



Математическое моделирование молекулярно-генетических систем

А.В. Ратушный

(лекция 3)

Кафедра информационной биологии ФЕНГУ



Цели лекции



- Обобщенный химико-кинетический метод моделирования (формальное представление моделей)
- Спецификации языка моделирования
- Процесс сборки математических моделей
- Моделирование многокомпарментных структур
- Моделирование ДНК- и РНК-специфичных процессов
- Моделирование полиплоидных генных сетей
- Развитые средства в рамках ОХКММ (заключение)



Обобщенный химико-кинетический метод моделирования формальное представление моделей

Динамика молекулярно-генетических систем в ОХКММ формально описывается моделями вида:

$$\frac{dx_i}{dt} = F_i(X, K) - x_i G_i(X, K)$$

где n – количество динамических переменных модели,

$X = (x_1, \dots, x_n)$ – вектор динамических переменных, как правило имеющих смысл концентраций,

$K = (k_1, \dots, k_m)$ – вектор параметров,

$F_i(X, K)$ и $G_i(X, K)$ – рациональные функции, принимающие неотрицательные значения при всех неотрицательных значениях переменных и параметров.

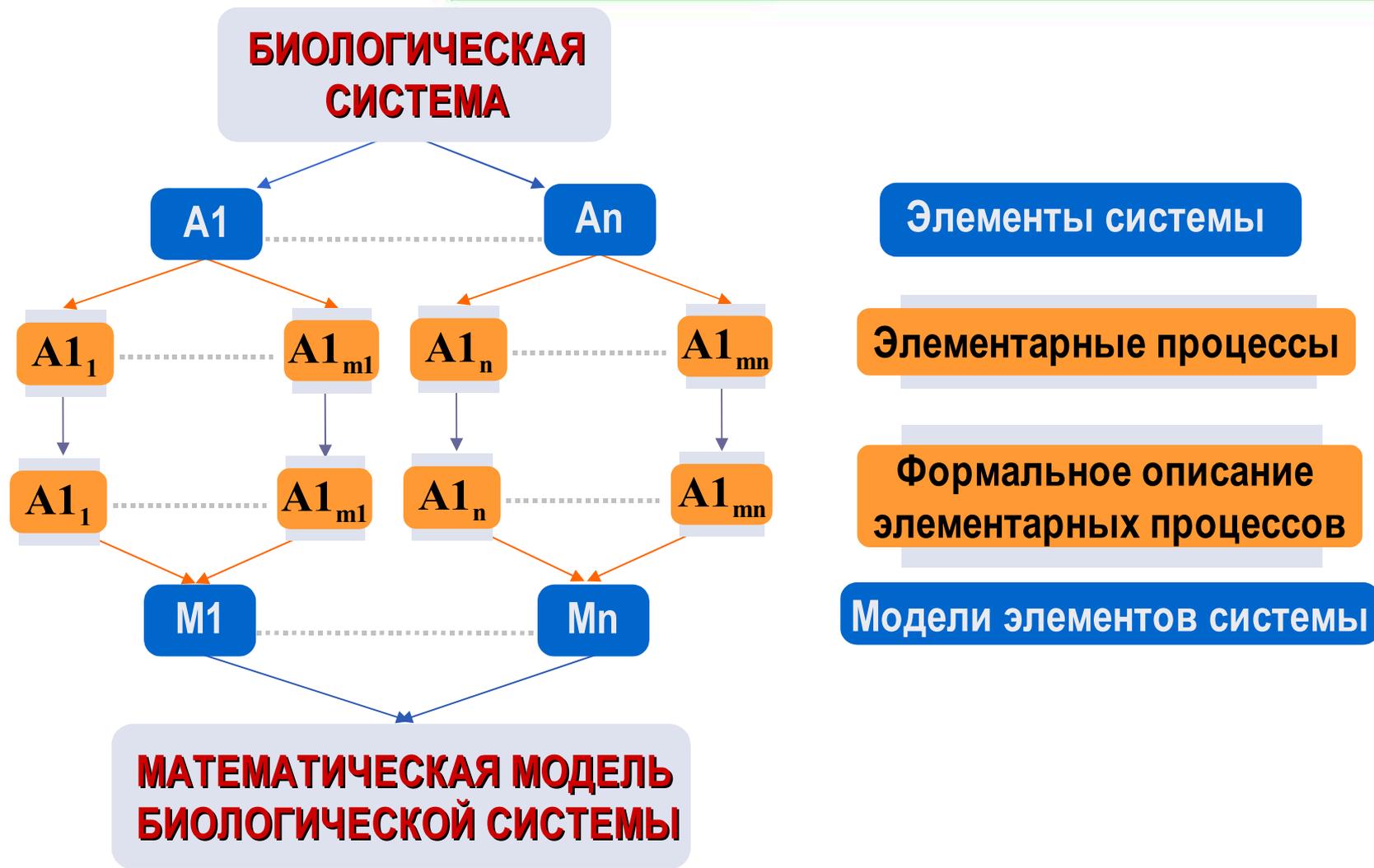


Обобщенный химико-кинетический метод моделирования формализация моделей

Формализация в рамках ОХКММ осуществляется на основе блочного подхода, который подразумевает проведение иерархической декомпозиции исходной системы до элементарных структур и процессов, описание элементарных процессов в терминах элементарных блоков с последующим конструированием из элементарных процессов моделей элементарных структур, из которых в дальнейшем проводится сборка моделей генетической системы и ее вариантов.



