

ГУРСКИЙ
Виталий Валериевич

**МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ ГЕННЫХ СЕТЕЙ
В РАННЕМ РАЗВИТИИ ДРОЗОФИЛЫ**

Специальность 05.13.18 — математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Работа выполнена в Секторе теории твердого тела Учреждения Российской академии наук
«Физико-технический институт им. А. Ф. Иоффе РАН»

Научный руководитель:

Доктор физико-математических наук, профессор

А. М. Самсонов

Официальные оппоненты:

Доктор физико-математических наук, профессор

А. П. Киселев

Кандидат физико-математических наук

С. А. Руколайне

Ведущая организация:

Федеральное государственное образовательное
учреждение высшего профессионального
образования «Московский государственный
университет им. М. В. Ломоносова»

Защита диссертации состоится «___» _____ 2009 г. в ___ часов на заседании
диссертационного совета Д 212.229.13 Санкт-Петербургского государственного
политехнического университета по адресу: 195251 Санкт-Петербург, Политехническая
ул., д. 29, корпус ___, аудитория ___ .

Автореферат разослан «___» _____ 2009 г.

Ученый секретарь диссертационного совета Д 212.229.13

доктор технических наук, профессор

Б. С. Григорьев

Актуальность. Математическое моделирование биологических процессов переживает период бурного развития, тесно связанный с возникновением новых междисциплинарных научных областей, таких как биоинформатика и системная биология. Современное развитие математических методов и алгоритмов в биологии необходимо для интерпретации и обработки огромного массива экспериментальных данных, накопленных в этих областях. Именно математические модели и методы обработки информации могут позволить собрать эти экспериментальные данные в согласованную картину развития и функционирования организмов и их подсистем.

Для построения реалистичных математических моделей активности генов на мезоскопическом уровне описания важны исследования, проясняющие степень применимости различных приближений при формулировке уравнений модели, чтобы модель была адекватной с одной стороны и не была излишне сложной с другой. Современные модели генных сетей содержат параметры, значения которых трудно определить экспериментально. Как следствие, актуальной становится разработка эффективных алгоритмов определения значений таких параметров в ходе решения обратной задачи моделирования, то есть путем подгонки решений модели к имеющимся экспериментальным данным. Другой важной задачей математического моделирования в биологии является исследование механизмов, ответственных за устойчивое развитие организмов (робастность). Корректное математическое описание механизмов биологической робастности необходимо для предсказательной силы моделей.

Цель работы заключается в разработке и исследовании математических моделей активности генов сегментации в раннем развитии эмбриона плодовой мушки дрозофилы (*Drosophila melanogaster*). С использованием численных и аналитических методов построены и исследуются модели типа реакции–диффузии, механизмы параметрической устойчивости решений, а также совершенствуются методы решения обратной задачи моделирования, заключающейся в определении значений параметров в модельных уравнениях по экспериментальным данным.

Научная новизна результатов исследования складывается из следующих фактов:

Впервые описана континуальная модель типа реакции–диффузии, описывающая активность генов сегментации в эмбрионе дрозофилы с высокой точностью.

Впервые описан математический механизм сокращения числа аттракторов в динамических системах, в которых размерность фазового пространства является функцией времени.

Предложена новая модификация модели активности генов, с помощью которой решения модели демонстрируют улучшенную параметрическую устойчивость.

Предложен новый способ подбора начальной точки в пространстве параметров для задачи оптимизации значений параметров в модели активности генов.

Научная и практическая ценность работы. Полученные результаты могут быть использованы для построения робастных моделей генных сетей в многоклеточных организмах, включая человека. Такие модели важны для решения многих практических проблем медицины, таких как создание лекарств нового поколения, осуществляющих контроль над специфической генной активностью.

На защиту выносятся следующие положения:

1. Построена континуальная модель, описывающая динамику концентраций белков, кодируемых генами сегментации в ходе раннего развития эмбриона дрозофилы. Учет модельных уравнениях процесса деления ядер в эмбрионе реализует математический механизм контроля над числом аттракторов в модели.
2. Показано, что параметрическую устойчивость решений в модели активности генов можно улучшить с помощью модификации функционала качества, определяющего близость решений к экспериментальным данным. Эта модификация эквивалентна требованию квазистационарности решений на больших временах.
3. Выведена система уравнений для вариаций пространственных положений точек перегиба решений в модели активности генов при вариации внешнего фактора в уравнениях модели. Решения уравнений для вариаций определяют параметрическую устойчивость стационарных решений в модели активности генов со ступенчатой функцией регуляции.
4. Получена система ограничений в виде неравенств на параметры в модели активности генов, позволяющая найти приближенные значения параметров, при которых решения в модели качественно совпадают с экспериментальными данными. Использование приближенных значений параметров в качестве начальной точки повышает эффективность локальных методов минимизации функционала качества.

Апробация работы. Основные результаты диссертации докладывались на Международных конференциях “Days on Diffraction” (С.-Петербург, 2000, 2001; 2003; 2004), на Европейских междисциплинарных школах по нелинейной динамике для систем и анализу сигналов “Euroattractor” (Варшава, Польша, 2000; 2002), на Международной междисциплинарной школе по моделированию транскрипционных регуляторных сетей (Амблетез, Франция, 2002), на Московской международной конференции по

