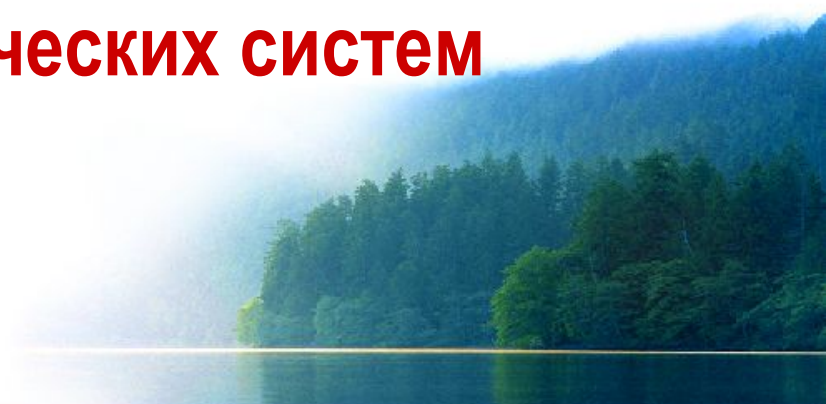




Математическое моделирование динамики молекулярно-генетических систем



Кафедра информационной биологии ФЕН НГУ



Что такое модель?



Модель, в широком смысле любой образ, аналог (мысленный или условный: изображение, описание, схема, чертеж, график, план, карта и т. п.) какого-либо объекта, процесса или явления («оригинала» данной модели), используемый в качестве его «заместителя», «представителя».

Энциклопедия Кирилла и Мефодия



Что такое моделирование?



Моделирование - исследование каких-либо явлений, процессов или систем объектов путем построения и изучения их моделей; использование моделей для определения или уточнения характеристик и рационализации способов построения вновь конструируемых объектов. Моделирование одна из основных категорий теории познания: на идее моделирования по существу базируется любой метод научного исследования как теоретический (при котором используются различного рода знаковые, абстрактные модели), так и экспериментальный (использующий предметные модели).

Энциклопедия Кирилла и Мефодия



Для чего нужна модель?



Модель нужна

Для понимания строения изучаемого объекта: его структуры, основных свойств, законов развития и взаимодействия с окружающим миром;

Для того чтобы научиться управлять объектом (или процессом) и определить наилучшие способы управления при заданных целях и критериях;

Для того чтобы прогнозировать прямые и косвенные последствия (побочные эффекты) реализации заданных способов и форм воздействия на объект.

Горстко А.Б. Познакомьтесь с математическим моделированием, 1991 г.

Приемы и способы моделирования

Моделирование

Материальное (предметное)

Способы моделирования, при которых исследование ведется на основе модели, воспроизводящей основные геометрические, физические, динамические и функциональные характеристики изучаемого объекта.

Физическое

(реальному объекту противопоставляется его увеличенная или уменьшенная копия)

- макет здания;
- модель самолета;
- планетарий.

Аналоговое

(основано на аналогии процессов и явлений, имеющих различную физическую природу, но одинаково описываемых формально)

- изучение механических колебаний с помощью электрической схемы.

Идеальное

Носит теоретический характер и основано не на материальной аналогии объекта и модели, а на аналогии идеальной, мыслимой.

Интуитивное

(основано на интуитивном представлении об объекте исследования)
- жизненный опыт (интуитивная модель окружающего мира).

Знаковое

(использует в качестве моделей знаковые преобразования какого-либо вида: схемы, графики, чертежи, формулы, наборы символов и т.д., а также включает совокупность законов, по которым можно оперировать с выбранными знаковыми образованиями и их элементами)



При разработке любой модели необходимо определить объект моделирования, цель моделирования и средства моделирования. В соответствии с объектом и целями математические модели в биологии можно подразделить на **три больших класса**.

Первый - **регрессионные модели**, включает эмпирически установленные зависимости (формулы, дифференциальные и разностные уравнения, статистические законы) не претендующие на раскрытие механизма изучаемого процесса.

Второй класс – **имитационные (портретные) модели** конкретных сложных живых систем, как правило, максимально учитывающие имеющуюся информацию об объекте. Имитационные модели применяются для описания объектов различного уровня организации живой материи - от биомакромолекул до моделей биогеоценозов.

Третий класс - **качественные модели**, разрабатываемые с целью изучения фундаментальных закономерностей функционирования ключевых элементов систем.



В соответствии с применяемыми подходами, используемыми для моделирования динамики биологических систем, знаковые модели классифицируются на

непрерывные - используют аппарат непрерывной математики, дифференциальных, интегро-дифференциальных уравнений

- **стохастические** - основаны на применении вероятностных методов
- **логические** - применяются методы дискретной математики, логики
- **гибридные** - применяются комбинации вышеперечисленных методов



Процесс построения моделей



- 1.** Формируются основные вопросы о поведении системы, ответы на которые планируется получить с помощью модели.
- 2.** Из множества законов, управляющих поведением системы, учитываются те, влияние которых существенно при поиске ответов на поставленные вопросы. При этом некоторые законы сворачиваются в один, обобщенный закон.
- 3.** В дополнение к этим законам, если необходимо, для системы в целом или отдельных ее частей формулируются определенные гипотезы о функционировании. Как правило, эти гипотезы правдоподобны в том смысле, что могут быть приведены некоторые теоретические доводы в пользу их принятия.
- 4.** Гипотезы так же, как и законы, выражаются в форме определенных математических соотношений, которые объединяются в некоторое формальное описание модели.



Одной из центральных проблем постгеномной молекулярной биологии и генетики является исследование принципов организации, функционирования и эволюции в живых системах генных сетей - молекулярно-генетических систем, обеспечивающих формирование и поддержание фенотипических (молекулярно-генетических, биохимических, физиологических, морфологических и т.д.) характеристик живых организмов на основе информации, закодированной в их геномах.



Генная сеть — группа координировано функционирующих генов, контролирующих физиологические, биохимические, молекулярные и др. функции организмов

Генные сети как правило являются большими системами со сложной структурно-функциональной организацией и сложным динамическим поведением.



Основные элементы генной сети

Ядро генной сети - группа координированно работающих генов

Пути передачи сигналов, обеспечивающие активацию или подавление экспрессии

РНК и белки, кодируемые генами

Низкомолекулярные компоненты, осуществляющие переключение функции генных сетей в ответ на внешние воздействия (гормоны и другие сигнальные молекулы), энергетические компоненты, различные метаболиты и т.д.

Отрицательные и положительные обратные связи, стабилизирующие параметры генной сети на определенном уровне или, напротив, отклоняющие их от исходного значения, обеспечивая переход к новому функциональному состоянию



Специфика объекта моделирования



Типичные генетические системы состоят из большого количества компонентов различной природы и структурно-функциональной организации, часто функционируют на нескольких иерархических уровнях организации живых систем

- Биохимический уровень** (биохимические процессы и реакции)
- Генетический уровень** (расположение и ориентация генов и регуляторных элементов, регуляция экспрессии генов, матричные ориентированные процессы, полиаллельность и т.д.)
- Субклеточный и клеточный уровни** (компарментализация, процессы межкомпарментного обмена веществами и энергией, активный и пассивный транспорт и т.д.)
- Органый, тканевой и организменный уровни** (надкомпарменты, взаимодействие органов, тканей, организмов между собой и с внешней средой)
- Популяционный уровень** взаимодействие между особями и с внешней средой, возрастная структура популяции, эволюционные процессы)

В большинстве случаев, генные сети обладают нелинейным динамическим поведением в

силу наличия в них отрицательных и положительных обратных связей



Специфика объекта моделирования



Биологические системы являются самовоспроизводящимися динамическими системами.

Особенностью функционирования биологических систем во времени является то, что у них, как правило, в процессе функционирования не остаются постоянными параметры пространственной организации, а также могут меняться характеристики генетического строения.

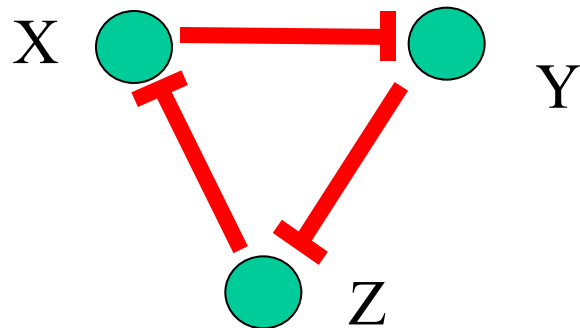
Факторы изменения пространственной организации строения биологической системы:
рост организмов, регенерация органов и тканей, деление и гибель клеток, органелл, размножение организмов, метаморфоз и т.д.