Обобщенный химико-кинетический метод моделирования формализация моделей





Элементарный процесс как единица моделирования



Спецификация структуры элементарного блока:

[*/]

[/*]

{bb} <имя блока>.

□

{pp} <имя продукта 1> □ [<имя продукта 2>] □ ... [<имя продукта n>].

□

{ee} <имя расширения 1> □ [<имя расширения 2>] □ ... [<имя расширения k>].

□

{cc} <имя параметра 1> □ [<имя параметра 2>] □ ... [<имя параметра m>].

Z1 Z2...Zm

□

{ee} <имя расширения 1> □ [<имя расширения 2>] □ ... [<имя расширения k>].

{cc} <имя параметра 1> □ [<имя параметра 2>] □ ... [<имя параметра m>].

Z1 Z2...Zm

{zz} <имя окончания записи блока>.



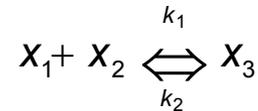
Элементарный процесс как единица моделирования



Пример структуры элементарного блока, описывающего бимолекулярную реакцию:

```
/*      k1
/* a + b <====> c
/*      k2

{bb} <B(<=>)J(4)>.
{pp} <l(a)> + <l(b)>;
{pp}           = <l(c)>.
{ee} <*type#(bimolecular_reaction)>.
{cc} <K(1)>;
      1.0e-05
{cc} <K(2)>.
      1.0e-00
{zz} <E(++++)>.
```



$$\bar{X} = (x_1, x_2, x_3), \quad \bar{P} = (k_1, k_2),$$

$$F: \frac{dx_1}{dt} = \frac{dx_2}{dt} = -\frac{dx_3}{dt} = k_2 \cdot x_3 - k_1 \cdot x_1 \cdot x_2$$



Элементарный процесс как единица моделирования



В общем случае формальные блоки могут реализовывать любой закон переработки входящей информации и банк блоков может пополняться по мере возникновения необходимости.

В настоящее время банк содержит 51 блок, которые в совокупности позволяют проводить очень широкий спектр действий. Например, имеется блок формирования точек запоминания динамических характеристик системы, имеется блок запуска новой задачи Коши, имеются блоки арифметических действий и т.д.



Элементарный процесс как единица моделирования

Процедура построения математической модели основана на локальной независимости элементарных процессов. Это, с одной стороны позволяет описывать элементарные процессы независимо друг от друга и создавать их постоянно пополняемый банк, а с другой - объединение элементарных процессов в единую модель выражается в суммировании скоростей изменения концентраций по каждому веществу по всем элементарным процессам, в которых участвует данный компонент модели.



Имена динамических переменных и параметров математических моделей



Имена динамических переменных и параметров математических моделей состоят из списка атрибутов и их значений:

`<Attribute_Name_1(Attribute_Value_1)...Attribute_Name_N(Attribute_Value_N)>`



Служебные атрибуты имен динамических переменных и констант математических моделей



Существуют служебные зарезервированные атрибуты, например:

<**B**(...)J(...)> – идентификатор блока и режим использования блока;

<...**Z**(...)> – расширение;

<...**O**(...)> – ориентация блока или гена, обрабатывается специальной программой;

<...**A**(...)> – аллельность, обрабатывается специальной программой;

<...**G**(...)> – принадлежность к определенному участку генетической карты, обрабатывается специальной программой;

<...**C**(...)> – принадлежность к определенному компартменту;

<...**S**(...)> – принадлежность к определенному сценарию;



Унификация имен динамических переменных и констант математических моделей



Для унификации имен считается, что:

ss(...) – атрибут сущности

Примеры:

ss(protein) – белок;

ss(complex) – комплекс;

ss(enzyme) – фермент;

ss(compound) – низкомолекулярное вещество;

l(...) – атрибут наименования

Пример:

<ss(enzyme)l(Citrate synthase)> – фермент цитрат синтаза;



Унификация имен динамических переменных и констант математических моделей



Для унификации имен считается, что:

Имена регуляторов транскрипции представляются следующим образом:

`<ss(complex)I(имя регулятора)>`

Пример:

`<ss(complex)I(IHF)>` - регулятор Integration Host Factor

Имена белков, синтезируемых с транскриптов представляются следующим образом:

`<ss(protein)I(имя гена с большой буквы)>`

Пример:

`<ss(protein)I(AcnB)>` – белок, синтезированный с гена `acnB`



Унификация имен динамических переменных и констант математических моделей



Для унификации имен считается, что:

Имена ферментов, синтезируемых с транскриптов и ферментативных комплексов представляются следующим образом:

<ss(enzyme)|(название фермента из исходного файла)>

Пример:

<ss(enzyme)|(Citrate synthase)> – фермент цитрат синтаза

Имена метаболитов представляются следующим образом:

<ss(compound)|(название вещества из исходного файла)>

Пример:

<ss(compound)|(PYR)> – пируват



Синонимы имен и значений атрибутов

Существует возможность задавать синонимы имен и значений атрибутов, например:

<...**A**(...)...> = <...**Allele**(...)...>;