вычислительной молекулярной биологии “MCCMB-2003” (Москва, 2003), на Международной ежегодной конференции BMES (США, 2003), на Политехнических симпозиумах «Молодые ученые – промышленности Северо-Западного региона» (С.-Петербург, 2004; 2007), на Международной конференции по системной биологии «From Molecules & Modeling to Cells (SysBio 2005)» (Гозау, Австрия, 2005), на Международной конференции по вычислительной системной биологии “WCSB 2005” (Тампере, Финляндия, 2005), на Международных симпозиумах по сетям в биоинформатике “ISNB” (Амстердам, Голландия, 2006; 2007), на Международной конференции по биоинформатике регуляции и структуры генома “BGRS 2006” (Новосибирск, 2006), на Международной конференции по нанобиотехнологиям (С.-Петербург, 2006), на Международных симпозиумах “Atomic Cluster Collisions: Structure and Dynamics from the Nuclear to the MesoBioNano scales” (ISACC’07, Дармштадт, Германия, 2007; ISACC’08, С.-Петербург, 2008), на Российско-Немецкой конференции по системной биологии (Москва, 2008), на Международной конференции по системной биологии «ICSB-2008» (Гетеборг, Швеция, 2008), на семинаре кафедры зоологии Кембриджского университета (Кембридж, Великобритания, 2008), на семинаре кафедры прикладной математики и статистики Университета Стони Брук (Стони Брук, США, 2003), на семинарах ФТИ им. А.Ф. Иоффе РАН (С.-Петербург, 2005–2006 гг.), на семинарах отдела компьютерной биологии СПбГПУ (С.-Петербург, 2002–2008 гг.).

Публикации. По теме диссертации опубликовано 6 статей в реферируемых научных журналах, одобренных ВАК, и 4 статьи в научных сборниках (без учета аннотаций докладов на конференциях).

Структура диссертации. Диссертация состоит из введения, четырех глав, заключения, списка цитируемой литературы и приложения. Основной текст работы содержит 148 страниц, включая 34 рисунка, 12 таблиц, заключение и список литературы. Приложение содержит 20 страниц, включая 8 рисунков.

Краткое содержание работы

Во введении обоснована актуальность исследуемой проблемы, сформулированы цели исследования, раскрыты новизна и практическая значимость полученных результатов, приведены основные положения, выносимые на защиту.

В первой главе для удобства прочтения даны основные биологические определения и описывается общий вид уравнений математических моделей, исследуемых в диссертации.

Рассматривается математическая модель биологического процесса, называемого «экспрессией генов», происходящего в клетках живых организмов. Геном является участок молекулы ДНК, содержащий определенную информацию, которая физически закодирована в виде последовательности нуклеотидов. Экспрессией гена называется процесс, в ходе которого эта информация считывается с помощью специальных молекулярных механизмов и преобразуется в специфическую молекулу РНК (этот шаг считывания информации называется транскрипцией гена), которая затем преобразуется другими молекулярными механизмами в специфическую молекулу белка (этот шаг называется трансляцией РНК). Конечным продуктом экспрессии гена является белок, который, таким образом, кодируется геном.

Белки, как продукты экспрессии генов, могут регулировать экспрессию этих и других генов. Присоединяясь к специфическим участкам ДНК, молекулы белка, кодируемого геном А, могут либо увеличить вероятность производства новых молекул РНК и белка с гена Б (в этом случае говорят, что «ген А активирует ген Б»), либо уменьшить («ген А репрессирует ген Б»), либо не оказывают никакого влияния. В этом смысле гены взаимодействуют друг с другом посредством своих продуктов экспрессии, образуя сложную систему с нелинейными взаимодействиями, называемую генной сетью. Генную сеть часто представляют в виде графа, в вершинах которого располагаются гены (или продукты их экспрессии), а ребрами обозначаются межгенные взаимодействия.

В первые часы жизни эмбрион дрозофилы состоит из ядер, вокруг которых позднее появляются клеточные мембраны, завершая тем самым процесс формирования клеток. Каждое ядро содержит одинаковый набор генов. В диссертации исследуется модель экспрессии генов сегментации, которые названы так в силу того, что их активность контролирует разделение эмбриона на сегменты. Экспрессия генов сегментации происходит в течение ограниченного промежутка времени в ходе раннего развития эмбриона дрозофилы. Пространственное распределение вдоль центральной оси эмбриона функций концентраций белков, кодируемых генами сегментации, несет важную биологическую функцию, определяя, в конечном счете, пространственные границы сегментов эмбриона.

Континуальные уравнения, описывающие динамику концентраций $u^a(x, t)$ белков, кодируемых генами a в момент t в точке x на центральной оси эмбриона, выводятся из закона сохранения массы и закона Фика для диффузии и имеют вид:

$$\frac{\partial u^a}{\partial t} = S^a(u, x, t) - P^a(u^a) + D^a \frac{\partial^2 u^a}{\partial x^2}, \quad a = 1, \dots, N, \quad (1)$$

где $u = u(x, t)$ — вектор концентраций $u^a = u^a(x, t)$ для N белков, S^a — скорость синтеза a -го белка, P^a — скорость его химического распада, и D^a — коэффициент диффузии a -го белка между ядрами. При постановке начально-краевой задачи система уравнений (1) снабжается начальными условиями и краевыми условиями Неймана.

Также исследуется дискретная модель для концентраций $u_i^a(t)$ белка от a -го гена в i -м ядре эмбриона:

$$\frac{du_i^a}{dt} = S_i^a(u_i, t) - P^a(u_i^a) + \tilde{D}^a \left[(u_{i-1}^a - u_i^a) + (u_{i+1}^a - u_i^a) \right], \quad a = 1, \dots, N, \quad i = 1, \dots, M(t), \quad (2)$$

где через $u_i = u_i(t)$ обозначен вектор концентраций белков от всех генов в ядре i . Коэффициенты \tilde{D}^a и D^a в (1) и (2) имеют разную размерность. Уравнения (2) не являются дискретной версией уравнений (1) по следующей причине. Клеточная структура эмбриона представлена в (2) явным образом через зависимость от номеров i ядер будущих клеток. Поэтому перестроение этой структуры в ходе деления ядер выражается в том, что число ядер M меняется со временем и, соответственно, число уравнений в (2). Таким образом, уравнения (2) представляют динамическую систему, размерность фазового пространства которой есть функция времени и удваивается в результате каждого деления ядер.

Во второй главе формулируется и исследуется континуальная модель вида (1), описывающая экспрессию пяти генов сегментации. Решается обратная задача моделирования, а именно, находятся значения параметров в уравнениях путем минимизации функционала качества, определяющего разницу между решением и экспериментальными данными. Исследуется роль, которую клеточная структура эмбриона и ее перестроение могут иметь с точки зрения моделирования.

