

НЕЧЕТКОЕ МОДЕЛИРОВАНИЕ РЕГУЛЯЦИИ ГЕНОВ СЕГМЕНТАЦИИ ДРОЗОФИЛЫ НА ЯЗЫКЕ ЛИНГВИСТИЧЕСКИХ ПРАВИЛ

© 2021 г. А.А. Макашов*, Е.М. Мясникова*, А.В. Спиров*, **

*Санкт-Петербургский политехнический университет Петра Великого, 195251, Санкт-Петербург, Политехническая ул., 29

**Институт эволюционной физиологии и биохимии им. И.М. Сеченова РАН, 194223, Санкт-Петербург, просп. Тореза, 44

***Институт научной информации по общественным наукам РАН, 117997, Москва, ул. Кржижановского, 15/2

E-mail: alexander.spirov@gmail.com

Поступила в редакцию 26.12.2019 г.

После доработки 12.08.2020 г.

Принята к публикации 21.09.2020 г.

Понимание молекулярных механизмов регуляции генов — одна из фундаментальных проблем, на которые нацелена системная биология. В принятых подходах к моделированию генной регуляции зачастую невозможно отделить свойства модели, адекватно отображающие характеристики биологического объекта, от обусловленных природой модели. В данной работе развит логический подход к моделированию регуляции экспрессии генов, позволяющий в перспективе максимально использовать все имеющиеся эмпирические данные и минимизировать использование априорных гипотез. При этом правила регуляции генов формулируются в виде лингвистических (сформулированных на языке близком человеческому) условий. Авторы используют более точно отображающий реальность нечеткий подход, известный как нечеткое лингвистическое моделирование. Впервые лингвистическое моделирование использовано для разработки моделей активности цис-регуляторных модулей и применено к проблемам анализа экспрессии генов биохимической разметки эмбрионов. Такой подход позволил прояснить существенные детали механизмов «считывания» первичных морфогенетических градиентов двумя или более цис-регуляторными модулями одного гена в их взаимодействии.

Ключевые слова: регуляция генов, паттерны экспрессии, энхансеры, грамматика энхансеров, логическое моделирование, нечеткое моделирование.

DOI: 10.31857/S0006302921010063

Понимание генной регуляции в целом и поведения генных регуляторных сетей (ГРС), в частности, это важнейшие фундаментальные проблемы, на которые нацелена современная системная биология. Для этих задач применяется подход, основанный на данных (data-driven approach), в котором имеющиеся данные используются для разработки модели, а затем выводы, полученные путем компьютерных тестов, проверяются новыми экспериментами, и эти новые результаты используются для улучшения модели [1]. Поэтому в системно-биологической парадигме равно важны как эксперименты, так и количественный анализ и моделирование полученных данных.

Сокращения: ГРС — генные регуляторные сети, ЦРМ — цис-регуляторные модули, ТФ — транскрипционные факторы, НЛМ — нечеткое лингвистическое моделирование.

В настоящее время несколько разных подходов используются для изучения работы гена посредством моделирования [2]. Основная методологическая проблема в этой области заключается в том, что зачастую трудно или даже невозможно различить свойства модели, адекватно отображающие некоторые характеристики исследуемого биологического объекта, от тех свойств модели, которые обусловлены природой подхода к моделированию. Прежде всего это относится к проблеме выбора априорных гипотез при моделировании. Например, модели генной регуляции, исторически развиваемые в парадигме математической физики (прежде всего, модели реакции-диффузии), характеризуются рядом черт, обусловленных природой подхода, достаточно далекого от биологических реалий.

Одной из самых развитых областей системно-биологического анализа генной регуляции явля-

ется моделирование динамики паттернов (биохимической разметки) эмбриона дрозофилы. Здесь наиболее популярными подходами являются нейросетевое (коннекционистское, connectionist) и термодинамическое моделирование [3, 4]. Коннекционистские модели (или модели генных схем, gene circuits) относятся к классу «огрубленных» подходов, тогда как термодинамическое моделирование можно рассматривать как более детальный уровень моделирования [5]. Помимо этих подходов следует упомянуть упрощенные логические модели, опубликованные в начале 2000-х годов [6, 7], а также комбинированные модели последних лет, в которых объединяются транскрипционная модель и дифференциальные уравнения [8–10].

Коннекционистские модели рассматривают ГРС как простой однослойный перцептрон (искусственная нейронная сеть). Подход описывает ГРС системой уравнений математической физики (обычно уравнениями реакции-диффузии), по одному уравнению для скорости экспрессии каждого гена, при этом члены реакции основаны на матрице межгенных регуляторных взаимодействий (gene-gene action) [3, 5]. Метод неявно основан на ожидании того, что организация связей у хорошо обученного перцептрона будет соответствовать функциональной структуре изучаемой ГРС [11]. Эти ожидания очень старые, и такие модели, как оказалось, хороши для моделирования поведения ГРС в норме (дикого типа), но подобное моделирование не всегда подходит для того, чтобы предсказать последствия мутаций, возмущений или шума (низкая прогностическая сила [12]). Нам известны только работы по моделированию мутантов по генам *Kr* [13] и *bcd* [14].

Термодинамическое моделирование основано на экспериментальных или биоинформационных знаниях об организации регуляторных областей (энхансеров или цис-регуляторных модулей, ЦРМ) данного гена [2, 4, 15]. Это количественное моделирование для прогнозирования экспрессии генов исходя из анализа их нуклеотидной последовательности, и оно основано на подходах статистической термодинамики. Статистическая термодинамика характеризует специфическое связывание транскрипционных факторов (ТФ) с сайтами связывания на ДНК в равновесных условиях [16]. Таким образом, подход явно включает наши знания о сайтах связывания для ТФ, их силе, ориентации и точной позиции. Это дает ложное впечатление, что подход рассматривает точную экспериментально выявленную регуляторную организацию энхансеров. Но на самом деле это не так: подход обычно игнорирует многие детали иерархической организации регуляторных взаимодействий. Он обычно игнорирует неаддитивный, синергетический характер интеграции регуляторного входящего сигнала при помощи

ДНК-матрицы и механизма транскрипции (см. работы [4, 15]). Это прежде всего обусловлено существенной ограниченностью наших знаний о природе и молекулярных механизмах этой интеграции [17–20].

Все эти представления побуждают исследователей рассматривать другие подходы к моделированию поведения ГРС, более ориентированные на данные и позволяющие непосредственно включать в модель весь доступный эмпирический материал и делать это простым, естественным путем. Мы считаем, что наилучшим способом сделать это является дальнейшее развитие семейства разнообразных подходов, обычно называемых нечетким лингвистическим моделированием (НЛМ) [2]. Эти подходы, в частности, используют конструкции «если ..., в таком случае..., еще ...» («if ..., then..., else...») для словесного описания разнообразных взаимоотношений между регулирующими входными сигналами и путями их интеграции в выходные сигналы. Это позволяет сформулировать с помощью простых правил известные детали разнообразной и иерархической регуляторной организации генов.

В этой работе мы развиваем подход НЛМ к регуляции экспрессии генов с точностью до отдельных ЦРМов. Конкретно мы сосредоточимся на примере регуляций генов сегментации эмбриона дрозофилы, как одном из самых исследованных модельных объектов.

Нечеткие логические модели регуляции генов. Поведение биологических систем по своей природе нечеткое. Несмотря на то что новые высокопроизводительные технологии предоставляют биологам огромное количество ценных данных, необходимы методы, которые включают в моделирование погрешность (неточность, изменчивость), позволяя иметь дело со многими генами с невыясненной функцией и с «зашумленными» экспериментальными данными. Для получения биологически значимых результатов информация должна объединяться из разных источников для конструирования моделей ГРС. Такие нечеткие экспертные знания включают базы данных о генах и их продуктах, а также информацию о взаимодействиях между генами.

Нечеткая логика, как фундаментальная составляющая нечеткого метода, представляет собой комбинацию различных математических принципов представления знаний об изучаемом явлении в нечетком, не бинарном виде (в отличие от Булева подхода). Метод нечеткой логики состоит из ряда процедур, которые используются для сопоставления определенного входящего и ему соответствующего выходящего сигнала: процесс, называемый «нечетким выводом». Двумя наиболее известными методами вывода являются те, которые были связаны с именами Е. Mamdani

и M. Sugeno [21, 22]. Эти модели зависят от использования простого языка правил «if-then», описывающего ответ системы как функции нескольких лингвистических переменных.

Лингвистические описания (или описания на «естественных языках», в отличие от, например, формализма математической физики) наблюдаемых процессов могут быть легко переформулированы в предложения типа «if-then», которые отражают поведение системы, без необходимости использовать математические функции. Как правило, такие описания знаний на основе «естественного языка» явно или косвенно включают дискретизированные описания объектов или относятся к ним (например, низкие, средние и высокие концентрации).

Таким образом, представляется весьма перспективным развивать компьютерный подход, позволяющий прямое преобразование качественных знаний и описаний в создаваемую рабочую модель. При этом первым шагом будет найти подходящее формальное представление дискретных, неточных естественных языковых терминов, которые описывают состояния объектов. Второй шаг – найти подходящее формальное представление функций, которое позволит непосредственно объединить качественные знания на основе естественного языка и тем самым облегчить интерпретацию функциональности.

Развиваемый подход имеет отношение к логическому моделированию генной регуляции – подходам к нахождению логик генной регуляции (gene-regulatory logic inferring [23]). Логические вентили (logic gates), которые используют более одного входа, уже давно применяются для описания сложных взаимодействий между ТФ, специфически связывающимися со своими сайтами на генно-регуляторном элементе. Регуляторная функция, которой описываются цис-регуляторные взаимодействия между ТФ, связанными со своими сайтами, и управляющими экспрессией их гена-мишени, может быть представлена в виде Булевых логик [21, 24]. Соответственно, переход от булевой логики к нечеткой представляется естественным обобщением.

В недавние годы опубликованы серии работ, реализующих моделирование на основе нечеткой логики (преимущественно, гибридные подходы) на уровне грубого моделирования ГРС (нечеткие модели генных сетей, fuzzy gene-circuit models) [25–29]. Вместе с тем в системной биологии продолжают разработки моделей детального уровня, включающих формулирование генно-регуляторной логики (с целью расшифровать логику действия ТФ на генно-регуляторные элементы при контроле транскрипции) [30, 31]. Параллельно продолжают работы по выяснению правил организации сайтов связывания для ТФ в функ-

циональные генно-регуляторные модули, ЦРМ [32–36]. Это можно назвать грамматическими правилами для энхансера или грамматикой цис-регуляторного модуля [30, 31, 37]. Это способ переформулировать экспериментальные выводы, выраженные на языке экспериментаторов, на язык, используемый для разработки моделей. В свою очередь, область грамматики генно-регуляторных элементов естественно отнести к более общей области проблем на стыке молекулярной биологии и лингвистики [38, 39].

Несмотря на относительную развитость ряда направлений НЛМ на уровне грубого моделирования, попытки развить подходы нечеткого моделирования на уровне детального (fine-grained) описания находятся еще в самом зачаточном состоянии (см. работу [40]). В этой статье мы предлагаем нашу версию НЛМ на уровне детального моделирования. Мы соединяем подходы нечеткого моделирования с подходами, исходящими из формулировки правил грамматики цис-регуляторных элементов и подходами к нахождению логик генной регуляции при действии ТФ.

Примечательно, что именно подход к моделированию динамики генных сетей, основанный на нечеткой логике [41], оказался наиболее эффективным для тестовых задач обратной инженерии функциональной организации ГРС по сравнению с рядом других подходов ([42], проект «The DREAM Challenges» (<https://www.synapse.org/#!Synapse:syn3049712/wiki/74631>)). Это побуждает нас развить аналогичный подход для моделирования паттернов экспрессии генов, с точностью до отдельных (полу)автономных ЦРМов.

Мы намереваемся интегрировать данные экспрессии генов и информацию о связывании ТФ с ДНК с конечной целью расшифровки логики регуляции экспрессии гена. Такой подход может количественно охарактеризовать закономерности взаимодействия между ТФ, объединяя цис-регуляторные логики и кинетику транскрипции в одной модели. Полученные логики цис-регуляций могут затем использоваться для установления функциональной организации ГРС. Мы иллюстрируем такой подход на тестовом примере одного из наиболее изученных наборов ЦРМов – энхансеров гена hunchback, hb [43; 44; 45; 46; 47; 48; 49; 50; 51]. Он служит экспериментальной моделью для изучения многих других цис-регуляторных модулей в биологии развития и, в более общем плане, в эволюционной биологии (evo-devo).

