

Math-Net.Ru

Общероссийский математический портал

Н. Н. Козлов, Е. И. Кугушев, Т. М. Энеев, Структурообразующие характеристики транскрипционного процесса,
Матем. моделирование, 1998, том 10, номер 6, 3–19

<https://www.mathnet.ru/mm1288>

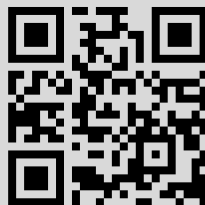
Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением

<https://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 18.97.14.84

20 апреля 2025 г., 00:58:31



МАТЕМАТИЧЕСКИЕ МОДЕЛИ И ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

СТРУКТУРООБРАЗУЮЩИЕ ХАРАКТЕРИСТИКИ ТРАНСКРИПЦИОННОГО ПРОЦЕССА

© Н.Н. Козлов, Е.И. Кугушев, Т.М. Энеев

Институт прикладной математики им. М.В. Келдыша РАН, Москва

Представлены результаты компьютерного моделирования образования вторичных структур молекул РНК, порождаемых в ходе транскрипции. Расчеты проводились в модели с возможностью любых локальных структурных перестроек, понижающих свободную энергию вторичной структуры РНК, и в модели с их запретом. Показано, что запрещение перестроек в модели процесса заметным образом понижает количество правильных структурных связей. Исследована зависимость процесса от значения параметра структуризации, позволяющая выдвинуть предположение о возможном прерывистом характере процесса транскрипции. Установлена также выделенность кодирующей цепи ДНК и естественного направления транскрипции.

TRANSCRIPTION PROCESS PARAMETERS AFFECTING ON RNA STRUCTURE FORMATION

N.N. Kozlov, E.I. Kugushev, T.M. Eneev

Keldysh Institute of Applied Mathematics of Russian Academy of Sciences, Moscow

The results of computer simulation of RNA secondary structure formation during transcription are described. Both cases have been investigated: when any structure reorganization decreasing RNA free energy is permitted, and when any structure reorganization is forbidden. It's proved that structure reorganization prohibition noticeably decreases the amount of correct secondary joints in RNA structure. The investigation of simulated process dependence on structurisation parameters makes possible to assume that transcription process has intermittent nature. Transcription natural direction (from 3'-end to 5'-end) preference was established.

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта: 96-01-01192).

Введение

Настоящие исследования лежат в русле работ по изучению структур и эволюции больших дискретных систем различной природы, проводившихся начиная с 1970 г. Среди них следует отметить известные исследования по математическому моделированию гравитационного взаимодействия галактик [1], работы по численному моделированию процесса формирования планет в протопланетном диске [2], проведение которых стало возможным благодаря разработке метода виртуальных контактов [3]. Опыт, накопленный при решении указанных задач, оказался чрезвычайно ценным.

Опираясь на этот опыт, Т.М.Энеев в 1987 году предложил использовать метод виртуальных контактов для построения эффективного алгоритма реконструкции сложных биологических макромолекул (белков) по аминокислотной последовательности, которая образуется на основе текста ДНК. Задача эта чрезвычайно сложна, не имеет надежной базы исходных данных, и поэтому начинать с нее опробование разработанных подходов оказалось преждевременно. Однако был найден весьма удачный объект для реализации указанной идеи – молекулы РНК. Именно эти молекулы стали своеобразным полигоном для дальнейшей отработки и модификации метода виртуальных контактов и других методов. Была предложена одна математическая модель процесса структуризации таких молекул.

В настоящее время завершен первый этап исследований, проведенный на основе разработанного специального алгоритма. Этот алгоритм позволил достичь среднего времени расчета процесса образования вторичной структуры РНК, на два порядка меньшего, чем при традиционном подходе, применявшемся за рубежом. Стало возможным проведение серии численных экспериментов для опубликованных международных каталогов генов тРНК. В дополнение к главному алгоритму был создан алгоритм расчета свободной энергии вторичных структур РНК различных типов, а также алгоритм визуализации на дисплее ЭВМ вторичных структур РНК. Все разработанные алгоритмы были опробованы при изучении указанной модели процесса сворачивания тРНК во вторичную структуру.

Целесообразность и эффективность применения методов прикладной и вычислительной математики в исследовательской работе по биологии вообще и по молекулярной биологии в частности в настоящее время не вызывает сомнений. За последние десятилетия в этом направлении были достигнуты определенные успехи. Эти успехи, однако, имеют пока относительно скромный характер, ввиду больших трудностей, с которыми столкнулись исследователи в ходе применения математических методов для анализа строения и функционирования биологических систем. Трудности эти связаны прежде всего с исключительной сложностью большинства биологических структур – сложностью, несравнимой со сложностью традиционных объектов применения математики в области естествознания и техники. Поэтому одной из проблем, возникающих на пути использования математических методов исследования в биологии, является проблема подбора математического аппарата, адекватного специфике изучаемых объектов.

Известно, что в естествознании при изучении тех или иных объектов или явлений могут быть применены два подхода – феноменологический и структурный. Феноменологический подход, основанный на анализе неких осредненных характеристик изучаемой системы или явления (как, например, давление газа, его плотность и температура в физике) нашел применение и в некоторых биологических задачах (например, в иммунологии). Представляется, однако, что структурный подход, основанный на изучении функционирования той или иной системы на уровне ее элементарных объектов, является предпочтительным для биологических задач. Это тем более относится к молекулярной биологии, дискретный характер объектов которой, как правило, выходит на первый план. Соответственно такому характеру объектов в данном случае, по-видимому, целесообразно использовать методы дискретной математики.

Настоящая работа посвящена изучению процессов формирования пространственных структур биологических макромолекул – рибонуклеиновых кислот (РНК) в ходе их транскрипции (копирования) с ДНК (дезоксирибонуклеиновой кисло-

ты). На рис.1 показан участок ДНК примерно в 60 пар нуклеотидов, который экранируется ферментом, называемым РНК-полимеразой. Синтез молекулы РНК, порождаемой полимеразой, называется транскрипцией. РНК, возникающая в ходе транскрипции, состоит из элементов (нуклеотидов) четырех типов, которые традиционно обозначаются как А, U, С, G. После транскрипции идет процесс структуризации РНК, не показанный на рисунке. В ходе этого процесса происходит Уотсон-Криковское соединение некоторых пар нуклеотидов по принципу, называемому принципом комплементарности: соединяться могут только А с U, U с А, С с G или G с С; любые другие пары считаются неподступимыми. Набор Уотсон-Криковских связей называется вторичной структурой РНК. Число потенциально возможных структур быстро растет с увеличением длины молекулы РНК (это число $\sim 10^8$ для молекул РНК, рассматриваемых в данной работе). Несмотря на это, в живой клетке образующиеся структуры РНК всегда одни и те же. Определение такой естественной структуры РНК из указанного набора потенциально возможных структур составляет главную проблему данного исследования. Наиболее простые по структуре и короткие молекулы РНК были исследованы в [4], где во вводной части на стр. 24-38 есть подробное описание важнейших понятий, имеющих отношение к рассматриваемым вопросам математического моделирования.