В рамках принятых предположений о функциях S^a и P^a , уравнения континуальной модели (1) предложены в следующем виде [3, 4, 5, 7, 10]:

$$\frac{\partial u^a}{\partial t} = \chi(t) R^a g \left(\sum_{b=1}^N T^{ab} u^b + m^a u^{\text{Bcd}}(x) + h^a \right) - \lambda^a u^a + D^a \frac{\partial^2 u^a}{\partial x^2}, \quad a = 1, \dots, N, \quad (3)$$

где $N = 5$ — число генов (белков), $R^a = \text{const}$, монотонно возрастающая нелинейная функция $g(y) = 0.5(1 + y / \sqrt{1 + y^2})$ описывает плавное переключение уровня синтеза от нулевого значения до максимального и называется «функцией регуляции», аргумент функции g в скобках в (3) описывает сумму всех факторов, регулирующих синтез a -го белка. В число этих факторов входят концентрации всех белков $u^b(x, t)$, которые влияют на синтез с весом T^{ab} (коэффициент взаимодействия между генами a и b : если ген b активирует ген a , то $T^{ab} > 0$; если репрессирует, то $T^{ab} < 0$; если не оказывает никакого влияния, то $T^{ab} = 0$), а также заданная концентрация материнского белка Bicoid $u^{\text{Bcd}}(x)$,

влияющая на синтез с весом m^a . Постоянная h^a описывает среднее влияние остальных, не учтенных явно, факторов, слагаемое $\lambda^a u^a$ в (3) представляет линейное приближение для скорости распада белка P^a , функция $\chi(t)$ описывает влияние деления ядер в эмбрионе, которое приводит к изменению общего числа копий генов и, следовательно, изменяет среднюю скорость синтеза белков от этих генов. Уравнения (3) снабжаются начальными условиями и краевыми условиями Неймана на ограниченном интервале центральной оси эмбриона. В качестве объекта исследования выбраны гены hb , Kr , gt , kni и eve поскольку экспрессия этих генов может считаться независимой от других генов в рассматриваемом интервале центральной оси эмбриона.

Значения пятидесяти параметров $p = \{R^a, T^{ab}, m^a, h^a, \lambda^a, D^a\}$, $1 \leq a, b \leq 5$, в модели (3) априори не известны и были найдены путем решения обратной задачи моделирования в ходе численной минимизации следующего функционала:

$$F = \sum_{a,k} \int_{\Omega} (u^a(x, t_k) - w^a(x, t_k))^2 dx, \quad (4)$$

где u^a есть решение уравнений (3), w^a суть экспериментальные функции концентраций белков, через Ω обозначен рассматриваемый интервал на центральной оси эмбриона, и суммирование производится по всем генам a и всем моментам времени k , в которые имелись экспериментальные данные. Минимизация функционала (4) выполнялась с помощью оптимизированного метода наискорейшего спуска.

Ранее активность описанных генов в эмбрионе дрозофилы исследовалась только в рамках дискретной модели (2), в которой явно учтена клеточная структура эмбриона. Однако оставалось неясным, насколько важна дискретная природа клеточной структуры эмбриона и деления ядер с точки зрения математического моделирования. С целью ответить на этот вопрос исследовалась континуальная модель (3), поскольку в такой модели клеточная структура эмбриона аппроксимируется континуальной средой. Были рассмотрены три варианта системы (3) с разными функциями $\chi(t)$, моделирующими деление ядер в эмбрионе с разной степенью точности. В каждой из этих моделей значения параметров были получены минимизацией функционала (4). В результате решения обратной задачи моделирования во всех моделях были получены решения, описывающие экспериментальные функции концентраций с удовлетворительной точностью (рис. 1).

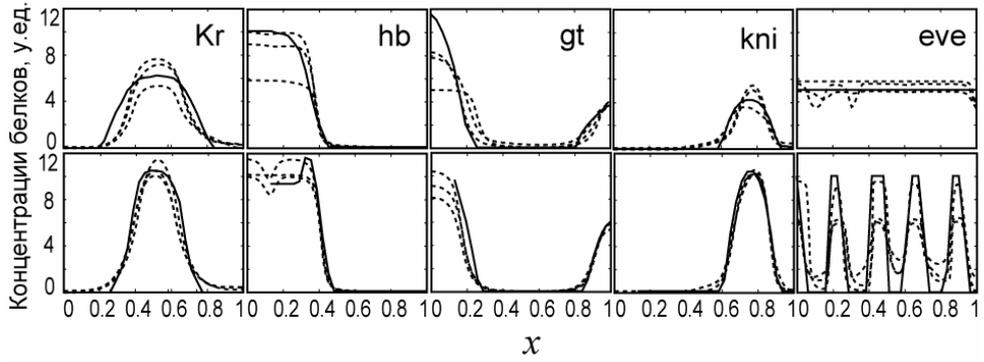


Рис. 1. Решения u^a (пунктирные кривые) в континуальной модели (3) для трех разных функций $\chi(t)$ в сравнении с экспериментальными данными (сплошные кривые). Названия белков указаны на графиках. Верхний ряд графиков соответствует раннему циклу развития эмбриона, нижний ряд — более позднему. Безразмерная пространственная координата изменяется вдоль центральной оси эмбриона.

Таким образом, показано, что явный учет клеточной структуры эмбриона и деления ядер не является необходимым при моделировании пространственно-временного распределения концентраций («паттернов») белков.

Для более детального исследования «неявных» эффектов, которые может давать учет деления ядер в модели, были изучены дискретные уравнения для «идеальной» системы двух генов в модели (2) с различным числом актов деления ядер. После каждого деления число ядер удваивается, и деления происходят до некоторого фиксированного момента времени, после которого число ядер в системе навсегда остается конечным.