НАШ ПОДХОД

Мы развиваем наш подход НЛМ к моделям генной экспрессии на уровне детального моделирования (уровень кластеров сайтов связывания

для ТФ). Этот уровень основан на детальном знании о функциональной организации генных регуляторных элементов и наиболее приближен к биологическим реалиям, в отличие от упрощенных подходов грубого уровня.

Мы используем анализ грамматики энхансеров (правила размещения сайтов связывания друг относительно друга, включая расстояния между ними), принадлежащих к одному семейству, для формулировки генно-регуляторной логики в виде наборов регуляторных функций [28, 29]. Знание этой грамматики позволяет в итоге сформулировать регуляторные функции для описания поведения ГРС. Эти правила могут быть использованы для развития НЛМ детального уровня.

Самое общее преимущество НЛМ по сравнению с другими подходами к моделированию поведения генных сетей заключается в том, что:

1) подход не использует априорных моделей механизмов генной регуляции и устроен как «черный ящик» (на входе регулирующие параметры, на выходе уровень экспрессии гена), так что зависимость выхода от входных параметров определяется системой нечеткого логического вывода,

2) нечеткий логический вывод в силу простоты организации находит конкретное частное решение простыми формальными процедурами и работает существенно быстрее по сравнению, например, с моделированием системами дифференциальных уравнений, что многократно ускоряет поиск при оптимизации эволюционными методами.

Развиваемый здесь подход, восходящий к классическим публикациям по нечеткому моделированию [52], можно трактовать как развитие подхода кусочно-линейных аппроксимаций нелинейных функций в приложении к разработке и анализу нелинейных динамических моделей генных и клеточных регуляторных сетей (см. работы [53, 54]).

Наш объект изучения. Почти 100 охарактеризованных энхансеров контролируют формирование паттерна раннего эмбриона дрозофилы, вероятно, наиболее изученный процесс эмбрионального развития ([45, 55–57]). Эти энхансеры регулируют экспрессию порядка 50 генов, контролирующих формирование передне-заднего и дорзо-вентрального паттернов на этапах сегментации и гаструляции. Такой массив экспериментальных и биоинформационных данных дает беспрецедентную возможность анализа регуляторной грамматики этих энхансеров, контролируемых примерно 30 ТФ. Результаты геномного секвенирования целого ряда различных видов дрозофилы делает возможным анализ эволюции этих энхансеров [58].

Морфогенетический градиент фактора Bicoid (Bcd) является одной из наиболее изученных моделей эволюционной биологии развития [48, 59]. Фактор Bcd активирует набор генов сегментации дрозофилы через набор энхансеров, управляемых Bcd. Было описано несколько десятков активирующихся при помощи Bcd энхансеров, действующих в начале развития дрозофилы, и некоторые из них (прежде всего проксимальный элемент *hb* и элемент 2-й полосы гена *even-skipped*) были подробно изучены [15, 45, 47]. Функциональным ядром таких элементов является кластер сайтов связывания фактора Bcd (рис. 1а,б). Bcd обычно требует несколько сайтов связывания для активации экспрессии генов, и эта активация обычно является кооперативной.

Энхансеры, активируемые Bcd, как и многие другие ЦРМ, интегрируют множество регуляторных входных сигналов от активаторов и ко-активаторов и репрессоров, и ко-репрессоров для обеспечения воспроизводимости и устойчивости раннего эмбрионального паттерна (рис. 1). Регулирование является иерархическим и включает в себя как минимум три уровня.

Первый уровень — это уровень «раскрытия» хроматина, а ключевым фактором, участвующим в открытии, является Zelda, Zld (рис. 1б). Zld представляет собой сайт-специфический ТФ, который также помогает связывать другие факторы, изменяя локальную доступность хроматина, тем самым играя главную роль в раскрытии конденсированного хроматина. Исследователи называют Zld-подобные факторы факторами-«пионерами», поскольку они контролируют активацию генома зиготы через их способность раскрывать конденсированный хроматин [60, 61].

Второй уровень — типичное кооперативное связывание ДНК, например, между факторами Bcd, Hb и другими ключевыми активаторами, и коактиваторами. Этот эффект увеличивает общее родство связывания, что приводит к согласованной «загрузке» сайтов при более низком общем уровне концентрации ТФ. Как результат, это обеспечивает и резкий регуляторный «скачок» в ответ на относительно небольшое увеличение концентрации активаторов. Далее, обсуждаемые энхансеры всегда содержат сайты связывания для факторов-репрессоров, как правило, вблизи сайтов для активаторов (рис. 1). Репрессоры обычно характеризуются как короткодействующие: они действуют репрессивно на соседние с ними связанные активаторы.

Третий уровень представляет собой транскрипционную синергию, в которой многочисленные специфически связанные ТФ взаимодействуют с компонентами транскрипционной машины, либо с базальными субъединицами, либо с общими ко-активаторами транскрипции [62].

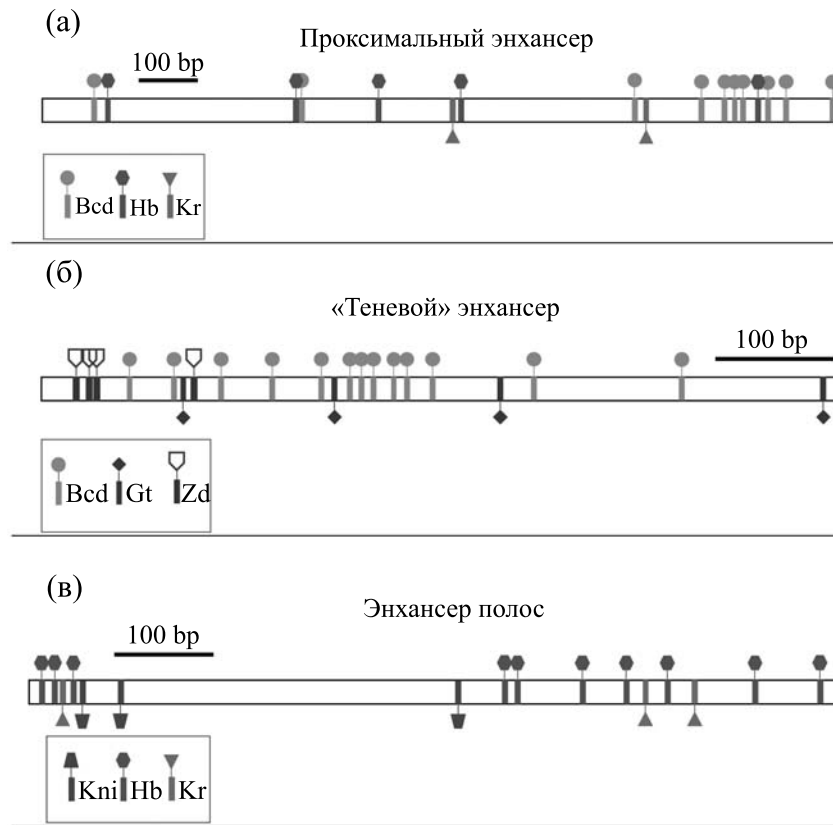


Рис. 1. Детали функциональной организации трех наиболее изученных энхансеров гена *hunchback* (*hb*): проксимальный энхансер (а), дистальный («теневой», distal anterior enhancer, DAE) энхансер (б) и энхансер «полос» (stripe-element) (в). Указаны позиции сайтов связывания ключевых транскрипционных факторов (найденных экспериментально или биоинформационными методами). Хорошо видна тенденция сайтов связывания находиться в кластерах. Сайты активаторов показаны над элементом, а сайты репрессии – под ним: Bcd – Bicoid, Kr – Kruppel, Kni – Knirps, Gt – Giant, Zd – Zelda (согласно [47] и [64], с изменениями).

Конкретно, например, показано, что Bcd активирует синергическую транскрипцию, взаимодействуя с TBP-ассоциированными факторами TAFII60 и TAFII110 [17, 18]. Если эти факторы отсутствуют, то активации транскрипции не происходит.

Цис-регуляторные модули гена *hunchback*. Согласно относительно недавним (и совсем недавним) публикациям, ранний паттерн экспрессии гена *hb* находится под контролем трех ЦРМ: проксимального, теневого и энхансера полос [46, 47, 63, 64], что иллюстрировано схемами на рис. 1 (данные сведены на нашем веб-ресурсе NOX.ro [65, 66]). При этом функции проксимального и теневого элементов весьма схожи (что ставит вопрос о роли такой функциональной избыточности) [46, 47]. Все три элемента работают на интересующей нас здесь стадии раннего эмбриогенеза – четырнадцатом цикле. Согласно недавним исследованиям, вклад этих элементов в суммарный уровень экспрессии гена *hb* можно трактовать как близкий к аддитивному [63].

Проксимальный элемент (рис. 1а) безусловно относится к одним из самых изученных. Тем не менее многие критические детали его функционирования по сию пору неясны. Наиболее изученная роль этого элемента – резко активировать ген *hb* при концентрации бикоида выше критической [47, 48, 67–73]. Сходным образом работает целый ряд ЦРМ ранних генов эмбриона дрозофилы, и этот энхансер считается для них прототипическим. Элемент имеет как несколько экспериментально охарактеризованных сайтов для фактора Bcd, так и ряд сайтов, обнаруживаемых биоинформационными методами. Резкий функциональный отклик (почти по типу «все или ничего») объясняется хорошо исследованной кооперативностью связывания (и действия) фактора Bcd с этим элементом [48, 74, 75]. Помимо этого, в проксимальном энхансере обнаружены (и частично охарактеризованы экспериментально) сайты для Hb и Kr (рис. 1а), как и некоторых других ТФ (не приведены на рисунке). Такая избыточность сайтов как для активации, так и для репрессии нуждается в дальнейшем анализе. Ос-

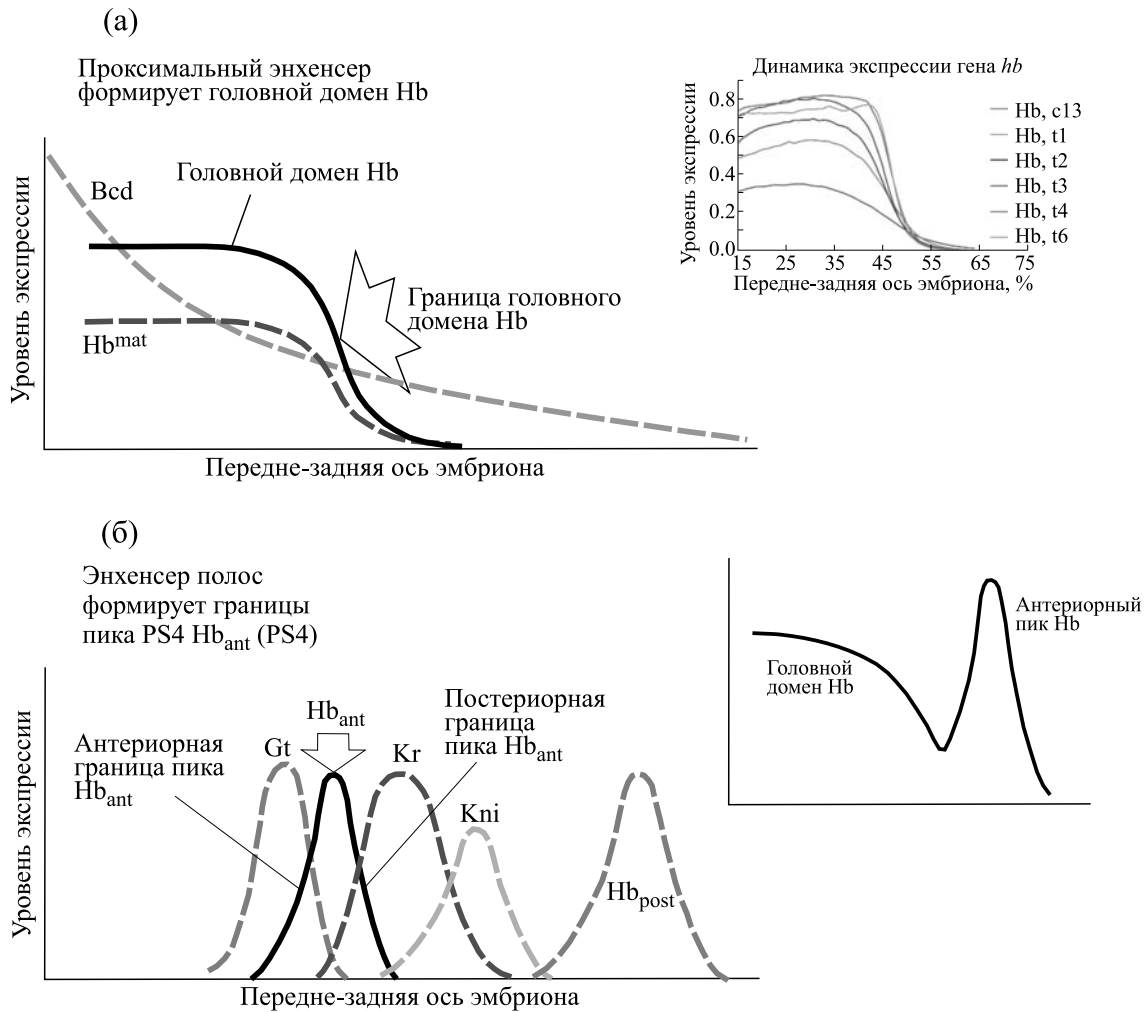


Рис. 2. Профиль экспрессии гена *hb* в головной половине раннего эмбриона дрозофилы включает головной домен и anteriорный пик (Hb_{ant}). Их формирование контролируется тремя (полу)автономными энхансерами. (а) – Роль проксимального энхансера. В самом раннем развитии и до начала четырнадцатого цикла ген *hb* находится под преимущественным контролем активатора Bicoid (Bcd) через проксимальный (и теневого) энхансеры. Это определяет появление, рост и формирование головного домена *hb* со все более крутой (постериорной) границей (склоном) этого домена. Приведен профиль материнского фактора Hb, также контролирующего головной домен. На врезке показано, как меняется геометрия головного домена в раннем эмбриогенезе: склон становится все более крутым. (б) – Роль энхансера полос. С начала четырнадцатого цикла контроль над *hb* переходит к элементу полос (см текст). Энхансер контролирует формирование anteriорного и постериорного (Hb_{post}) пиков (полос). На врезке иллюстрируются вклады проксимального энхансера и энхансера полос в итоговую картину экспрессии гена *hb* в четырнадцатом цикле (головной домен + anteriорный пик). Экспрессия измеряется в относительных условных единицах от 0 до 1.

