МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ 2013 год, том 25, номер 4, стр. 126-134

УДК 519.72

КОМПЬЮТЕРНЫЙ АНАЛИЗ ПРОЦЕССОВ СТРУКТУРООБРАЗОВАНИЯ НУКЛЕИНОВЫХ КИСЛОТ

© *2013 г.* ***Н.Н. Козлов, Е.И. Кугушев, Т.М. Энеев***

Институт прикладной математики им. М.В. Келдыша РАН, Москва
gencod@keldysh.ru

Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (коды проектов 10-01-00145,11-01-00110).

Многолетний опыт математического моделирования сложных космологических систем (звездные системы, протопланетные облака) позволил организовать исследования процессов структурообразования макромолекул рибонуклеиновых кислот. Принципиально новым в нашем подходе является моделирование не только структурообразования как отдельного явления, но процесса рождения макромолекулы в целом. Представляются результаты исследования процессов структурообразования макромолекул РНК. Применение многопроцессорных вычислительных комплексов позволило провести серию вычислений для существенно более длинных молекул РНК – ферментов, длина которых превышает несколько сотен нуклеотидов. Полученные результаты не только подтверждают гипотезу о прерывистости процесса удлинения молекулярной цепи, но и позволяют сделать оценку периода этого процесса.

Ключевые слова: процессы структуризации РНК, первичная, вторичная, третичная структуры, транскрипция.

COMPUTER ANALYSIS OF STRUCTURIZATION PROCESSES OF NUCLEIC ACIDS

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

Keldysh Institute of Applied Mathematics of RAS, Moscow

The long-term experience of mathematical modeling complex cosmological systems (star systems, protoplanets cloud) has allowed organizing researches of structurization processes of macromolecules of ribonucleic acids. Essentially new in our approach the modeling not only structurization as separate phenomenon, but process of macromolecule birth as a whole is. The results of research of structurization processes of RNA macromolecules are represented. The application of multiprocessor computer complexes has allowed to carry out a series of calculations for much longer molecules RNA – ferments, which length exceeds some hundreds nucleotides. The received results not only confirm a hypothesis about intermittence of process of lengthening of a molecular circuit but also allow making an estimation of the period of this process.

Key words: interacting structuring processes of RNA, primary, secondary, tertiary structures, transcription.

Определение пространственной структуры биологических макромолекул по известной первичной структуре – одна из важнейших фундаментальных проблем современной молекулярной биологии, поскольку пространственная структура молекулы определяет в основном ее биохимические свойства. Опираясь на опыт моделирования аккумуляционного процесса формирования планетных систем, академик Т.М. Энеев предложил использовать метод виртуальных контактов (метод расчета указанного аккумуляционного процесса) для построения эффективного алгоритма реконструкции сложных биологических макромолекул. 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 основе многих компьютерных методов определения пространственных структур биомолекул лежит поиск глобальных или локальных минимумов свободной энергии биомолекулы. Главные трудности, возникающие здесь – большой объем вычислений, резко растущий с размером молекулы, и большое количество локальных минимумов, среди которых нужно выбрать минимум, соответствующий реальной пространственной структуре молекулы. Наш подход состоит в использовании моделей процессов последовательного возникновения и роста биомолекул в живой клетке (трансляция, репликация, транскрипция). Это, с одной стороны, приводит к значительному понижению числа возникающих локально устойчивых конфигураций биомолекулы, с другой стороны, оказалось, что данные модели допускают эффективное распараллеливание вычислительных процессов, происходящих в них, и это позволяет достигать разумного вычислительного времени на современных параллельных вычислительных системах. Пространственная структура молекулы строится в два этапа. Сначала определяется ее вторичная структура, а затем – третичная. Нами были предложены принципиально новые подходы к решению указанных задач. Основная идея определения вторичной структуры заключается в моделировании последовательного процесса ее формирования в ходе постепенного роста молекулярной цепи. По мере роста молекулы строится цепочка межструктурных переходов от состояния, когда вторичная структура еще отсутствует, к состоянию, когда молекула обладает полной локально устойчивой вторичной структурой. Во время каждого межструктурного перехода происходит локальная минимизация свободной энергии молекулы. Последовательность этих переходов определяется тем, каким образом образуется молекулярная цепь РНК в ходе транскрипционного процесса. Подход этот был назван последовательным. Его применение дало заметное повышение качества предсказания вторичных структур РНК и позволило выдвинуть гипотезу о прерывистом характере транскрипции.