Факторы изменения генетического строения биологической системы:
репликация, рекомбинация, мутации, делеции, вставки, деление клеток, слияние половых клеток, трансгенез, привнесение новых геномов или частей генома путем внедрения в биологическую систему других биологических систем, как то вирусов, бактерий, паразитов, параллельный перенос генетической информации и т.д.



Логические подходы основаны на применении для описания механизмов в терминах логики и дискретной математики.

Приведем пример описания работы генной сети в терминах булевой логики.



Считается, что гены могут находиться в двух состояниях: “экспрессируется” (true,1), либо “не экспрессируется” (false,0). Текущее состояние генной сети описывается списком экспрессирующихся и не экспрессирующихся генов в дискретные моменты времени.

$X(\text{следующее}) = \text{NO } Z(\text{текущее})$

$Y(\text{следующее}) = \text{NO } X(\text{текущее})$

$Z(\text{следующее}) = \text{NO } Y(\text{текущее})$

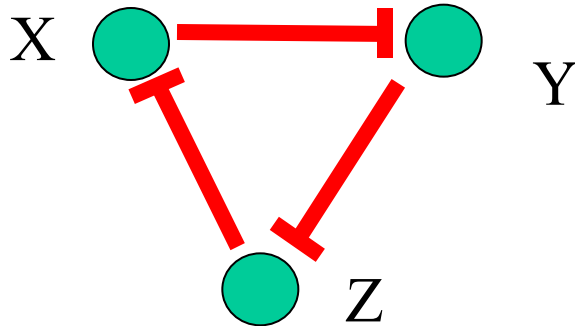


Таблица состояний:

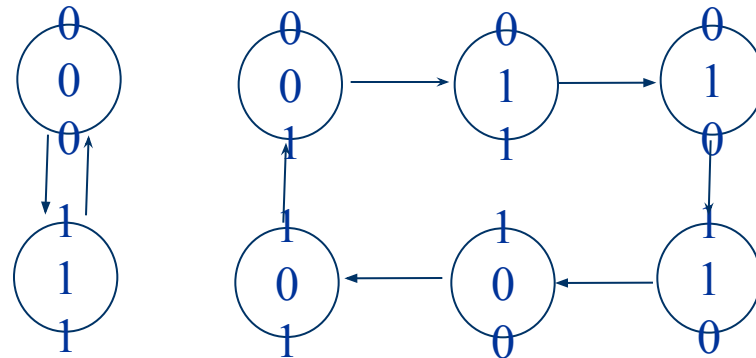
<u>текущее</u>	<u>следующее</u>
000	111
001	011
010	110
011	010
100	101
101	001
110	100
111	000

$X(\text{следующее}) = \text{NO } Z(\text{текущее})$

$Y(\text{следующее}) = \text{NO } X(\text{текущее})$

$Z(\text{следующее}) = \text{NO } Y(\text{текущее})$

Схема переходов состояний





Стохастическое моделирование



В стохастических моделях молекулярные события описываются как случайные процессы. В них каждое элементарное событие рассматривается как некоторый вероятностный процесс. Базовая процедура вычисления элементарного события состоит в выбрасывании датчиком случайных чисел некоторого числа в интервале $[0,1]$. Затем полученное число интерпретируется на основании вероятностного закона функционирования рассматриваемого события. Существуют два метода стохастического моделирования. С дискретным и непрерывным временем. В стохастическом подходе с дискретным временем весь событийный отрезок разбивается на конечное число временных отрезков. Считается, что система в текущий момент времени находится в определенной точке времени. Чтобы узнать как изменится состояние системы в следующей точке, для каждой элементарной подсистемы подсчитываются сколько за текущий отрезок времени произойдет тех или иных событий (сколько распадется молекул, сколько синтезируется, сколько модифицируется, переместится из одного компартмента в другой и т.д.) Все произошедшие события добавляются (с положительным или отрицательным знаком) к текущему состоянию. Полученное новое состояние и будет таковым для следующей временной точки.



В стохастическом подходе с непрерывным временем считается, что система из текущего момента времени переходит в следующую временную точку, которая вычисляется как минимальное время изменения состояния системы. Чтобы узнать величину этого отрезка для каждой элементарной подсистемы с помощью датчика случайных чисел вычисляется время наступления в нем одного события при условии что система имеет текущее состояние. Минимальный интервал и будет тем самым, на величину которого передвинется система в процессе временного развития. При этом текущее состояние изменится только на одно элементарное событие. Например, если элементарное событие состоит в том, что разрушилась одна молекула, то в системе общий пул данных молекул уменьшится на одну единицу, если же элементарным событием является перемещение одной молекулы из ядра в цитоплазму, то в текущем состоянии в цитоплазме станет на одну молекулу данного вещества меньше, а в ядре - на одну молекулу больше.



Динамика молекулярно-генетических систем в непрерывных моделях формально описывается системами дифференциальных уравнений, в которых могут присутствовать запаздывающие аргументы и члены, описывающие пространственную неоднородность и диффузию.

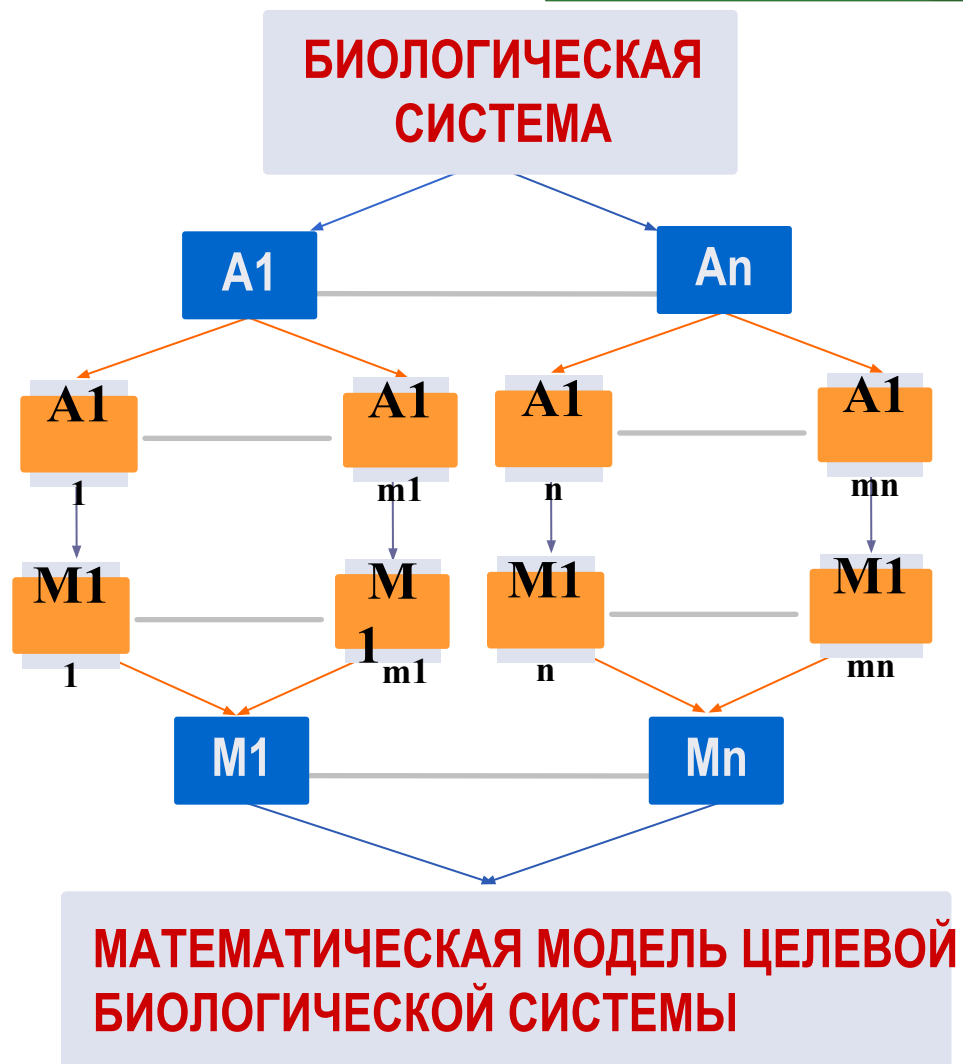


Иерархический метод моделирования



ИММ использует в качестве основного подхода принцип иерархической декомпозиции целевой биологической системы до элементарных подсистем, описание элементарных подсистем отдельными элементарными моделями и конструирования моделей целевой системы, а также ее различных модификаций из элементарных моделей, как из строительных блоков, объединяя их между собой на основе закона суммирования локальных скоростей синтеза/утилизации веществ, с учетом заданной структурно-функциональной организации моделируемой системы.