новная проблема заключается в том, что крутизна отклика гена *hb* в целом слишком велика, чтобы объяснить ее только кооперативностью действия Bcd (врезка на рис. 2а иллюстрирует этот рост крутизны границы домена *hb*). Вопрос о том, как и какие именно другие ТФ участвуют в этих процессах, дающих в итоге отклик «все или ничего», требует дальнейшего изучения и объяснения.

Энхансер полос включает в себя сайты связывания для факторов Kr, Kni, Hb [47, 76] (рис. 1в). Под контролем энхансера полос ген-мишень дает два четких острых пика экспрессии в передней и

задней частях эмбриона (антериорный и постериорный пики или полосы, Hb_{ant} и Hb_{post} , как на рис. 2б). (Hb_{ant} известен также под именем пик PS4).

Полагают, что для элемента полос ген *hb* активируется какими-то малоисследованными общими активаторами, а роль этого элемента состоит именно в репрессии *hb* всюду, за исключением двух позиций, где и формируются пики экспрессии. При этом, в частности, склоны anteriорного пика «справа» определяются доменом репрессора Kr, а «слева» (как полагают) – доменом



Рис. 3. Диаграмма системы нечеткой логики в приложении к проблеме моделирования экспрессии генного регуляторного элемента (энхансера). Машина вывода (inference engine) использует нечеткие функции генной регуляции, которые формулируются на основе анализа грамматики энхансера (см. текст).

головного Hb (саморепрессия), как иллюстрируется рис. 26.

Детали и механизмы действия энхансера полос все еще мало исследованы. В частности, неясны механизмы репрессии и степень их кооперативности.

Моделирование активности цис-регуляторных модулей с использованием нечеткой логики. Мы используем стандартное общее представление для НЛМ, приведенное на рис. 3. Специфика нашего подхода заключается в приложении общей стратегии НЛМ к конкретике деталей функционирования анализируемого семейства энхансеров.

Взаимодействия между биологическими объектами представляют собой процессы, которые влияют на будущее состояние объектов-мишеней исходя из текущего состояния объектов-эффекторов. Компьютерные модели имитируют взаимодействия при помощи функций, которые оперируют вычислительными представлениями о состояниях. Эти функции отображают текущие состояния эффекторов (входящие данные) в новые состояния или изменения состояния мишеней (выходные данные), как на рис. 3.

В общем случае механизм логического вывода включает фаззификацию, нечеткий вывод (fuzzy inference), и дефаззификацию (рис. 3). На входе нашей НЛМ находится набор уровней экспрессии регуляторных факторов, контролирующих экспрессию гена-мишени через анализируемый нами энхансер этого гена (сравни рис. 1 и 2), на выходе — уровень экспрессии (т. е. уровень продукции мРНК) нашего гена-мишени. (Для единообразия экспрессия генов приведена и в экспериментальных данных, и на выходе эксперимента к относительным условным единицам от 0 до 1.)

Процедуру фаззификации в нашем случае естественно выполнять исходя из треугольных термов с включением краевых трапецевидных термов [77] (см. рис. 4б). Для стадии дефаззификации мы воспользуемся широко употребляемой

и простой процедурой дефаззификации по высоте (height defuzzification) [78].

Для нашего набора энхансеров (тест-объектов) мы формулируем набор грамматических правил, позволяющий определять вид и характеристики регуляторных функций как кусочно-линейных средствами НЛМ (см. работы [52–54]). Конкретно мы делаем такую аппроксимацию для функций активации и репрессии гена-мишени транскрипционным фактором-эффектором.

Мы исходим из того, что каждый ген, контролирующий формирование эмбрионального паттерна, состоит из нескольких (полу-) автономных регуляторных элементов (ЦРМ), так что каждый такой элемент описывается своей нечеткой моделью. На выходе такой модели уровень экспрессии гена, определяемый рассматриваемым ЦРМ, тогда как на входе — уровни концентраций сайт-специфических ТФ, способных специфически связываться с рассматриваемым ЦРМ и действовать на ген через него.

Нечеткие функции генной регуляции. Даже самый поверхностный анализ динамики паттернов экспрессии гена *hb* показывает, что кинетика действия и активаторов и репрессоров нелинейна. Мы ниже приходим к заключению, что для наших целей достаточно использовать признанную и достаточно простую модель регуляции рассматриваемого гена данным фактором (через данный ЦРМ) — модель Хилла. Модель Хилла описывает или активацию или реессию. Мы, естественно, используем нечеткую формулировку этого закона. НЛМ могут быть сформулированы таким образом, чтобы они аппроксимировали сигмоидные профили функций Хилла, описывающие активирующие и ингибирующие эффекты, что иллюстрируется на рис. 4.

Достаточно четырех нечетких множеств и набора простых правил для создания кусочно-линейных аппроксимаций функций Хилла для активации. Качество аппроксимации зависит от количества нечетких множеств, но оно уже

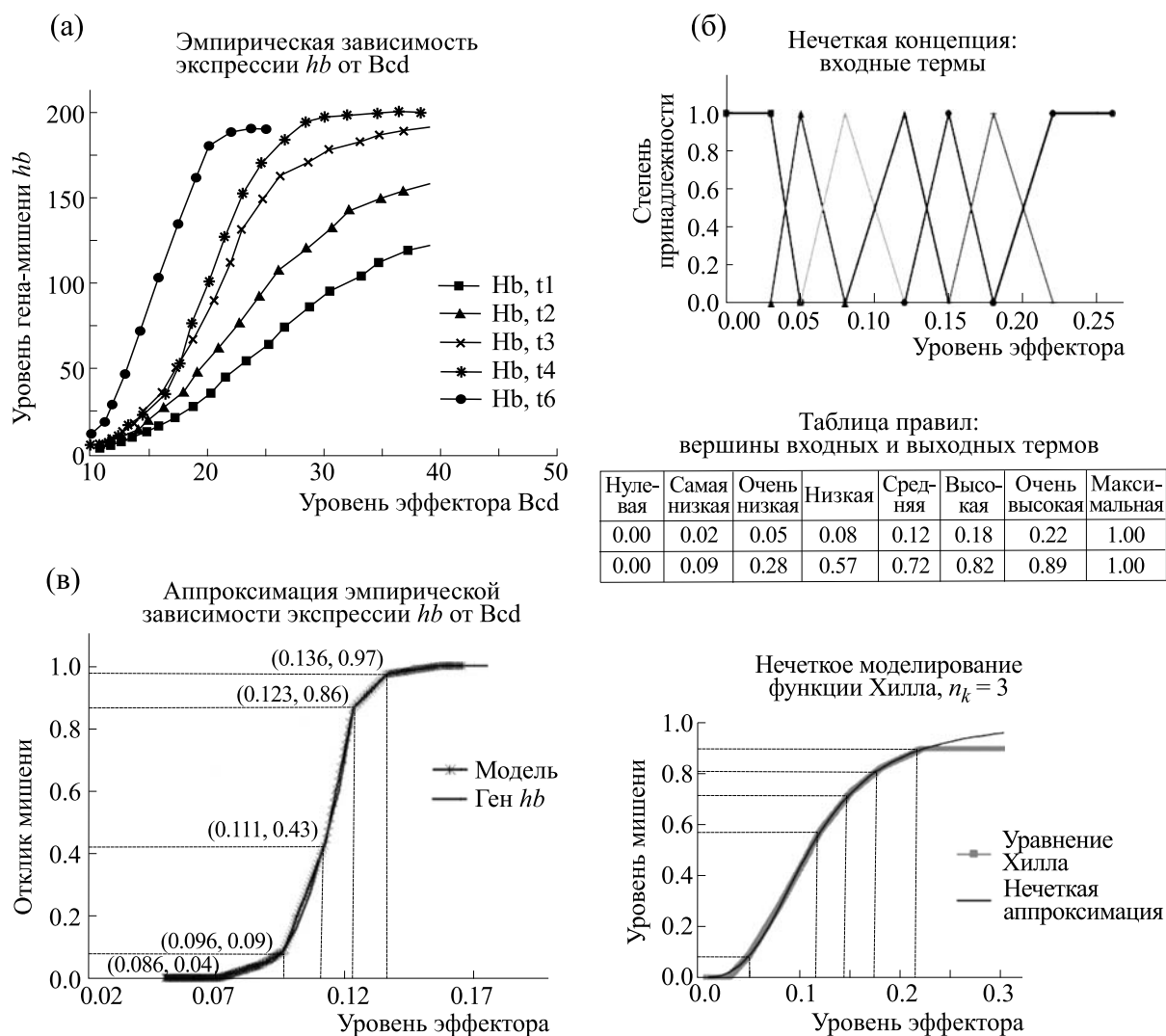


Рис. 4. Нечеткое представление функции Хилла для задач моделирования кооперативного действия активатора на ген-мишень. (а) — Эмпирические зависимости экспрессии гена *hb* от уровней кооперативного активатора *Bcd* на ранних стадиях эмбриогенеза дрозофилы (по материалам веб-ресурса FlyEx [91]). (б) — Детали подхода к нечеткому представлению функции Хилла (коэффициент Хилла $n_k = 3$) пятью термами: два крайних, имеющих форму трапеции, и три средних — треугольные; представлены входные термы нечеткой концепции, таблица правил и результат кусочно-линейной аппроксимации этой функции. (в) — Пример конкретного разбиения эмпирической зависимости *hb* от *Bcd* (для временного класса t1 четырнадцатого цикла; сравни с панелью (а)) на пять термов. Эмпирический график аппроксимирован отрезками прямых. Отмечены координаты точек перелома аппроксимации. Координаты границ отрезков (по оси эффектор) соответствуют центрам входных термов, тогда как их координаты по оси «ген-мишень» определяют значение центров выходных термов (эти пары значений приведены в скобках; см текст). Экспрессия измеряется в относительных условных единицах от 0 до 1.

удовлетворительно, если даже используются только четыре нечетких множества (ср. [79]). Мы в нашем моделировании использовали пять или шесть термов для аппроксимации функций Хилла (рис. 4).

Используя только три нечетких множества, можно задать удовлетворительную кусочно-линейную аппроксимацию функций Хилла для кооперативной репрессии [79]. В общем случае вопрос параметров фазификации, как и вопрос па-

раметров самой функции Хилла, должен решаться оптимизацией при адаптации модели.

Наши алгоритмы процедур нечеткой логики. В описании подхода мы будем пользоваться терминами эффектор и ген-мишень. Под эффектором типично понимается сайт-специфический ТФ и предполагается, что в простейшем случае он действует на ген-мишень (через ЦРМ этого гена) или как активатор, или как репрессор. Ниже мы часто будем использовать понятие терм для лингвисти-

ческой переменной, определяющей уровни экспрессии эффектора и мишени.

Входные (и выходные) термы. Входные и выходные термы определяются (строятся) по такому алгоритму.

Строим эмпирический график зависимости уровня экспрессии гена-мишени от уровня эффектора (на примере активатора, как на рис. 46). Делим профиль зависимости гена-мишени от эффектора на области, в которых профиль близок к линейному. Таким образом, профиль аппроксимируется ломаной. Области не должны пересекаться и в совокупности давать весь отрезок $(0, 1)$. Количество областей равно количеству входных термов.

