП. Для эффективной инициации трансляции 40S субъединицы рибосом должны легко проходить 5'НТП, поэтому такие негативные характеристик несвойственны мРНК генов высокого уровня экспрессии. В то же время, сайт внутренней инициации трансляции предъявляет совсем другие требования к 5'НТП - а именно, к вторичной структуре и сайтам связывания некоторых белков.

Существование двух механизмов инициации трансляции предоставляет широкие возможности как для частного, так и для общего контроля экспрессии генов. На СЛАЙДЕ 11 изображен механизм контроля экспрессии нескольких генов, регулируемых поступлением железа. Сенсорная молекула - Iron Responsive Protein - изменяет конформацию в зависимости от наличия в клетке железа. В мРНК генов, контролируемых этим фактором, есть специальный сайт связывания, размещенный в непосредственной близости от кэпа. Если IRP активен, он связывается с этим сигналом и ингибирует связывание фактора инициации трансляции 4F и кэпа. После этого мРНК выводится из трансляционного процесса, так как рибосомы перестают считать ее матрицей. Если железо появляется, IRP диссоциирует и трансляция возобновляется.

СЛАЙД 12

На следующем рисунке изображен механизм общего контроля трансляции. Под воздействием стресса, гормонов и т.п. вработывают регуляторные каскады, активирующие белок 4E-BP. Этот белок обладает высокой аффинностью к компоненту фактора инициации трансляции 4F - белку 4E, чья роль заключается в распознавании кэпа. Если этот белок выведен из оборота, трансляция большинства мРНК останавливается. В этой связи следует отметить, что мРНК, содержащие сайт внутренней инициации трансляции не подвержены этому механизму, так как не содержат кэпа и их трансляция не зависит от этого белка.

Наконец, резимируя, можно сказать следующее - и это уже имеет отношение к тому, что будет говорить Юрий Георгиевич.

Известно (СЛАЙД 13), что эукариотически мРНК различаются по интенсивности трансляции - так, если внести в систему трансляции *in vitro* две мРНК, то скорость наработки белка может различаться во много раз. Почему так происходит? - эффективность трансляции зависит от характеристик матрицы. Каждый из многочисленных этапов трансляции происходит с определенной скоростью - например, взаимодействие фактора 4F и кэпа, связывание с малой субъединицей рибосомы, движение рибосомы по лидерной последовательности, распознавание AUG кодона - это процессы, скорость которых зависит от нуклеотидной последовательности и ее характеристик. Аналогичным образом варьирует скорость элонгации и терминации трансляции.

Тот факт, что далеко не все такие характеристики известны, не позволяет предсказывать трансляционную активность мРНК с абсолютной точностью, хотя некоторые такие подходы уже разработаны.

Часть 2

ТЕОРЕТИКО-КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ ПРОЦЕССА ТРАНСЛЯЦИИ
Лектор Зав. сектором молекулярной эволюции ИЦиГ СО РАН, к.б.н.,
с.н.с. Ю.Г. Матушкин.

На рис.2 изображена схема процесса трансляции прокариотической мРНК. Трансляция начинается с инициации - сборки рибосомы на AUG-кодоне. В процессе элонгации - движения рибосомы вдоль мРНК, рибосоме может мешать возникновение локальных вторичных структур - шпилек, возникающих при наличии близко расположенных

инвертированных повторов. Кроме того, скорость элонгации зависит от скорости прохождения рибосомой конкретного кодона, которая зависит от количества тРНК, специфичных к этому кодону. Существуют "быстрые" кодоны, для которых есть много тРНК, несущих нужную аминокислоту, и "медленные", для которых тРНК мало. Для некоторых организмов "быстрые" кодоны значительно чаще встречаются в высокоэкспрессирующихся последовательностях, чем в низкоэкспрессирующихся.

Наличие в генетическом коде (рис.3) синонимичных серий кодонов (кодирующих одну и ту же аминокислоту), позволяет организму использовать разные по "скорости" кодоны для разных по степени экспрессии фракций мРНК.

Однако, далеко не для всех организмов "быстродействие" составляющих мРНК кодонов однозначно определяет ее экспрессивность (рис.4).