Блок схема моделирования



Элементы системы (компар-
менты, клетки, органы и т.д.)

Элементарные подсистемы
(гены, промоторы, ферменты и т.
л.)

модели
элементарных подсистем

модели элементов системы



Какие подсистемы считаются элементарными?



Элементарной подсистемой моделируемой системы считается подсистема, которая может быть рассмотрена автономно, вне строения исходной системы, и для описания ее свойств не требуется знания свойств остальных подсистем и их взаимодействий. Элементарными подсистемы понимаются только в контексте развиваемой модели. Подразумевается, что для элементарных подсистем необходимо и можно построить математические модели и в дальнейшем они могут служить элементарными строительными блоками при моделировании целевой биологической системы и ее различных модификаций.

Моделирование динамики функционирования элементарных подсистем является начальным этапом создания модели биологической системы. Процесс конструирования модели из элементарных подсистем основан на законе суммирования локальных скоростей и допускает эффективную компьютерную автоматизацию. Однако следует заметить, что разработка адекватной модели, описывающей динамику функционирования генной сети, не заканчивается этапом разработки базы элементарных моделей. После того как разработаны элементарные модели, наступает важный этап проверки поведения моделей, описывающих динамику поведения самой системы, ее подсистем и различных модификаций.

В общем случае **процесс декомпозиции биологической системы до уровня элементарных подсистем является не формализуемым процессом** и в каждом случае решается индивидуально и во многом зависит от целей моделирования и степени изученности моделируемой системы. Тем не менее, можно указать несколько групп клеточных подсистем, с которыми в основном приходится иметь дело при моделировании генных сетей.



Примерный список элементарных подсистем генных сетей.



№	Название группы
1	деградация макромолекул
2	формирование комплексов, обладающих биологически-значимыми функциональными активностями (активные формы ферментов, транскрипционные факторы, РНК-полимеразы, рибосомы и др.)
3	Ферментативные реакции с учетом регуляции их активности различными факторами.
4	генетические процессы, в том числе процессы инициации и терминации репликации, транскрипции и трансляции, с учетом их регуляции
5	элонгация репликации, транскрипции и трансляции
6	Пути передачи сигналов
7	сплайсинг
8	рекомбинация
9	Репарация,
10	транспорт
11	Созревание, модификация белков

При этом следует помнить, что обычно элементарные подсистемы с биохимической точки зрения отнюдь не являются элементарными. Таковыми мы их считаем только в рамках разрабатываемой модели. Исходя из логики и целей моделирования, одни и те же подсистемы могут быть элементарными в одних случаях и не являться таковыми в других.



В таблице мы привели примерный список элементарных подсистем. Список не претендует на полноту. Не все приведенные подсистемы обязательно присутствуют в конкретной геномной сети. Некоторые подсистемы, даже если они и входят в геномную сеть, при моделировании могут игнорироваться, так как выполняют роль обеспечивающих нормальное функционирование рассматриваемой геномной сети. Например, при моделировании часто не рассматриваются системы транскрипции, сплайсинга, транспорта, трансляции, созревания белков или описываются очень упрощенно. Решение о том, что учитывать, а чем можно поступить, всегда принимается специалистом, проводящим моделирование.

Уровень сложности, детальности описания элементарных подсистем также существенно зависит от целей моделирования.



Общий вид элементарных моделей



$$\frac{dX}{dt} = V(Y, K)$$

X - вектор управляемых переменных, Y - вектор управляющих переменных, K – список параметров. Допускается вхождение одних и тех же переменных в оба списка, но, в общем случае, списки X и Y не совпадают и могут вообще не пересекаться. Переменные обычно имеют смысл концентраций веществ или вероятностей реализации выделенных состояний веществ. Переменные из списка Y , не входящие в список X являются параметрами для текущей элементарной модели. V - описывает закон изменения скоростей концентраций (вероятностей) веществ из списка X . Функцию V будем называть управляющим законом (модели). Вид V определяется на основании данных о структурной организации моделируемой подсистемы с учетом имеющихся экспериментальных данных о кинетике ее функционирования. Определение вида функций V является основной задачей моделирования.



Методы реконструкции моделей элементарных подсистем



Биохимическая природа молекулярно-генетических процессов - основа для вывода функции V



Фундаментальной основой для вывода функции V является биохимическая природа молекулярно-генетических процессов. Поэтому для определения вида функций V могут быть применены формулы, вытекающие из биохимического подхода и кинетического закона действующих масс.



КИНЕТИЧЕСКИЙ ЗАКОН ДЕЙСТВУЮЩИХ МАСС



Кинетический закон действующих масс выводится на основе теории соударений.

Пусть имеется биохимическая реакция



Тогда скорость $V_{A+B \rightarrow C}$ формирования комплекса C в текущий момент времени t равна $V_{A+B \rightarrow C}(t) = k_1[A][B]$, а скорость $V_{A+B \rightarrow C}$ распада комплекса C на составляющие A и B в текущий момент времени t равна $V_{C \rightarrow A+B}(t) = k_2[C]$.



Если имеется конкретная биохимическая схема реакций вида (1), то мгновенная скорость изменения концентрации любого вещества равняется сумме локальных скоростей изменения концентрации данного вещества в каждой реакции, в которой данное вещество участвует. Это простое правило позволяет легко выписать итоговую систему дифференциальных уравнений, описывающую целевую биохимическую схему.



Теорема Корзухина – теоретическая основа для применения химико-кинетического подхода к моделированию молекулярно-генетических систем (не строгая формулировка)

Будем считать, что имеется некоторая совокупность веществ X_i , для которых известны временные кривые изменения их концентраций x_i в некотором фиксированном объеме. Поставим вопрос – можно ли описать данную совокупность кривых некоторой «химической» моделью. Ответ на этот вопрос дает теорема Корзухина.

Не ограничивая общности мы можем считать, что правые части дифференциальных уравнений системы произвольного порядка являются полиномами

$$\mathbf{dx}_i/dt = \mathbf{P}_i(\mathbf{X}) \quad (2).$$

Теорема Корзухина утверждает, что для любого фиксированного интервала времени $[0, T]$ и любой заданной точности существует биохимическая схема, построенная из би- и мономолекулярных реакций, которая приближает траектории системы (2) с заданной точностью.

Доказательство теоремы является конструктивным, т.е. вид системы точно выписывается.



Теоретически обосновывает применение о химико-кинетического метода для моделирования живых систем. Алгоритм доказательства может быть применен для непосредственного построения модели.

Теорема Корзухина показывает, что химико-кинетический подход является самодостаточным. В принципе для построения модели достаточно использовать только моно- и бимолекулярные реакции.

При использовании химико-кинетического подхода V является полиномом степени не выше двух. Однако на практике химико-кинетический подход в чистом виде применяется редко, так как приводит к большой избыточности переменных и параметров. Для их оценки как правило не хватает данных. Поэтому чаще химико-кинетический метод комбинируют с различными дополнительными предположениями. В частности применяется гипотеза о быстрых и медленных переменных. Применение данной гипотезы является основой для большинства упрощений моделей.