<...**C**(...)...> = <...**Compartment**(...)...>

<...**ss**(compound)...> = <...**ss**(substrate)...>;

<...**I**(HMG-CoA reductase)...> = <...**I**(3-hydroxy-3-methylglutaryl-CoA reductase)...>



Элементарный процесс как единица моделирования

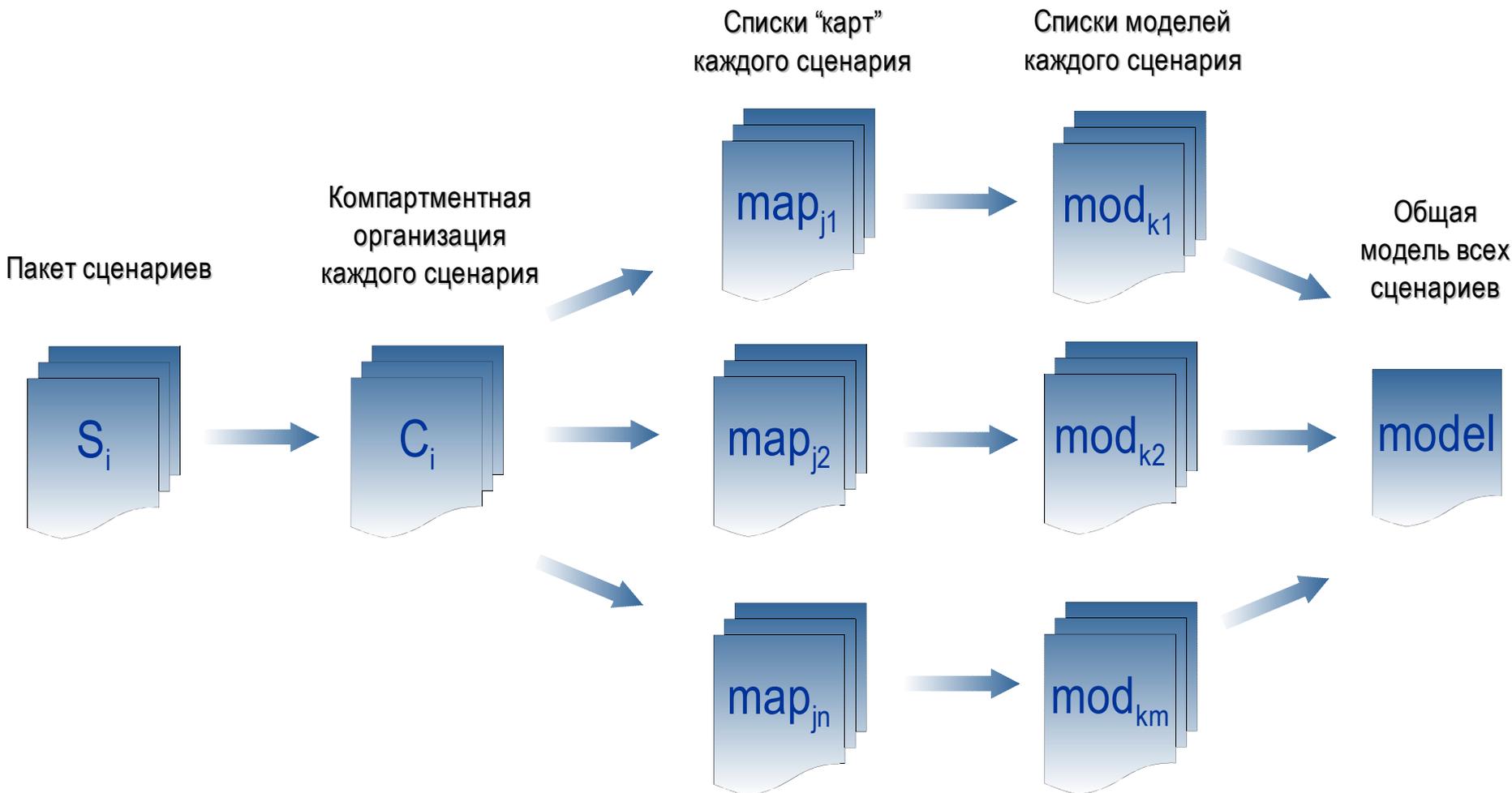
В модели продукты считаются различными, если между их именами нет детального посимвольного совпадения.

В модели имена продуктов имеют глобальный смысл.

В модели параметры также считаются различными, если между их именами нет детального посимвольного совпадения, но в отличие от продуктов, имена и значения параметров каждого блока хранятся отдельно от имен и значений параметров других блоков. Т.е. параметры имеют локальный смысл. В двух различных блоках модели параметры могут иметь одинаковые имена и различные числовые значения.