Строим треугольные термы, вершины которых совпадают со значениями уровня гена-активатора в звеньях ломаной на графике (рис. 46, внизу). Координаты вершин занесены в верхнюю строку таблицы на рис. 46. Пусть всего имеется n термов (в примере $n = 5$). Тогда элементы таблицы в общем виде заданы $[0, B_1, B_2, \dots, B_n, 1]$, где B_i – координата вершины i -го терма. Построение термов проиллюстрировано на рис. (46, вверху). Треугольный терм имеет вершину в точке B_i со значением 1 и вершины в точках B_{i-1} и B_{i+1} со значением 0. Два крайних терма имеют вид трапеции, т.е. в вершинах с координатами 0 и B_1 у первого терма и с координатами B_n и 1 у последнего терма имеют значения, равные единице.

Определяем значения уровней гена-мишени в звеньях ломаной (рис. 46, внизу). Эти значения, являющиеся центрами выходных термов, занесены во вторую строку таблицы на рис. 46. По этим координатам строятся выходные термы, следуя тем же правилам, что и входные.

Генерация термов в виде треугольников или трапеций на основе их центров – в общем случае является непростой задачей, так как необходимо учитывать расстояния между центрами и следить за тем, чтобы центр масс фигуры совпадал с используемым центром.

База правил: После определения термов конструируется база правил. Для этого на каждом линейном участке ломаной, начиная с левого, ставится соответствие между номером входного терма, соответствующего участку, и номером выходного терма, чей центр совпадает с уровнем

экспрессии гена-мишени в правом конце этого участка.

Наши алгоритмы для модели двух эффекторов одной мишени. Здесь мы рассмотрим реализацию нашего подхода на примере активации генной экспрессии для случая, когда на ген-мишень действует не один (как выше), а два эффектора. Например, у нас имеются два фактора-активатора, $A1$ и $A2$, значения каждого из которых находятся в интервале $[0, 1]$, определяющие вместе уровень экспрессии мишени (который тоже определяется на интервале $[0, 1]$). Пусть для простоты уровень фактора $A1$ описывается двумя термами («отсутствует» и «присутствует»), тогда как фактора $A2$ – тремя: «низкий», «средний» или «высокий».

Далее мы определяем нечеткое множество для гена-мишени как три терма для скорости транскрипции: «нулевая», «низкая» и «высокая». Соответственно, система нечетких правил (П1–П6) для нашей модели может быть определена как набор правил общего вида «if-then» («если-тогда»):

П1: ЕСЛИ $x1$ = отсутствует И $x2$ = низкий, ТОГДА y = нулевая;

П2: ЕСЛИ $x1$ = отсутствует И $x2$ = средний, ТОГДА y = нулевая;

П3: ЕСЛИ $x1$ = отсутствует И $x2$ = высокий, ТОГДА y = нулевая;

П4: ЕСЛИ $x1$ = присутствует И $x2$ = низкий, ТОГДА y = нулевая;

П5: ЕСЛИ $x1$ = присутствует И $x2$ = средний, ТОГДА y = медленная;

П6: ЕСЛИ $x1$ = присутствует И $x2$ = высокий, ТОГДА y = быстрая.

Такой набор правил нагляднее представить таблицей (табл. 1).

Рассмотрим в заключение более общий случай, когда оба эффектора имеют по три уровня (рассмотрим на примере активации). Скорости транскрипции теперь характеризуются пятью термами. Поскольку запись базой правил становится слишком громоздкой, то приведем только табличный вид (табл. 2).

Сходные построения используются для случая пары репрессоров, $R1$ и $R2$.

Пример базы правил для трех термов приведен в табл. 3.

Таблица 1. Скорости транскрипции при разных значениях активаторов $A1$ и $A2$ (2×3)

		Уровень фактора $A1$		
		Низкий	Средний	Высокий
Уровень фактора $A2$	Отсутствует	Нулевая	Нулевая	Нулевая
	Присутствует	Нулевая	Медленная	Быстрая

Таблица 2. Скорости транскрипции при разных значениях активаторов A1 и A2 (3 × 3)

		Уровень фактора A1		
		Низкий	Средний	Высокий
Уровень фактора A2	Низкий	Нулевая	Самая медленная	Медленная
	Средний	Самая медленная	Медленная	Быстрая
	Высокий	Медленная	Быстрая	Самая быстрая

Таблица 3. Скорости транскрипции при разных значениях активаторов R1 и R2 (3 × 3)

		Уровень фактора R1		
		Низкий	Средний	Высокий
Уровень фактора R2	Низкий	Самая быстрая	Быстрая	Медленная
	Средний	Быстрая	Медленная	Самая медленная
	Высокий	Медленная	Самая медленная	Нулевая

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Известно, что Vcd имеет тенденцию действовать как активатор кооперативно, вместе с другими (ко)-активаторами и кооперативность диктует свои правила для организации энхансеров. Мы можем назвать это грамматическими правилами [32, 33, 37]. В наиболее изученных случаях можно сформулировать грамматические правила, в частности, для силы, порядка и длины спейсеров для комплекса кластеров сайтов связывания (см. работы [30, 35, 45]).

Энхансеры, управляемые морфогенами. Для лингвистического описания функционирования энхансера было бы разумно определить функциональную организацию таких ЦРМ. Для этого необходимо сформулировать грамматические правила, описывающие энхансер.

Главные правила (на примере энхансеров, управляемых Vcd, как на рис. 1 (ср. с рис. 2):

1) Типичный энхансер, управляемый Vcd, включает несколько сайтов связывания (рассеянных по последовательностям энхансера) для пионерных факторов, таких как Zld или другие кофакторы, способные «открывать» хроматин. В противном случае энхансер будет слишком слаб, чтобы активироваться Vcd.

2) Ядро энхансера состоит из кластера сайтов связывания Vcd. Если два или более соседних сайта расположены слишком близко друг к другу или слишком далеко друг от друга, то они не действуют кооперативно.

3) Энхансер включает в себя несколько сайтов связывания для других активаторов, и эти активаторы имеют тенденцию действовать кооперативно с Vcd.

4) Энхансер включает в себя несколько сайтов связывания для близкодействующих репрессоров, и каждый сайт активации должен находиться в близости (> 100 п.н.) от сайта репрессора. В противном случае энхансер будет работать в областях эмбриона, где он должен быть репрессируван.

Полное детальное моделирование действия энхансера должно в явном виде описывать три иерархических уровня энхансера, управляемого морфогеном, с помощью пионерных факторов (i), кооперативного связывания с матрицей активаторов (ii) и синергии с выходным механизмом транскрипции (iii). В этой статье мы сосредоточимся на нечетком моделировании только уровня ii в сравнении кооперативной активации с кооперативной репрессией.

Нечеткие правила грамматики энхансеров. Как неоднократно отмечалось, в основе одного из наиболее изученных ЦРМ — проксимального энхансера — кластер сайтов связывания фактора Vcd (как иллюстрирует рис. 1a). Более того, функционально близкий (и функционально избыточный) теневой элемент также имеет в своей основе кластер сайтов связывания Vcd (рис. 1b).

Кооперативность активации гена *hb* фактором Vcd вообще и через проксимальный элемент в частности многократно обсуждалась в литературе (см. работу [48] и ссылки в ней). Поэтому у нас есть все основания сравнивать нашу нечеткую модель активации гена *hb* фактором Vcd (через проксимальный элемент) с моделью Хилла для активации.

С целью реализовать НЛМ обсуждаемых в статье энхансеров мы выполняем анализ грамматических правил их конструирования из сайтов свя-

зывания активаторов и репрессоров (см. рис. 1). В итоге мы пришли к таким заключениям.

Во-первых, ядро энхансера состоит из кластера сайтов связывания фактора Bcd. Эти сайты, как правило, разделяются расстояниями, равными или пропорциональными шагу спирали ДНК (10–11 п.н.) [44, 45, 80]]. Это то, что называют периодичностью в распределении сайтов связывания на ДНК энхансера. Именно такое расположение делает возможным кооперативную активацию таких энхансеров [80]. Специфическое связывание с матрицей ДНК первой молекулы Bcd облегчает связывание второй, эти две связанные молекулы облегчают связывание третьей молекулы Bcd и так далее. Такие процессы определяют как кооперативное связывание с матрицей ДНК (template binding cooperativity). Более того, анализируемый нами класс энхансеров включает сайты связывания для некоторых других активаторов, и эти сайты также имеют тенденцию располагаться относительно сайтов Bcd согласно правилу периодичности [45, 80].

Во-вторых, энхансер включает в себя несколько сайтов связывания для «близкодействующих» репрессоров, действующих посредством квенчинга (quenching), и каждый сайт активации должен находиться поблизости (> 100 п.н.) от такого сайта репрессора. Типичный энхансер имеет тенденцию отвечать на связывание репрессора с его сайтом (в этом энхансере) неаддитивно. (Полагают, что при этом задействуется вся молекулярная машина транскрипции, включающая общие и специфические факторы.) Такую неаддитивность (в терминах химической кинетики) описывают как кооперативную или синергетическую репрессию. У нас есть основания (по аналогии с кооперативными активаторами) связать эту кооперативность с кластеризацией сайтов связывания этих ТФ.

Модель проксимального элемента ген *hb*. Два весьма схожих по характеристикам и функциям элемента гена *hb* — проксимальный и теневой энхансеры — относятся к одним из самых изученных, как обсуждалось выше. Принято считать, что первичная функция проксимального элемента — это обеспечивать активацию гена *hb* морфогеном Bcd. Тем не менее обширные исследования до сих пор не принесли ясности и однозначности в сложную картину экспериментальных наблюдений и выводов. Первая проблема заключается в том, что крутизна границы домена гена-мишени (*hb*) во второй половине четырнадцатого цикла слишком велика (см. рис. 2а), чтобы ее можно было объяснить только кооперативным действием Bcd [47, 48, 70–72]. В биохимических экспериментах показано, что кооперативность действия Bcd соответствует коэффициенту Хилла $n_k = 3$ [68, 70], тогда как эмпирические наблюдения по-

казывают, что во второй половине четырнадцатого цикла эмпирическая зависимость экспрессии *hb* от концентрации Bcd достигает $n_k = 12$ (как иллюстрируется графиком на рис. 4а), что не может быть объяснено кооперативностью действия Bcd [48]. В качестве альтернативных гипотез резонно рассматривать (кооперативное) действие других активаторов (см., например, работу [81]). Еще одно популярное альтернативное объяснение — это самоактивация гена *hb* кодируемым им фактором Hb [48, 75, 82].

Наконец, еще одно возможное объяснение — что высокая крутизна склона домена экспрессии *hb* достигается не кооперативной активацией, а является результатом кооперативной репрессии также крутых склонов соседствующих доменов репрессоров (прежде всего K_g и K_{ni}) [47, 83, 84].

Вторая проблема состоит в том, что до сих пор неясно, как взаимодействуют обсуждаемые элементы с еще одним ЦРМ этого гена — энхансером полос (будет рассмотрен в следующем разделе).

Здесь следует отметить, что склон переднего домена экспрессии *hb* является, в свою очередь, вторичным морфогенетическим градиентом репрессивного действия и, как полагают, контролирует позиции границ доменов минимум восьми генов-мишеней [44]. Поэтому исчерпывающие знания и строгие заключения о природе и механизмах контроля крутизны и точности позиции этого склона экспрессии *hb* имеют существенное биологическое значение.

Соответственно, в этом разделе мы рассмотрим средствами НЛМ эти проблемы гена *hb* на примере проксимального элемента.

Кооперативная активация фактором Bcd. Мы начнем наш анализ с простейших нечетких моделей действия единственного активатора на ген-мишень (имея в виду кооперативное действие Bcd на ген *hb*). Мы зададим уровни экспрессии и фактора (активатора *e*) и гена-мишени (*t*) тремя одинаковыми треугольными термами каждый (см. раздел «Наш подход»). Сила активатора (т.е. нелинейность, кооперативность его действия на мишень) заданы тремя возможными вариантами таблицы нечетких правил, как показано на рис. 5 для слабого, умеренного и сильного активатора.