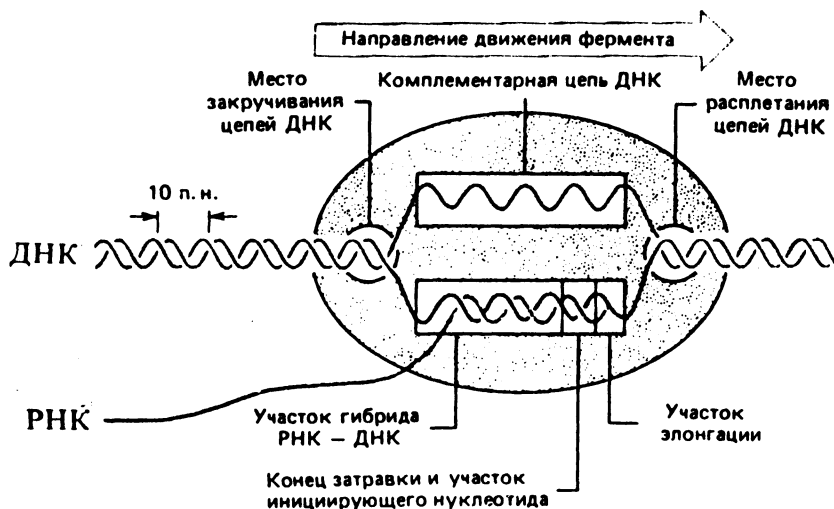


Рис. 1. Синтез молекулы РНК, порождаемой полимеразой в ходе транскрипции

Процессы формирования пространственных структур существенно дискретны. В ходе транскрибирования нуклеотиды, из которых состоит первичная структура РНК, попарно "склеиваются" по принципу их комплементарности. Существенно при этом, что комплементарность нуклеотидов имеет одно-однозначный характер. Данное обстоятельство упрощает анализ процесса "склеивания", однако, количество вариантов конечных структур, которое может быть получено только за счет комплементарности, весьма велико даже для самых простых молекул РНК. Выбор варианта из всего их множества, сформированного только по принципу удовлетворения условия комплементарности, может быть осуществ-

лен на основе минимизации приращения свободной энергии формируемой молекулы. При этом под свободной энергией понимается та часть внутренней энергии химической системы, за счет которой может быть совершена работа над окружающей средой.

Итак, по принципу комплементарности нуклеотидов может быть сформировано множество вариантов структур той или иной молекулы РНК, из которого необходимо выбрать вариант, который обеспечит минимум свободной энергии этой молекулы. Такой выбор может быть осуществлен только путем перебора всех возможных вариантов указанных конечных структур РНК. Методы классического математического анализа по поиску экстремума функций здесь неприменимы. В этом состоит одна из трудностей задачи, так как указанный перебор в общем случае трезвычайно трудоемок и связан с большим объемом вычислений.

Другая трудность связана с многоэкстремальностью задачи. Среди множества вариантов пространственных структур молекулы РНК, получаемых путем перебора, имеются локальные минимумы свободной энергии по независимым переменным. В связи с этим возникает проблема отбора из множества локальных экстремумов такого, который реализуется в физической реальности. В качестве такого экстремума можно было бы взять экстремум глобальный, что и делалось некоторыми исследователями [5-7]. Однако этот вариант решения проблемы представляется весьма спорным, так как трудно представить в данном конкретном случае реальный физический механизм, который мог бы осуществить поиск и идентификацию глобального экстремума. Более перспективным представляется путь, согласно которому реализация локального минимума свободной энергии РНК непосредственно связана с самим механизмом процесса транскрипции и осуществляется в ходе последовательного роста молекулярной цепи. Такой именно путь и положен в основу исследования, которое обсуждается в настоящей работе.

Подводя итог изложенному, выше сформулируем постановку задачи.

Имеется линейная, так называемая первичная структура РНК, выдвигающаяся из РНК-полимеразы, связанной с ДНК. Данная первичная структура РНК представляет собой копию соответствующего гена ДНК. По мере выдвигания РНК из полимеразы, происходит ее свертывание в некую пространственную структуру с помощью сцепления ("склеивания") тех или иных комплементарных нуклеотидов. Принято различать вторичную и третичную (фактическую) пространственные структуры РНК, причем, вторичная структура представляет собой третичную, собственно пространственную структуру, развернутую на плоскости. Вторичная структура является промежуточной и вспомогательной конструкцией, однако, в нахождении ее и содержится, по-видимому, главная трудность и ключ к решению всей задачи. Настоящая работа посвящена поиску вторичной структуры РНК.

Как уже было отмечено ранее, учет одной лишь комплементарности нуклеотидов недостаточен для решения задачи. Необходимо построение самой модели процесса, который должен состоять в общем случае из отдельных различных этапов и, по-видимому, быть циклическим. Построение такой модели является одной из главных целей работы.

Отмеченная выше существенная дискретность изучаемого объекта естественным образом привела к теоретико-множественным приемам исследования. Необходимость перебора большого количества потенциальных вариантов решения задачи потребовала разработки соответствующей комбинаторики такого перебо-

ра, а также разработки методов тщательной организации и редактирования массивов данных, хранящихся в памяти компьютера. Здесь был использован опыт работы авторов в области астрофизических проблем, где также приходилось иметь дело со сложными дискретными системами, анализом их структур и функционирования.

Ниже представлены результаты исследований по математическому моделированию транскрипционного процесса [4]. Этот процесс соответствует важнейшему этапу функционирования ДНК в клетке живого организма. Как известно, ДНК является собой как бы хранилище генетической информации; в клетке непосредственно функционируют различные по типу и назначению молекулы РНК (рибонуклеиновой кислоты). Возникновение молекул РНК происходит при считывании определенных участков ДНК, которые после некоторой модификации (при замене основания Т в тексте ДНК на основание У в тексте РНК [4]) образуют линейные молекулы РНК. В последующем происходит структуризация молекул РНК. В данной работе речь идет о некоторых параметрах транскрипционного процесса, которые могут определяющим образом влиять на формирование структур РНК. Формирование структур анализируется на основе функционирования математической модели, предложенной и опробованной авторами.

Специфика настоящей статьи предполагает использование большого количества терминов современной молекулярной биологии. Краткое представление необходимых терминов с объяснением их взаимосвязи применительно к задачам математического моделирования было дано ранее во вводной части работы [6], см. также рис. 1.

