Math-Net.Ru

Общероссийский математический портал

Л. С. Бекасов, О стратегиях моделирования генетических структур, *Вестн. Сам. гос. техн. ун-та. Сер. Физ.-мат. науки*, 2005, выпуск 38, 158–162

DOI: <https://doi.org/10.14498/vsgtu394>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 52.87.193.239

26 апреля 2023 г., 14:45:00

Р и с. 4. Зависимость интегральной функции распределения $F(k)$ от числа фактически хранящихся контейнеров при разной средней вместимости площадки

Построенные графики дают возможность определить требуемую вместимость терминала с заданной доверительной вероятностью. Из графиков видно, что средний запас грузов на складе соответствует наиболее вероятному запасу, однако он может быть превышен. С увеличением размера терминала абсолютная величина возможного отклонения действительного количества хранящегося груза от среднего значения увеличивается. При доверительной вероятности $[p]=0,95$ относительное отклонение не превышает $15 \div 20\%$.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Маликов О. Б. Деловая логистика. СПб: Политехника, 2003. 223 с.
2. Маликов О. Б. Склады гибких производств. Л.: Машиностроение, 1986. 183 с.
3. Вентцель Е. С., Овчаров Л. А. Теория вероятности и ее инженерные приложения. М.: Наука, 1991. 383 с.
4. Тихоненко О. М. Модели массового обслуживания в информационных системах. Минск: Техпринт, 2003. 327 с.

Поступила 9.02.2005 г.

УДК 57.01+577.4

Л. С. Бекасов

О СТРАТЕГИЯХ МОДЕЛИРОВАНИЯ ГЕНЕТИЧЕСКИХ СТРУКТУР

Представлены основные сведения из начал генетики, касающиеся построения дезоксирибонуклеиновой кислоты (ДНК) и синтеза белков. Описаны основные особенности построения этой сверхсложной системы, касающиеся её формы и сущности её элементарных компонент — нуклеотидов. Предложены некоторые подходы формализации нуклеотидов и их характера связей в биоструктурах.

Элементарной единицей биологических систем является органическая клетка. Внутри таких клеток у эукариот (растения, животные, люди) содержится молекула дезоксирибонуклеиновой кислоты (ДНК). ДНК, как считают биологи, хранит в себе всю информацию о строении того организма, которому она принадлежит [1, 2]. Конструктивно эта молекула представляет собой «веревочную лестницу», скрученную в спираль. Ступеньки этой лестницы образуются за счет химических структур, четырех основных органических соединений — нуклеотид, именуемых как аденин (А), тимин (Т), гуанин (G) и цитозин (С). Эти конструкции в ДНК соединяются строго попарно на комплементарной основе (аденин — тимин, гуанин — цитозин). Особенность в их соединении определяется тем, что оно реализуется только при помощи ато —

мов водорода. Причем со стороны А и Т участвуют по одному атому, следовательно цепочка АТ скреплена двумя атомами водорода, тогда как другая цепочка GC скреплена тремя атомами водорода: один атом со стороны цитозина и два атома со стороны гуанина. Число ступенек (оснований) в ДНК очень большое и в организмах эта молекула свёрнута в клубок, т.е. она имеет двойной характер упаковки, что обуславливает различие нуклеотидов с точки зрения пространственной ориентации, а тем самым различие в силах взаимного влияния. Эти особенности в совокупности с характером расположения оснований в «лестнице» и составляют информационную сущность ДНК. Длина ДНК внутри какого-либо типа эукариот всегда постоянна, и у человека составляет, примерно, 3×10^9 пар нуклеотидов (п. н.).

Каждая клетка какого-либо функционального органа животного постоянно генерирует белки *одного типа*. Это её главная задача. Иногда тип белка может корректироваться [2], но не существенно.