Рис. 2а демонстрирует, как выглядят в динамике профили экспрессии *hb* вместе с профилем Bcd. Сравнение результатов простой нечеткой модели и реальных данных приводит к заключению, что модели слабой и средней (но не сильной) активации ближе всего к экспериментальным данным. Более детальный анализ с большим числом термов и более детальными таблицами правил показывает, что такая модель слишком проста (сверхупрощение) и требуются более детальные нечеткие модели. При этом модель можно развивать просто увеличением числа одинако-

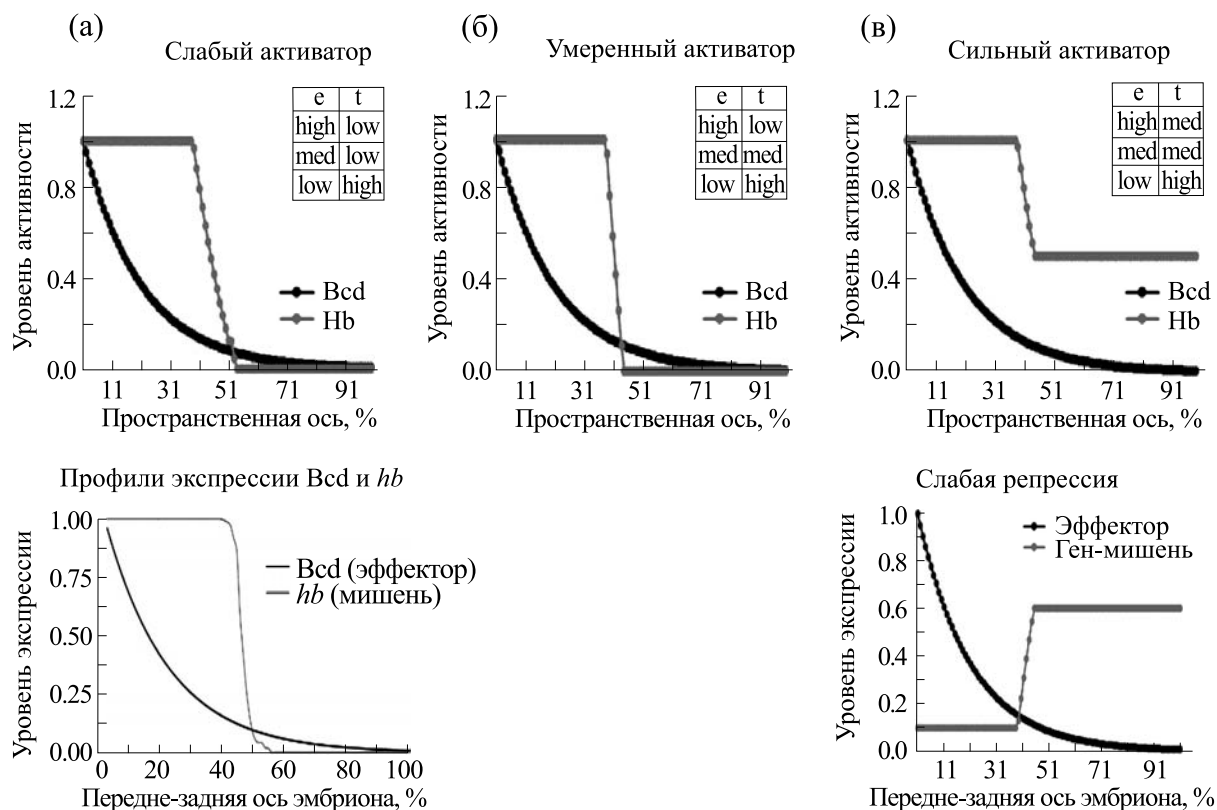


Рис. 5. Зависимость отклика (уровня экспрессии) простейшей нечеткой модели проксимального энхансера гена *hb* (здесь *t* — ген-мишень) от меры кооперативности эффектора-активатора (определяемой таблицей нечетких правил). Фактор-эффектором (*e*) выступает Bcd. Таблицы зависимости *t* от *e* соответствуют слабой (а), умеренной (б) и сильной (в) модели активации (и для эффектора и для мишени задано по три уровня — high, med, low). Модель сопоставляется с экспериментальными данными: на панели (а) внизу приведены эмпирические профили экспрессии фактора Bcd и его мишени гена *hb* для эмбриона дрозофилы на стадии середины четырнадцатого цикла (по материалам веб-ресурса FlyEx [91]). Можно заключить, что эмпирический профиль *hb* схож и с моделью (а), и с моделью (б), но не с моделью (в). На панели (в) внизу приведен вид зависимости экспрессии гена-мишени при альтернативной модели, когда эффектор является не активатором, а репрессором. Это соответствует таблице правил, отличающейся от таковой на (в) тем, что столбец *t* перевернут (сверху вниз: high, med, med).

вых термов и тщательно подобранных больших по размеру таблиц правил. А можно использовать относительно небольшое (пять-шесть) число термов и для эффектора, и для мишени и простую таблицу правил. Но при этом нужно подбирать размеры каждого термина индивидуально (вручную или автоматизированно, как рассматривается в разделе «Наш подход»). Мы ниже воспользуемся последним подходом к НЛМ.

Рассмотрим теперь нечеткое моделирование кооперативного действия активатора на ген-мишень в рамках модели Хилла и с неодинаковыми термами. Это проиллюстрирует принцип развиваемого подхода и позволит сравнить эмпирические профили гена-мишени (*hb*) с результатами нечеткого моделирования такой моделью.

Нечеткое моделирование функции Хилла выполняли, как описано в разделе «Наш подход». Моделирование выполняли для серии величин n_k ($= 1, 2, 3, 4$) и константой «диссоциации» (K_d) по-

добранной так, чтобы склон домена экспрессии находился (примерно) на одном месте при тестируемых значениях n_k . Использовали шесть термов для эффектора и для мишени (см. рис. 6).

Вместе с нечеткой моделью Хилла в координатах «доза—эффект» («активатор—мишень») результаты представлены также в виде профилей в зависимости от кривой активатора, экспоненциальной вдоль пространственной оси (т. е. профиль активатора аппроксимировали отрезком экспоненты). Это позволяло сравнить эмпирические профили паттернов экспрессии гена *hb* вдоль главной (передне-задней) оси раннего эмбриона дрозофилы. На рис. 6 модельные профили сопоставляются с экспериментальным профилем гена *hb* на той самой ранней стадии четырнадцатого цикла, когда, как полагают (см. работу [48]), экспрессия гена-мишени *hb* находится под преимущественным контролем Bcd.

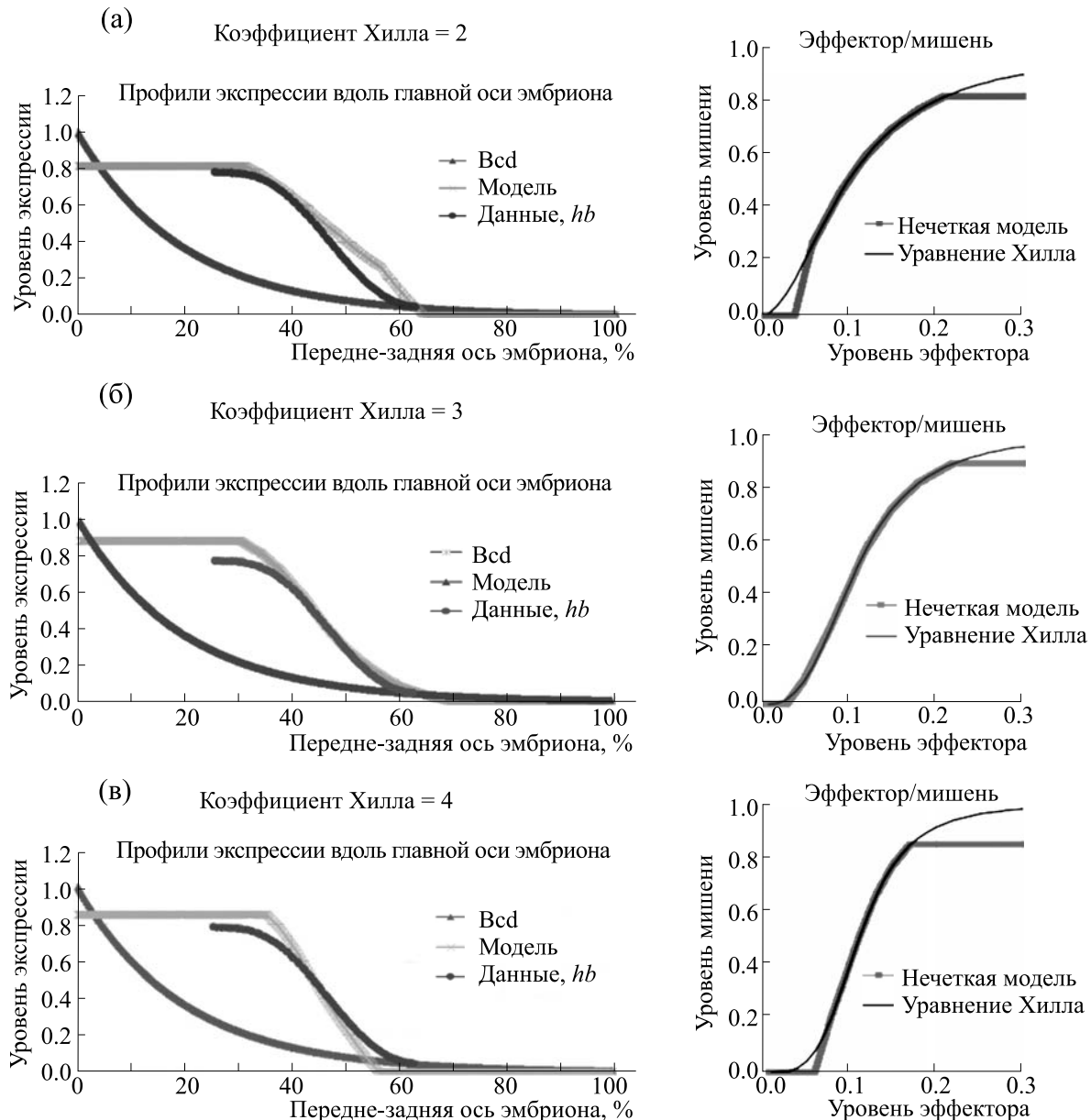


Рис. 6. Нечеткое моделирование действия активатора Bcd на проксимальный элемент гена *hb* демонстрирует, что активация гена *hb* морфогеном Bcd в начале четырнадцатого цикла может быть аппроксимирована законом Хилла при $n_k = 3$. (а)–(в) – Модель профиля экспрессии гена-мишени вдоль передне-задней оси эмбриона под контролем экспоненциального пространственного градиента Bcd при разных уровнях кооперативности действия фактора Bcd (величина коэффициента Хилла $n_k = 2, 3, 4$) и соответствующая кривая зависимости уровня экспрессии гена-мишени от уровня активатора (на врезках справа). Теоретические модели профиля экспрессии гена-мишени сопоставляются с экспериментальным профилем гена *hb* в начале четырнадцатого цикла.

Рис. 6 демонстрирует, что эмпирический профиль экспрессии гена *hb* (самое начало четырнадцатого цикла) хорошо аппроксимируется нечеткой моделью функции Хилла с параметрами $n_k = 3$ и $K_d = 0.0013$ (рис. 6б). На основании наших результатов мы приходим к заключению, что простой гипотезы кооперативного действия ТФ Bcd на проксимальный энхансер достаточно, чтобы объяснить геометрию профиля Hb в начале че-

тырнадцатого цикла. Это сопоставимо с заключениями других авторов [68, 69]. Вместе с тем существенный рост крутизны границы домена *hb* в ходе четырнадцатого цикла требует альтернативных гипотез для своего объяснения.

Кооперативная активация фактором Hb^{mat}. Как отмечалось выше, основная проблема с количественным моделированием динамики профиля

Hb заключается в том, что крутизна границы домена экспрессии *hb* (т.е. именно то, что мы здесь моделируем) очень быстро, в ходе четырнадцатого цикла, нарастает и становится столь крутой, что в рамках модели Хилла это соответствует $n_k = 12$. Столь высокая кооперативность активации сомнительна из общих соображений кинетики и молекулярных механизмов действия активатора Vcd. Тщательные количественные исследования кооперативности кинетики активации кластеров сайтов связывания Vcd в биохимических экспериментах также ставят под сомнение столь высокую кооперативность.

Возможное и широко обсуждаемое объяснение состоит в том, что проксимальный элемент интегрирует активирующее действие не только от Vcd, но и от других ТФ-активаторов (например Hb [81]). Прежде всего, обращает на себя внимание наличие в проксимальном элементе как предполагаемых, так и экспериментально подтвержденных сайтов связывания фактора Hb. Вместе с тем, как хорошо известно, фактор Hb поставляется в яйцеклетку-зиготу материнским организмом еще до активации эмбрионального гена *hb* [85], так что он формирует домен в передней части раннего эмбриона с пологой границей (напоминая более поздний домен эмбрионального Hb). Важность этого материнского домена в предопределении раннего развития эмбриона хорошо известна [86]. Поэтому мы выполнили здесь альтернативное нечеткое моделирование, приняв в качестве активатора материнский Hb^{mat}.