Численное исследование аттракторов и их областей притяжения в фазовом пространстве динамической системы (2) выявило следующую закономерность:

$$S_{\tilde{M}} > A_{\tilde{M}}^1 > A_{\tilde{M}}^2 > A_{\tilde{M}}^3 > \dots \quad \forall \tilde{M}, \quad (5)$$

где \tilde{M} — заданное конечное число ядер в системе, $S_{\tilde{M}}$ — число линейно устойчивых стационарных точек уравнений (2) при данном \tilde{M} , $A_{\tilde{M}}^k$ — число всех стационарных аттракторов, реализующихся в системе, в которой начальное число ядер выбиралось равным $\tilde{M}/2^k$. Таким образом, $A_{\tilde{M}}^k$ описывает число аттракторов в системе с k последовательными делениями и вычисляется путем решения уравнений (2) с большим количеством случайных начальных условий.

Из (5) можно сделать вывод, что описанная зависимость размерности фазового пространства от времени играет динамическую роль в модели, сокращая общее число возможных стационарных состояний в динамической системе. Чем больше актов делений происходит в модели, тем меньшее число стационарных состояний реализуется в качестве

аттракторов. Этот вывод является следствием переменной размерности системы (2) во времени; он проиллюстрирован на рис. 2 в частном случае.

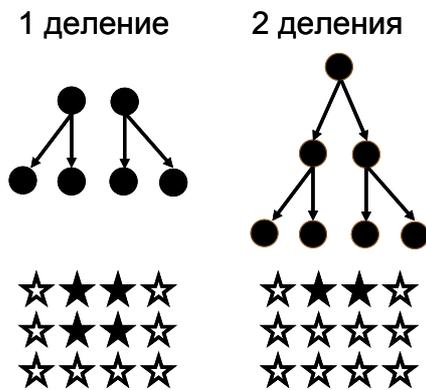


Рис. 2. Иллюстрация эффекта сокращения числа аттракторов в модели (2) для случая $\tilde{M} = 4$. Показаны два случая: в первом система приходит к \tilde{M} ядрам в результате одного деления, во втором — двух делений. В обоих случаях число линейно устойчивых стационарных точек составляет $S_{\tilde{M}} = 16$ (звездочки на рисунке). В первом случае из 16 стационарных точек только 4 реализуются как аттракторы (закрашенные звездочки) при варьировании начальных условий в модели; во втором — только 2. Остальные стационарные состояния (не закрашенные звездочки) недостижимы для динамики в модели при любых начальных условиях в обоих случаях.

В третьей главе в математических моделях (1) и (2) исследуются механизмы, ответственные за параметрическую устойчивость решений.

Рассмотрена дискретная модель экспрессии шести генов, основанная на следующих уравнениях [9]:

$$\frac{du_i^a(t)}{dt} = R^a g \left(\sum_{b=1}^N T^{ab} u_i^b + m^a u_i^{\text{Bcd}} + h^a \right) - \lambda^a u_i^a + \tilde{D}^a [(u_{i-1}^a - u_i^a) + (u_{i+1}^a - u_i^a)], \quad (6)$$

где белки нумеруются через $a = 1, \dots, 6$, ядра в эмбрионе нумеруются через $i = 1, \dots, M(t)$. Решения уравнений (6) биологически обоснованны во временном интервале $0 \leq t \leq \tau$, где τ определяется экспериментально измеренной длительностью биологически важного промежутка времени. Уравнения (6) снабжаются фиксированными начальными условиями.

Значения параметров в (6) получались путем решения обратной задачи моделирования, то есть в ходе минимизации функционала, определяющего квадрат разности между решением уравнений (6) и экспериментальными данными. С целью улучшения параметрической устойчивости решений в модели (6) минимизация функционала осуществлялась с дополнительным требованием квазистационарности решений на временах $t > \tau$. Устойчивость решений оценивалась вычислением отклонения возмущенных решений от экспериментальных данных в норме L_2 .

В результате многократного применения известного стохастического метода «численного отжига» (Metropolis et al., 1953) для численной минимизации функционала качества были получены три набора (p_1, p_2, p_3) значений параметров p , дающие

правильную динамику в биологически важный период времени $0 \leq t \leq \tau$ (рис. 3). Для большинства параметров их значения в наборах p_s , $s = 1,2,3$, близки друг к другу, однако значения некоторых параметров отличаются в p_s довольно значительно.

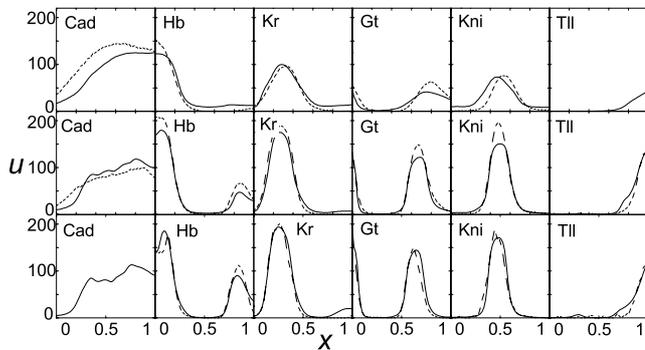


Рис. 3. Решение u_i^a (сплошные кривые) в модели (6) для одного из трех правильных наборов значений параметров p , найденных в результате минимизации функционала качества, в сравнении с экспериментальными данными (пунктирные кривые) для трех моментов времени. По оси абсцисс отложены масштабированные

положения x ядер i на единичном интервале, по оси ординат — концентрации белков, указанных на графиках, в условных единицах.

Была исследована устойчивость решений при найденных значениях параметров к возмущениям трех типов: возмущениям начальных условий, возмущениям решений u_i^a в нескольких фиксированных моментах времени $t > 0$, и возмущениям значений параметров. Характеристики возмущений первых двух типов были оценены по разбросу экспериментальных данных в ходе анализа концентраций всех белков во многих эмбрионах.

Возмущенные решения в модели (6) с $p = p_s$, $s = 1,2,3$, сравнивались с возмущенными решениями в модели (6) со значениями параметров p , полученными ранее в результате решения обратной задачи без дополнительного требования квазистационарности решения на больших временах. Для возмущений всех трех типов решения в модели с $p = p_s$ демонстрировали более устойчивое поведение.