Здесь необходимо оговориться, что белок генерируется при помощи 21 разновидности аминокислот. Механизм же генерации аминокислот сложный, и протекает с использованием *системы* рибонуклеиновых кислот (РНК), каждая из которых выполняет свои функции (транспортную, матричную, рибосомную). Сами же молекулы РНК синтезируются на базе ДНК. Осуществление синтеза происходит после того как ДНК расщепляется т.е. «лестница» расплетается вдоль, в строго определенных местах и на одной из ее частей начинают «наращиваться» нуклеотиды, однако основанием, на котором наращиваются комплементарные нуклеотиды, служит молекула РНК, которая называется матричной (мРНК). Более того, комплементарным нуклеотидом для гуанина теперь уже является не цитозин, а его модифицированный аналог — урацил (U). Строго определенная последовательность генерируемых аминокислот аддитивно объединяется рибосомной, матричной и транспортной РНК в полипептидную цепь, образуя непосредственно белок, который затем выводится из клетки. Такова общая структурная модель генератора белков, как одного из самых главных компонентов живого организма.

Особенность упаковки нуклеотидов в молекуле ДНК состоит в том, что некоторые множества последовательно расположенных нуклеотидов участвуют в синтезе аминокислот, т. е. являются информативно востребуемыми компонентами гена, и в структурном отношении организованы довольно сложно. Эта сложность проявляется в том, что такой фрагмент ДНК является обособленной единицей, снабжённой элементами *управления, запуска, непосредственно информационной частью* — экзонами, которая перетасована с «дремлющими» секторами гена — интронами. Протяженность экзонов и интронов теорией генетики четко не определена. Считается [2], что интроны — «дремлющая» часть молекулы ДНК в некоторых случаях, могут модифицировать генерируемый ею белок, т. е. интрон может выполнять информационные функции и становиться рабочей частью цепочки нуклеотидов. Структура гена как совокупность его регуляторных, промоторных и информационных составляющих представлена на рисунке. Звено, именуемое промотором осуществляет распознавание начала информационной части отрезка ДНК и инициацию транскрипции информационной части гена. Исследования этой части ДНК [3] показывают, что она содержит как детерминированные сайты, выполняющие сигнальные функции, так и вставки, назначение которых не определено. Авторы высказывают гипотезу о том, что они предназначены для сохранения периодичности кодов в промоторах, но могут также выполнять какие-то информационные задачи.

Структурные особенности эукариотического гена, кодирующего белок

В работах [4, 5] представлены методы моделирования биоструктур, в частности С. В. Петухов [4] предложил оригинальную теорию, основанную на тензорном анализе, моделирующую водородные связи в АТ и ГС основаниях, (только такой совокупности), которая выявляет динамику процессов в ДНК применительно к временным поясам суток. М. Я. Эйнгорин [5] рассматривает синтез белка, исходя из представления АТГС — алфавита конъюнктивной нормальной формой, что однозначно определяет механизм его организации. Однако такая теория не позволит точно смоделировать ген полностью, поскольку все его сайты содержат совокупности нуклеотидов, назначение которых остаётся невыясненным. Кроме того у про- и эукариотов некоторые стандартные функции в регуляторных и промоторных частях гена выполняются различными кодами [2]. Язык, как средство общения человека с человеком, предназначен для обмена некоторыми сведениями. Такой обмен позволяет накапливать знания, корректировать действия человека, способствует динамике эмоций и т. д. Принципиально язык состоит из двух составляющих, — синтаксиса Ξ и семантики Φ . Синтаксис включает в себя множество $\{\xi\}$ различных символов, — букв и специальных знаков, тогда как семантика задаётся множеством $\{\phi\}$ заранее оговоренных правил интерпретации как отдельных символов, так и их совокупностей. Элементы ξ_i синтаксиса определяются той предметной областью, к которой принадлежит язык, как средство передачи каких-либо сообщений, об объектах, явлениях и ощущениях в различных формах этих сообщений (текстовых, акустических, видовых, жестовых, физических, химических, биологических и т.д.). Однако следует отметить одну важную характеристику этих элементов, касающуюся их содержательности, иначе, в зависимости от типа языка элемент ξ_i синтаксиса может быть:

- чисто абстрактным символом,
- обозначать какой-то объект (предмет, действие), либо ощущение.

В первом случае значимость символа ξ_i определяется его местом в слове, а также совокупностью других символов, расположенных, перед ним и после него, т. е. всей системой символов (слова).

Во втором случае символ — ξ_i (например, иероглиф) может иметь свойство адекватности некоторому объекту (возможно действию).