1. Исходные положения

Данное исследование представляет собой развитие работы [4]. Исследуется два типа молекул РНК: множество молекул тРНК и множество молекул 5S рРНК. Основные характеристики молекул первого множества приводятся в [4]. Рассмотрим другое множество молекул. Детальное изучение 5S рРНК молекул началось в 1967 году, когда Ф. Сенгер впервые установил первичную структуру одной из таких молекул. В настоящее время известно уже несколько сот нуклеотидных последовательностей 5S рРНК, и установлены наиболее вероятные их естественные вторичные структуры. В таких структурах характерное число спиральных участков может достигать 5 и более, в отличие от молекул тРНК, где характерные 4 спиральных участка иногда вырождаются в 3 [4]. Указанные спиральные участки 5S рРНК образуют так называемую Y-форму (рис.2, см. также [8]). На рис.3 представлены 8 изображений естественных вторичных структур этих молекул, которые были отобраны из 238 изображений, приведенных в [9] и полученных на компьютере по данным из последнего опубликованного каталога указанных молекул [10]. Изображения вторичных структур приведены в том же порядке, что и в [10]. "Руки" Y-формы находятся в нижней части изображения, корневой стебель – в верхней. Движение по молекулярной цепи от 5'-конца к 3'-концу соответствует обходу контура изображения против часовой стрелки. Под каждым изображением дан текст – сопровождение из [10]: порядковый номер и краткий идентификатор организма – хозяина 5S рРНК. В приводимых изображениях строго выдержаны все характерные размеры, в отличие от сглаженных изображений рис.2, где указанные размеры искажены на ряде участков вторичной структуры. На рис.3 дается изображение вторичной структуры человеческой 5S рРНК (SO71), а также некоторые типы структур из [9], где в частности просматриваются структуры с шестью четкими спиральными

ми участками (SO61) (ср. с рис.2, где две вторичные структуры имеют 4 и 5 спиральных участков соответственно).

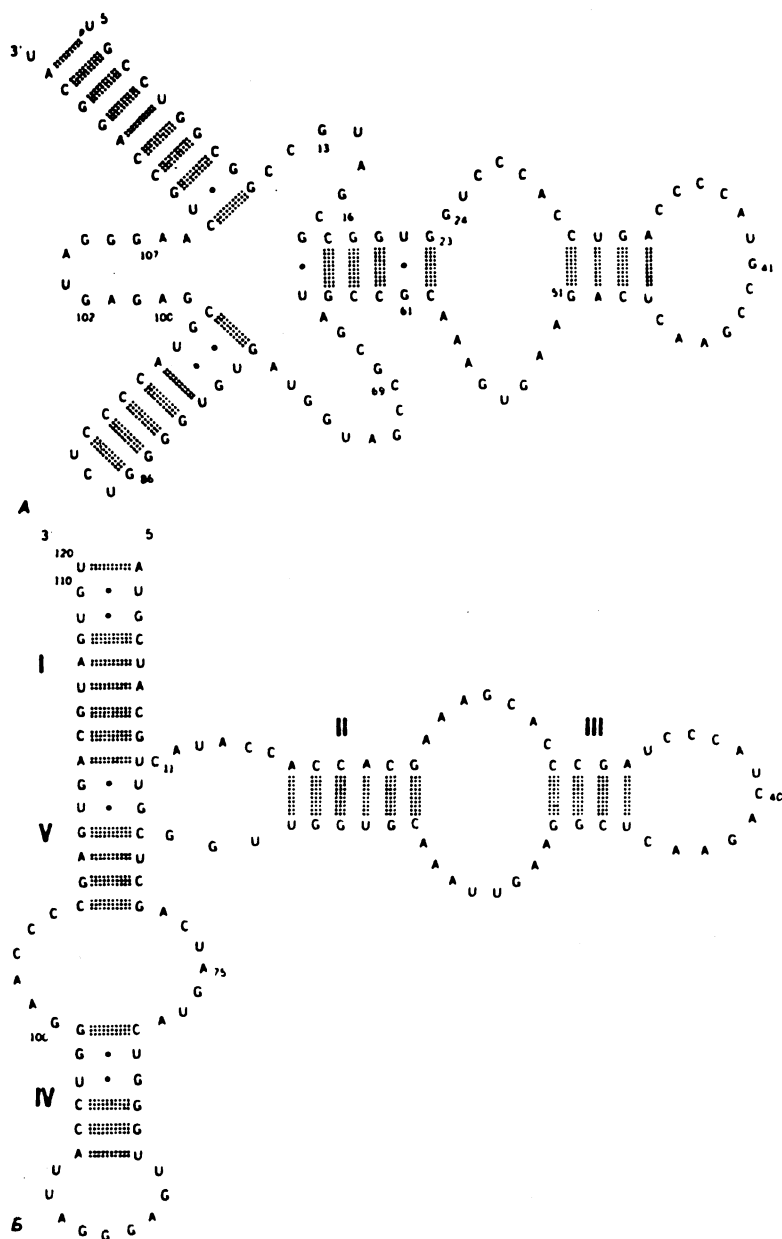


Рис. 2. Во вторичной структуре прокариотической 5S-РНК имеется 4 спиральных участка (А), тогда как эукариотическая 5S-РНК образует 5 спиральных участков (Б). [8]

Определение пространственной структуры рибонуклеиновых кислот является одной из фундаментальных проблем молекулярной биологии, поскольку форма, которую приобретают эти молекулы в пространстве, в значительной степени определяет их функциональные свойства. Энергетически самая сильная составляющая пространственной структуры РНК — это ее вторичная структура. Прямые или косвенные биохимические методы определения вторичных структур молекул РНК хуже работают при росте их длины (если вообще остаются возможными).

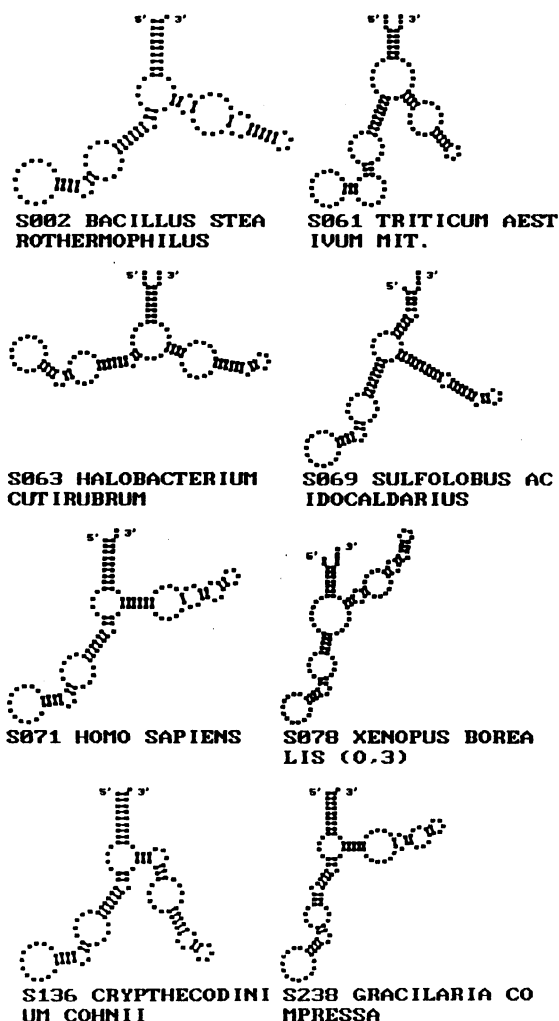


Рис. 3. Изображения естественной вторичной структуры восьми 5S рРНК, полученные по данным [10]. Набор изображений для 23S рРНК приводится в [9].

В связи с этим в последнее время интенсивно развиваются компьютерные методы определения (предсказания) этих структур по известному составу молекулярной цепи РНК, т.е. по ее первичной структуре. Основная проблема, которую здесь приходится решать – это выбор реальной вторичной структуры среди громадного множества потенциально возможных энергетически локально устойчивых вторичных структур.