В качестве примера исследования параметрической устойчивости опишем схему и результаты возмущений решений на временах $t > 0$. В интервале $0 \leq t \leq \tau$ фиксировались 24 момента времени t_s , в которые решения возмущались следующим образом:

$$(u_i^a)_{\text{pert.}}(t_s) = u_i^a(t_s) + \sqrt{u_i^a(t_s)} \varepsilon, \quad (7)$$

где u_i^a — невозмущенное решение, $(u_i^a)_{\text{pert.}}$ — возмущенное решение, ε — случайная переменная с нормальным распределением $\varepsilon \sim N(0, \sigma)$ с нулевым средним и фиксированным стандартным отклонением σ . Способ возмущения решений в (7) был

выбран в качестве простой модели логнормального распределения концентраций белков, наблюдаемого в эксперименте. Анализ экспериментальных кривых дал приблизительный разброс $0.5 < \sigma < 1$ для разных белков. В каждый момент t_s и для каждого фиксированного σ производилось 500 случайных возмущений согласно формуле (7). Результаты вычислений представлены на рис. 4.

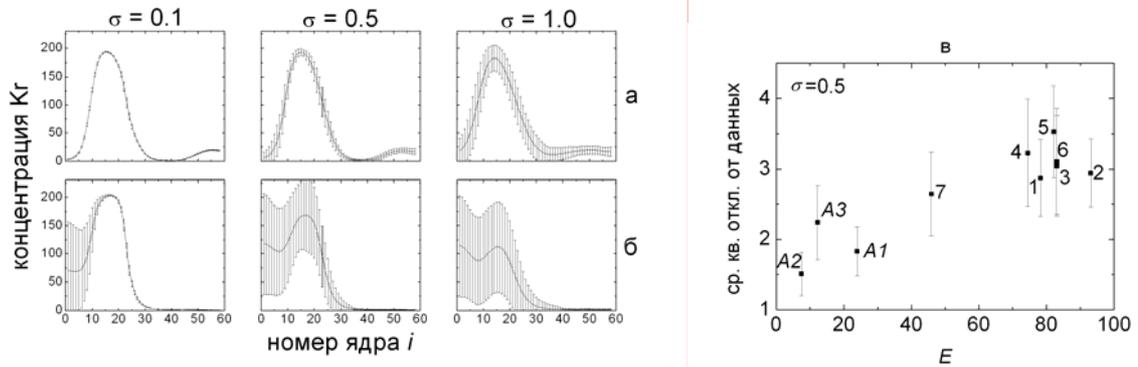


Рис. 4. Сравнение решений, возмущенных по формуле (7), в моделях двух типов. (а): Возмущенное решение для белка Kγ в модели (6) со значениями параметров из наборов p_1 – p_3 для трех значений σ . На графиках показаны средние значения концентрации $(u_i^a)_{\text{pert}}$ для Kγ и стандартные отклонения в каждом ядре в момент $t = \tau$. (б): То же, что в (а), но для значений параметров, полученных решением обратной задачи без дополнительного требования квазистационарности решений на временах $t > \tau$. (в): Зависимость среднеквадратичного отклонения возмущенных решений от экспериментальных данных (в долях от такого же отклонения невозмущенных решений) от степени E квазистационарности невозмущенных решений на временах $t > \tau$ при $\sigma = 0.5$ в моделях двух типов. E вычисляется как среднеквадратичное отклонение невозмущенных решений при $t = \tau$ от аттрактора в модели. Через A1–A3 обозначены результаты для значений параметров p_1 – p_3 ; цифрами 1–7 на графике обозначены результаты для семи известных наборов значений параметров, полученных решением обратной задачи без дополнительного требования квазистационарности решений на временах $t > \tau$.

Таким образом, использование дополнительного ограничения на поведение решений на больших временах при решении обратной задачи моделирования позволило существенно улучшить параметрическую устойчивость решений в модели.

Во второй части главы исследуется позиционная вариабельность решений уравнений вида (3) при вариации функции $u^{\text{Bcd}}(x)$. Позиционной вариабельностью решений называется разброс пространственных положений точек перегиба решений уравнений вида (3) при различных функциях $u^{\text{Bcd}}(x)$. Математически моделируется экспериментально наблюдаемый эффект, иллюстрированный на рис. 5.

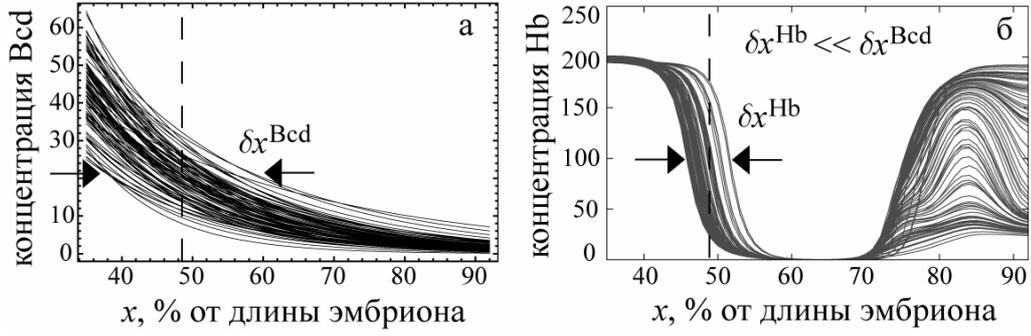


Рис. 5. Эффект малой позиционной вариабельности решений на примере белка Hb: $\delta x^{\text{Hb}} \ll \delta x^{\text{Bcd}}$. (а): Экспериментальные функции $u^{\text{Bcd}}(x)$, полученные из 45 эмбрионов дрозофилы. Позиционная вариабельность функции $u^{\text{Bcd}}(x)$, обозначенная через δx^{Bcd} , вычисляется как разброс точек пересечения фиксированного значения концентрации с данными функциями, где фиксированное значение равно значению «усредненной» функции $u^{\text{Bcd}}(x)$ в точке, отмеченной пунктирной линией. (б): Решения для белка Hb при $t = \tau$ в модели (6), в которой в качестве $u^{\text{Bcd}}(x)$ поочередно взяты 45 функций из (а). Величина δx^{Hb} определяется как разброс пространственных положений точек перегиба решений из (б). Везде вместо дискретного индекса i (номер ядра) используется непрерывная переменная x , равная положению ядра на центральной оси эмбриона (в % от длины эмбриона).