Теорема Тихонова

Пусть дана система дифференциальных автономных уравнений

$$\begin{cases} \dot{x}_i = f_i(x, y), & i = 1, \dots, p, \\ \varepsilon \dot{y}_j = g_j(x, y), & j = 1, \dots, q. \end{cases} \quad (1.13)$$

Уравнения

$$\dot{x}_i = f_i(x, y) \quad (1.14)$$

назовем системой медленных движений, а уравнения

$$\varepsilon \dot{y}_j = g_j(x, y) \quad (1.15)$$

— системой быстрых движений. Пусть далее

$$y_j = \varphi_j(x) \quad (1.16)$$

является решением системы

$$g_j(x, y) = 0. \quad (1.17)$$

Это решение называется устойчивым в области D пространства (x) , если для каждой точки (x_1^0, \dots, x_p^0) из D решение $y_j = \varphi_j(x^0)$ является устойчивым положением равновесия системы (1.15). Областью притяжения решения системы (1.16) при фиксированных (x_1^0, \dots, x_p^0) называется совокупность всех точек (y_1^0, \dots, y_q^0) , для которых при $y_i(0) = y_i^0$ и $t \rightarrow \infty$ траектории системы (1.15) стремятся к $y_j = \varphi_j(x^0)$. Система

$$\begin{cases} \dot{x}_i = f_i(x, y), \\ 0 = g_j(x, y), \end{cases} \quad (1.18)$$

получающаяся из (1.13) при $\varepsilon = 0$, называется вырожденной.



· Теорема. Решение системы (1.13) стремится к решению системы (1.18) при $\varepsilon \rightarrow 0$, если соответствующее решение системы быстрых движений устойчиво и начальные условия лежат в области притяжения этого решения. Эта теорема выполняется при $t \geq \tau$, где $\tau \sim |\varepsilon \ln \varepsilon|$.



Значение Теоремы Тихонова



Теорема Тихонова позволяет понижать порядок системы дифференциальных уравнений. Для того чтобы применить этот метод к биохимической схеме необходимо разделить переменные системы на быстрые и медленные. Быстрые переменные рассмотреть к стационаре. В результате эти переменные могут быть исключены из системы. Платой за это становится усложнение вида формул, стоящих в правой части оставшейся системы. Общего алгоритма разделения переменных на быстрые и медленные не существует.



Пример применения теоремы Тихонова

Простая каталитическая реакция



Формула Михаэлиса-Ментен

Простейшие преобразования и переход к исходным обозначениям дает знаменитую формулу

$$V = \frac{k_2 E_0 S}{K_m + S}$$

Разделение процессов на быстрые и медленные.



Как уже говорилось, общего алгоритма разделения переменных на быстрые и медленные не существует. Однако иерархическая организация больших по размерности систем основывается на существовании в них быстрых и медленных процессов. Теорема Тихонова основана на разделении переменных системы на быстрые и медленные. Но в сложных системах одни переменные могут одновременно участвовать и в быстрых и в медленных процессах. В процессе динамического развития систем, быстрые процессы протекают вблизи от своего стационарного состояния. Это позволяет рассматривать быстрые процессы изолированно (локально независимо) от других процессов, что позволяет проводить упрощение исходных систем, построенных на основании биохимических схем.



Квазистационарный метод исключения переменных (метод Кинга-Альтмана)



Важный частный случай применения принципа квазистационарности для исследования механизмов ферментативных реакций



Основу принципа квазистационарности составляет рассмотрение биохимической схемы в равновесии и применение предположения об избыточности субстратов, эффекторов, кофакторов в среде по сравнению с концентрацией фермента. На этом основании их изменением в реакции пренебрегают. В результате для поиска вероятностей пребывания системы в выделенных состояниях получается система линейных уравнений. Для нее были развиты эффективные методы поиска аналитических формул, которые являются рациональными полиномами, зависящими от концентраций субстратов, эффекторов, кофакторов, продуктов, как управляющих переменных состояниями фермента.



Метод Кинга-Альтмана



Пусть имеется фермент E, который может пребывать в n состояниях: e_i , $i=1, \dots, n$. Пусть из i-го состояния в j-е фермент переходит, следуя кинетике первого порядка, с константой скорости перехода k_{ij} . Тогда общее уравнение на i-е состояние выглядит следующим образом.

$$\frac{de_i}{dt} = k_{1i}e_1 + k_{2i}e_2 + \dots + k_{ni}e_n + \dots - \sum k_{ij}e_i$$

Метод Кинга-Альтмана

Это выражение равно нулю, поскольку мы считаем, что выполняется условие стационарности. Число выражений подобного типа равно, очевидно, n (по одному для каждой из n форм), однако независимыми являются только $(n - 1)$ из них. Действительно, одно из уравнений равно сумме остальных $(n - 1)$ уравнений. Чтобы решить систему уравнений относительно n неизвестных, необходимо иметь еще одно уравнение; его можно получить исходя из условия, что сумма концентраций всех форм должна быть равна e_0 , общей концентрации фермента:

$$e_1 + e_2 + \dots + e_n = e_0. \quad (3.2)$$

Какое именно из n исходных уравнений заменяется уравнением (3.1), неважно, однако при разрешении системы относительно e_m удобно заменять m -е уравнение. В таком случае мы получаем следующую систему уравнений:

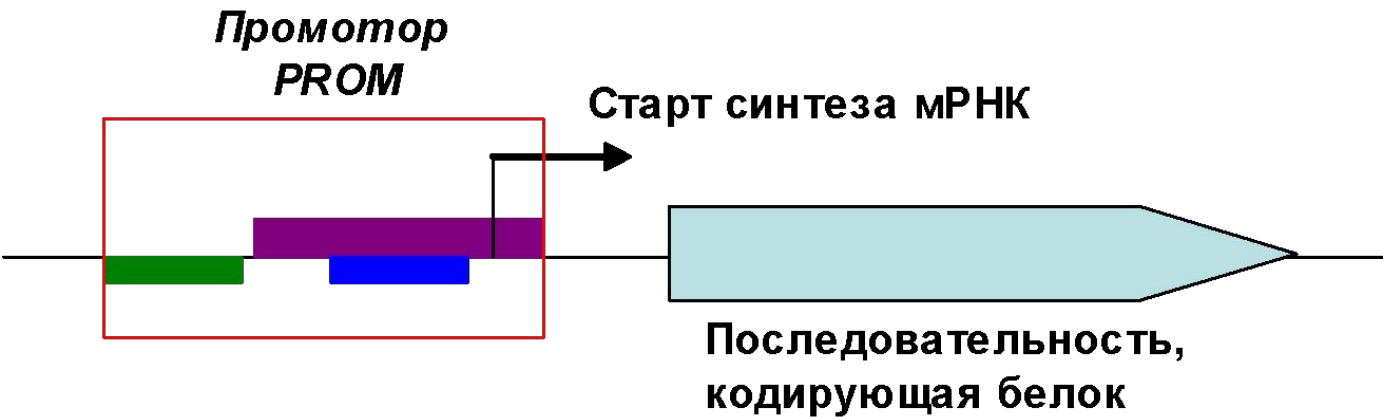
$$\begin{aligned} - \sum k_{1j} e_1 + k_{21} e_2 + \dots + k_{m1} e_m + \dots + k_{n1} e_n &= 0, \\ k_{12} e_1 - \sum k_{2j} e_2 + \dots + k_{m2} e_m + \dots + k_{n2} e_n &= 0, \\ \vdots & \\ e_1 + e_2 + \dots + e_m + \dots + e_n &= e_0, \\ \vdots & \\ k_{1n} e_1 + k_{2n} e_2 + \dots + k_{mn} e_m + \dots - \sum k_{nj} e_n &= 0. \end{aligned}$$






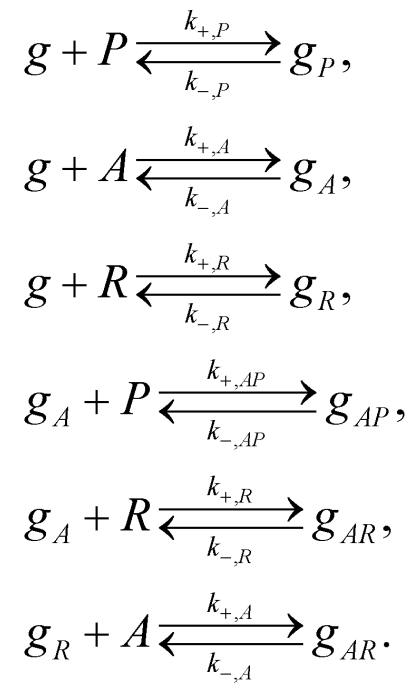
Эту систему уравнений можно решить, используя правило Крамера,

$$e_m = \frac{\begin{vmatrix} -\sum k_{1j} & k_{21} & \dots & 0 & \dots & k_{n1} \\ k_{12} & -\sum k_{2j} & \dots & 0 & \dots & k_{n2} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & e_0 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ k_{1n} & k_{2n} & \dots & 0 & \dots & -\sum k_{nj} \end{vmatrix}}{\begin{vmatrix} -\sum k_{1j} & k_{21} & \dots & k_{m1} & \dots & k_{n1} \\ k_{12} & -\sum k_{2j} & \dots & k_{m2} & \dots & k_{n2} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ 1 & 1 & \dots & 1 & \dots & 1 \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ k_{1n} & k_{2n} & \dots & k_{mn} & \dots & -\sum k_{nj} \end{vmatrix}} \quad (3.2)$$