Нами разработан индекс эффективности элонгации (ЕЕI) (рис.5), учитывающий как кодонный состав мРНК (первое слагаемое), так и эффект локальных комплементарностей - локальных шпилек (второе слагаемое). Анализ 78 видов одноклеточных показал, что они разделяются на 5 групп, в зависимости от того, какой процесс вносит наибольший вклад в скорость элонгации. И только в одной группе адекватно работает "частотно-кодонная" разновидность ЕЕI, аналогичная по своему биологическому смыслу индексу адаптивности кодонов (САI), который разработал в 80-х годах Поль Шарп.

На рис.6 показана формула "частотно-кодонной" части ЕЕI. Фактически, значение $T_a(i)$ будет средним рассчитанным временем, которое рибосома потратила на ожидание соответствующих тРНК с аминокислотами в процессе элонгации i -ой мРНК. А $\beta_{\alpha(i,j)}$ является в некотором смысле эквивалентом времени, которое рибосома провела на j -ом кодоне, дожидаясь соответствующей тРНК с аминокислотой. Чем больше молекул тРНК для этого кодона, тем быстрее какая-нибудь из них свяжется с рибосомой и несомая ею аминокислота будет присоединена к полипептидной цепи.

На рис.7 показана формула "локально-комплементарной" части ЕЕI. $T_e(i)$ оценивает по уровню автокомплементарности i -й мРНК среднее время, затрачиваемое рибосомой на стадию транслокации: $T_e(i) = t_{\min} \cdot (1-p(i)) + t_{\max} \cdot p(i)$. При этом $p(i)$ - это сигмоидная функция, интеграл которой стремится к 1 при увеличении $LCI(i)$, значение которого определяется количеством и энергопотенциалом вторичных структур (шпилек), которые могут образовываться на i -й мРНК.

На рис.8 показана формула, определяющая значение $LCI(i,j)$, где i - это номер рассматриваемой мРНК, а j - номер нуклеотида, для которого будут подсчитаны все вторичные структуры (шпильки) заданной конфигурации (с учетом энергии и без оного), в которые он входит. Например, конфигурация шпилек может быть такой: длины инвертированных повторов от 3 до 9 нуклеотидов, расстояние между повторами от 3 до 50. На рисунке изображены 2 инвертированных повтора длиной 9, расстояние между повторами равно 18. Соответственно, они образуют шпильку со стеблем, длиной 9 нуклеотидов и петель, длиной 18. Если индекс локальной комплементарности подсчитывается с учетом энергии вторичных структур, то используются значения энергий связи динуклеотидов, также показанных на рисунке.

На рис.9 показана формула, определяющая значение так называемую ζ -форму индекса локальной комплементарности $LCI^{\zeta}(i)$ для i -й мРНК. Без учета энергии вторичных структур подсчитывается количество шпилек со стеблем и пельей определенных длин (не более и не менее заданных параметров). Все это суммируется по нуклеотидам, а потом нормируется на

их количество. В результате получается число, характеризующее количество потенциальных шпилек, могущих образоваться на данной мРНК.

На рис.10 показана формула, определяющая значение так называемую Ψ -форму индекса локальной комплементарности $LC\Psi(i)$ для i -й мРНК. От предыдущей она отличается тем, что при подсчете учитывается энергия вторичных структур, и число, характеризующее мРНК становится тем больше, чем "крепче" окажутся шпильки могущих на ней образоваться.

На рис.11 показана формула критерия адекватности индекса эффективности элонгации. После того, как мы упорядочили гены по значению EI , необходимо оценить, насколько этот порядок адекватен реальной эффективности элонгации. В качестве критерия мы используем гены рибосомных белков, поскольку известно, что именно эти гены экспрессируются в любой клетке очень активно. Т.е., если в полученном нами распределении генов гены рибосомных белков будут статистически достоверно опознаваться как высокоэкспрессирующиеся, то использованная форма индекса EI адекватно работает в данном организме. Фактически подсчитывается "сдвиг" выборки рибосомных генов от случайного распределения. Если величина сдвига $d(\xi)$ близка к 0, это означает, что рибосомные гены распределены близко к случайному образу, и использованная форма индекса EI неадекватно отражает зависимость эффективности элонгации от нуклеотидного состава. Если величина сдвига $d(\xi)$ близка к 100, это, наоборот, означает адекватность индекса - рибосомные гены распознаны как высокоэкспрессирующиеся. Если величина сдвига $d(\xi)$ близка к -100, это означает, что рибосомные гены распознаны как низкоэкспрессирующиеся, что неправильно, но, тем не менее, можно сделать вывод, что индекс отражает какие-то закономерности.