Позиционная вариабельность решений была исследована в непрерывном приближении и в пределе ступенчатой функции регуляции: $g = H$, где $H(y)$ — функция Хевисайда. Получены аналитические стационарные решения в непрерывной модели экспрессии четырех генов, а именно решения следующей системы уравнений [8]:

$$R^a H \left(\sum_{b=1}^N T^{ab} u^b(x) + m^a u^{\text{Bcd}}(x) + V^a(x) + h^a \right) - \lambda^a u^a(x) + D^a \frac{d^2 u^a(x)}{dx^2} = 0, \quad (8)$$

где индекс $a = 1, \dots, 4$ нумерует концентрации белков от рассматриваемых генов, $V^a(x)$ — внешний фактор, отличный от $u^{\text{Bcd}}(x)$, остальные обозначения описаны выше. Уравнения (8) снабжаются краевыми условиями Неймана на ограниченном интервале на центральной оси эмбриона (интервал $0 \leq x \leq 1$ для безразмерной пространственной переменной).

Аналитические решения системы уравнений (8) разыскивались в виде суперпозиции «стоячих кинков» (волн переключения) с точками перегиба, имеющими заданные координаты q_s . В общем виде решения можно записать следующим образом:

$$u^a(x, \vec{q}) = c_0^a(x, \vec{q}) + c_1^a(x, \vec{q}) \exp(\gamma^a x) + c_2^a(x, \vec{q}) \exp(-\gamma^a x), \quad a = 1, \dots, 4, \quad (9)$$

где через $\vec{q} = (q_1, \dots, q_S)$ обозначен вектор положений q_s всех S точек перегиба решений для всех a , $\gamma^a = \sqrt{\lambda^a / D^a}$, и коэффициенты $c_k^a(x, \vec{q})$ суть кусочно-постоянные функции от x .

Значения q_s были фиксированы равными пространственным положениям точек

перегиба функций концентраций белков из эксперимента. Из графиков для белков Hb, K₁, Gt и K_{n1} из нижнего ряда на рис. 3 можно определить $S = 10$ таких положений, нумеруемых в порядке их следования на единичном интервале переменной x . Можно показать, что по построению решений (9) положения q_s должны удовлетворять равенствам:

$$\sum_b T^{a(s),b} u^b(q_s, \bar{q}) + m^{a(s)} u^{\text{Bcd}}(q_s) + V(q_s) + h^{a(s)} = 0, \quad s = 1, \dots, 10, \quad (10)$$

где $a(s)$ есть однозначная функция $s \rightarrow a(s)$, связывающая номер s положения q_s с номером (или названием) белка a , чье стационарное решение имеет точку перегиба в q_s (например, $a(3) = \text{“Hb”}$; положение пунктирной линии на рис. 5б совпадает с q_3). Функции $u^b(x, \bar{q})$ в (10) суть решения (9).

При фиксированных q_s равенства (10) можно интерпретировать как уравнения для значений параметров $p = \{R^a, T^{ab}, m^a, h^a, \lambda^a, D^a\}$. Были найдены несколько наборов p_k значений параметров, удовлетворяющих равенствам (10), с помощью метода разностной эволюции. Таким образом, при значениях параметров p_k стационарные решения (9) существуют с точками перегиба в q_s , то есть, эти решения качественно совпадают с экспериментальными функциями концентраций рассматриваемых белков.

Варьирование (10) приводит к линейной алгебраической системе уравнений для вариаций δq_s вида:

$$f_s^1(\bar{q}, p, A, l) \delta q_s + f_s^0(\bar{q}, p, A, l, \delta A, \delta l) = 0, \quad s = 1, \dots, 10; \quad (11)$$

где A и l — параметры аппроксимации функции $u^{\text{Bcd}}(x)$: $u^{\text{Bcd}}(x) = A \exp(-lx)$; f_s^0 и f_s^1 — известные функции.

Система (11) была решена при найденных значениях параметров $p = p_k$ и при $\delta A = \sigma^A$, $\delta l = \sigma^l$, где σ^A и σ^l суть стандартные отклонения во множестве значений A и l , соответственно, полученных из 45 экспериментальных функций $u^{\text{Bcd}}(x)$. Полученные решения δq_s были использованы для вычисления так называемых коэффициентов позиционной вариабельности K_s для каждой точки перегиба: $K_s = \delta q_s / \delta x^{\text{Bcd}}(q_s)$, $s = 1, \dots, 10$, где δx^{Bcd} оценивалось из эксперимента.

Из всех наборов значений параметров p_k три набора ($p_1 - p_3$) соответствовали очень малым значениям коэффициентов K_s . В частности, для положения q_3 точки перегиба решения для белка Hb (пунктирная линия на рис. 5б) соответствующий коэффициент фильтрации при трех наборах значений параметров $p_1 - p_3$ равен: $K_3 = 0.009, 0.112$ и 0.217 . Эти результаты согласуются с экспериментальными оценками.

Таким образом, полученная система уравнений для вариаций положений точек перегиба стационарных решений в модели дает решения, правильно описывающие экспериментально наблюдаемую пространственную вариабельность концентраций белков.

В четвертой главе предложен комбинированный метод решения обратной задачи моделирования, который состоит в нахождении приближенных значений параметров в моделях вида (3) и (6) и последующего использования приближенных значений параметров в качестве стартовой точки для стандартного метода наискорейшего спуска.