Наиболее широко распространен метод, основанный на поиске вторичной структуры, обеспечивающей глобальный минимум ее свободной энергии. Первые попытки применения этого метода делались еще в 60-е годы [11-12], где в качестве свободной энергии вторичной структуры бралось количество вторичных связей в структуре с отрицательным знаком. С развитием и уточнением моделей свободной энергии [13-17] выяснилось, что прямой поиск глобального минимума далеко не всегда приводит к вторичным структурам, наблюдаемым в природе. Данное обстоятельство вызвало развитие методов субоптимизации [17-18], последовательно-статистических [7,19], третичного взаимодействия [20], сравнительного [21,22], локальной минимизации свободной энергии [23, 24] и др.

В предложенном нами [4,25,26] параметрическом подходе число допустимых структур резко ограничивается, и проблема выбора среди них реальной вторичной структуры облегчается. Вторичная структура здесь постепенно образуется в ходе моделируемого транскрипционного процесса, порождающего молекулярную цепь данной РНК. При заданных параметрах модели образуется только одна структура. Конечно, здесь необходимо иметь модель и параметры транскрипционного процесса, адекватные реальности.

В данной работе мы описываем новые результаты, полученные на основе параметрического подхода. Исследования процесса структуризации проводились для 238 молекул 5S рРНК и 1146 молекул тРНК. Расчеты проводились в модели с возможностью любых локальных структурных перестроек, понижающих свободную энергию вторичной структуры РНК, и в модели с их запретом. Исследована также зависимость решения от выбора транскрибируемой цепи ДНК и от направления транскрипции.

2. Методы

Следует отметить, что при выборе модели транскрипционного процесса нас прежде всего интересуют те его детали, которые могут оказывать заметное влияние на структурообразование в порождаемой молекулярной цепи РНК. В связи с этим в предыдущих работах [4,25,26] использовалась математическая модель, которая отражала два основных процесса, имеющих место в реальной транскрипции. Это элонгация – процесс постепенного удлинения молекулярной цепи и структуризация – процесс постепенного наращивания структуры на уже образовавшемся участке молекулярной цепи. В ходе структуризации происходит возникновение вторичных связей, обеспечивающих локальную минимизацию свободной энергии структуры. При этом запрещаются перестройки структуры, т.е. разрыв старых связей, препятствующих возникновению новых.

Модель с запретом структурных перестроек показала весьма хорошие результаты. Ее использование позволяет правильно определять в среднем 87% структурных связей для молекул тРНК. Тем не менее известно [27], что в ходе реальной транскрипции возможны перестройки шпильчатых элементов вторичной структуры. Явление это также обсуждалось в [7,20,28]. Поэтому моделирование структурных перестроек может оказаться весьма важным моментом,

приближающим математическую модель процесса структурообразования к реальности. Однако вопрос о том, какие структурные перестройки возможны и с какой скоростью они протекают, остается до сих пор открытым.

В связи с этим была разработана математическая модель структурообразования с возможностью любых локальных структурных перестроек, понижающих свободную энергию вторичной структуры РНК. В настоящей работе представляются результаты математического моделирования, проведенного в рамках этой модели. Обратимся теперь к более полному описанию математической модели исследуемого процесса.

Описание модели. Процесс моделировался как пошаговый, отражающий последовательный рост молекулярной цепи РНК. На каждом этапе фазы структуризации и элонгации моделировались последовательно. Таким образом, нуклеотиды, присоединяющиеся к молекулярной цепи на следующем шаге, не участвуют в структуризации на данном шаге до тех пор, пока структура не стабилизируется, и не произойдет переход к следующему шагу процесса.

Процесс элонгации определяется двумя параметрами. Первый параметр, период структуризации T – это отношение V_L/V_S скорости V_L удлинения молекулярной цепи РНК к скорости V_S структуризации. В нашей модели, когда элонгация и структуризация происходят последовательно, параметр T по своему смыслу – это число нуклеотидов, на которое удлиняется цепь РНК за время структуризации. Второй параметр, начальная фаза L_0 – это длина молекулярной цепи РНК на первом шаге процесса. Наличие этого параметра отражает то, что рассматриваемая РНК может быть составной частью транскрипта большего размера. В этом случае L_0 – это длина остатка молекулярной цепи от предыдущих шагов транскрипции. По своему смыслу L_0 неотрицательно и не превосходит T . При численном моделировании параметры T , L_0 могли принимать только целые неотрицательные значения.

Процесс структуризации также моделировался пошаговым образом. К началу структуризации молекулярная цепь имеет некую длину и вторичную структуру (может быть, пустую), получившуюся на предыдущих шагах транскрипции. Исходя из этого формируется множество допустимых вариаций структуры (ДВС), т.е. множество неких допустимых подструктур, которые потенциально могут быть добавлены к уже имеющейся структуре. Затем выбирается допустимая вариация, дающая наибольшее падение свободной энергии вторичной структуры, и добавляется к ней. Тем самым вторичная структура модифицируется. Определение множества ДВС и модификация вторичной структуры повторяется до тех пор, пока находится хотя бы одна допустимая вариация, понижающая ее свободную энергию. Как только на множестве допустимых вариаций минимальный прирост свободной энергии оказывается неотрицательным, процесс структуризации оканчивается.

Процесс структуризации определяется двумя внутренними факторами: правилами выбора множества ДВС и моделью свободной энергии, т.е. правилами ее вычисления для данной вторичной структуры. В рассматриваемых нами моделях множество ДВС формируется из стеблей. Стебель – это двуспиральный участок РНК, в обоих нитях которого нуклеотиды попарно комплементарны и спарены последовательно без пропусков. Любая вторичная структура может быть описана как семейство составляющих ее стеблей. Поэтому с конструктивной точки зрения указанный выбор множества ДВС позволяет получать любые вторичные структуры. В рассмотренной ранее модели без структурных перестроек в мно-

жество ДВС допускались только такие стебли, которые не пересекались по своим вторичным связям со связями, уже имеющимися во вторичной структуре. При добавлении таких стеблей в структуре возникают новые связи, а старые всегда сохраняются. В предложенной модели со структурными перестройками в множество ДВС допускаются любые стебли. При добавлении нового стебля в структуре удаляются старые связи, противоречащие новым. В результате структура может не только наращиваться, но и реорганизовываться, перестраиваться (если это энергетически выгодно). Принципиальным моментом данного подхода является то, что в ходе такой структуризации на каждом шаге достигается локальный минимум свободной энергии. Это определяется тем, что множество ДВС ограничено возникновением только одиночных новых стеблей, а не парами, тройками и т.д., что неизбежно привело бы к глобальному минимуму свободной энергии. Как было указано в п. 1, глобальный экстремум не дает удовлетворительных результатов при сопоставлении с экспериментальными данными.

В [4] было проведено сравнение четырех опубликованных моделей свободной энергии вторичной структуры. В качестве основной была выбрана модель [15], которая достаточно проста и точна. Все результаты, рассматриваемые в настоящей работе, получены в рамках этой модели энергии.