Иерархия структур нуклеиновых кислот

С точки зрения наиболее энергетически сильных связей структуру молекулы РНК (ДНК) можно описывать иерархически. Это, прежде всего, первичная структура молекулы, описывающая ее как цепочку нуклеотидов, последовательно соединенных наиболее сильными фосфодиэфирными связями. Некоторые нуклеотиды в этой цепочке связаны попарно уотсон-криковскими связями, которые также достаточно сильны. Структура этих связей называется вторичной структурой РНК (ДНК). В силу конечности множества возможных уотсон-криковских связей число вторичных структур, которые может

принимать данная молекула, конечно, но весьма значительно. Определение реальной вторичной структуры молекулы по ее известной первичной структуре также является важной фундаментальной проблемой молекулярной биологии. Под третичной структурой молекулы понимается пространственная форма, которую принимает ее молекулярная цепочка в пространстве под воздействием уотсон-криковских и других более слабых потенциалов. Четвертичной структурой называется форма молекулы, которую она приобретает, связываясь в комплекс с другими биомолекулами. Первичная структура у молекулы РНК одна, а возможных вторичных (третичных, четвертичных) структур много. Опишем элементы вторичной структуры (рис.1).

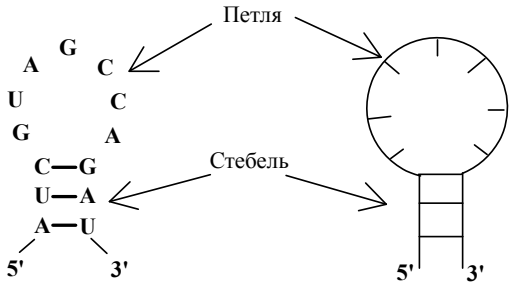


Рис.1. Однонитевый (петля) и двухнитевый (стебель) элементы вторичной структуры РНК.

С точки зрения вторичной структуры все нуклеотиды в молекуле можно разбить на два класса – спаренные (т.е. образующие уотсон-криковскую связь с каким-либо другим нуклеотидом) и свободные (неспаренные). Формально вторичная структура РНК – это описание всех спаренных и свободных оснований в молекулярной цепи. Двухнитевые или двуспиральные участки во вторичной структуре РНК называются стеблями. Длиной стебля называется число пар оснований в нём. Однонитевые участки (и семейства таких участков) называются петлями. Таким образом, вторичная структура РНК – это совокупность стеблей и петель.

Петлей называется замкнутая последовательность однонитевых участков РНК, концы которых соединены уотсон-криковскими вторичными связями. При этом начало каждого следующего участка соединено с концом предыдущего, а конец последнего участка соединен с началом первого. Однонитевые участки, входящие в состав петли, называются ее ветвями или звеньями. Длиной петли называется число свободных нуклеотидов, входящих в ее состав. Выделяют следующие типы петель (рис.2).

Шпилечная петля – соединяет первую и вторую нить в одном стебле. Это простейшая однозвенная петля (она состоит из одного однонитевого участка). Считается, что шпилечная петля всегда содержит не менее трех нуклеотидов. Боковая петля содержит два однонитевых участка, причем один из которых вырожден или имеет нулевую длину (не содержит ни одного несвязанного нуклеотида). Длина же второго участка называется длиной боковой петли. Внутренняя петля –содержит два однонитевых участка. Длины этих участков являются параметрами, определяющими петлю. Многозвенная петля – содержит несколько однонитевых участков. Число этих участков и их длины являются параметрами, определяющими петлю.