В континуальной модели (3) на первом шаге находятся приближенные значения параметров в рамках следующих упрощающих предположений:

- исходя из экспериментальных данных, пространственный интервал разбивается на отрезки (Q_s, Q_{s+1}) преимущественной локализации одного белка (рис. 6);

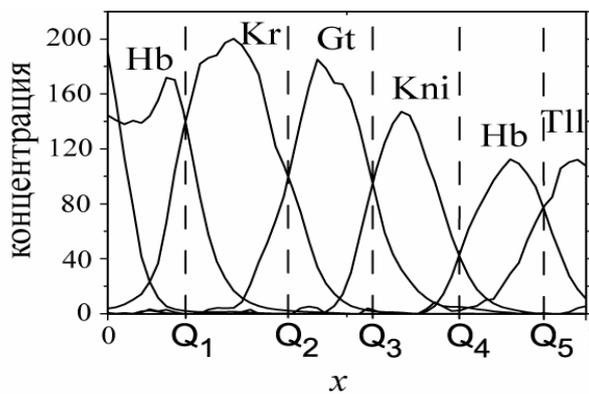


Рис. 6. Экспериментальные функции концентрации белков Hb, Kr, Gt, Kni и TII в момент $t = \tau$. Пространственный интервал условно разбит на интервалы (Q_s, Q_{s+1}) преимущественной локализации только одного белка, название которого приведено для каждого интервала.

- исходя из анализа литературных данных, предполагаются следующие ограничения на параметры T^{ab} , характеризующие тип взаимодействия между генами:

$$T^{aa} > 0, T^{ab} < 0, \forall a, b; \quad (12)$$

- в уравнениях (3) полагается $D^a = 0$ и $g = H$, где H — функция Хевисайда.

В принятых предположениях система (3) расцепляется, и x становится параметром. В такой системе для каждого фиксированного x может быть 2^N различных аттракторов с компонентами либо 0, либо R^a/λ^a , где N — число генов (белков). Для построения решений, качественно сходных с рис. 6, необходимо найти значения параметров, при которых в каждом интервале (Q_s, Q_{s+1}) реализуется аттрактор с компонентой R^a/λ^a (максимальное значение концентрации) для соответствующего белка и 0 для остальных белков. Доказано, что для того, чтобы в модели реализовывались аттракторы только с одной компонентой, равной R^a/λ^a , и нулевыми остальными компонентами, достаточно выполнения следующих ограничений на параметры:

$$T^{aa} \frac{R^a}{\lambda^a} + T^{ab} \frac{R^b}{\lambda^b} + m^a u^{Bcd}(x) + h^a < 0 \quad \forall a \neq b, \forall x. \quad (13)$$

Обозначим через $S^a(x,t) = \sum_b T^{ab} u^b(x,t) + m^a u^{Bcd}(x) + h^a$ аргумент функции регуляции в модели (3), и $S_0^a(x) = S^a(x,0)$. С учетом (12) доказывается, что для того, чтобы в каждом (Q_s, Q_{s+1}) выполнялось $u^a \rightarrow R^a/\lambda^a$ при $t \rightarrow \infty$ для соответствующего белка a и $u^b \rightarrow 0$ для всех остальных b , достаточно потребовать, чтобы $S_0^a(x) > 0$ внутри (Q_s, Q_{s+1}) и $S_0^a(x) < 0$ вне этого интервала. Например, для белка Hb ($a = 2$) соответствующие ограничения имеют следующий вид (см. также рис. 6):

$$S_0^2(x) > 0, x \in (0, Q_1) \cup (Q_4, Q_5); \quad S_0^2(x) < 0, x \in (Q_1, Q_4) \cup (Q_5, 1). \quad (14)$$

Аналогичные ограничения выписываются для остальных белков.

Приближенные значения параметров R^a , λ^a и D^a вычислялись из требования равенства максимальных значений экспериментальных концентраций значениям R^a/λ^a и путем подгонки стационарных решений вида (9) к экспериментальным функциям концентраций (рис. 6) в локальных пространственных областях. Приближенные значения остальных параметров (T^{ab} , m^a , h^a) фиксировались согласно ограничениям (12), (13), (14) и аналогам (14) для других белков.

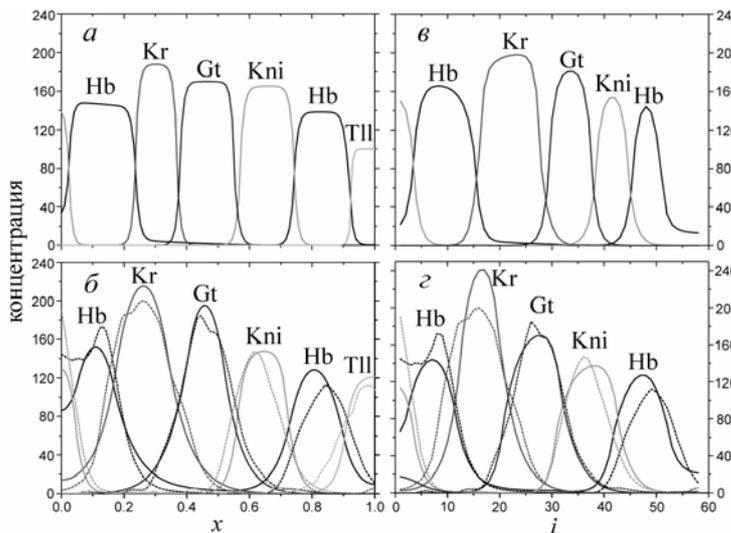


Рис. 7. (а): Решение в модели (3) при приближенных значениях параметров и $t = \tau$. (б): Решение в той же модели (сплошные кривые), но при значениях параметров, полученных в результате применения метода наискорейшего спуска с приближенными значениями параметров в качестве начальной точки, в сравнении с экспериментальными функциями концентраций (пунктирные кривые).