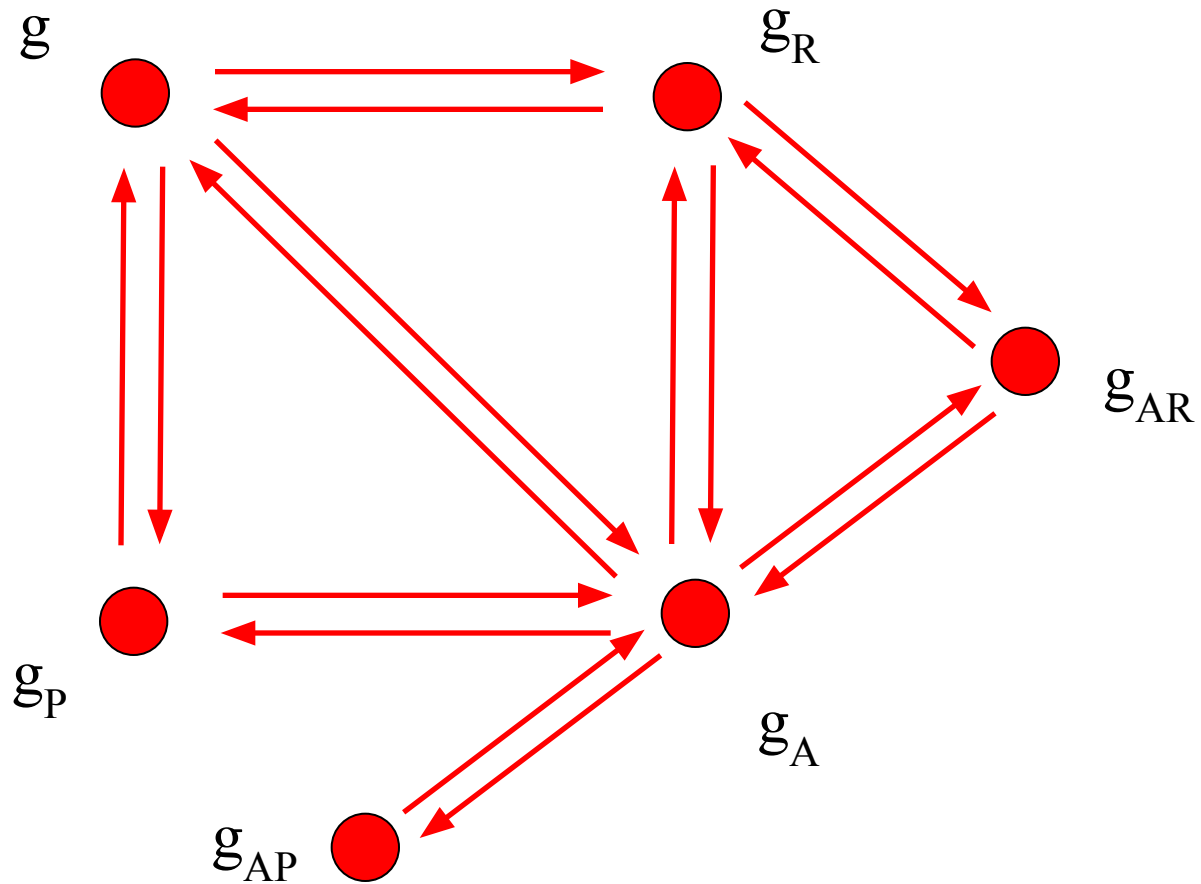
Равновесный подход к моделированию функций промоторов



-  - Сайт взаимодействия с РНК-полимеразой
-  - Сайт взаимодействия с активатором
-  - Сайт взаимодействия с репрессором



Равновесный подход к моделированию функций промоторов



Равновесный подход к моделированию функций промоторов

$$\left\{ \begin{aligned} \frac{dg}{dt} &= (k_{-,R}g_R - k_{+,R}Rg) + (k_{-,A}g_A - k_{+,A}Ag) + (k_{-,P}g_P - k_{+,P}Pg) \\ \frac{dg_R}{dt} &= -(k_{-,R}g_R - k_{+,R}Rg) + (k_{-,A}g_{AR} - k_{+,A}Ag_R) \\ \frac{dg_A}{dt} &= -(k_{-,A}g_A - k_{+,A}Ag) + (k_{-,R}g_{AR} - k_{+,R}Rg_A) + (k_{-,AP}g_{AP} - k_{+,AP}Pg_A) \\ \frac{dg_P}{dt} &= -(k_{-,P}g_P - k_{+,P}Pg) \\ \frac{dg_{AR}}{dt} &= -(k_{-,R}g_{AR} - k_{+,R}Rg_A) - (k_{-,A}g_{AR} - k_{+,A}Ag_R) \\ \frac{dg_{AP}}{dt} &= -(k_{-,AP}g_{AP} - k_{+,AP}Pg_A) \end{aligned} \right.$$

Приравняем к нулю правую часть системы и учтем уравнение баланса

$$g + g_A + g_R + g_P + g_{AR} + g_{AP} = g_0$$

Итоговые формулы элементарной модели функционирования гипотетического промотора

$$g_R = \frac{k_{+,R}}{k_{-,R}} Rg, g_A = \frac{k_{+,A}}{k_{-,A}} Ag, g_{AR} = \frac{k_{+,R}}{k_{-,R}} \frac{k_{+,A}}{k_{-,A}} RA g, g_P = \frac{k_{+,P}}{k_{-,P}} Pg, g_{AP} = \frac{k_{+,AP}}{k_{-,AP}} \frac{k_{+,A}}{k_{-,A}} APg$$

$$g = \frac{g_0}{1 + \frac{k_{+,R}}{k_{-,R}} R + \frac{k_{+,A}}{k_{-,A}} A + \frac{k_{+,P}}{k_{-,P}} P + \frac{k_{+,R}}{k_{-,R}} \frac{k_{+,A}}{k_{-,A}} RA + \frac{k_{+,AP}}{k_{-,AP}} \frac{k_{+,A}}{k_{-,A}} AP}$$



Аппроксимация механизмов в терминах обобщенных функций Хилла



Теперь перейдем к рассмотрению подхода, который используется для моделирования динамики функционирования элементарных подсистем в условиях недостатка данных об их структуре. Недостаток структурных данных приводит к тому, что построение модели не может вестись от биохимической схемы. Тогда конструктивным инструментом моделирования является приближение экспериментальных данных аппроксимирующими функциями. В математике наиболее разработан аппарат аппроксимации сплайн функциями. Но сплайн функции обладают серьезным недостатком: они не принимают в расчет биохимическую природу процессов, протекающих в ГС и, следовательно, никак не отражают ее в своих формулах и значениях параметров. Тем не менее, на промежуточных этапах построения моделей элементарных процессов, например, на стадии предобработки кинетических данных, использование сплайн функций может быть весьма полезным и эффективным. Для построения же аппроксимирующих моделей элементарных подсистем ГС целесообразнее использовать обобщенные функции Хилла (ОФХ).