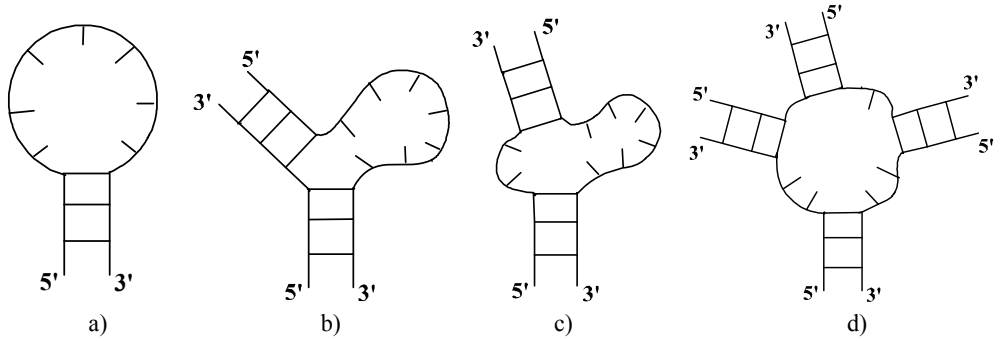


Рис.2. Типы петель (однонитевых участков вторичной структуры): a) шпилечная петля; b) боковая петля; c) внутренняя петля; d) многозвенная петля.

Модель процесса структурообразования

Наш подход состоит в математическом моделировании двух основных черт транскрипционного процесса: последовательного роста молекулярной цепи РНК в ходе транскрипции и постепенного последовательного формирования ее вторичной и третичной структуры. В основу модели образования вторичной структуры РНК положено взаимодействие двух основных процессов, влияющих на структурообразование. Первый – элонгация – последовательный рост молекулярной цепи в ходе транскрипции. Второй – структуризация – последовательное возникновение и формирование вторичной структуры РНК на том участке ее молекулярной цепи, который уже образовался к данному времени. Структурные перестройки при этом обеспечивают локальную минимизацию свободной энергии сформировавшегося участка молекулы. Оба этих процесса рассматриваются как дискретные. Молекулярная цепь удлиняется на целое число нуклеотидов, вторичная структура изменяется путем возникновения или разрыва целого числа вторичных связей. Отметим, что основным параметром в такой модели является относительная скорость элонгации T, определяемая как отношение скорости роста молекулярной цепи к скорости структурообразования (параметр структуризации). В рамках нашей модели параметр T можно рассматривать как количество новых нуклеотидов, на которое удлинится молекулярная цепь РНК за то время, пока на старом ее участке будет происходить формирование и стабилизация вторичной структуры. На рис.3. представлены конкурирующие стебли, из которых необходимо выбрать естественную структуру.

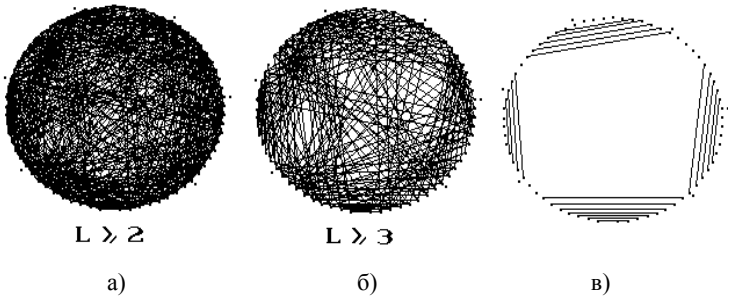


Рис.3. Транспортные РНК. Конкурирующие стебли длины L для аспарагиновой тРНК человека (а-б) и ее естественная вторичная структура – клеверный лист (в).