Результаты НЛМ для кооперативной активации фактором Hb^{mat} (с коэффициентами Хилла $n_k = 2, 3, 4$) приведены на рис. 7а. Рисунок наглядно иллюстрирует, что действие пространственно более крутого (чем экспонента фактора Vcd) активатора может объяснить большую крутизну склона домена экспрессии зиготного гена *hb*. Там же приведены реальный профиль экспрессии гена *hb* для середины четырнадцатого цикла. Видно, что кооперативное действие Hb^{mat} может частично объяснить наблюдаемую более высокую крутизну склона *hb* (чем модель для Vcd, см. рис. 7б). (Заметим также, что больший, чем в эмпирических данных, сдвиг границы домена экспрессии *hb* в модели объясним отсутствием деградации в нашей простой НЛМ.) Однако крайне высокая и все нарастающая крутизна склона домена экспрессии *hb* во второй половине четырнадцатого цикла по-прежнему нуждается в дальнейшем исследовании.

Коактивация двумя факторами. В развитие подхода к НЛМ генно-регуляторных элементов необходимо уметь включать в модель действие на данный элемент нескольких регулирующих ТФ. Случай нечеткого действия двух эффикторов на

мишень особенно нагляден, поскольку набор правил представим таблицей (раздел «Наш подход», табл. 1–3).

В общем же совокупное действие может вычисляться из отдельных акций (активации и репрессии) согласно операторам AND, OR или MEAN (см. работу [41]).

В развитие вышеприведенных моделей нам резонно рассмотреть НЛМ с двумя активаторами, Vcd и Hb^{mat}. Vcd мы моделируем двумя или тремя термами (как фактор A2 в табл. 1 и 2), тогда как Hb^{mat} — тремя термами (фактор A1). Результат работы такой НЛМ приведен на рис. 7в. Наше моделирование в рамках табл. 1 и 2 дало сходные результаты.

Таким образом, (неаддитивное) действие пары активаторов (Vcd и Hb^{mat}) более реалистично моделирует активацию гена мишени. Поэтому мы можем предположить, что коактивация этих двух факторов существенна для объяснения крутизны профиля *hb* в более позднем четырнадцатом цикле.

Самоактивация гена. Наконец, рассмотрим последнюю из простых и распространенных альтернативных объяснений динамики склона домена экспрессии *hb*. А именно, исследуем поведение такой НЛМ, в которой исходный, начальный профиль экспрессии Hb (в результате экспрессии гена *hb*), полученный, например, кооперативной активацией фактором Hb^{mat} ($n_k = 3$), на втором шаге моделирования сам становится активатором для экспрессии своего гена, как на рис. 7г. В представленной серии тестов мы считаем механизм самоактивации кооперативным ($n_k = 3$). В результате такой серии шагов самоактивации профиль экспрессии очень быстро становится очень крутым.

Следовательно, в результате тестов с нашими НЛМ мы приходим к заключению, что (кооперативная) самоактивация гена фактором, который этот ген кодирует, — это самое простое объяснение крутизны домена экспрессии гена *hb*. Эти результаты согласуются с заключениями некоторых предыдущих публикаций [48, 75, 82]. Тем не менее проблема еще далека от своего окончательного разрешения, поскольку самоактивация гена *hb* не исследована во всех деталях и у исследователей нет единого мнения по этому вопросу (см. работу [47]).

Возможное альтернативное объяснение существенного нарастания крутизны границы домена — это действие элемента полос, рассмотренное ниже.

Модель элемента полос гена *hb*. Опубликованные результаты анализа активности элемента полос свидетельствуют, что его функция — это ре-

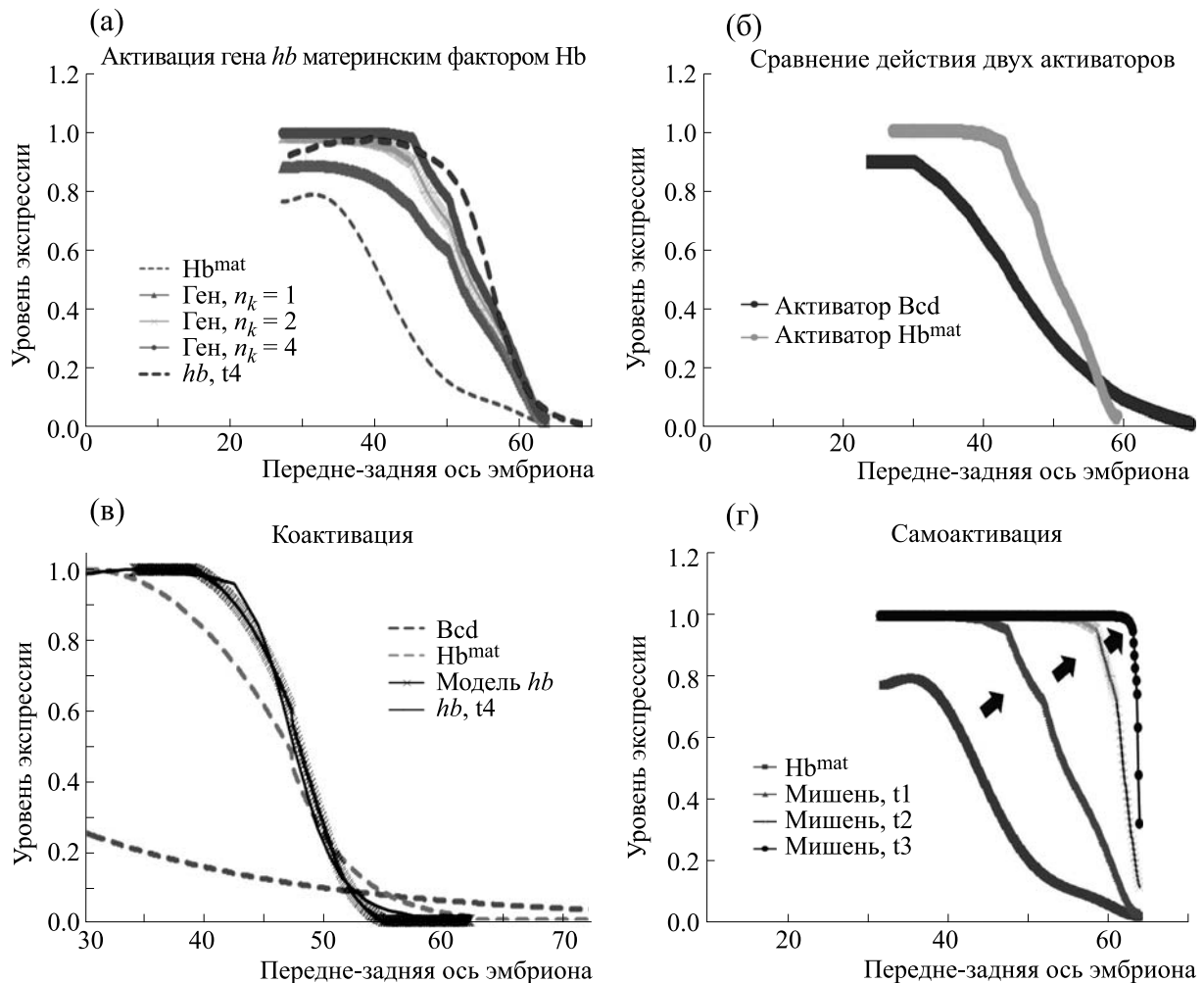


Рис. 7. Возможные более общие случаи механизмов действия проксимального энхансера гена hb . (а) и (б) – Нечетко-логическое моделирование действия активатора Hb^{mat} в рамках модели Хилла с коэффициентами Хилла $n_k = 2, 3, 4$. (а) – Профили экспрессии гена hb как результат действия активатора Hb^{mat} в рамках модели Хилла с $n_k = 2, 3, 4$ в сравнении с эмпирическим профилем экспрессии hb в середине четырнадцатого цикла (t_4). (б) – Сопоставление результата нечеткого моделирования отдельно для активатора Bcd и для активатора Hb^{mat} (при одном и том же значении коэффициента Хилла $n_k = 3$). Видно, что действие материнского Hb способно объяснить большую крутизну склона домена экспрессии гена hb . (в) – Нечеткое моделирование совместного действия пары активаторов (коактивация) Bcd и Hb^{mat} . Также приведен профиль экспрессии в середине четырнадцатого цикла (t_4). Видно, что совместное действие двух активаторов в рамках нечеткой модели дает более крутой склон домена экспрессии гена hb , чем действие каждого из факторов по отдельности. (г) – Динамика профиля экспрессии гена hb для случая нечеткой модели самоактивации. Исходный профиль материнского фактора Hb^{mat} действует как активатор на зиготный ген hb , давая профиль экспрессии hb к моменту t_1 (мишень, t_4). Далее, с момента t_1 этот профиль зиготного фактора Hb действует как активатор (само-активация), давая, в свою очередь, профиль экспрессии hb к моменту t_2 (мишень, t_2). Наконец, с момента t_2 этот профиль зиготного фактора Hb действует как активатор, давая в итоге очень крутой профиль экспрессии hb к моменту t_3 (мишень, t_3).

прессирование гена-мишени на всем протяжении главной оси раннего эмбриона, за исключением двух небольших областей, как на рис. 26 [47]. Это и дает в итоге два четких пика экспрессии.

Мы начнем наш анализ кооперативной репрессии с общих результатов, аналогично таковым, описанным в параграфе «Кооперативная

активация фактором Bcd » и на рис. 5. Рис. 5в(1) иллюстрирует то общее соображение, что для крутых склонов доменов экспрессии требуется как активация, так и репрессия по неаддитивному (кооперативному) механизму.

Продолжим изучение средствами НЛМ кооперативного действия репрессоров на ген-мишень в рамках модели Хилла с достаточным числом тер-

мов и разными уровнями кооперативности (с различными значениями коэффициента n_k). Мы сосредоточимся на анализе поведения домена PS4 (антериорный пик, Hb^{ant} , см рис. 26). Этот домен представляет собой четкий и крутой пик локальной экспрессии. Его задний склон, как полагают, находится под преимущественным контролем «S-образного» профиля фактора Kг (а также фактора Kni) [50, 87]. Передний склон домена PS4, как полагают, находится под преимущественным контролем профиля фактора-репрессора Gt [88]. Есть также экспериментальные наблюдения, что этот склон находится под контролем собственно фактора Hb, кодируемого геном hb [47], что труднее представить с точки зрения кинетики.

Сначала рассмотрим результаты моделирования кооперативности репрессии гена-мишени репрессором Kг, исходя из простейшей гипотезы, что именно профиль Kг отвечает за геометрию профиля пика экспрессии hb (его постериорный склон), как иллюстрируется на рис. 8а–в. Моделирование выполняли для ряда величин n_k ($= -1, -2, -3, -4$) и константой «диссоциации» подобранной так, чтобы склон домена экспрессии находился (примерно) на одном месте при тестируемых значениях n_k . Профили репрессоров для ранней экспрессии генов сегментации естественно аппроксимировали отрезком S-образной (сигмоидной) функции (см. рис. 2). Использовано пять или шесть термов для эффектора и для мишени (рис. 8).

Результаты численного решения уравнения Хилла (относительно участка сигмоидного профиля репрессора) сопоставляются с результатами нечеткого моделирования и с эмпирическим профилем экспрессии гена hb (на уровне мРНК) в середине четырнадцатого цикла, как представлено на рис. 8а–в. Рис. 8 демонстрирует, что эмпирический профиль экспрессии hb примерно соответствует решению Хилла со значением n_k между -3 и -4 .

Найденная нами существенная кооперативность действия Kг на hb находится в некотором несогласии с известным количеством и распределением экспериментально подтвержденных сайтов связывания для Kг на энхансере полос (пара в одной половине энхансера и всего один — в другой, как на рис. 1в). В этой связи можно предложить, что не все сайты для Kг найдены. Здесь существенно, что биоинформационный анализ демонстрирует наличие большего числа сайтов для Kг, расположенных плотными кластерами, как в случае теневого элемента (см. рис. 1б). Альтернативная модель кооперативной корепрессии рассмотрена в следующем подразделе.

Далее, аналогичным образом, антериорный склон домена PS4, как полагают, находится под

преимущественным контролем «S-образного» профиля фактора-репрессора Gt [88]. Результаты нашего нечеткого моделирования контроля эффектором Gt переднего склона пика PS4 в сопоставлении с профилем модели Хилла с $n_k = -3, -4$ демонстрируют, что эмпирическая кривая весьма сходна с результатами нечеткого моделирования (не показано). Отметим здесь, что знак действия Gt на ген hb (активатор или репрессор) все еще обсуждается [64]. В качестве альтернативы некоторые авторы рассматривают механизм саморепрессии hb через энхансер полос [47]. Поэтому вопрос нуждается в дальнейшем экспериментальном и теоретическом анализе.