Аппроксимация механизмов в терминах обобщенных функций Хилла



ОФХ являются естественным расширением класса рациональных полиномов, которые возникают по методу Кинга-Альтмана. Они не только воспроизводят динамические свойства системы, но также несут в структуре своих формул информацию о механизмах протекания биохимических процессов. Следует отметить, что в обобщенных функциях Хилла информация о биохимических механизмах содержится в скрытой, сжатой форме, которую не легко расшифровать. Проблему восстановления биохимической модели по обобщенной функции Хилла еще предстоит решить. Применение аппроксимирующего подхода, вообще говоря, не требует привлечения знаний о механизмах рассматриваемого процесса. Реконструкция управляющей функции V может вестись непосредственно по кинетическим кривым, а переменные и параметры управляющей функции V соответствуют только тем сущностям, которые измеряются в эксперименте. В результате разрабатываемая модель оперирует минимальным количеством переменных. Прежде чем дать определение обобщенных функций Хилла, рассмотрим примеры.



Аппроксимация механизмов в терминах обобщенных функций Хилла



Регуляция экспрессии гена *purC* в клетке *E.coli*

Экспрессия гена, кодирующего фермент биосинтеза пиримидинов – дигидрооротазу, негативно регулируется пиримидинами (Wilson *et al.*, 1992). Мутационный анализ регуляторного района гена *purC* показал, что для ингибирования экспрессии этого гена необходимо формирование шпильки на 5'-конце транскрипта *purC*, перекрывающейся с сайтом связывания рибосомы (Wilson *et al.*, 1992). Образование шпильки контролируется нуклеотидчувствительным выбором сайта инициации транскрипции гена *purC*. При высокой концентрации СТР основная часть транскриптов *purC* иницируется с нуклеотида, позиция которого на 17 пар оснований ниже старта трансляции гена *purC*. Такие транскрипты способны формировать стабильную шпильку на 5'-конце мРНК в районе сайта связывания рибосомы, препятствуя нормальной инициации трансляции. При низкой концентрации СТР и высоком уровне GTP основная часть транскриптов *purC* иницируется с нуклеотида, позиция которого ниже на 2 пары оснований. В этой ситуации, укороченный транскрипт не способен образовывать стабильную шпильку на 5'-конце мРНК, что обеспечивает нормальную инициацию трансляции.



Аппроксимация механизмов в терминах обобщенных функций Хилла



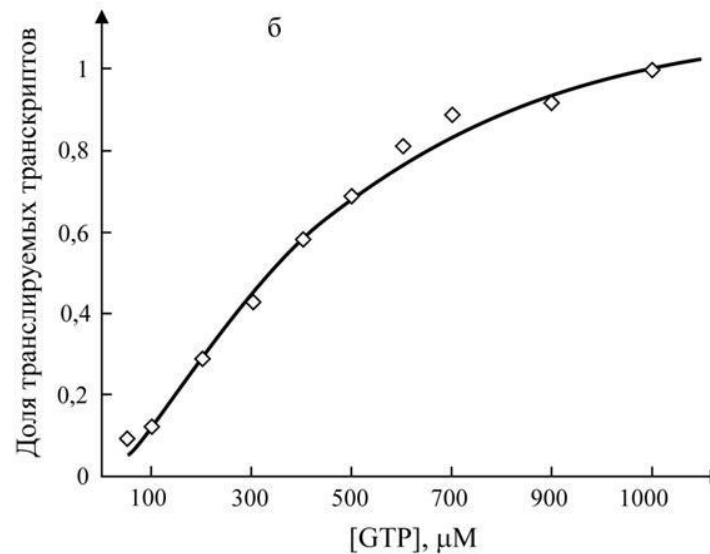
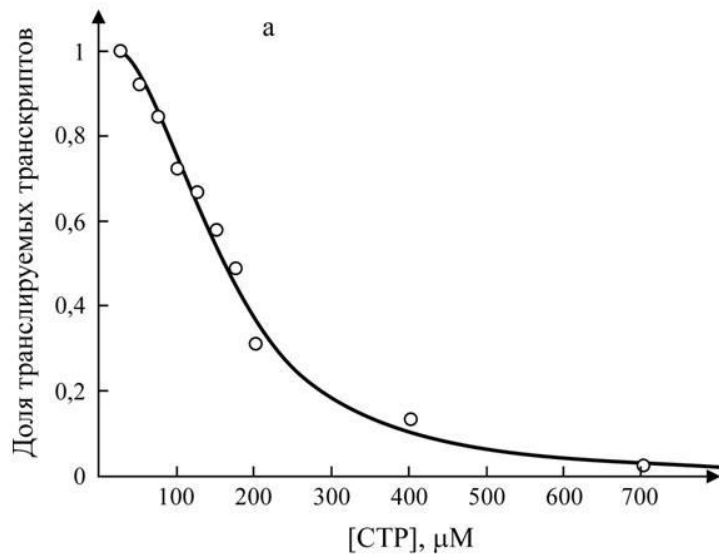
Регуляция экспрессии гена *purC* в клетке *E.coli*

Таким образом, уникальность рассматриваемой ситуации состоит в том, что регуляция синтеза белка *PurC* осуществляется по механизму отрицательной обратной связи путем выбора старта транскрипции, влияющего на возможность инициации трансляции транскрибируемой мРНК. В условиях избытка пиримидинов преимущественно синтезируются транскрипты, образующие шпильку в сайте посадки связывания рибосомы, и происходит блокирование инициации трансляции. В условиях недостатка пиримидинов в клетке синтезируются укороченные транскрипты, которые могут транслироваться, тем самым, повышая концентрацию дигидрооротазы в клетке (Wilson *et al.*, 1992).

Не вдаваясь в детальное описание механизма генетической регуляции экспрессии гена *purC*, который, как можно видеть из описанного выше, характеризуется исключительно высокой сложностью, обобщенная функция Хилла, которая дает очень компактное описание всей совокупности рассматриваемых процессов, выглядит следующим образом:



Аппроксимация механизмов в терминах обобщенных функций Хилла



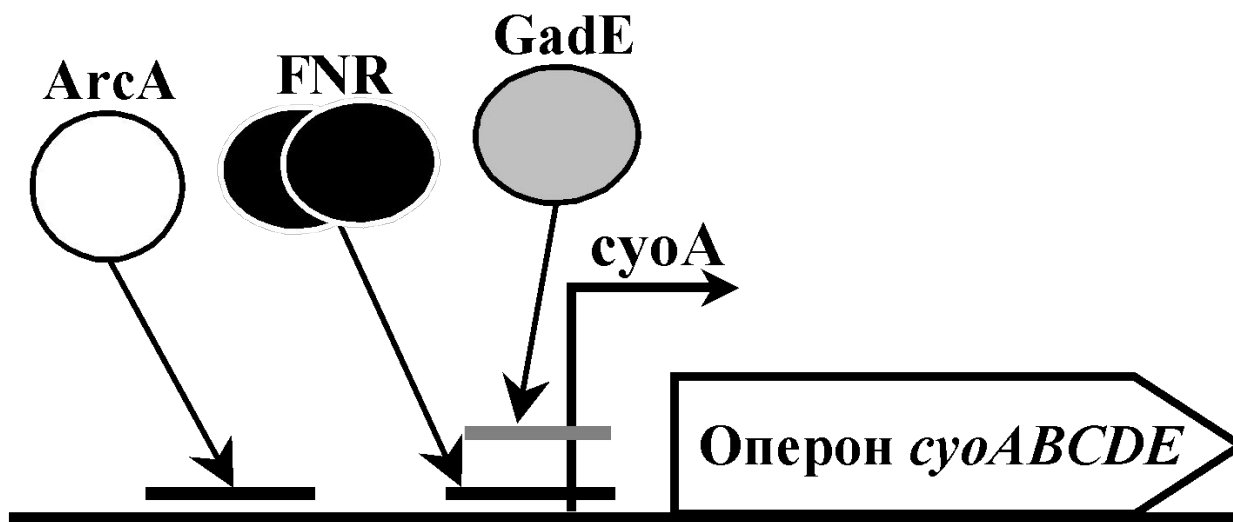
$$V = V_{max} \frac{\left(\frac{GTP}{k1_{GTP}}\right)^{n_g}}{1 + \left(\frac{GTP}{k2_{GTP}}\right)^{n_g} + \left(\frac{CTP}{k_{CTP}}\right)^{n_c} + \frac{GTP^{n_g} \cdot CTP^{n_c}}{k_{GTP,CTP}^{n_g+n_c}}}$$



Аппроксимация механизмов в терминах обобщенных функций Хилла



Модель регуляции экспрессии оперона *суoABCDE*



Структура промотора оперона *суoABCDE*, кодирующего цитохром во-оксидазу, из БД Ecocyc (<http://ecocyc.org>). Факторы ArcA и FNR являются ингибиторами транскрипции, фактор GadE – активатор транскрипции.