Сборка математической модели





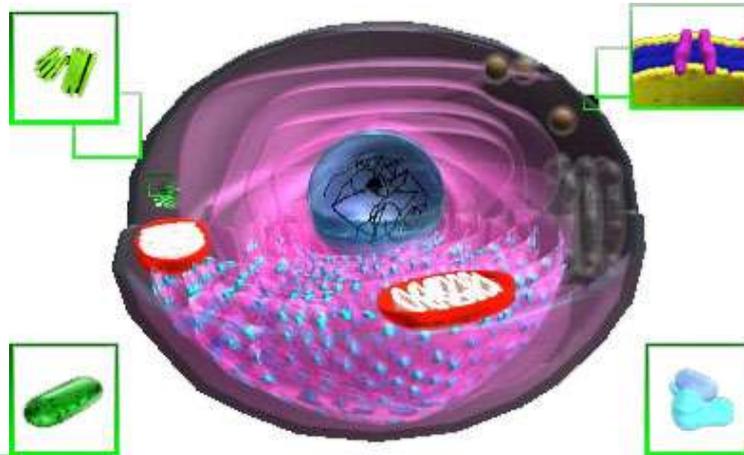
Моделирование многокомпарментных структур

Проблема 1

Многие процессы, протекающие в компартаментах являются идентичными процессами, а при рассмотрении сложных организмов, идентичными могут являться также и целые компартменты. Это приводит к проблеме описания идентичных процессов протекающих в разных компартментах.

Проблема 2

Между компартментами организма происходит обмен веществами – т.е. существует активный и пассивный транспорт. При этом, типичны ситуации транспорта одного и того же вещества из одного компартмента, сразу в несколько других, не обязательно идентичных, компартментов. Это приводит к проблеме описания процессов транспорта.



<http://www.ibiblio.org/virtualcell/tour/cell/cell.htm>



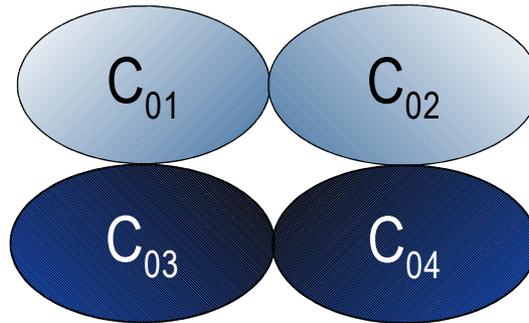
Моделирование многокомпарментных структур



Банк моделей
элементарных
процессов

- 1) $\Rightarrow [A]_{\text{comp_in}}$ 2) $\Rightarrow [B]_{\text{comp_in}}$ 3) $[A]_{\text{comp_in}} \Rightarrow$ 4) $[B]_{\text{comp_in}} \Rightarrow$
- 5) $[A]_{\text{comp_in}} + [B]_{\text{comp_in}} \Leftrightarrow [A,B]_{\text{comp_in}}$
- 6) $[A]_{\text{comp_in}} \Rightarrow [A]_{\text{comp_out}}$ 7) $[B]_{\text{comp_in}} \Rightarrow [B]_{\text{comp_out}}$

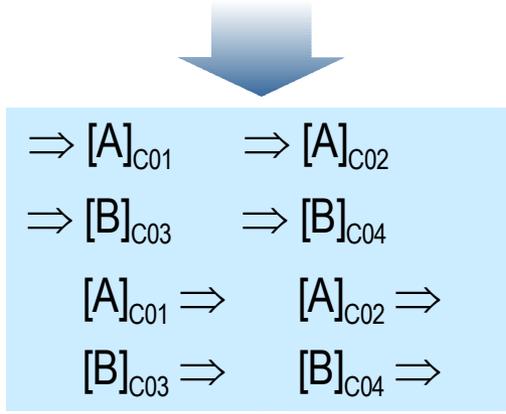
Компарментная
структура
организма



1,3) $\text{Comp_in} = \{C_{01}; C_{02}\}$

2,4) $\text{Comp_in} = \{C_{03}; C_{04}\}$

Модель
функционирования
организма





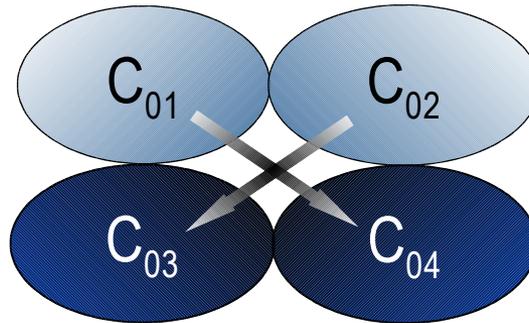
Моделирование многокомпарментных структур



Банк моделей
элементарных
процессов

- 1) $\Rightarrow [A]_{comp_in}$ 2) $\Rightarrow [B]_{comp_in}$ 3) $[A]_{comp_in} \Rightarrow$ 4) $[B]_{comp_in} \Rightarrow$
- 5) $[A]_{comp_in} + [B]_{comp_in} \Leftrightarrow [A,B]_{comp_in}$
- 6) $[A]_{comp_in} \Rightarrow [A]_{comp_out}$ 7) $[B]_{comp_in} \Rightarrow [B]_{comp_out}$

Компарментная
структура
организма



- 1,3) $Comp_in = \{C_{01}; C_{02}\}$
- 2,4) $Comp_in = \{C_{03}; C_{04}\}$
- 5) $Comp_in = \{C_{03}; C_{04}\}$
- 6) $Comp_in = \{C_{01}\}, Comp_out = \{C_{04}\}$
- 7) $Comp_in = \{C_{02}\}, Comp_out = \{C_{03}\}$