Корепрессия двумя факторами. Альтернативная гипотеза состоит в том, что сайты связывания Kг оказывают репрессивное действие кооперативно с сайтами связывания другого репрессора — Kni, которых охарактеризовано тоже три, так что получается плотный маленький кластер из трех сайтов связывания в одной половине энхансера и трех — в другой (рис. 1в). В целом этот вопрос требует дальнейших исследований.

С целью проанализировать возможности кооперативной репрессии двумя факторами (Kг и Kni) мы выполнили соответствующее нечеткое моделирование, как объяснено в разделе «Наш подход», табл. 3. Результаты приведены на рис. 9. Видно, что совместная кооперативность двух репрессоров объясняет крутизну профиля домена. Так что такая модель лучше подходит для интерпретации грамматики элемента полос.

На основании наших результатов мы приходим к заключению, что простой гипотезы кооперативного действия эффекторов-репрессоров на энхансер полос достаточно, чтобы объяснить геометрию склонов антериорного (PS4) пика hb в четырнадцатом цикле. При этом каждый склон пика PS4 независимо контролируется своим репрессором (или репрессорами). Примечательно, что кооперативность репрессии достаточно высока и сравнима с кооперативностью активации этого гена в том же четырнадцатом цикле. Мы не знаем публикаций, где бы обсуждалась кооперативность репрессии для генов сегментации.

Аддитивность действия регуляторных элементов данного гена. В заключение отметим, что при анализе генной регуляции для распространенных случаев нескольких (автономных и полуавтономных) регуляторных элементов возникает общая проблема: как интегрируется действие этих элементов, если они активны одновременно (и в данном клеточном массиве).

Недавние исследования, проведенные, в частности, и на рассматриваемых здесь регуляторных элементах, показали целый спектр суммирования действия двух элементов одного гена: супераддитивность, аддитивность и субаддитивность [89].



Рис. 8. Репрессия гена hb фактором Kt через элемент полос подчиняется закону Хилла. (а) — Крутой склон домена экспрессии репрессора Kt формирует постериорную границу пика PS4 (Hb_{ant}). (б)–(г) — Модель профиля экспрессии гена-мишени вдоль участка передне-задней оси эмбриона под контролем «S-образного» профиля репрессора Kt при трех разных уровнях кооперативности (величина коэффициента Хилла, $n_k = -2, -3, -4$) и соответствующая кривая зависимости уровня экспрессии гена-мишени hb от уровня репрессора Kt (на врезках справа). Теоретические модели профиля экспрессии гена-мишени сопоставляются с экспериментальным профилем гена hb в середине четырнадцатого цикла.

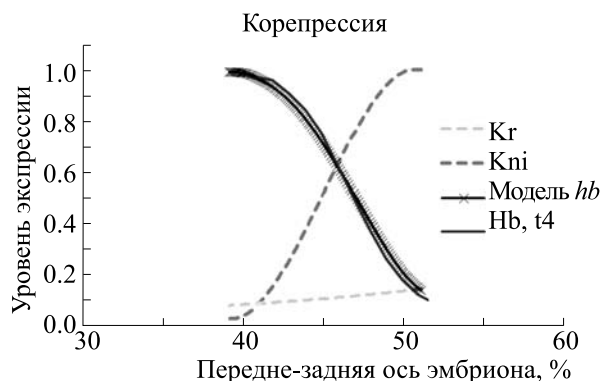


Рис. 9. Нечеткое моделирование действия пары репрессоров ($K\bar{g}$ и Kni , корепрессия) через элемент полос на крутизну постериорного склона пика PS4. Совместное действие двух репрессоров в рамках нечеткой модели дает более крутой склон домена экспрессии гена *hb*, чем действие каждого из факторов по отдельности (сравни с рис. 7в).

Исходя из простой гипотезы аддитивности действия элемента полос и проксимального элемента, мы наблюдаем, что суммарная картина экспрессии гена-мишени *hb* хорошо аппроксимирует эмпирические профили мРНК *hb* к середине четырнадцатого цикла, как показано на рис. 10.

Полученные здесь результаты являются существенным дополнением к нашим результатам детального детерминистического моделирования экспрессии гена *hb* (системой детальных уравнений в частных производных) [51], где акцент был сделан на наличии двух промоторов у этого гена. Для получения исчерпывающей картины деталей динамики этого гена требуется свести воедино оба подхода.

Сходно организованные гены. Тщательный анализ результатов исследования поведения ЦРМ генов сегментации эмбриона дрозофилы показывает, что схема динамики паттерна, анализируемая здесь на примере *hb*, может быть характерна для целого ряда генов. А именно, имеются ЦРМ, которые контролируют формирование достаточно обширных доменов экспрессии (аналогично проксимальному элементу гена *hb*). И имеются ЦРМ, которые контролируют формирование более коротких острых и четких пиков экспрессии в пределах начального широкого домена, как это имеет место в случае элемента полос гена *hb*. Прежде всего это относится к таким генам семейства *gap* (к которым относится и *hb*), как *gt*, $K\bar{g}$ и *tailless*. Эти ЦРМ были исследованы в ряде публикаций [55–57, 90]. Их функциональная близость к *hb* делает такие наблюдения по деталям их регуляции (сходство с *hb*), в достаточной мере ожидаемым. Все эти факты, вместе взятые, свидетельствуют о том, что моделируемые в этой статье детали динамики экспрессии под контролем пар ЦРМ с весьма различающейся стратегией контроля паттерна могут быть общим принципом ре-

гуляции генов, вовлеченных в процессы сегментации.

ВЫВОДЫ

Нами впервые нечеткое моделирование применяется для разработки моделей активности цис-регуляторных модулей и впервые применяется к проблемам экспрессии генов биохимической разметки ранних эмбрионов. Подход представлен в приложении к анализу конкретной проблемы перехода от самых ранних механизмов паттерна эмбриональной сегментации к более поздним. По такому сценарию охарактеризована экспрессия целого ряда генов эмбрионального паттерна, но наиболее изучена она для гена *hb*. Сценарий заключается в том, что самый ранний паттерн формируется как активация гена-мишени (например, ранний *hb*) морфогенетическими градиентами (прежде всего, *Vcd*) в форме широких доменов экспрессии. На этом этапе действуют свои, ранние механизмы контроля границы экспрессии (типично — кооперативная активация гена-мишени морфогеном). Примечательно, что эти процессы находятся под контролем соответствующих ранних полуавтономных регуляторных элементов (например, проксимальный элемент *hb*). На следующем этапе (на фоне этих широких доменов) быстро формируются четкие и узкие полосы (пики) экспрессии. Они контролируются своими регуляторными механизмами (например, контролем соседствующими доменами факторов-репрессоров). Эти процессы регулируются своими полуавтономными регуляторными элементами (например, элемент полос гена *hb*). Результирующий паттерн экспрессии анализируемых генов высоко динамичен и определяется наложением двух процессов контроля паттерна экспрессии.

С помощью подхода НЛМ проанализирована регуляция проксимальным энхансером крутизны

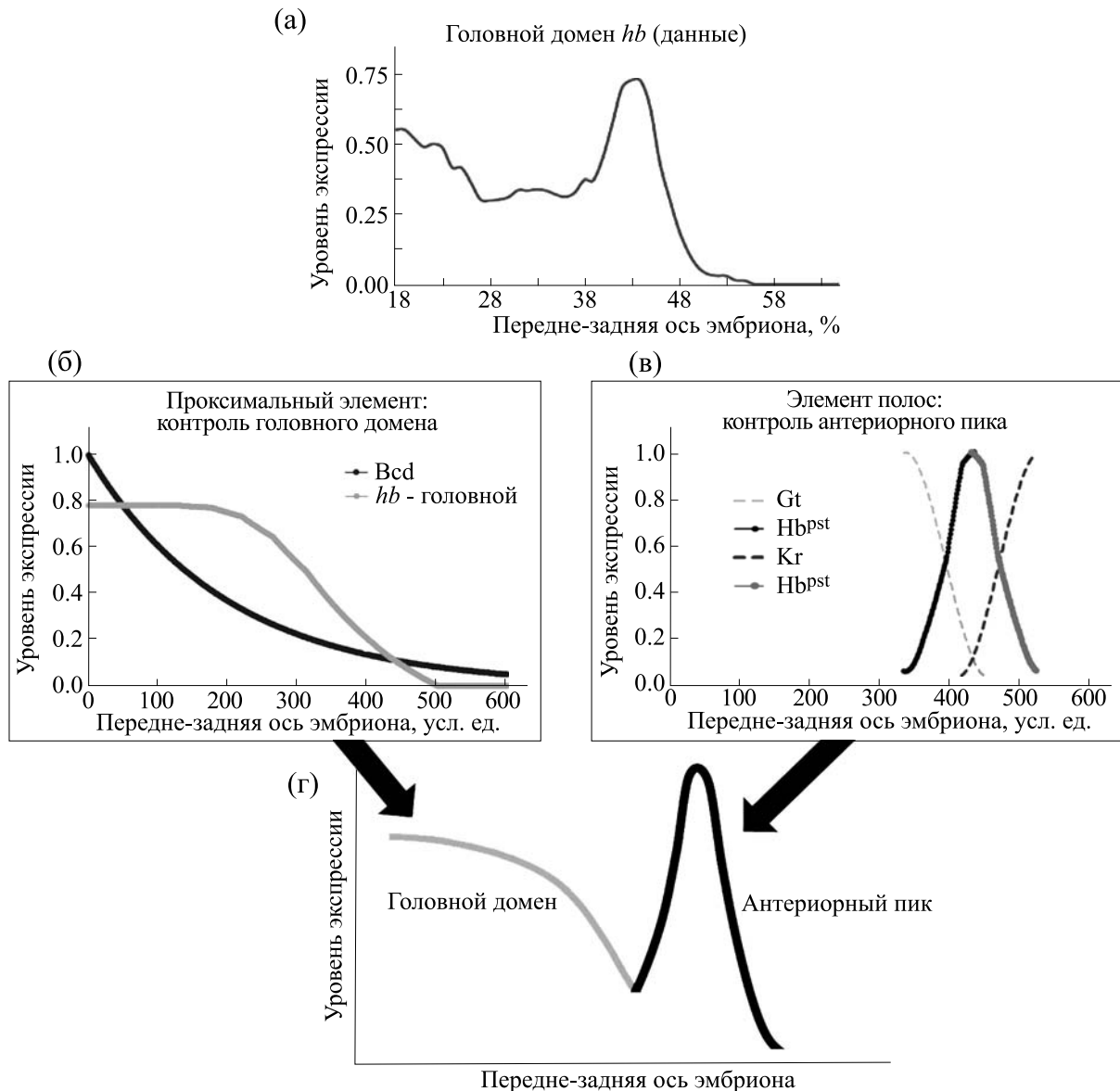


Рис. 10. Аддитивность действия двух основных регуляторных элементов гена *hb* в четырнадцатом цикле. Сумма профилей нечетких моделей действия проксимального элемента и элемента полос хорошо аппроксимируют эмпирический профиль экспрессии гена *hb*. (а) — Экспериментально наблюдаемый профиль экспрессии гена *hb* в начале четырнадцатого цикла (на уровне мРНК; график получен исходя из данных, доступных на веб-ресурсе BID BDTNP [92]). (б) — Профиль головного домена экспрессии *hb* как результат кооперативной ко-активации Bcd и другими активаторами через проксимальный элемент. (в) — Антериорный пик экспрессии *hb^{ant}* (домен PS4) как результат кооперативной корепрессии пар (или большего числа) репрессоров (независимо для переднего и заднего склона пика *hb^{ant}*). (г) — Суммарный профиль головной экспрессии, полученный нечетким моделированием и хорошо аппроксимирующий экспериментальный профиль (а).

постериорного склона переднего домена экспрессии *hb*. Показано, что рост этой крутизны может быть объяснен кооперативной коактивацией двух активаторов (Bcd и Hb^{mat}), а также самоактивацией фактором Hb его гена. Это соответствует количеству, силе и размещению сайтов для ТФ этого энхансера (грамматика ЦРМ) и, по крайней мере отчасти, объясняет парадоксально

высокую крутизну исследуемого склона, необъяснимую только кооперативной активацией морфогеном Bcd.

Подходом НЛМ также проанализирована регуляция энхансером крутизны постериорного склона anteriорного пика экспрессии гена *hb* (домен PS4). Показано, что крутизна этого склона может быть объяснена (кооперативной) коре-

прессией минимум двух факторов (K_г и K_{ни}). Это соответствует грамматике энхансера полос.