На рис.12 изображена графическая иллюстрация к смыслу "сдвига" распределения рибосомных генов.

На рис.13, 14, 15, 16, 17 приведены результаты исследования эффективности экспрессии генов ряда организмов. Эти результаты представлены также в Таблице 1. Видно, что ни одна из пяти принципиально возможных групп (A, ξ , Ψ , A ξ , A Ψ) не является пустой. В первую входит 32 организма. У них критичной для элонгации является только стадия присоединения тРНК. Вторичные структуры на эффективность элонгации в данных организмах не влияют.

Вторую группу составляют 8 организмов. В них только стадия транслокации является критичной. Ее скорость падает при повышении уровня локальных комплементарностей. При этом энергия связей не играет роли. Третья группа состоит из одного организма *P. aeruginosa PA01*. У данного организма эффективность элонгации, также определяется только стадией транслокации, однако энергия локальных вторичных структур является важным фактором, влияющим на скорость движения рибосомы. Во второй и третьей группе стадия присоединения тРНК не критична для эффективности элонгации и в этом смысле эти группы являются противоположными первой группе.

В четвертую группу входит 30 организмов, в пятую - три. В организмах четвертой и пятой групп обе стадии совместно влияют на эффективность элонгации. Но в четвертой, как и во второй группе, эффективность стадии транслокации зависит только от наличия комплементарных участков, но не от энергии формирования вторичных структур, связанных с ними. В пятой, также как и в третьей группе, эффективность элонгации зависит и от энергии вторичных структур.

Отметим некоторые закономерности, вытекающие из анализа результатов, представленных в Таблице 1. Во-первых, бросается в глаза, что самыми многочисленными группами, являются

А- и А^ζ-группы. В общей сложности в них входит 62 организма из 74 проанализированных. Если к ним прибавить организмы ^ζ-группы, то окажется, что в 70 организмах энергия вторичных структур не лимитирует процесс элонгации и только для четырех организмов энергия вторичных структур играет существенную роль в определении эффективности элонгации. Во-вторых, все археобактерии попадают в А^ζ-группу. Эволюционно близкие виды эубактерий также в подавляющем большинстве попадают в одни и те же группы. Эукариоты, также находятся в одной А-группе. Следовательно, эволюционно близкие виды имеют тенденцию к размещению в одной группе. Полученные результаты позволяют предложить, несколько биологически значимых интерпретаций.

Таблица 1. Разбиение одноклеточных организмов по группам.