Модель
функционирования
организма

$\Rightarrow [A]_{C01}$	$\Rightarrow [A]_{C02}$	$[A]_{C03} + [B]_{C03} \Leftrightarrow [A,B]_{C03}$ $[A]_{C04} + [B]_{C04} \Leftrightarrow [A,B]_{C04}$ $[A]_{C01} \Rightarrow [A]_{C04}$ $[A]_{C02} \Rightarrow [A]_{C03}$
$\Rightarrow [B]_{C03}$	$\Rightarrow [B]_{C04}$	
$[A]_{C01} \Rightarrow$	$[A]_{C02} \Rightarrow$	
$[B]_{C03} \Rightarrow$	$[B]_{C04} \Rightarrow$	



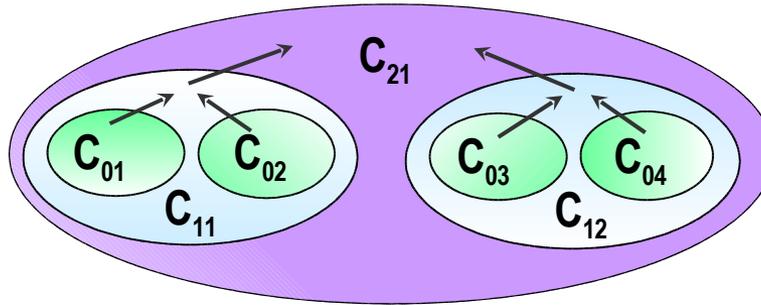
Моделирование многокомпарментных структур



Банк моделей
элементарных
процессов

- 1) $\Rightarrow [A]_{\text{comp_in}}$ 2) $\Rightarrow [B]_{\text{comp_in}}$ 3) $[A]_{\text{comp_in}} \Rightarrow$ 4) $[B]_{\text{comp_in}} \Rightarrow$
- 5) $[A]_{\text{comp_in}} + [B]_{\text{comp_in}} \Leftrightarrow [A,B]_{\text{comp_in}}$
- 6) $[A]_{\text{comp_in}} \Rightarrow [A]_{\text{comp_out}}$ 7) $[B]_{\text{comp_in}} \Rightarrow [B]_{\text{comp_out}}$

Компарментная
структура
организма



- 1) $\text{Comp_in}=\{C_{01}<C_{11}<C_{21}, C_{02}<C_{11}<C_{21}\}$
- 2) $\text{Comp_in}=\{C_{03}<C_{12}<C_{21}, C_{04}<C_{12}<C_{21}\}$
- 3,4,5) $\text{Comp_in}=\{C_{21}\}$
- 6) $\text{Comp_in}=\{C_{01}<C_{11}<C_{21}\}, \text{Comp_out}=\{C_{11}<C_{21}\}$
- 7) $\text{Comp_in}=\{C_{03}<C_{12}<C_{21}\}, \text{Comp_out}=\{C_{12}<C_{21}\}$
- 8) $\text{Comp_in}=\{C_{11}<C_{21}\}, \text{Comp_out}=\{C_{21}\}$
- 9) $\text{Comp_in}=\{C_{21}<C_{21}\}, \text{Comp_out}=\{C_{21}\}$



Модель
функционирования
организма

$$\begin{aligned} &\Rightarrow [A]_{C_{01}<C_{11}<C_{21}} && \Rightarrow [A]_{C_{02}<C_{11}<C_{21}} \\ &\Rightarrow [B]_{C_{03}<C_{12}<C_{21}} && \Rightarrow [B]_{C_{04}<C_{12}<C_{21}} \\ &[A]_{C_{21}} \Rightarrow && [B]_{C_{21}} \Rightarrow \\ &[A]_{C_{21}} + [B]_{C_{21}} \Leftrightarrow [A,B]_{C_{21}} \end{aligned}$$

$$\begin{aligned} &[A]_{C_{01}<C_{11}<C_{21}} \Rightarrow [A]_{C_{11}<C_{21}} \\ &[A]_{C_{02}<C_{11}<C_{21}} \Rightarrow [A]_{C_{11}<C_{21}} \\ &[B]_{C_{03}<C_{12}<C_{21}} \Rightarrow [B]_{C_{12}<C_{21}} \\ &[B]_{C_{04}<C_{12}<C_{21}} \Rightarrow [B]_{C_{12}<C_{21}} \\ &[A]_{C_{11}<C_{21}} \Rightarrow [A]_{C_{21}} \\ &[B]_{C_{12}<C_{21}} \Rightarrow [B]_{C_{21}} \end{aligned}$$



Моделирование многокомпарментных структур



1. Все рассмотренные модели построены на основе единого банка элементарных процессов, которые изначально не привязаны ни к какому конкретному компартменту.
3. Для построения конкретной модели необходимо задать компартментную структуру и указать принадлежность процессов и веществ к компартментам.
5. Построение модели состоит в многократном экстрагировании элементарных процессов из банка с учетом информации из п.2.
7. Привязка каждого вещества к компартменту состоит во включении в его имя наименования компартмента. В результате каждое имя конкретного вещества становится составным и однозначно характеризует его в контексте конкретной модели.



Моделирование ДНК- и РНК-специфичных процессов



Основная проблема:

Описание процессов, протекающих на отдельных участках матрицы.