(в) и (г): То же, что в (а) и (б), соответственно, но для дискретной модели (6). Концентрация Tll не показана в силу того, что в дискретной модели она (вместе с концентрацией Cad) использовалась в качестве заданного внешнего фактора.

Решение в полной (без описанных выше приближений) модели (3) при приближенных значениях параметров показано на рис. 7а. Видно, что графики концентраций качественно схожи с экспериментальными кривыми из рис. 6. Решение на рис. 7а уточнялось путем применения оптимизированного метода наискорейшего спуска, при этом приближенные значения параметров использовались в качестве стартовой точки

в процессе оптимизации. Результирующее решение показано на рис. 7б в сравнении с экспериментальными данными. Метод также применялся к решению уравнений дискретной модели (6); были получены удовлетворительные результаты (рис. 7в,г).

В Заключении сформулированы результаты и основные выводы работы.

Выводы

Математические модели, основанные на континуальном подходе при описании пространственно протяженных структур в раннем развитии эмбриона дрозофилы, дают адекватные решения в рамках мезоскопического подхода к моделированию экспрессии генов сегментации. При формулировке уравнений континуальной модели клеточная структура эмбриона может не учитываться для описания динамики функций концентраций белков. Однако учет в дискретной модели процесса деления ядер в эмбрионе ведет к изменению размерности фазового пространства модели во времени. Такой математический механизм может играть важную динамическую роль, связанную с эффективным контролем числа возможных аттракторов, и данная роль проявляется только в дискретной модели. Эти факторы частично определяют область применимости моделей двух типов.

Требование квазистационарности решений на больших временах при решении обратной задачи моделирования позволяет улучшить параметрическую устойчивость решений. Такое ограничение может использоваться для построения более робастных моделей в тех случаях, когда моделируемый процесс происходит в ограниченный промежуток времени и заканчивается до момента установления решений в модели.

В согласии с экспериментом разброс пространственных положений точек перегиба решений в модели экспрессии генов оказывается значительно меньше, чем аналогичная позиционная вариабельность функции концентрации белка V_{cd} , являющейся внешним фактором в модели. Такой эффект малой ошибки позиционирования может быть математически описан на уровне стационарных решений в модели со ступенчатой функцией регуляции.

Предложенный метод нахождения приближенных значений параметров в модели экспрессии генов сегментации позволяет строить решения, качественно близкие к экспериментальным данным. Использование такого метода делает более эффективными локальные методы оптимизации в пространстве параметров при решении обратной задачи моделирования, такие как метод наискорейшего спуска, поскольку использование приближенных значений параметров в качестве начальной точки значительно повышает вероятность нахождения глобального минимума функционала качества.

Основные результаты диссертации опубликованы в следующих работах:

1. **Samsonov, A.M. Exact solutions to a nonlinear reaction-diffusion equation and hyperelliptic integrals inversion** [Текст] / A.M. Samsonov, V.V. Gursky // Journal of Physics A: Mathematical and General. - 1999. - V. 32. - С. 6573-6588.
2. **Gursky, V.V. How gap genes make their domains: An analytical study based on data driven approximations** [Текст] / V.V. Gursky, J. Reinitz, A.M. Samsonov // Chaos. - 2001. - V. 11. - № 1. - С. 132-141.
3. **Gursky, V.V. Mathematical modeling of pattern formation due to the segmentation gene expression in Drosophila** [Текст] / V.V. Gursky, J. Reinitz, A.M. Samsonov // Proceedings of the 1st European Interdisciplinary School on Nonlinear Dynamics for Systems and Signal Analysis "Euroattractor 2000". – Germany: Pabst Science Publishers, 2001. - С. 540-544.
4. **Gursky, V.V. Pattern formation and nuclear divisions are uncoupled in Drosophila segmentation: Comparison of spatially discrete and continuous models** [Текст] / V.V. Gursky, J. Jaeger, K.N. Kozlov, J. Reinitz, A.M. Samsonov // Physica D: Nonlinear Phenomena. - 2004. - V. 197. - С. 286-302.
5. **Samsonova, M.G. A survey of gene circuit approach applied to modelling of segment determination in fruit fly** [Текст] / M.G. Samsonova, A.M. Samsonov, V.V. Gursky, C.E. Vanario-Alonso // Multiple aspects of DNA and RNA: from Biophysics to Bioinformatics. - Elsevier, 2005. - С. 305-323.
6. **Gursky, V.V. Cell divisions as a mechanism for selection in stable steady states of multi-stationary gene circuits** [Текст] / V.V. Gursky, K.N. Kozlov, A.M. Samsonov, J. Reinitz // Physica D: Nonlinear Phenomena. - 2006. - V. 218. - № 1. - С. 70-76.
7. **Самсонова, М.Г. Системный подход к исследованию развития организмов** [Текст] / М.Г. Самсонова, В.В. Гурский, К.Н. Козлов, А.М. Самсонов // Научно-технические ведомости СПбГТУ. - 2006. - № 2. - С. 222-234.
8. **Gursky, V.V. Approximate stationary attractors in Drosophila gap gene circuits in the limit of steep-sigmoid interactions** [Текст] / V.V. Gursky, A.M. Samsonov, J. Reinitz // Proceedings of the 5th International Conference on Bioinformatics of Genome Regulation and Structure. - Novosibirsk: Institute of Cytology and Genetics, 2006. - V. 2. - С. 125-127.
9. **Гурский, В.В. Модель с асимптотически устойчивой динамикой для сети генов gap в дрозофиле** [Текст] / В.В. Гурский, К.Н. Козлов, А.М. Самсонов, Дж. Рейниц // Биофизика. - 2008. - Т. 53. - № 2. - С. 235-249.
10. **Samsonov, A.M. On modelling of gene expression patterns in Drosophila embryo by gene circuit method** [Текст] / A.M. Samsonov, V.V. Gursky, K.N. Kozlov, J. Reinitz // Proceedings of the 2nd International Symposium "Atomic Cluster Collisions: structure and dynamics from the nuclear to the biological scale". - London: Imperial College Press, 2008. - С. 426-440.