Наименования организмов	
<i>A. tumefaciens</i> C58, <i>B. halodurans</i> C-125, <i>B. subtilis</i> , <i>B. melitensis</i> , <i>C. muridarum</i> , <i>C. trachomatis</i> , <i>C. pneumoniae</i> , <i>C. pneumoniae</i> AR39, <i>C. pneumoniae</i> J138, <i>E. coli</i> K12, <i>E. coli</i> O157 H7, <i>E. coli</i> O157 H7 EDL933, <i>H. influenzae</i> , <i>L. innocua</i> , <i>L. monocytogenes</i> EGD-e, <i>M. loti</i> , <i>M. leprae</i> , <i>P. multocida</i> , <i>S. typhi</i> , <i>S. Typhimurium</i> LT2, <i>S. meliloti</i> , <i>S. aureus</i> Mu50, <i>S. aureus</i> N315, <i>S. pneumoniae</i> R6, <i>S. pneumoniae</i> TIGR4, <i>S. pyogenes</i> , <i>S. pyogenes</i> strain MGAS8232, <i>Synechocystis</i> PCC6803, <i>V. cholerae</i> , <i>Y. Pestis</i> , <i>S. cerevisiae</i> , <i>S. pombe</i>	группа А (критичным для элонгации является кодонный состав гена)
<i>B. burgdorferi</i> , <i>Buchnera</i> sp APS, <i>C. jejuni</i> , <i>H. pylori</i> 26695, <i>H. pylori</i> J99, <i>M. genitalium</i> , <i>M. pulmonis</i> , <i>U. urealyticum</i>	группа ^ζ (критичным для элонгации является уровень локальных комплементарностей, энергия не играет значения)
<i>P. aeruginosa</i> PA01	группа Ψ (критичным для элонгации является уровень локальных комплементарностей с учетом энергии вторичных структур)
<i>A. aeolicus</i> , <i>C. crescentus</i> , <i>C. perfringens</i> , <i>C. acetobutylicum</i> , <i>D. radiodurans</i> R1, <i>F. nucleatum</i> ATCC 25586, <i>L. lactis</i> , <i>M. tuberculosis</i> H37Rv, <i>M. tuberculosis</i> CDC1551, <i>M. tuberculosis</i> H37Rv, <i>M. pneumoniae</i> , <i>Nostoc</i> sp PCC 7120, <i>R. conorii</i> Malish 7, <i>R. prowazekii</i> , <i>T. maritime</i> , <i>T. pallidum</i> , <i>A. pernix</i> K1, <i>A. fulgidus</i> , <i>Halobacterium</i> sp NRC-1, <i>M. thermoautotrophicum</i> , <i>M. jannaschii</i> , <i>M. kandleri</i> strain AV19, <i>M. acetivorans</i> str.C2A, <i>P. aerophilum</i> , <i>P. abyssi</i> , <i>P. horikoshii</i> , <i>S. solfataricus</i> , <i>S. tokodaii</i> , <i>T. acidophilum</i> , <i>T. volcanium</i>	группа А ^ζ (критичными для элонгации, являются и кодонный состав и уровень локальных комплементарностей, энергия не играет значения)

N. meningitidis MC58, *N. meningitidis* Z2491, *X. fastidiosa*

группа АΨ (*критичными для элонгации, одновременно, являются кодонный состав и уровень локальных комплементарностей, с учетом энергии*)

Создается впечатление, что в ходе эволюции организмы использовали разные стратегии оптимизации стадии элонгации. В организмах самой многочисленной А-группы, выработались настолько эффективные средства физического перемещения мРНК в рибосоме (или рибосомы вдоль мРНК), что стадия транслокации перестала быть (если вообще была когда-либо) лимитирующим фактором элонгации. В этих организмах отбор, направленный на повышение эффективности протекания стадии элонгации достигался за счет оптимизации кодонного состава генов.

Напротив, для организмов ζ- и Ψ- групп можно предположить, что стадия транслокации является чувствительной к локальным препятствиям, которые могут формироваться, в частности, и за счет локальных комплементарностей. В то же время размещение тРНК в А-сайте рибосомы организмов данных групп либо происходит приблизительно с одинаковой эффективностью для всех кодонов, либо протекает параллельно с процессом, обеспечивающим удаление препятствий перед рибосомой (предподготовка стадии транслокации), и второй процесс медленнее первого, так что он как бы экранирует первый процесс и, тем самым, обеспечивает эволюционную нейтральность кодонных мутаций и эволюционную чувствительность к мутациям, обеспечивающим пониженный уровень локальных комплементарностей. В остальных группах обе стадии существенно влияют на эффективность элонгации и поэтому оптимизации подверглись обе характеристики: кодонный состав и локальные комплементарности.

Наличие ζ- и Аζ-групп, с одной стороны, Ψ- и АΨ-групп, с другой, позволяет предположить, что за этим могут стоять различные механизмы преодоления рибосомой на стадии транслокации помех, создаваемых локальными вторичными структурами. В порядке гипотезы можно предположить, что в организмах ζ- и Аζ-групп наличие препятствия перед рибосомой запускает механизм, который каждый раз затрачивает определенную порцию ресурсов (временных, энергетических), удаляя с участка мРНК определенной длины все помехи, независимо от их энергии и количества. Напротив, в организмах из Ψ- и АΨ-групп механизм преодоления помех каким-то образом чувствует "мощность" помехи и затрачивает на ее преодоление пропорциональное количество времени (и, возможно, энергии). Следует отметить, что общее количество организмов, в которых выявлена чувствительность стадии транслокации к энергии вторичных структур оказалось незначительным, всего 4.