Одной из целей исследования была оценка равноценности двух цепей ДНК с точки зрения структуризации транскрибируемых с них РНК. Для этого помимо транскрипции с плюс-цепи (кодирующей цепи) рассмотрены варианты, в которых транскрибируется минус-цепь. Кроме того в вариантах расчета варьируется направление роста молекулярной цепи: прямое от 5'-конца к 3'-концу и обратное - от 3'-конца к 5'-концу.

3. Результаты

Математическое моделирование проводилось на двух семействах молекул РНК. Первое - из 1146 молекул тРНК, второе - из 238 молекул 5s рРНК. Тексты и предполагаемые реальные вторичные структуры молекул РНК этих семейств даны соответственно в [29] и [10]. Расчеты проводились для всех возможных значений параметров T и L_0 с шагом 1. Оказалось, что для всех рассматриваемых молекул РНК длина молекулярной цепи $L \leq 130$. Поэтому для унификации результатов численное моделирование проводилось в области $1 \leq T \leq 130$. Всего было промоделировано 16 вариантов (по 8 для каждого множества РНК). Мы их будем обозначать T_{ij} - для множества молекул тРНК, и S_{ij} - для множества молекул 5S рРНК. Индекс i может принимать два значения: 0 - расчет с запретом структурных перестроек и 1 - расчет с их возможностью. Индекс j может принимать четыре значения. Значение $j=1$ соответствует транскрипции с плюс-цепи ДНК (кодирующей цепи) в естественном направлении от 5'-конца к 3'-концу. Значение $j=2$ соответствует транскрипции с плюс-цепи ДНК в направлении от 3'-конца к 5'-концу, противоположном естественному. Значения $j=3$ и 4 соответствуют транскрипции с минус-цепи ДНК в прямом ($j=3$) и обратном ($j=4$) направлении.

В таблицах 1 и 2 приведены некоторые данные, полученные для вариантов T_{ij} и S_{ij} соответственно. Дадим описание обозначений. $R(T)$ - это средняя доля спаренных оснований естественной структуры молекул РНК, присутствующих в вычисленной финальной структуре. Индексы А, В, С, D, Е отражают способ выбора начальной фазы L_0 . Ее значение неизвестно и может меняться от моле-

кулы к молекуле. Индекс А означает выбор такого значения L_0 , которое (при данном значении T) дает наилучшее совпадение реальной и вычисленной вторичных структур; при индексе В полагается $L_0=0$; при индексе С L_0 выбирается так, чтобы при данном T свободная энергия молекулы была минимальна; при индексе D L_0 выбирается случайным образом; при индексе Е L_0 выбирается из условия наибольшей частоты возникновения данной вторичной структуры при всех возможных значениях L_0 для данного T . Помимо величин $\max RA - \max RE$ в таблицах приведены значения NF — это максимальное по T среднее число различных вторичных структур, которые получаются при всех значениях L_0 .

Таблица 1.

Некоторые характеристики вариантов расчета для множества 238 молекул 5s рРНК. S01-S04 - модель с запретом структурных перестроек. S11-S14 - модель с их разрешением.

цепь ДНК направление	плюс цепь				минус цепь			
	5'→3'		3'→5'		5'→3'		3'→5'	
	s01	s11	s02	s12	s03	s13	s04	s14
NF	22.3	16.8	20.0	13.8	20.3	15.1	21.2	16.4
A: $\max R$	61.4	67.8	57.3	63.4	50.1	55.0	52.7	57.3
B: $\max R$	44.4	50.9	39.0	46.1	33.5	40.9	38.1	42.3
C: $\max R$	41.0	44.5	39.5	42.6	31.8	33.2	34.2	36.5
D: $\max R$	25.6	37.5	26.3	37.4	22.9	31.9	23.1	32.2
E: $\max R$	28.4	38.6	30.0	38.1	23.4	31.3	22.9	31.3

Таблица 2

Некоторые характеристики вариантов расчета для множества 1146 молекул тРНК. T01-T04 - модель с запретом структурных перестроек. T11-T14 - модель с их разрешением.

цепь ДНК направление	плюс цепь				минус цепь			
	5'→3'		3→5'		5'→3'		3'→5'	
	T01	T11	T02	T12	T03	T13	T04	T14
NF	9.4	5.1	9.0	4.4	8.5	4.6	8.9	5.1
A: $\max R$	87.6	87.5	86.6	86.3	80.7	81.3	82.0	82.8
B: $\max R$	74.2	76.6	74.0	74.0	67.5	67.8	67.4	71.8
C: $\max R$	77.0	76.8	76.8	77.9	71.3	72.2	71.3	72.3
D: $\max R$	60.5	72.8	57.6	72.2	52.2	66.4	56.1	68.1
E: $\max R$	66.7	75.3	64.4	74.3	58.9	68.0	62.2	69.9

На рис. 4 приведены зависимости RA от T для вариантов S01-S04 и T01-T04. На рис. 5a,b также даны графики зависимостей величины RA от периода структуризации T для вариантов T01, T11, S01, S11, которые дают наглядное представление о характере изменения качества возникающих вторичных структур как при изменении периода структуризации T , так и при переходе от мо-

дели с запретом структурных перестроек к модели с их разрешением. Часть других данных приводится непосредственно в тексте.

Обратимся теперь к анализу полученных данных. Рассмотрим сначала влияние структурных перестроек на порождаемые вторичные структуры в основном случае транскрипции с кодирующей цепи ДНК в прямом направлении (варианты T01, S01 и T11, S11). Из рис. 5 можно видеть, что для обоих типов молекул РНК разрешение структурных перестроек резко улучшает качество порождаемых структур для малых значений относительной скорости транскрипции T . При минимальном значении $T=1$ процент правильных вторичных связей (RA) возрастает с 24.2% до 64.9% для молекул тРНК и с 9.1% до 28.4% для молекул 5S рРНК. С ростом относительной скорости транскрипции T эта разница уменьшается: сходит на нет для тРНК (87.6% и 87.5%) и остается заметной для 5S рРНК (61.4% и 67.8%). Это явление имеет место и для всех других способов выбора начальной длины L_0 . Следует отметить, что результат 67.8% для 5S рРНК превышает результат 60%, полученный в [17] для улучшенной модели свободной энергии, и является наиболее высоким для методов, в которых сравнению подвергается только одна порожденная структура.