Аппроксимация механизмов в терминах обобщенных функций Хилла



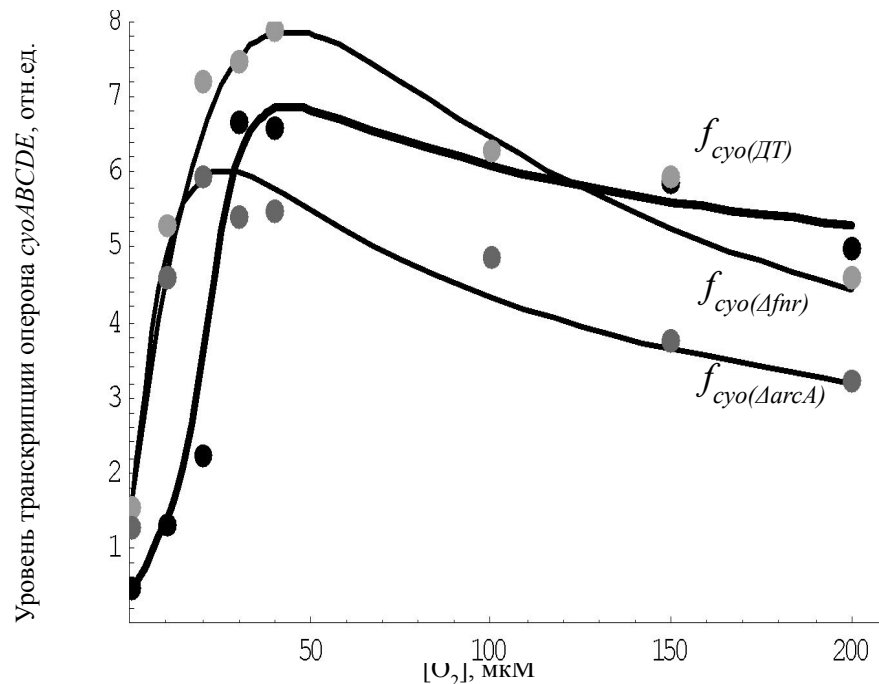
Регуляция экспрессии гена $purC$ в клетке $E.coli$

$$V_{суoABCDE} = \frac{k_0 + k_{Z_{GadE}} \cdot Z_{GadE}}{1 + Z_{ArcA} + Z_{GadE} + Z_{Fnr} + Z_{ArcA,Fnr} + Z_{ArcA,GadE}}$$



Аппроксимация механизмов в терминах обобщенных функций Хилла: Модель регуляции экспрессии оперона *cyoABCDE*

Зависимость уровня транскрипции оперона *cyoABCDE* от концентрации кислорода в среде для клеток *E.coli* дикого типа ($f_{cyo(DT)}$) и для мутантных штаммов ($f_{cyo(\Delta arcA)}$ и $f_{cyo(\Delta fnr)}$). Точки на графике – экспериментальные данные (Tseng et al., 1996). Сплошные линии – расчеты по моделям (16, 17, 18). Значения параметров, при которых получены расчетные кривые: $k_0=0.4$; $ka_0=1.4$; $ka_1=2.8$ мкМ; $ha_1=1$; $ka_2=40$ мкМ; $ha_2=1.7$; $kf_0=1.2$; $kf_1=1.8$ мкМ; $hf_1=1$; $kf_2=19$ мкМ; $hf_2=1.5$; $kaf_1=10$ мкМ; $haf_1=5.5$; $kaf_2=20$ мкМ; $haf_2=5.7$; $kfa_1=0.54$ мкМ; $kfa_2=0.1$ мкМ.



$$f_{cyo}([O_2]) = k_0 \frac{1 + \left(\frac{O_2}{ka_1}\right)^{ha_1} + \left(\frac{O_2}{kf_1}\right)^{hf_1} + \left(\frac{O_2}{kaf_1}\right)^{haf_1}}{1 + \left(\frac{O_2}{ka_2}\right)^{ha_2} + \left(\frac{O_2}{kf_2}\right)^{hf_2} + \left(\frac{O_2}{kaf_2}\right)^{haf_2}} \cdot \frac{1 + \frac{O_2}{kfa_1}}{1 + \delta_{fa} + \frac{O_2}{kfa_2}}$$



Аппроксимация механизмов в терминах обобщенных функций Хилла



Сделаем несколько комментариев к рассмотренной модели экспрессии оперона *cydAB*. По совокупности имеющихся сведений о структурно-функциональной организации промоторной области и механизмах действия можно построить структурный каркас модели, однако дальше продвинуться не представляется возможным в виду отсутствия данных о механизмах функционирования транскрипционных факторов. Мы попытались использовать данные по влиянию кислорода на активность транскрипционных факторов Fnr и ArcA для вывода модели на основе кинетических данных, так как двигаться от структурного каркаса модели не представляется возможным. Интерпретация полученной модели в терминах биохимических механизмов оказывается затрудненной. Однако, это вынужденная жертва, которая вызвана незнанием важных деталей механизма функционирования данных промоторов. Мы видим, что в случаях, когда структурных данных явно недостаточно для генерации биохимических схем, или структурные данные не обеспечиваются необходимым пакетом кинетических данных, при выводе моделей целесообразно непосредственно исходить из имеющихся кинетических данных. При этом моделирование ведется в терминах тех сущностей, от изменения которых изучена активность экспрессии промотора. Т.е. применение обобщенных функций Хилла позволяет избежать излишней детализации в описании механизмов, оставив задачу построения детальной модели на будущее. В то же время применение обобщенных функций Хилла не является просто аппроксимацией, не имеющей биологического смысла. Данный класс функций является наиболее подходящим для построения аппроксимаций кинетических кривых биологического происхождения, так как в их структуре в неявном виде отражены биологические механизмы. Например, мы знаем, какие члены в моделях (16, 19), ответственны за действие кислорода через транскрипционный фактор ArcA, а какие – через Fnr.

Итерационное определение обобщенных функций Хилла.