На рис.18,19 изображен профиль значений индекса локальной комплементарности (LCI) вдоль мРНК *E. Coli* и интерпретация показанных результатов. Последовательности были упорядочены по значению индекса эффективности элонгации (EEI) и из них были выделены 2 подгруппы: высокоэкспрессирующиеся (Hi) и низкоэкспрессирующиеся (Low). Для каждой группы был подсчитан профиль индекса локальной комплементарности в окрестности иницирующего и терминирующего кодонов. Как видно из рисунка имеет место "яма" вокруг иницирующего кодона - т.е. в этой области в результате отбора первичная структура образует существенно меньше шпилек, что может быть связано с необходимостью легкой посадки рибосомы.

Характерно, что для высокоэкспрессирующихся последовательностей "яма" глубже. Также характерно, что пик на 3'-конце в районе терминирующего кодона существенно выше для высокоэкспрессирующихся последовательностей - возможно, это помогает рибосоме быстрее освободиться от мРНК. Этот огромный пик на 3'-конце формируется реальными

3'-концами мРНК, т.к. на 3'-концах генов внутри оперонов он отсутствует. Возможно, это связано со стабильностью мРНК по отношению к нуклеазам или с усилением терминации трансляции и/или транскрипции.

Яма на 5'-конце своими размерами коррелирует со стерическими размерами рибосомы. Также это связано с оперонной структурой генома *E.coli*, т.к. стартовый кодон одного гена может находиться в кодирующей части другого.

E. Coli принадлежит к группе А, т.е. для адекватности индекса эффективности экспрессии в кодирующих районах нет нужды использовать данные о локальной комплементарности. Однако, как можно видеть, вторичная структура может играть роль в изменении эффективности экспрессии гена, связанной с процессами инициации и терминации.

На рис.20 изображена схема оперонной структуры генов прокариот, в частности *E.coli*. Вся мРНК может считываться единой молекулой в процессе транскрипции от промотора до терминатора транскрипции. Однако, в одной молекуле мРНК могут быть закодированы несколько полипептидных цепей: от иницирующего кодона до терминирующего. При этом кодирующие последовательности могут пересекаться.

На рис.21 изображена генетическая карта фага λ , умеренного двухцепочечного ДНК-вого бактериофага паразитирующего на *E.coli*. На карте указаны опероны, считающиеся с разных промоторов, расположенные на обеих цепях ДНК. Как видно, регуляция жизненного цикла даже столь маленького организма весьма сложна.

На рис.22-31 изображены графики усредненных профилей индекса локальной комплементарности LCI в окрестности иницирующего и терминирующего кодонов для различных выборок открытых рамок считывания (ОРС) из оперонов *E.coli*. ОРС - кусок мРНК, начинающийся с иницирующего кодона и кончающийся терминирующим кодоном, который может кодировать белок. Любой ген в прокариотах - ОРС, но не любая ОРС кодирует белок. Видно, что профиль LCI зависит от места ОРС в опероне.

На рис.32 изображена обобщенная схема процесса трансляции в эукариотических клетках. мРНК синтезируется в ядре в результате транскрипции ДНК. Затем мРНК претерпевает процесс сплайсинга и затем транспортируется в цитоплазму, где происходит трансляция, т.е. синтез белка по РНК-овым матрицам.

На рис.33 изображена структурно-функциональная схема устройства мРНК. Собственно полипептидную цепь кодирует только внутренняя транслируемая область, все остальное служит для регуляции процесса трансляции.

На рис.34-36 изображена обобщенная схема сканирующей модели инициации трансляции. Смысл ее в том, что 40S субъединица рибосомы связывается с мРНК далеко слева от сайта инициации трансляции (ориентация мРНК "слева-направо" - "5'-конец -3'-конец").