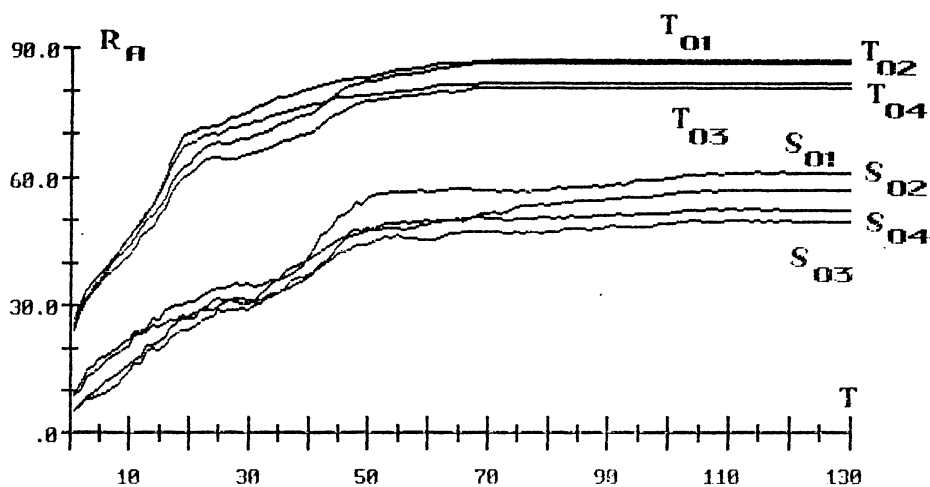


Рис. 4. Зависимости RA от T - среднего процента правильных связей в возникающей вторичной структуре РНК для вариантов S01-S04 (нижнее семейство кривых) и T01 - T04 (верхнее семейство). Наилучшие значения дают варианты T01 и S01, соответствующие транскрипции с плюс-цепи ДНК в естественном направлении (5'-конец — 3'-конец). Наихудшие значения дают варианты T03 и S03, соответствующие транскрипции с минус-цепи в естественном направлении.

Важным моментом является то, что при разрешении структурных перестроек существенно уменьшается разнообразие порождаемых структур РНК. Например, максимальное число NF различных финальных структур при фиксированном значении T и всевозможных значениях L_0 падает для тРНК с 9.4 в варианте T01 до 5.1 в варианте T11, аналогично для 5S рРНК с 22.3 в варианте S01 до 16.8 в варианте S11. Явление понижения NF имеет место при переходе от любого варианта без структурных перестроек 0j к варианту со структурными пе-

рестройками 1j как для тРНК, так и для 5S рРНК. Улучшение качества порожденных структур при уменьшении их разнообразия является косвенным подтверждением наличия и важной роли структурных перестроек в ходе формирования вторичной структуры РНК. При этом для более длинных молекул 5S рРНК влияние структурных перестроек существенней, чем для молекул тРНК.

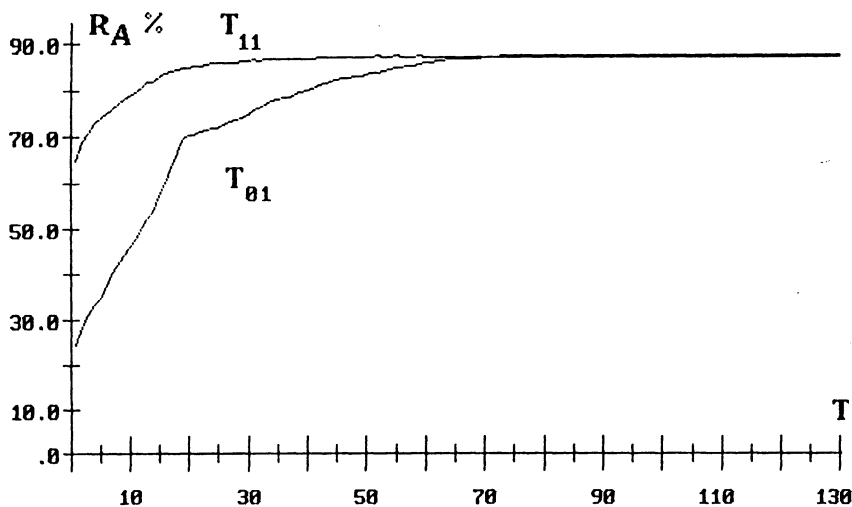
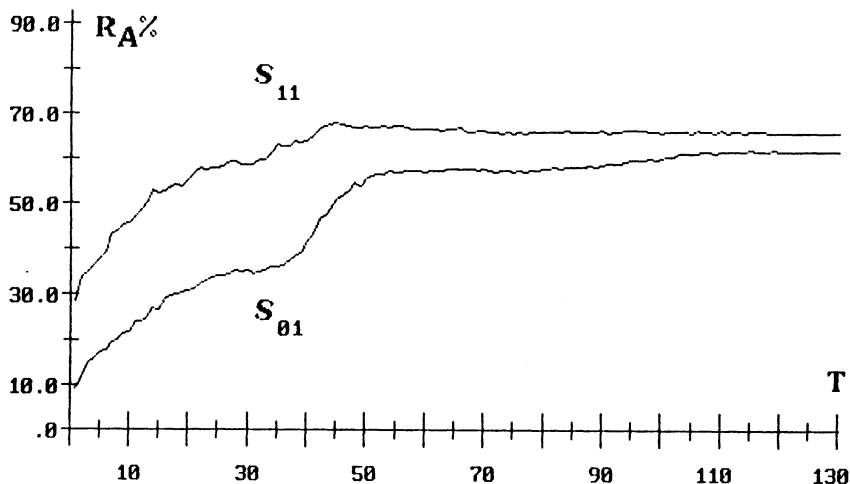


Рис. 5. Сравнение качества порождаемых вторичных структур для моделей с запретом структурных перестроек (варианты S_{01} и T_{01}) и с их полным разрешением (варианты S_{11} и T_{11}). Изменение R_A - доли правильных (естественных) вторичных связей, возникающих структур, в зависимости от значения относительной скорости транскрипции (параметр структуризации T), равной отношению скоростей удлинения молекулярной цепи и образования вторичной структуры. Приводятся средние величины, полученные на множестве 238 молекул 5S рРНК (а) и на множестве 1146 молекул тРНК (б). Можно видеть, что разрешение перестроек вторичной структуры заметно улучшает качество получающихся вторичных структур.

Рассмотрим теперь вопрос о влиянии относительной скорости элонгации T на процесс структурообразования. В [4] этот вопрос рассматривался для случая, когда структурные перестройки запрещены. Оказалось, что с ростом T качество формирования структур заметно улучшается. Наихудшие результаты получались при минимальном значении $T=1$, соответствующем случаю, когда скорость структурообразования велика по отношению к скорости удлинения молекулярной цепи. Это можно было объяснить тем, что при запрете структурных перестроек неверные вторичные связи, образовавшиеся на ранних стадиях транскрипции, не дают образовываться тем реальным связям, которые, например, соединяют начальные и конечные участки молекулярной цепи. Следовало ожидать, что разрешение структурных перестроек изменит эту ситуацию. Действительно, при $T=1$ количество правильных образующихся вторичных связей возросло для молекул тРНК с 24.2% до 64.9% (варианты T01 и T11), а для 5S рРНК с 9.1% до 28.4% (варианты S01 и S11). Значения эти, однако, заметно ниже наилучших, которые достигаются при $T=55$ для тРНК (87.5%) и при $T=46$ для 5S рРНК (67.8%) – параметр $\max RA$ в вариантах T11 и S11. При максимальных значениях $T=130$, соответствующих относительно медленному структурообразованию, качество структур несколько ниже максимального: 87.2% для тРНК и 65.3% для 5S рРНК.