Взаимодействие между указанными процессами моделируется в простейшей форме. Считается, что фазы роста молекулярной цепи и структуризации разделены во времени и циклически повторяются одна за другой. Скорости обоих процессов полагаются постоянными и не зависящими от нуклеотидного состава молекулярной цепи РНК, от сложности образующейся структуры и других обстоятельств. При таких допущениях модель роста вторичной структуры РНК в ходе транскрипции выглядит следующим образом (см. табл.1).

Таблица 1. Точность описания структуры падает при увеличении минимально допустимой длины стеблей *L* (приведены средние значения для транспортных РНК).

Длина	Число стеблей	Способность описывать структуры
≥1	893.3	100%
≥2	270.5	98.7%
≥3	109.9	95.1%

На рис.4. представлено разнообразие структурных элементов при увеличении точности описания вторичной структуры.

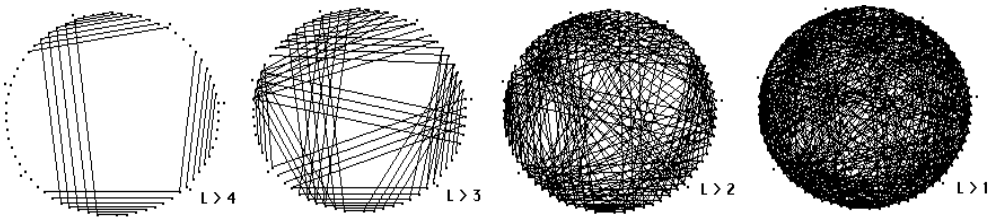


Рис.4.

Разнообразие структурных элементов резко растет при увеличении точности описания вторичной структуры, т.е. при уменьшении минимально допустимой длины стеблей *L*.

Процесс описывается как пошаговый. На каждом шаге сначала происходит структуризация, при которой вторичная структура РНК формируется в пределах того участка молекулярной цепи, который уже образовался к данному шагу процесса. После образования и стабилизации вторичной структуры происходит удлинение молекулярной цепи на постоянное число нуклеотидов, равное параметру структуризации процесса *T*. После этого происходит переход к следующему шагу процесса. Процесс заканчивается после того, как молекулярная цепь вырастет полностью и структура полностью сформируется.

Известно, что некоторые молекулы РНК могут образовываться в составе более длинного фрагмента (транскрипта), включающего их. Чтобы отразить этот факт, в модели используется еще один параметр L_0 – начальная длина молекулярной цепи. Это число нуклеотидов молекулярной цепи, уже образовавшихся к началу первого шага процесса. Таким образом, на шаге процесса с номером k молекулярная цепь содержит $L(k) = L_0 + (k-1)T$ нуклеотидов. На последнем шаге процесса величина $L(k)$ ограничивается полной длиной молекулярной цепи, обозначаемой $Lrna$. Поскольку оба параметра процесса T и L_0 целочисленные, то исходя из физического смысла мы имеем $T \geq 1$,

$0 \le L_0 \le T-1$. Если $T > Lrna$, то полное развитие процесса происходит уже на первом шаге и результат совпадает с результатом для случая $T = Lrna$. Поэтому поведение процесса образования вторичной структуры РНК моделируется в области $1 \le T \le Lrna$, $0 \le L_0 \le T-1$ на дискретной решетке $T = 1,2,3,...$; $L_0 = 0,1,2,...$.