Для ДНК-специфичных процессов:

- описание процессов связывания веществ с выделенными участками ДНК;
- описание ориентированных матричных процессов (репликация, транскрипция);
- описание взаимодействия матричных процессов с различными состояниями участков ДНК;
- описание взаимодействия ориентированных матричных процессов между собой;
- описание матричных процессов с учетом замкнутости/линейности геномов;
- описание рекомбинации, кроссинговера, делеций и вставок.

Для РНК-специфичных процессов:

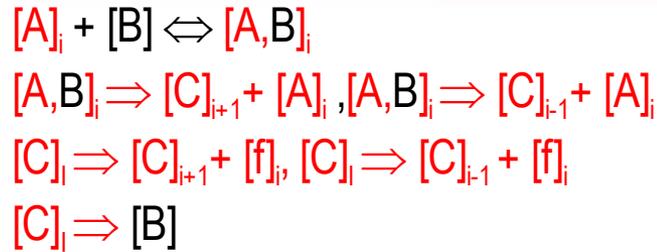
- описание трансляции как матричного процесса;
- описание сплайсинга.



Моделирование ДНК- и РНК-специфичных процессов

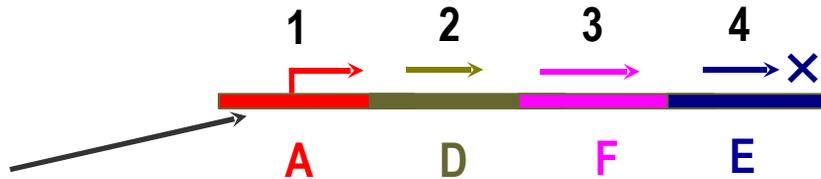


Банк моделей
ДНК-специфичных
процессов

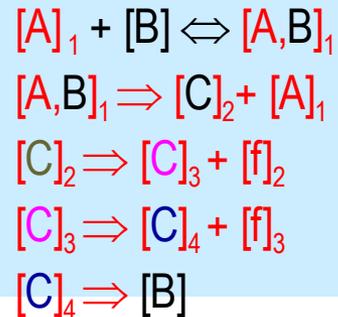


Геном

Генетический
элемент



Модель ДНК-
специфичных
процессов

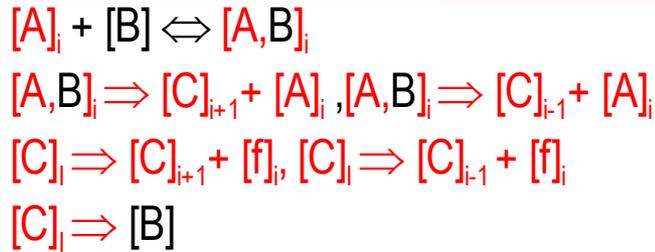




Моделирование ДНК- и РНК-специфичных процессов

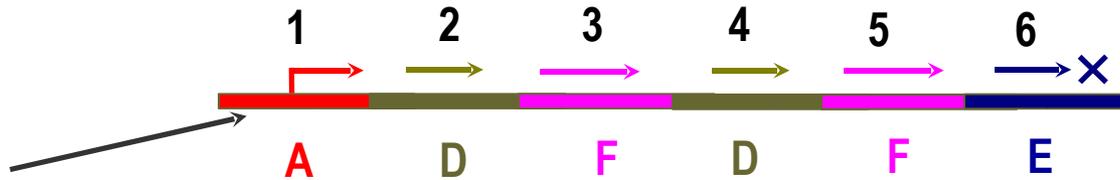


Банк моделей
ДНК-специфичных
процессов

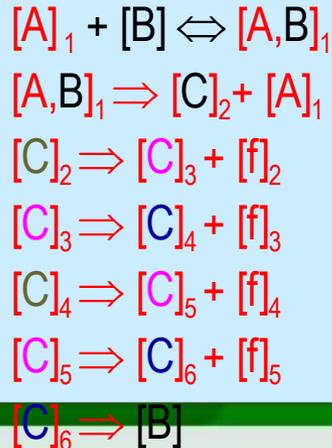


Геном

Генетический
элемент



Модель ДНК-
специфичных
процессов

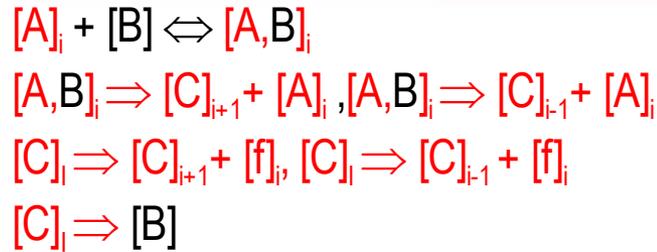




Моделирование ДНК- и РНК-специфичных процессов

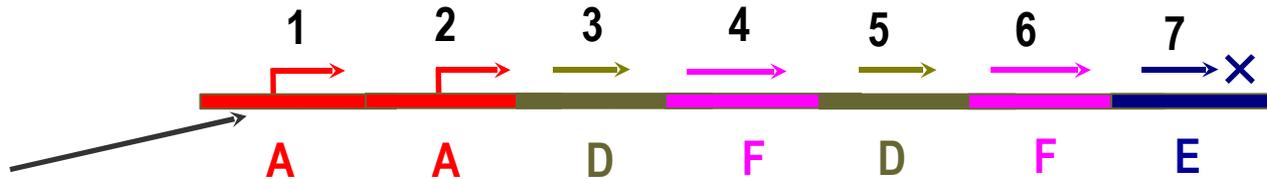


Банк моделей
ДНК-специфичных
процессов

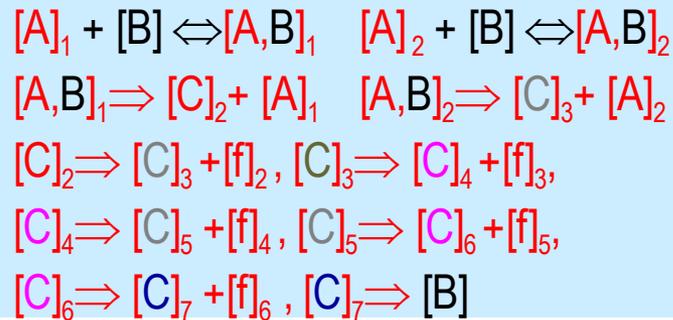


Геном

Генетический
элемент



Модель ДНК-
специфичных
процессов

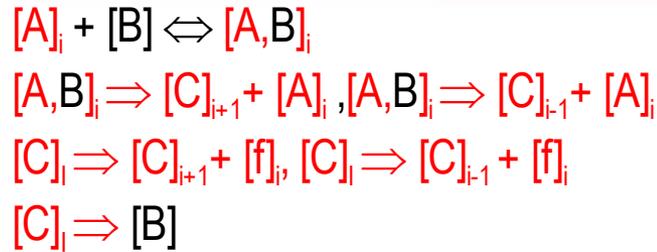




Моделирование ДНК- и РНК-специфичных процессов

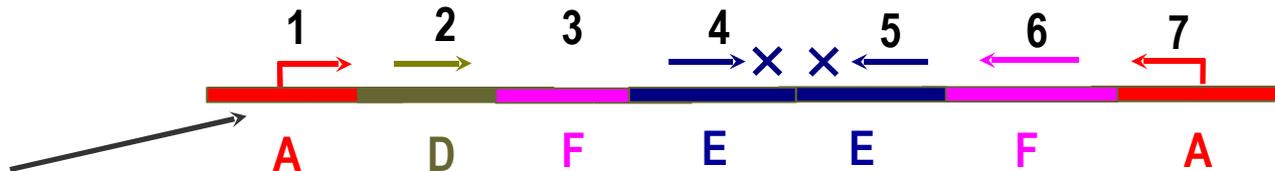


Банк моделей
ДНК-специфичных
процессов

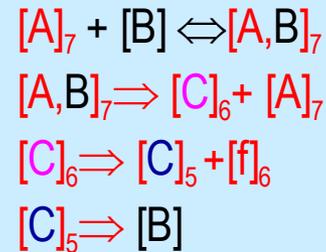
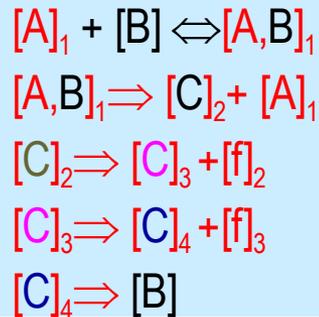


Геном

Генетический
элемент



Модель ДНК-
специфичных
процессов





Моделирование ДНК- и РНК-специфичных процессов



1. Все рассмотренные модели генетических процессов построены на основе единого банка элементарных матричных процессов.