Указанные данные говорят в пользу того, что реальные значения параметра структуризации T не являются малыми: молекулярная цепь РНК может удлиняться на десятки нуклеотидов за то время, пока образуется вторичная структура на уже синтезированном участке цепи. Здесь наблюдается явное противоречие. Реальная скорость транскрипции оценивается примерно в 40 нуклеотидов в секунду [18], стр.144), в то же время скорость структурообразования имеет порядок 0,001 сек и менее [30]. Это соответствует случаю $T=1$. Одно из возможных объяснений этого явления – это прерывистое удлинение молекулярной цепи синтезируемой РНК сразу на несколько десятков нуклеотидов. Прерывистое удлинение РНК может достигаться либо за счет предварительного накопления синтезируемых элементов ее цепи внутри РНК-полимеразы, либо за счет прерывистой работы и сдвига вдоль ДНК самой РНК-полимеразы. В пользу первого говорит то, что внутри полимеразы может находиться 50-60 последних нуклеотидов синтезируемой РНК [8]. В пользу второго говорит то, что по всей видимости во время транскрипции двойная спираль ДНК раскрывается на участке длиной порядка двух витков [8], поэтому в принципе возможен синтез одновременно примерно 20 нуклеотидов, после которого РНК-полимераза могла бы продвигаться по ДНК сразу на два витка вперед.

Особо следует отметить топологические препятствия, могущие возникнуть в случае непрерывной транскрипции. Если бы РНК полимеразы перемещалась по ДНК с шагом в один нуклеотид, последовательно "растегивая" ее двойную спираль (как это обычно и подразумевается), то она неминуемо должна была бы вращаться вокруг оси ДНК с периодом одного оборота, равным числу нуклеотидов в одном витке спирали ДНК. Это, при указанной выше скорости транскрипции, составило бы примерно 4 оборота в секунду и могло бы вызвать намагнивание синтезируемой цепи РНК на ДНК. Таких топологических коллизий не было бы в процессе, в котором двойная спираль ДНК "растегивается" на целое число витков, происходит наращивание РНК на длину, соответствующую этим виткам, а затем ДНК "застегивается", и РНК-полимераза сдвигается по ДНК на это же число витков. В целом это явление заслуживает дальнейшего изучения.

Рассмотрим теперь вопрос об эквивалентности структурообразующих свойств плюс- и минус- цепей ДНК и направлении транскрипции. Проведем сравнение вариантов счета i_j при фиксированной модели процесса i . При анализе как вариантов $S_{i1}-S_{i4}$, так и $T_{i1}-T_{i4}$ прежде всего наблюдается определенная устойчивость по важнейшим параметрам $\max NF$ и $\max RA$. Среднее значение $\max NF$ равно 21.0 для вариантов S_{0j} и 15.5 для вариантов S_{1j} . Максимальное отклонение от среднего равно соответственно 4.5% и 10.6%. Аналогично, для вариантов T_{0j} , T_{1j} имеем средние $\max NF$ 9.0 и 4.8, а максимальные отклонения 5.0% и 8.3% соответственно. Для характеристики $\max RA$ имеем средние 55.4 и 63.4 и максимальные отклонения 11.0% и 13.2% для вариантов S_{0j} и S_{1j} соответственно, а для вариантов T_{0j} и T_{1j} - средние 84.2 и 84.3 при максимальных отклонениях 4.0% и 3.8%.

Такая устойчивость вполне объяснима комплементарностью двух цепей ДНК, что позволяет возникать одним и тем же комплементарным связям при структуризации РНК, а также почти симметрией изучаемых структур РНК. Однако полной эквивалентности не наблюдается. Можно наблюдать только некую квазиэквивалентность. Это вполне объяснимо, так как современные нуклеотидные последовательности РНК могли быть измененными по своему нуклеотидному составу по сравнению с первичными последовательностями, для которых полная эквивалентность [31] как плюс-, так и минус-цепей, а также направлений транскрипции могли иметь место.

Разделим вопрос об квазиэквивалентности цепей ДНК и направлении транскрипции. Из расчетов следует, что изменение цепи более существенно влияет на результаты структуризации, чем направление транскрипции. Так, для перехода от плюс-цепи к минус-цепи при естественном направлении транскрипции $5' \rightarrow 3'$ параметр RA уменьшается на 11.3% (варианты S_{01} , S_{03}) и на 12.8% (варианты S_{11} , S_{13}). Переход от направления $5' \rightarrow 3'$ к $3' \rightarrow 5'$ приводит к изменению RA только на 4.1% (варианты S_{01} , S_{02}) и 4.4% (варианты S_{11} , S_{12}). Это данные по $5S$ рРНК. Для тРНК имеем уменьшение на соответственно 6.9% (варианты T_{01} , T_{03}) и 6.2% (варианты T_{11} , T_{13}) при изменении цепи ДНК для естественного направления транскрипции $5' \rightarrow 3'$ и на 1.0% (варианты T_{01} , T_{02}) и 1.2% (варианты T_{11} , T_{12}) при изменении направления транскрипции.

Приведенный краткий анализ показывает, что кодирующая цепь и направление $5' \rightarrow 3'$ (направление реальной транскрипции) выделяются в данных расчетах (как для $5S$ рРНК, так и для тРНК) прежде всего по наилучшему предсказанию вторичных связей ($\max RA$). Подтверждением этого утверждения может служить также то, что общая характеристика NF , характеризующая разнообразие вторичных структур, возникающих в ходе структуризации (при разных значениях параметров T и L_0), принимает наибольшее значение для основного варианта транскрипции с плюс-цепи ДНК в естественном направлении (варианты S_{01} , S_{11} , T_{01} , T_{11}). При этом уменьшение разнообразия структур в остальных вариантах совпадает с уменьшением количества правильных вторичных связей в возникающих структурах. Этот интересный и неожиданный факт, обнаруженный в расчетах, по-видимому говорит о том, что при указанных отклонениях процесса транскрипции от реального происходит наиболее вероятное исчезновение структур, максимально приближенных к естественным.

Сравним полученные результаты по параметру RA с результатами других исследователей. Авторам неизвестны расчеты по структуризации молекул $5S$ рРНК, сделанные для модели свободной энергии [15]. Из четырех моделей [13-16] лишь по модели [14] есть данные в [32,17]. В связи с этим был проведен

дополнительный расчет варианта S11A, у которого исходные данные те же, что и у S11, но модель свободной энергии взята из [14]. Оказалось, что в этом варианте $\max RA=63.4\%$. Для сравнения укажем, что адекватный параметр в [32] равен 18.0% , в [17] – 45.0% (данные для варианта, в котором в ходе моделирования порождается только одна вторичная структура для каждой молекулы РНК).

В заключение сравним различные способы выбора начальной фазы L0 для вариантов транскрипции с плюс-цепи ДНК в естественном направлении (S01, S11, T01, T11). Среди всех способов лучшим естественно является способ A по своему определению (выбор наилучшей структуры), поэтому мы его учитывать не будем. Если расположить остальные четыре способа в порядке возрастания $\max R$ (см. табл.1, табл.2: величины $\max RB - \max RE$), т.е. возрастания близости порождаемых вторичных структур к естественным структурам, то получим порядок D, E, B, C для обоих вариантов 5S рРНК (S01, S11) и порядок D, E, C, B для обоих вариантов тРНК (T01, T11).