Опишем теперь более подробно, как моделируется структуризация, т.е. процесс формирования вторичной структуры РНК на том ее участке, который уже образовался. Это тоже пошаговый процесс. К началу структуризации молекулярная цепь может уже обладать вторичной структурой, возникшей на предыдущих шагах основного процесса. В нашей модели структура наращивается путем добавления к ней элементарных структурных элементов – стеблей. При добавлении к структуре нового стебля в ней возникают новые вторичные связи и могут разрываться старые. Модель структуризации включает в себя правила формирования множества допустимых стеблей и правила определения межструктурного перехода. В наиболее широком варианте множество допустимых стеблей включат в себя все стебли, оба спаренных участка которых находятся в пределах той части молекулярной цепи, которая имеется к данному моменту времени. В этом случае возможно не только наращивание вторичной структуры (т.е. возникновение новых вторичных связей), но и ее перестройка, т.е. разрушение старых связей, препятствующих возникновению новых. В более узком варианте множество допустимых стеблей содержит только стебли, наращивающие структуру без необходимости разрыва старых связей. В этом случае вторичная структура формируется без перестроек. Заметим, что множество допустимых стеблей можно рассматривать как вариант множества допустимых межструктурных переходов. В настоящей работе мы будем рассматривать только модель процесса с разрешением любых структурных перестроек. Некоторые результаты расчетов представлены на рис.5.

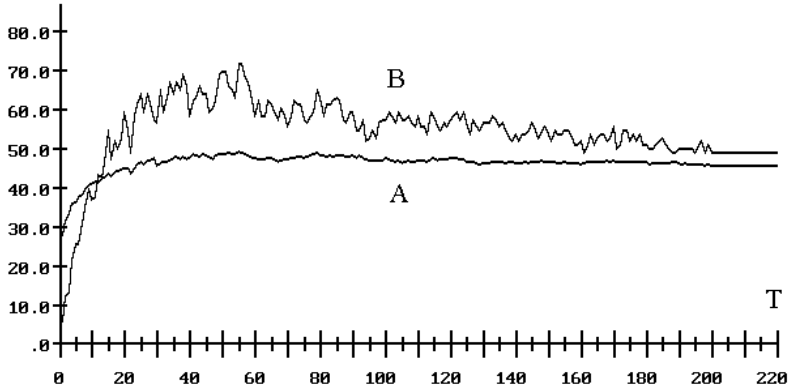


Рис.5. Расчет вторичной структуры 106 молекул РНК-ферментов. Качество предсказания в зависимости от скорости приращения молекулярной цепи. Плавный график (А) – средний процент правильно предсказанных связей вторичной структуры. Прерывистый график (В) –средняя доля молекул, у которых вычисленная вторичная структура не менее чем на 50% совпадает с натуральной. В диапазоне скорости роста молекулярной цепи от 20 до 60 нуклеотидов за раз наблюдается заметный рост качества предсказания вторичной структуры.

Как уже упоминалось, структуризация моделируется как пошаговый процесс. На каждом ее шаге формируется множество допустимых стеблей. Затем выбирается такой стебель, добавление которого к структуре даст наибольшее понижение ее свободной энергии. Этот стебель добавляется к структуре, после чего повторяется следующий шаг структуризации. Процесс кончается после того, как структура стабилизируется. Это означает, что множество тех допустимых стеблей, добавление которых способно понизить свободную энергию структуры – пусто. Иначе говоря, процесс кончается, когда мы попадаем в точку локального минимума свободной энергии вторичной структуры на множестве допустимых межструктурных переходов. После окончания процесса структуризации молекулярная цепь удлиняется и цикл шагов основного процесса повторяется.

Пространственная структура

Для перехода к пространственной структуре обратимся к рис.6.