3. Для построения конкретной модели генетической регуляции и матричного синтеза необходимо задать генетическую структуру (генетическую карту) и указать принадлежность процессов и веществ к генам.
5. Построение модели состоит в многократном экстрагировании процессов из банка с учетом информации из п.2.
7. Привязка процессов к генам заключается в добавлении к наименованию продукта идентификатора местоположения гена в составе генома. В результате имя каждого продукта становится составным и однозначно характеризует его в контексте конкретной модели.



Моделирование полиплоидных генных сетей



Основная проблема

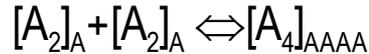
Описание реакций и процессов, которые осуществляются комплексами, состоящими из белков, кодируемых разными полиаллельными вариантами одного и того же гена, порождает проблему комбинаторного перебора всех полиаллельных вариантов веществ. Комбинаторных вариантов тем больше, чем больше глубина реакций.



Моделирование полиплоидных генных сетей



Банк моделей формирования комплексов



Моноаллельный локус: $A = \{\emptyset\}$



Модель формирования комплекса для гаплоидной системы

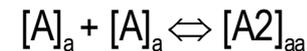
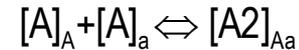


2 реакции

Диаллельный локус: $A = \{A, a\}$



Модель формирования комплекса для диплоидной системы



9 реакций

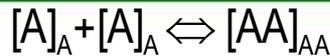




Моделирование полиплоидных генных сетей



Банк моделей формирования комплексов



Моноаллельный локус: $A = \{\emptyset\}$



Модель формирования комплекса для гаплоидной системы

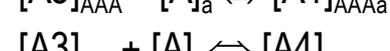
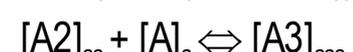
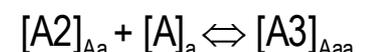
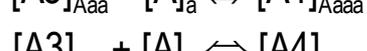
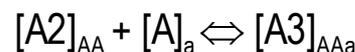
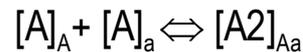