Элементы ϕ_j семантики также определяются предметной областью или принадлежностью самого языка популяциям (национальностям) людей или иным живым существам. Что касается всей совокупности семантических правил того или иного языка, то ответ на этот вопрос вряд ли может быть однозначным. Безусловно то, что во всех языках существуют некоторые множества $\{\phi_j\}$, такие, которые регламентируют сочетания синтаксических элементов ξ_i в слове. Именно порядок расположения элементов синтаксиса и их количество отдельно взятого подмножества определяют семантику слова в целом.

Однако текстовая форма языка появилась много позже её акустического аналога. В этой связи следует обратить внимание на количественную форму синтаксических элементов в языке. Так, если рассматривать алфавит и специальные знаки европейских письменных языков, то можно констатировать что их количество имеет один и тот же порядок, а конкретнее это 28–40 символов. Синтаксис языков азиатских стран и стран Ближнего Востока базируется на иероглифах. Это существенное отличие в синтаксисе обуславливает также существенное различие в семантике. Акустический аналог отражения ощущений, мыслей и иных чувств (разговорная речь человека) имеет такую же структурную организацию однако его синтаксис определен анатомическими особенностями. В частности, речевые звуки [7] генерируются при помощи *четырёх формант* (четырёх различных несущих частот). Их значение определено физиологическими особенностями, что обуславливает неповторимость человеческого голоса. Сходство четырёх формант с четырьмя нуклеотидами — основными «буквами» генетического кода и структурой организации многих белков, (например иммуноглобулин [2]), состоящую из четырёх полипептидных цепочек, порождает адекватную интерпретацию совокупностей этих букв и наталкивает на мысль о том что в генетике может доминировать четверичная система счисления. Другими словами, если сущность обычного слова зависит как от самих букв, так и от того, с какими буквами рядом (справа и слева) находится рассматриваемая буква, то этот принцип целесообразно применить и при анализе генетических кодов, используя, например, аппарат исчисления предикатов. Предикат в формальных языках отражает либо свойство (одноместный предикат) объекта $P(x)$, либо

отражает либо свойство (одноместный предикат) объекта $P(x)$, либо отношение (многоместный предикат) между двумя и более объектами $Q(x, z)$, где P, Q — предикаты; x, z — объекты.

Такую форму представления можно осуществить к сочетаниям генетических букв А, Т, G, C, (U) и отдельно взятых их параметров.

На следующем этапе необходимо интерпретировать значимость конкретного нуклеотида либо основания в их протяжённости по двуспиральной цепочки ДНК (в полной мере, либо фрагментарно). Каждый нуклеотид это структура, состоящая из гетероциклического азотистого основания (одного из четырёх) и дезоксирибофосфата, которые на комплементарной основе соединяются с себе подобными, образуя основания молекулы ДНК. При этом какие-либо указания на то, что эта структура неустойчива, отсутствуют. Следовательно, необходимо искать некий количественный эквивалент корреляции между двумя соседними гетероциклическими азотистыми основаниями.

Анализ химической структуры [1] фрагмента ДНК побуждает использовать максимально возможное число физических и химических параметров. В частности, В.И. Полтев с соавторами [8] получил подробные результаты расчетов минимума энергии взаимодействия между основаниями в разных взаимных положениях исходя из атом-атомных потенциальных функций, которые состоят из двух компонент: кулоновской и потенциала, зависящего от межатомного расстояния (потенциал Леонарда-Джонса). Если же учитывать мутационные процессы т. е. изменение химической формулы основания, то следует учитывать значение молекулярных масс μ_i элементов, входящих в нуклеотид, количество элементов в основании и их суммарную молекулярную массу $M = \sum_i \mu_i$, где i — число элементов в основании, либо удельное (по фосфату) усреднённое значение $\bar{\mu}$ этой величины.