1. Рациональный полином

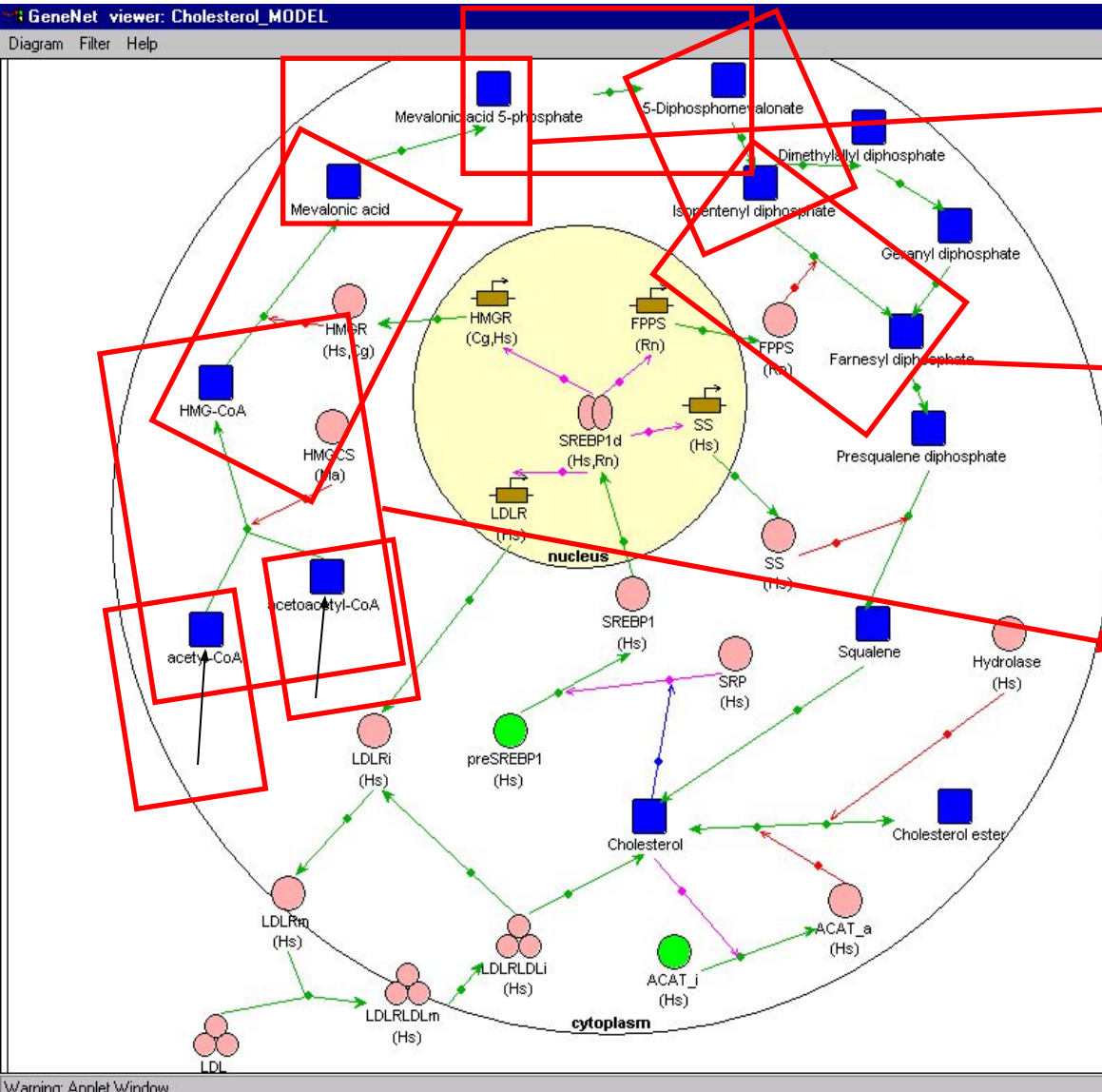
$$h(X) = \sum_{\alpha \in A} \delta_{\alpha} \prod_{x \in X_{\alpha}} (x/k_{\alpha})^{h_{\alpha,x}} / \sum_{\alpha \in A} \prod_{x \in X_{\alpha}} (x/k_{\alpha})^{h_{\alpha,x}}$$

Является обобщенной функцией Хилла (ОФХ)

2. Сумма произведение ОФХ является ОФХ

3. Функция, которая может быть получена из ОФХ, путем подстановки в нее вместо переменных и параметров обобщенных функций Хилла, является обобщенной функцией Хилла.

Закон суммирования локальных скоростей элементарных подсистем



$$A \xrightarrow{k} B, \quad a = [A], \quad b = [B]$$

$$-da/db = db/dt = ka$$

$$E + S \xrightleftharpoons[k_{-1}]{k_1} ES \xrightarrow{k_2} E + P,$$

$$e = [E], \quad s = [S], \quad e_a = [ES], \quad p = [P]$$

$$ds/dt = -dp/dt = -k_3 e_o s / (K_m + s),$$

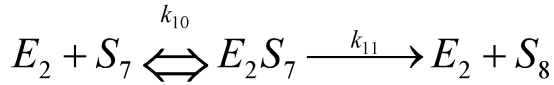
$$K_m = (k_{-1} + k_2) / k_1$$

$$e_o = e + e_a,$$

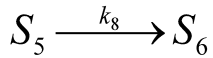
$$E + S_1 + S_2 \xrightleftharpoons[k_{d2}]{k_{d1}} ES_1S_2 \xrightarrow{k_3} E + P$$

$$\frac{dP}{dt} = -\frac{dS_1}{dt} = -\frac{dS_2}{dt} = V, \quad V = k_3 \frac{ES_1S_2}{(k_{d1} + S_1)(k_{d2} + S_2)}$$

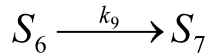
Закон суммирования локальных скоростей элементарных подсистем



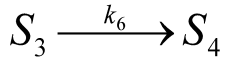
$$\frac{dS_8}{dt} = -\frac{dS_7}{dt} = k_5 \frac{E_2 S_7}{(k_{10} + S_7)}$$



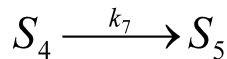
$$\frac{dS_6}{dt} = -\frac{dS_5}{dt} = k_8 S_5$$



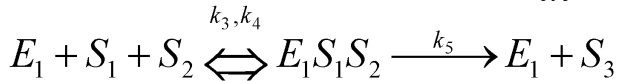
$$\frac{dS_7}{dt} = -\frac{dS_6}{dt} = k_9 S_6$$



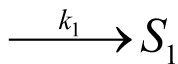
$$\frac{dS_4}{dt} = -\frac{dS_3}{dt} = k_6 S_3$$



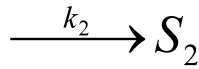
$$\frac{dS_5}{dt} = -\frac{dS_4}{dt} = k_7 S_4$$



$$\frac{dS_3}{dt} = -\frac{dS_1}{dt} = -\frac{dS_2}{dt} = V, \quad V = k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)}$$



$$\frac{dS_1}{dt} = k_1$$



$$\frac{dS_2}{dt} = k_2$$

$$\frac{dS_1}{dt} = k_1 - k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)}$$

$$\frac{dS_2}{dt} = k_2 - k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)}$$

$$\frac{dS_3}{dt} = k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)} - k_6 S_3$$

$$\frac{dS_4}{dt} = k_6 S_3 - k_7 S_4$$

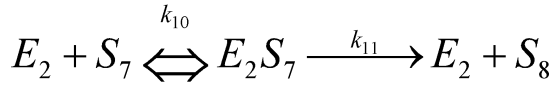
$$\frac{dS_5}{dt} = k_7 S_4 - k_8 S_5$$

$$\frac{dS_6}{dt} = k_8 S_5 - k_9 S_6$$

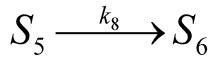
$$\frac{dS_7}{dt} = k_9 S_6 - k_{11} \frac{E_2 S_7}{(k_{10} + S_7)}$$

$$\frac{dS_8}{dt} = k_{11} \frac{E_2 S_7}{(k_{10} + S_7)}$$

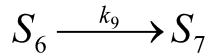
Закон суммирования локальных скоростей элементарных подсистем



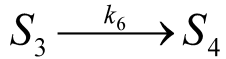
$$\frac{dS_8}{dt} = -\frac{dS_7}{dt} = k_5 \frac{E_2 S_7}{(k_{10} + S_7)}$$



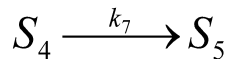
$$\frac{dS_6}{dt} = -\frac{dS_5}{dt} = k_8 S_5$$



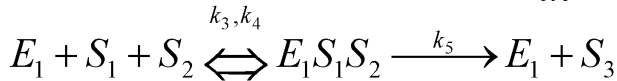
$$\frac{dS_7}{dt} = -\frac{dS_6}{dt} = k_9 S_6$$



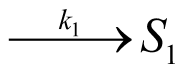
$$\frac{dS_4}{dt} = -\frac{dS_3}{dt} = k_6 S_3$$



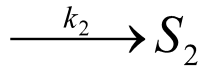
$$\frac{dS_5}{dt} = -\frac{dS_4}{dt} = k_7 S_4$$



$$\frac{dS_3}{dt} = -\frac{dS_1}{dt} = -\frac{dS_2}{dt} = V, \quad V = k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)}$$



$$\frac{dS_1}{dt} = k_1$$



$$\frac{dS_2}{dt} = k_2$$

$$\frac{dS_1}{dt} = k_1 - k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)}$$

$$\frac{dS_2}{dt} = k_2 - k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)}$$

$$\frac{dS_3}{dt} = k_5 \frac{E_1 S_1 S_2}{(k_3 + S_1)(k_4 + S_2)} - k_6 S_3$$

$$\frac{dS_4}{dt} = k_6 S_3 - k_7 S_4$$

$$\frac{dS_5}{dt} = k_7 S_4 - k_8 S_5$$

$$\frac{dS_6}{dt} = k_8 S_5 - k_9 S_6$$

$$\frac{dS_7}{dt} = k_9 S_6 - k_{11} \frac{E_2 S_7}{(k_{10} + S_7)}$$

$$\frac{dS_8}{dt} = k_{11} \frac{E_2 S_7}{(k_{10} + S_7)}$$



Конец лекции