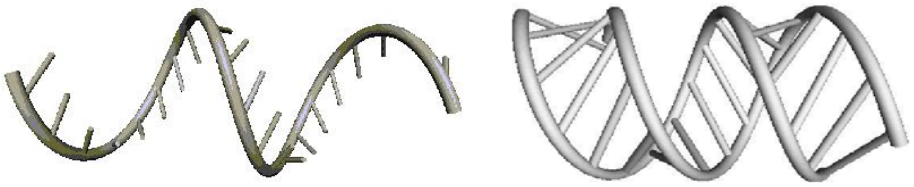
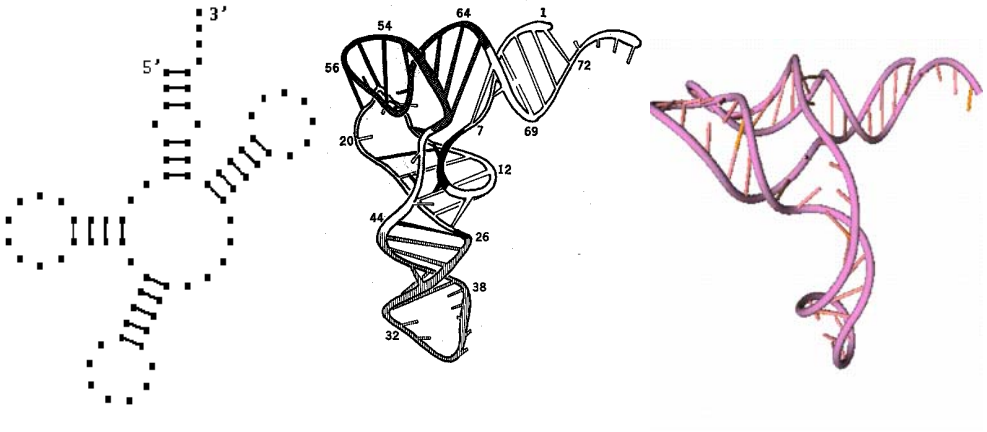


Рис.6. Свободный тонкий упругий стержень и двойная спираль ДНК.



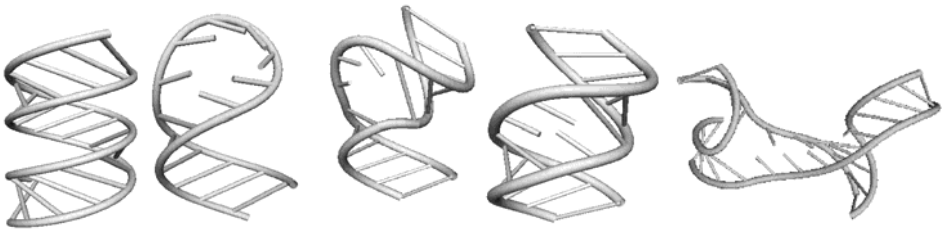
a) b) c)

Рис.7. Пример пространственной структуры молекулы тРНК Phenilalanine Yeast, полученной методами математического моделирования: a) вторичная структура; b) рентгеноструктурный анализ [5]; c) компьютерные вычисления.

Математическая модель двухспиральной молекулы ДНК на основе тонкого упругого прямолинейного однородного симметричного стержня стала уже классическим средством для изучения ее пространственных форм [8]. В [4] была предложена математическая модель пространственной структуры молекулы РНК, в которой молекулярная

цепь рассматривается как тонкий упругий стержень, имеющий в ненапряженном состоянии форму винтовой линии. С шагом, равным длине одного нуклеотида, в стержень вделаны жесткие перемычки, равные по длине половине уотсон-криковской связи. В простейшей модели параметры винтовой линии и ориентация перемычек выбираются так, что два свободных стержня одинаковой длины, будучи правильно расположены, образуют двойную спираль в А-форме. Пример пространственной структуры дан на рис.7.

Пространственная структура молекулы собирается из стеблей и петель в соответствии с заданной вторичной структурой. Каждая петля состоит из семейства тонких упругих стержней с взаимно-согласованными краевыми условиями равновесия. В соответствии с краевыми условиями концы стержней ориентируются так же, как концы нитей в стебле А-формы ДНК. Пространственная структура молекулы строится в два этапа. Сначала определяется ее вторичная структура, а затем третичная. При построении третичной структуры молекулярная цепь рассматривается как тонкий упругий стержень с поперечными стяжками, соответствующими вторичной структуре. При соответствующем выборе параметров стержня его пространственная форма приближенно описывает пространственную структуру молекулы РНК. В нашей работе РНК описывается как упругий стержень с абсолютно твердыми поперечными связями. Пространственная структура молекулы собирается из базовых элементов (рис.8.), каждый из которых –замкнутый контур, состоящий из семейства упругих стержней, последовательно соединенных между собой жесткими перемычками. Определение статически устойчивой пространственной формы базовых элементов сводится к решению системы краевых задач, каждая из которых состоит в определении конфигурации тонкого упругого стержня, удовлетворяющей геометрическим условиям на его концах. Пространственная форма стержня определятся на основе уравнений равновесия. В положении равновесия достигает экстремума упругая энергия.