Что касается характера связанности элементарных последовательностей цепочки (двух букв генетического алфавита, расположенных рядом), то наиболее целесообразным является представление этого параметра через некоторую функцию $W(\{\mu_i\})$. На начальном этапе можно применить непосредственно либо коэффициент

$$\gamma = \frac{\mu_{i+1}}{\mu_i}, \quad (1)$$

либо иной предикат. С другой стороны возможно использование молекулярной массы всего основания через сумму его компонент, т. е.

$$\mu^* = \mu_A + \mu_T; \mu^{**} = \mu_G + \mu_C, \quad (2)$$

где μ^* — молекулярная масса цепочки АТ, μ^{**} — молекулярная масса цепочки GС, а также через отношение $\frac{\mu^*}{\mu^{**}}$ этих величин. Поскольку сами компоненты А, Т, G, C, (U) могут при транскрипции (трансляции) или других процедурах, модифицироваться [2], то следует использовать их дифференциальные показатели, т. е.

$$\mu_{AT} = \mu_A - \mu_T; \mu_{GC} = \mu_G - \mu_C, \quad (3)$$

тогда соответствующие предикаты $S(\mu_{AT})$, $U(\mu_{GC})$ позволят получить дифференциальный характер модели распределения комплементарных компонент в ДНК.

Формулы (1)–(3) представляют собой основу для четырёх моделей генетических структур. Так, использование формулы (1) в матричном виде для 16-ти сочетаний А, Т, G, C –алфавита позволит определить характер взаимосвязи нуклеотидов как в ДНК, так и в РНК всех модификаций. Условие (2) целесообразно применить для определения характера связи всех комплементарных компонент в ДНК. Представляет интерес соотношение (3), поскольку с его помощью можно получить (используя только лишь водородные связи в пуринах и пиримидинах) две кодовые последовательности. Одна из этих последовательностей даёт возможность получить двоичный код, отражающий характер расположения компонент раздельно для 5⁺-лидерной и 3⁺-трейлерной частей ДНК, тогда как объединение этих двух кодовых конструкций делает возможным получение некоторой количественной оценки всей ДНК, но уже в троичном коде (1, 0, –1). Анализ белковых конструкций с использованием коррекции четверичного кода двумя коэффициентами, один из которых $K_{en}(E)$, где E — энергия взаимодействия между двумя осно-

ваниями, базируется на энтальпии атом-атомных связей [8], тогда как другой $K_{\mu}(S(\mu_{AT}), U(\mu_{GC}))$ напрямую связан с молекулярной массой нуклеотидов в аминокислоте. Такой подход способствует большей гибкости при анализе как трёх-, так и двунуклеотидных последовательностей, которые имеют существенное влияние в регуляторных частях генов и способен наиболее полно учитывать мутации в этих структурах. Матричный аппарат, используемый для построения математических моделей в генетических структурах также получает большие функциональные возможности.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *Кемп П., Армс К.* Введение в биологию. М.: Мир, 1988. 672 с.
2. *Сингер М., Берг П.* Гены и геномы. В двух томах. М.: Мир, 1998. Т.1. 375 с, Т. 2. 392 с.
3. *Озолин О. Н., Деев А. А., Масулис И.С., Часов В. В., Костяницина Е. Г., Пуртов Ю. А., Архипов И. В., Брок-Волчанский А. С.* Уровни структурной организации промоторной ДНК *Escherichia coli* // Биофизика, 2002. Т. 47, Вып. 5. С. 809–819.
4. *Петухов С. В.* О теории бинарных языков генетического кода и генетической алгебры. Депонировано в ВИНТИ РАН 02.02.2004 г. № 182-В2004.
5. *Эйнгорин М. Я.* Основы кодирования и управления в молекулярной биологии. Н. Новгород: Изд-во Нижегородского медицинского академии, 2001. 116 с.
6. *Вахитов Я. Ш.* Теоретические основы электроакустики и электроакустическая аппаратура. М.: Искусство, 1982. 416 с.
7. *Темников Ф. Е.* Теоретические основы информационной техники. М.: Энергия, 1979. 512 с.
8. *Полтев В. И., Дерябина А. С., Гонзалес Э., Грохлина Т. И.* Взаимодействия между основаниями нуклеиновых кислот. Новые параметры потенциалов и новые минимумы энергии // Биофизика, 2002. Т. 47, Вып. 6. С. 996–1004.

*Поступила 28.07.2005 г.
После переработки 7.09.2005 г.*