Худшим во всех случаях, как и следовало ожидать, оказывается случайный выбор начальной фазы L0 (способ D). Несколько лучше случайного оказывается способ E (выбор наиболее часто возникающей структуры). Однако и он во всех случаях заметно ниже остальных двух способов. Это может означать, что реальную вторичную структуру не отличает частота возникновения при варьировании параметров процесса транскрипции.

Оставшиеся способы – это способ B (L0=0) и способ C (наименьшее значение свободной энергии структуры). Из таблиц 1 и 2 (величины $\max RB$ и $\max RC$) видно, что для 5S рРНК построение вторичной структуры с начальной фазой L0=0 несколько лучше, чем выбор структуры с наименьшей свободной энергией. Для молекул тРНК ситуация противоположная. Это может объясняться тем, что гены молекул тРНК значительно чаще входят в состав более длинных транскриптов, чем гены молекул 5S рРНК.

СПИСОК ЛИТЕРАТУРЫ

1. Eneev T.M., Kozlov N.N., Sunyaev R.A. Tidal Interaction of Galaxies. *Astron. and Astroph.*, 1973, v.22, №1, p.41-60.
2. Eneev T.M., Kozlov N.N. The problem of simulation of planetary systems accumulation processes. -*Advances Space Res.*, 1981, v.1, p.201-215.
3. Энеев Т.М., Козлов Н.Н. О новом методе численного моделирования эволюции сложных дискретных систем. Доклады АН СССР, 1982, т.263, №4, с. 820-824.
4. Козлов Н.Н., Кугушев Е.И. Изучение модели процесса сворачивания тРНК во вторичную структуру // Математическое моделирование, 1993, т.5, №6, с. 24-55.
5. Туманян В.Г., Сотникова Л.Е., Холопов А.Е. Об определении вторичной структуры РНК по последовательности нуклеотидов. Доклады АН СССР, 1966, т.166, №6, с.1465-1468.
6. Jacobson A.B., Good L., Simonetti J., Zuker M. *Proc. Natl. Acad. Sci. USA*, 1984, v.86, p.7706-7710.
7. Nussinov R., Tinoco I. Sequential folding of a messenger RNA molecule // *J. Mol. Biol.*, 1981, v.151, p.519-533.
8. Lewin B. *Genes*. - John Wiley and Sons, New York, Chichester, Brisbane, 1983.
9. Козлов Н.Н., Кугушев Е.И. Изображения вторичных структур 5s рРНК по Берлинскому каталогу 1985 г.; Козлов Н.Н., Кугушев Е.И. Pictures of 5s rRNA secondary structures from Berlin catalogue 1985. Препринт ИПМ им. М.В. Келдыша АН СССР. -М., 1991, №111.
10. Erdmann V.A., Wolters J., Huysmans E., De Wachter R. Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences // *Nucleic Acids Res.*, 1985, v.13, p.105-153.
11. Fresco J.R., Alberts B.M., Doty P. Some molecular details of the secondary structure of ri-

- bonucleic acid. // *Nature*, 1960, v. 188, p. 98-101.
12. *Tumanian V.G., Sotnikova L.E., Kholopov A.V.* Determination of secondary RNA structure from the succession of nucleotides // *Doklady Biochemistry*, 1966, v. 166, № 6.
 13. *Tinoco I., Uhlenbeck O. and Levine M.D.* Estimation of secondary structure in ribonucleic acids // *Nature*, 1971, v. 230, p. 362-367.
 14. *Salsler W.* Globin mRNA sequences: Analysis of base pairing and evolutionary implications // *Cold Spring Harbor Symp. Quant. Biol.*, 1977, v. 42, p. 985-1002.
 15. *Jacobson A.B., Good L., Simonetti J., Zuker M.* Some simple computational methods to improve the folding of large RNAs // *Nucleic Acids Res.*, 1984, № 12, p. 45-52.
 16. *Freier S.M., Kierzek R., Jalger J.A., Sugimoto N., Caruthers M.H., Neilson T., Turner D.H.* Improved free-energy parameters for predictions of RNA duplex stability // *Proc. Natl. Acad. Sci. USA*, 1986, v. 8, p. 9373-9377.
 17. *Jaeger J.A., Turner D.H., Zuker M.* Improved prediction of secondary structure for RNA // *Proc. Natl. Acad. Sci. USA*, 1989, v. 86, p. 7706-7710.
 18. *Zuker M.* On finding all suboptimal foldings of an RNA molecule // *Science*, 1989, v. 244, p. 48-52.
 19. *Le S.-Y., Chen J.-H., Nussinov R., Maizel J.V.* An improved secondary structure computation method and its application to intervening sequences in the human alpha-like globin mRNA precursors // *Comput. Applic. Biosci.*, 1988, v. 4, p. 337-344.
 20. *Abrahams J.P., van den Berg M., van Batenburg E., Pleij C.* Prediction of RNA secondary structure, including pseudoknotting, by computer simulation // *Nucleic Acids Res.*, 1990, v. 18, p. 3035-3044.
 21. *Konings D.A.M., Hogeweg P.* Pattern analysis of RNA Secondary Structure // *J. Mol. Biol.*, 1989, v. 207, p. 597-614.
 22. *Le S.-Y., Zuker M.* Predicting common foldings of homologous RNAs // *J. Biomol. Struct. Dyn.*, 1991, v. 8, p. 1027-1044.
 23. *Martinez H.M.* An RNA folding rule // *Nucleic Acids Res.*, 1984, v. 12, p. 323-333.
 24. *Martinez H.M.* An RNA secondary structure workbench // *Nucleic Acids Res.*, 1988, v. 16, p. 1789-1798.
 25. *Козлов Н.Н., Кугушев Е.Н.* Процесс сворачивания РНК во вторичную структуру. Компьютерное моделирование на 219 и 1146 генах тРНК // *ДАН*, 1992, т. 324, № 1, с. 200-205.
 26. *Kozlov N.N., and Kugushev E.I.* Computer simulation of tRNA secondary structure folding // *Comput. Applic. Biosci.*, 1993, v. 9, p. 253-258.
 27. *Kramer F.R., Mills D.R.* Secondary structure formation during RNA synthesis // *Nucleic Acids Res.*, 1981, v. 9, p. 5109-5124.
 28. *Mironov A.A., Kister A.* RNA secondary structure formation during transcription // *J. Biomolec. Struct. Dyn.*; 1986, v. 4, p. 1-9.
 29. *Sprinzl M., Dank N., Nock S., Schon A.* Compilation of tRNA sequences and sequences of tRNA genes // *Nucleic Acids Res.*, 1991, v. 19, suppl., p. 2127-2171.
 30. *Cantor C.R., Schimmel P.R.* Biophysical chemistry. Part III. - W.H. Freeman and Company, San Francisco, 1980.
 31. *Eigen M., Schuster P.* Hypercycle: principles of natural self organization, Berlin, 1979.
 32. *Papanicolaou C., Gouy M., Ninio J.* An energy model, that predicts the correct folding of both the tRNA and the 5S RNA molecules // *Nucl. Acids. Res.*, 1984, v. 12, p. 31-44.

Поступила в редакцию 01.12.95
Окончательный вариант 12.12.97