Затем она двигается по мРНК до иницирующего кодона AUG, здесь связывается с 60S субъединицей рибосомы и далее начинается процесс синтеза белка, кончающийся, когда рибосома встречает терминирующий кодон.

Эффективность трансляции эукариотических мРНК может быть снижена ложными стартами трансляции и/или стабильными шпильками.

На рис.37 изображена принципиальная схема стохастической модели трансляции. Процесс элонгации индивидуальной молекулы мРНК описан в модели с учетом конкретного

кодонного контекста. Учтено то, что среднее время экспонирования изоакцепторной мРНК в А-сайте рибосомы зависит от конкретного вида экспонированного в нем кодона, точнее от частоты его использования, а также учтен стерический размер рибосомы, запрещающий сближение рибосом меньше чем на некоторое фиксированное количество нуклеотидов. Данный вариант модели позволяет рассчитывать многие важные характеристики элонгации индивидуальных эукариотических мРНК с учетом их кодонного состава и интенсивности поступления рибосом на мРНК и интенсивности их схода с мРНК: подсчитывать среднюю скорость элонгации, оценивать среднее расстояние между рибосомами в составе полисомы, выявлять места сгущения рибосом, рассчитывать элонгационные профили индивидуальных мРНК, сравнивать элонгацию различных мРНК между собой, в том числе проводить оценку влияния синонимичных замен, выявлять лимитирующие звенья и т.д.

За один такт в модели осуществляется одно трансляционное событие. Чтобы вычислить это событие сначала определяются все события, которые в текущий момент не запрещены. Например, инициация трансляции считается не запрещенной, если сайт инициации стерически не занят ранее стартовавшей рибосомой. Предполагается, что если предыдущая рибосома не продвинулась по мРНК на определенное количество кодонов, то следующая инициация запрещена. Продвижение транслирующей рибосомы на один кодон вправо считается разрешенным только, если соседняя к ней со стороны 3'-конца рибосома расположена не ближе, чем на определенное количество кодонов. Это минимально допустимое сближение определяется стерическим размером рибосомы. Процесс терминации в данной версии модели ничем не блокируется. Далее для потенциально разрешенных процессов датчиком случайных чисел осуществляется прогноз времени его осуществления. Для каждого такого процесса время оценивается на основании модели его протекания описанного в терминах химических реакций. В настоящей версии все процессы описаны как мономолекулярные события. Константы скоростей процессов в этих моделях определяются особенностями процессов инициации трансляции, индивидуальными особенностями кодонного состава, контекста сайта терминации трансляции и являются внешними параметрами модели. После этого определяется минимальное оцененное время. Оно объявляется следующим тактом модели. Текущее время моделирования процесса увеличивается на данную величину, а произошедшим считается процесс, для которого оцененное время оказалось минимальным. Затем процесс повторяется.

На рис.38 -- результаты расчетов движения рибосом, которые выдаются в виде статистических характеристик:

- средняя скорость элонгации каждого кодона мРНК,
- средняя заселенность кодонов рибосомами,
- среднее количество рибосом, находящихся в стерическом контакте с 3'-проксимальной рибосомой. Также вычисляются дисперсии этих показателей, и оцениваются другие усредненные показатели трансляции мРНК

На рис.38 хорошо видно, что 24-ый кодон САС (кодирующий гистидин) имеет наименьшую скорость прохождения рибосомы. Для повышения скорости элонгации мы можем заменить его на более "быстрый" синонимичный кодон САТ или на какой-либо "быстрый" кодон, кодирующий аминокислоту близкую по своим физико-химическим свойствам к гистидину, что не мешает функции белка, закодированного в этой нуклеотидной последовательности. С другой стороны, на 9-ом кодоне ААА рибосома проводит много времени из-за "стопки" рибосом впереди. Модифицируя (используя синонимичные кодоны) последовательность после 9-го кодона, можно добиться "разблокировки" этого сайта мРНК.

Известно, что для опознавания сайта инициации трансляции очень существенны контекстные свойства окружающего района (рис.39). Однако, многие эукариотических мРНК, содержащиеся в банках данных, имеют стартовый AUG кодон в неоптимальном

контексте. Мы предположили, что структурные свойства района старта трансляции также влияют на эффективность инициации трансляции.