a) b) c) d) e)

Рис.8. Элементы третичной структуры РНК: a) стебель; b) шпилечная петля; c) боковая петля; d) внутренняя петля; e) многозвенная петля.

Пространственная форма молекулы, рассчитанная на основе наших моделей, воспроизводит топологию пространственной структуры, полученной методами рентгеноструктурного анализа. Предлагаемый подход позволяет вычислять пространственные структуры молекул РНК, для которых не проводился рентгеноструктурный анализ (рис.9).

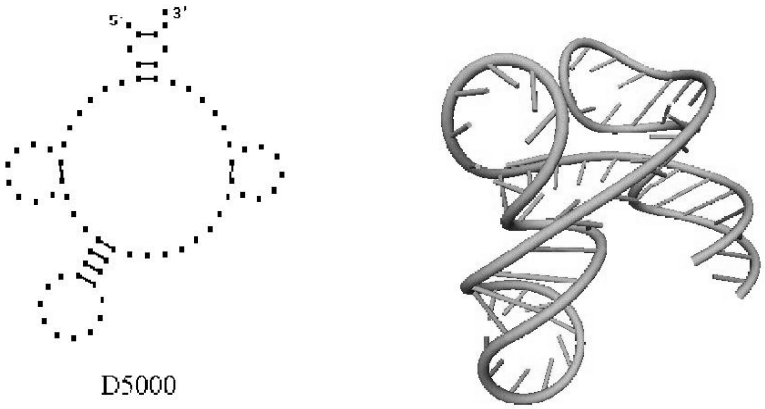


Рис.9. Пример построенной третичной структуры транспортной РНК аспарагиновой кислоты Asterina Pectini.

СПИСОК ЛИТЕРАТУРЫ

1. *Энеев Т.М., Козлов Н.Н., Кугушев Е.И.* Процессы структуризации биомолекул. Результаты математического моделирования. – М.: ИПМ им. М.В. Келдыша РАН, 1995, препринт №69, 22 с.
2. *Козлов Н.Н., Кугушев Е.И., Энеев Т.М.* Структурообразующие характеристики транскрипционного процесса // Математическое моделирование, 1998, т.10, №6, с.3-19.
3. *Козлов Н.Н., Кугушев Е.И., Энеев Т.М.* Параллельные вычисления при решении некоторых задач астрофизики и молекулярной биологии // Математическое моделирование, 2000, т.12, №7, с.65-70.
4. *Кугушев Е.И., Старостин Е.Л., Пирогова Е.Е.* Математическая модель образования трехмерной структуры РНК. – М.: ИПМ им. М.В. Келдыша РАН, 1997, препринт №77, 24 с.
5. *Kim S.H., Suddath F.L., Qugley G.J., McPherson A., Sussman J.L., Wang A.H.J., Seeman N.C., Rich A.* (1974) Three-Dimensional Tertiary Structure of Yeast Phenylalanine Transfer RNA. Science, 185, N 4149, 435-440.
6. *Lee S.-Y. and Zuker M.* // Journ. of Biomol. Struct. and Dynamics, 1991, v.8, №5, p. 1027-1044.
7. *Sprinzl M., Dank N., Nock S., Schon A.* Compilation of tRNA sequnces and sequences of tRNA genes // Nucleic Acids Res., 1991, v.19, suppl., p. 2127-2171.
8. *Benham C.J.* Geometry and mechanics of DNA superhelicity // Biopolymers, 1983, v.22, №11, p.2477-2495.

Поступила в редакцию 30.10.12