3 реакции

Диаллельный локус: $A = \{A, a\}$



Модель формирования комплекса для диплоидной системы



17 реакций



Моделирование полиплоидных генных сетей



1. Все рассмотренные особенности строения живых систем модели построены на основе единого банка элементарных процессов, в которых полиплоидность в явном виде не описывается.
3. Для построения модели необходимо задать какие белки, кодируются полиплоидными вариантами одного и того же гена.
5. Построение модели состоит в экстрагировании элементарных процессов из банка и комбинаторном построении полиплоидной модели на основе информации из п.2.
7. Полиплоидный состав вещества фиксируется путем добавления к его наименованию идентификаторов аллелей. В результате каждое имя, характеризующее вещество, становится составным и однозначно характеризует его в контексте конкретной полиплоидной модели.



Заключение

Рассмотренные примеры показывают, что специфика строения биологических систем действительно порождает ряд проблем, которые накладывают определенные требования к технологии моделирования компартментных структур организма, генетических процессов и полиаллельности. Одновременно примеры подсказывают и направление развития адекватных средств моделирования. Во-первых, во всех рассмотренных примерах модели строились методом блочного конструирования имен веществ. Иными словами, чтобы зафиксировать местоположение вещества в конкретном компартменте, на конкретном участке ДНК, или, его аллельный состав к собственному имени вещества добавляются соответствующие идентификаторы. Немаловажным при этом оказывается возможность разделить процедуры описания процессов и процедуры привязки этих процессов. Элементарные процессы в примерах приводились вне связи с конкретной структурой живых систем. Сама привязка процессов и веществ к конкретным структурам моделируемой системы выносилась на более поздней стадии, на которой задавалась информация о составе компартментов моделируемой системы, о составе процессов в каждом компартменте, о генетическом составе каждого компартмента, если таковой имеется, а также давалась информация об аллельном составе генов.



Заключение

Следовательно построение модели живой системы может быть организовано в виде алгоритма последовательного разворачивания элементарных процессов, в ходе выполнения которого происходит конструирование имен веществ и параметров. Все этапы конструирования модели: разбор аллельности, генетическая привязка, привязка к компартментам, могут быть выполнены последовательно друг за другом в определенном порядке. В результате выполнения такого алгоритма полученная модель автоматически учтет все перечисленные особенности строения биологической системы.





Заключение



К настоящему времени разработано программное обеспечение, позволяющее строить модели многокомпарментных биологических систем, вплоть до организменного уровня.

Развитое к настоящему времени программное обеспечение позволяет:

5. Моделировать генетические процессы с учетом взаимного расположения генов и генетических элементов относительно друг друга и их ориентации в составе генома;
7. Рассматривать линейные, кольцевые и фрагментированные геномы;
9. Строить модели с учетом полиаллельности, отдельных генов, которая может быть любого порядка от 1 и выше;
11. Имеет удобные средства привязки процессов к отдельным компартаментам и описания процессов транспортировки веществ из одного компартамента в другие;



Заключение

1. Процессы могут описываться как в терминах биохимических реакций, так и в более общей форме (например, законы взаимопревращения веществ могут описываться уравнениями практически любой природы: полиномами, рациональными функциями, формулами, получаемыми в результате упрощения сложных химических процессов, могут применяться алгебраические уравнения, булевы и пороговые функции, вероятностной и детерминированной природы). В принципе арсенал законов, которые могут быть применены для описания элементарных процессов и действий не ограничен, так как блочный принцип, положенный в основу метода моделирования, позволяет легко наращивать арсенал средств практически без каких либо ограничений.
3. Моделирование может проводиться в терминах функциональных активностей, что значительно расширяет доступный арсенал моделирования.
5. Модели могут снабжаться произвольным количеством средств динамического мониторинга любых интегральных характеристик функционирования системы, что важно при определении фенотипических и морфологических характеристик моделируемого объекта, так как они, как правило носят комплексный характер.
7. Программное обеспечение позволяет строить математические модели, воспроизводящие условия экспериментов, что важно при сопоставлении данных полученных в натуральных экспериментах с данными численных расчетов.



Заключение

1. Моделирование является распределенным. Т.е. модели процессов могут храниться в разных файлах в банке моделей элементарных процессов. Банк может постоянно пополняться и любой процесс, размещенный в банке, доступен для использования при построении моделей. Также могут накапливаться банки генетических карт и банки карт сопряжения компартментов.
3. Все модели, разрабатываемые средствами языка SiBML, являются совместимыми между собой. Поэтому процесс построения математических моделей очень сложных систем может проводиться блочно, с последующим объединением блоков в единую математическую модель.
5. Разработанный стандарт языка предусматривает наличие синонимов для наименований веществ, аллелей, генов, компартментов. Это также расширяет возможности эффективного моделирования, в том числе позволяет разрабатывать эффективные средства автоматизированного построения математических моделей на основании информации, содержащихся в базах, в которых накапливается информация о структурно-функциональной организации генетических программ, контролирующих различные аспекты функционирования организмов, баз данных, в которых накапливаются данные о генетической регуляции экспрессии генов, баз данных, в которых накапливаются данные о метаболических реакциях.