Для всех имеющихся мРНК дрожжей был подсчитан индекс эффективности экспрессии для кодирующей части, т.е. получен вектор значений EEI. Для всех мРНК для каждой позиции был рассчитан индекс локальной комплементарности LCI. Оказалось, что для окрестностей AUG кодона имеет место достоверная отрицательная корреляция вектора EEI индексов и вектора LCI индексов для каждой позиции.

I. *Анализ 5' райнов генов.* Все гены разбили по возрастанию значения EEI на 12 групп. Первую группу составили 500 генов, имеющих самые низкие значения EEI-индексов. Вторую группу составили следующие 500 генов и т.д. Двенадцатая группа состояла из 304 генов с максимальными индексами (результат существенно не менялся, если мы брали другое разбиение). Затем мы выбросили из групп все гены, имеющие в рамке трансляции меньше 400 нуклеотидов. Для оставшихся генов для каждой группы были вычислены средние значения EEI и усредненные LCI-профили.

Затем были подсчитаны коэффициенты корреляции Пирсона между вектором значений EEI-индексов и векторами LCI-индексов для каждого нуклеотида и подсчитаны по критерию Стьюдента значимости найденных корреляций. Наиболее протяженная неразрывная область с достоверной отрицательной корреляцией ($p < 0.01$), устойчивой к параметрам расчета LCI, наблюдается для (-19,+13)- участка 5'-области (рис.4а,с). В случае, если уровень значимости снизить до 98.8% ($p < 0.012$) границы участка расширились до (-20,+25). На рис.40 границы данного участка отмечены вертикальными линиями.

II. *Анализ 3' райнов генов.* Известно, что окрестность стоп-кодона также влияет на эффективность трансляции. Мы обнаружили, что на участке (-55,+1) также наблюдается понижение LCI, однако оно выражено значительно слабее (Рис.40b,d). В области правее терминаторного кодона профиль значений LCI зависит от параметров расчета. При учете длинных шпилек (стем до 12 нуклеотидов) выявляется область с высоким усредненным значением LCI (Рис.40d). При учете шпилек до 6 нуклеотидов область с высоким усредненным значением LCI отсутствует (Рис.40b). Между LCI профилем и индексом EEI нет продолжительного участка с достоверной корреляцией. Более того, внутри 3' райнов или кодирующих частей (не соседствующих с 5'-концом), достоверные промежутки корреляции были относительно короткими и зависимыми от параметров подсчета LCI (сравнить рис.40а с рис.40с и рис.40b с рис.40d, соответственно).

В белок кодирующих областях уровень LCI имеет тенденцию к положительной коореляции с уровнем экспрессии. Однако возможно, что в значительной мере эта тенденция определяется отбором кодонов. Численные эксперименты по генерации случайных последовательностей с определенным кодонным составом подтверждают данное предположение.

На рис.41-45 изображены графики усредненных профилей индекса локальной комплементарности LCI в окрестностях иницирующего и терминирующего кодонов для полных наборов последовательностей мРНК различных организмов. Растение *Arabidopsis* (19700 генов) демонстрирует резкий изгиб профиля в области иницирующего кодона, высокое и относительно ровное значение LCI в кодирующей части и резкое падение после терминирующего кодона. Аналогичным образом ведет себя червь нематода (*C. Elegans*, 17035 генов). Весьма отличный профиль демонстрирует *Archaea Pyrobaculum aerophilum* (2605 генов) - изгиб в области иницирующего кодона плохо заметен на фоне "ямы", а после терминирующего кодона профиль возрастает столь же резко, как падает перед. Два следующих рисунка относятся к млекопитающим: *Homo sapiens* (34485 генов) и *Mus Musculus* (3480 генов). Они очень похожи друг на друга и отличаются от предыдущих

резкими "отрицательными" пиками в областях иницирующего и терминирующего кодонов. Несколько большая "пилообразность" графика для мыши объясняется меньшей статистикой (генов в 10 раз меньше, чем у человека).