Следует подчеркнуть, что кооперативные механизмы действия ЦРМ абсолютно необходимы в становлении эмбрионального паттерна для получения резкого ответа по типу «вкл/выкл» при формировании резкой границы экспрессии вдоль морфогенетических градиентов.

Мы уверены, что нечеткое лингвистическое моделирование поведения ГРС способно естественным образом охарактеризовать очерченные признаки типичных энхансеров. Таким образом, мы можем включить все уровни синергии энхансеров: кооперативное действие пионерного фактора (типа Zld), кооперативное связывание ТФ с матрицей ДНК и синергические взаимодействия связанных ТФ с машиной транскрипции. Мы убеждены, что модели активности генных сетей, основанные на натуральных языках и нечеткой логике, имеют внутреннюю функциональную организацию, более близкую к молекулярной машинерии генно-регуляторных механизмов, чем модели в парадигме математической физики. А это должно давать нечетким моделям большую предсказательную силу.

ИСТОЧНИКИ ФИНАНСИРОВАНИЯ

Результаты раздела «Наш подход» были получены в рамках государственного задания Министерства образования и науки Российской Федерации (№ 16.8549.2017 / 8.9). Остальные результаты получены в рамках гранта Российского научного фонда (проект № 17-18-01536).

КОНФЛИКТ ИНТЕРЕСОВ

Автор заявляет об отсутствии конфликта интересов.

СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания исследований с использованием людей и животных в качестве объектов.

СПИСОК ЛИТЕРАТУРЫ

1. J. Linde, S. Schulze, S. G. Henkel, and R. Guthke, *EXCLI J.* **14**, 346 (2015).
2. H. de Jong, *J. Comput. Biol.* **9** (1) 67 (2002).
3. J. Jaeger, Manu, and J. Reinitz, *Curr. Opin. Genet. Dev.* **22**, 533 (2012).
4. M. A. H. Samee, B. Lim, N. Samper, et al., *Cell Systems* **1** (6) 396 (2015).
5. A. Spirov, and D. Holloway, in *Evolutionary Algorithms in Gene Regulatory Network Research*, Ed. by N. No-
6. man, and H. Iba (Wiley Interscience, 2015), pp. 240–268.
7. L. Sanchez and D. Thieffry, *J. Theor. Biol.* **211** (2), 115 (2001).
8. D. Thieffry and L. Sanchez, *Curr. Opin. Genet. Dev.* **13** (4), 326 (2003).
9. K. Kozlov, V. Gursky, I. Kulakovskiy, and M. Samsonova, *BMC Genomics* **15** (12), S6 (2014).
10. K. Kozlov, V. V. Gursky, I. V. Kulakovskiy, et al., *BMC Genomics* **16** (13), S7 (2015).
11. V. V. Gursky, K. N. Kozlov, I. V. Kulakovskiy, et al. *PLoS One* **12** (9), e0184657 (2017).
12. M. A. Gibson and E. Mjolsness, in *Computational Modeling of Genetic and Biochemical Networks*, Ed. by J. M. Bower and H. Bolouri (MIT Press, Cambridge, MA, 2001), pp. 1–48.
13. E. Myasnikova and A. Spirov, *J. Bioinform. Comput. Biol.* **16** (2), 1840008 (2018).
14. K. Kozlov, S. Surkova, E. Myasnikova, et al., *PLoS Comput. Biol.* **8** (8), e1002635 (2012).
15. С. А. Андреев, М. Г. Самсонова и В. В. Гурский, *Биофизика* **60** (2), 225 (2015).
16. H. Janssens, S. Hou, J. Jaeger, et al., *Nat Genet.* **38** (10), 1159 (2006).
17. N. E. Buchler, U. Gerland, and T. Hwa, *Proc. Natl. Acad. Sci. USA* **100**, 5136 (2003).
18. F. Sauer, S. K. Hansen, and R. Tjian, *Science* **270**, 1825 (1995).
19. F. Sauer, D. A. Wassarman, G. M. Rubin, and R. Tjian, *Cell* **87**, 1271 (1996).
20. K. Gupta, D. Sari-Ak, M. Haffke, et al., *J. Mol. Biol.* **428** (12), 2581 (2016).
21. E. Nogales, R. K. Louder, and Y. He, *Annu. Rev. Biophys.* **46**, 59 (2017).
22. E. H. Mamdani and S. Assilian, *Int. J. Man-Machine Studies* **7** (1), 1 (1975).
23. M. Sugeno, *Industrial applications of fuzzy control* (Elsevier Science Pub. Co, 1985).
24. S. Istrail and E. H. Davidson, *Proc. Natl. Acad. Sci. USA* **102**, 4954 (2005).
25. R. D. Zeigler, J. Gertz, and B. A. Cohen, *BMC Bioinformatics* **8**, 272 (2007).
26. C.-P. Lee, Y. Leu, W.-N. Yang, *Applied Soft Computing* **12** (3), 1115 (2012).
27. J. Bordon, M. Moskon, N. Zimic, and M. Miha, *IEEE/ACM Trans. Comput. Biol. Bioinform.* **12** (5), 1199 (2015).
28. B. H. Wang, J. W. Li, and J. S. Lim, *Genetics Mol. Res.* **15** (3), gmr.15039002 (2016).
29. J. Liu, Y. Chi, C. Zhu, and Y. Jin, *BMC Bioinformatics* **18** (1), 241 (2017).
30. K. Wu and J. Liu, *IEEE Transactions on Fuzzy Systems* **25** (6) 1546 (2017).
31. S. Ben-Tabou de-Leon and E. H. Davidson, *Dev Biol.* **325**, 317 (2009).

31. B. Yan, D. Guan, C. Wang, et al., *Nat. Commun.* **8** (1), 1044 (2017).
32. S. Rastegar, I. Hess, T. Dickmeis, et al., *Dev Biol.* **318**, 366 (2008).
33. K. J. Won, A. Sandelin, T. T. Marstrand, and A. Krogh, *Bioinformatics* **24**, 1669 (2008).
34. J. Gertz, E. D. Siggia, and B. A. Cohen, *Nature* **457**, 215 (2009).
35. L. Li and Z. Wunderlich, *Front. Genet.* **8**, 63 (2017).
36. J. O. Yáñez-Cuna, E. Z. Kvon, and A. Stark, *Trends Genet.* **29**, 11 (2013).
37. J. Grice, B. Noyvert, L. Doglio, and G. Elgar, *PloS One* **10** (7), e0130413 (2015).
38. С. Т. Золян и Р. И. Жданов, *Философия науки и техники* **23** (1), 88 (2018).
39. S. T. Zolyan, and R. I. Zhdanov, *Semiotica* 2018 (225), 1 (2018). DOI: 10.1515/sem-2016-0214
40. X. Li, Y. Li, Y. Liu, and L. Wang, *Int. J. Pattern Recognit. Artif. Intell.* **31** (10), 1 (2017).
41. R. Küffner, T. Petri, L. Windhager, and R. Zimmer *PloS One* **5** (9), e12807 (2010).
42. D. Marbach, T. Schaffter, C. Mattiussi, and D. Floreano, *J. Comput. Biol.* **16** (2), 229 (2009).
43. D.M. Holloway, J. Reinitz, A. Spirov, and C. E. Vanario-Alonso, *Trends Genet.* **18**, 385 (2002).
44. D. Yu and S. Small, *Curr. Biol.* **18**, 868 (2008).
45. D. Papatsenko, Y. Goltsev, and M. Levine, *Nucl. Acids Res.* **37** (17), 5665 (2009).
46. M. W. Perry, A. N. Boettiger, and M. Levine, *Proc Nat Acad Sci USA* **108**, 13570 (2011).
47. M. W. Perry, J. P. Bothma, R. D. Luu, and M. Levine, *Curr. Biol.* **22**, 2247 (2012).
48. F. J. P. Lopes, A. V. Spirov, and P. M. Bisch, *Dev. Biol.* **370** (2), 165 (2012).
49. E. A. Zagrijchuck, M. A. Sabirov, D. M. Holloway, and A. V. Spirov, *J. Bioinform. Comput. Biol.* **12**, 1441009 (2014).
50. D. M. Holloway and A. V. Spirov, *PloS One* **10** (3), e0118450 (2015).
51. A. V. Spirov, E. M. Myasnikova, and D. M. Holloway, *J. Bioinform. Comput. Biol.* **14** (2), 1 (2016).
52. T. Takagi and M. Sugeno, *IEEE Trans. System Man Cybernet.* **15**, 116 (1985).
53. M. Storace and O. De Feo, *IEEE Trans. Circuits and Systems I: Regular Papers* **51** (4), 830 (2004).
54. Y. Park, K. M. Shaw, H. J. Chiel, and P. J. Thomas, *Eur. J. Appl. Math.* **29** (5), 905 (2016).
55. M. D. Schroeder, M. Pearce, J. Fak, et al., *PLoS Biol.* **2**, E271 (2004).
56. E. Segal, T. Raveh-Sadka, M. Schroeder, et al., *Nature* **451** (7178), 535 (2008).
57. M. Kazemian, C. Blatti, A. Richards, et al., *PloS Biol.* **8** (8), e1000456 (2010).
58. M. J. Bronski, C. C. Martinez, H. A. Weld, and M. B. Eisen, *Genes, Genomes, Genetics* **10**, g3.400959 (2019). DOI: 10.1534/g3.119.400959
59. D. Lebrecht, M. Foehr, E. Smith, et al., *Proc. Natl. Acad. Sci. USA* **102** (37), 13176 (2005).
60. Z. Xu, H. Chen, J. Ling, et al., *Genes & Development* **28** (6), 608 (2014).
61. S. M. Foo, Y. Sun, B. Lim, et al., *Curr. Biol.* **24**, 1341 (2014).
62. C. P. Verrijzer, and R. Tjian, *Trends Biochem. Sci.* **21**, 338 (1996).
63. M. V. Staller, B. J. Vincent, M. D. J. Bragdon, et al., *Proc. Natl. Acad. Sci. USA* **112** (3), 785 (2015).
64. X.-Y. Li, and M. B. Eisen, *BioRxiv* 2018-07-3010.1101/3808571 (2018).
65. A. V. Spirov, T. Bowler, and J. Reinitz, *Nucl. Acids Res.* **28**, 337 (2000).
66. A. V. Spirov, M. Borovsky, and O.A. Spirova, *Nucl. Acids Res.* **30**, 351 (2002).
67. W. Driever, G. Thoma, and C. Nüsslein-Volhard, *Nature* **340**, 363 (1989).
68. X. G. Ma, D. Yuan, K. Diepold, et al., *Development* **122**, 1195 (1996).
69. D. S. Burz, R. Rivera-Pomar, H. Jäckle, and S. D. Hanes, *EMBO J.* **17**, 5998 (1998).
70. T. Gregor, D.W. Tank, E. W. Wieschaus, and W. Bialek, *Cell* **130**, 153 (2007).
71. F. He, Y. Wen, D. Cheung, et al., *BMC Dev. Biol.* **10**, 80 (2010).
72. F. He, J. Ren, W. Wang, J. Ma, *PLoS One* **6** (4), e19122 (2011).
73. D. M. Holloway, F. J. P. Lopes, L. da Fontoura Costa, et al., *PLoS Comput. Biol.* **7** (2), e1001069 (2011).
74. G. Struhl, K. Struhl, and P.M. Macdonald, *Cell* **57** (7), 1259 (1989).
75. F. J. P. Lopes, F. M. C. Vieira, D. M. Holloway, et al., *PLoS Comput. Biol.* **4** (9), e1000184 (2008).
76. J. S. Margolis, M. L. Borowsky, E. Steingrímsson, et al., *Development* **121** (9), 3067 (1995).
77. J. M. Mendel, *Uncertain Rule-Based Fuzzy Logic Systems: Introduction and New Directions* (Prentice-Hall, Upper Saddle River, NJ, 2001).
78. Ю. И. Митюшин, Б. И. Мокин и А. П. Ротштейн, *Soft Computing: идентификация закономерностей нечеткими базами знаний* (Універсум-Вінниця, 2002).
79. L. Windhager, Ph.D. Dis. (der Fakultät für Mathematik, Informatik und Statistik, der Ludwig-Maximilians-Universität, München, 2013).
80. V. J. Makeev, A. P. Lifanov, A. G. Nazina, and D. A. Papatsenko, *Nucl. Acids Res.* **31**, 6016 (2003).
81. M. Simpson-Brose, J. Treisman, and C. Desplan, *Cell* **78** (5), 855 (1994).
82. E. M. Myasnikova, and A.V. Spirov, *BioSystems* **166**, 50 (2018).
83. D. E. Clyde, M. S. Corado, X. Wu, et al., *Nature* **426** (6968), 849 (2003).
84. X. Y. Li, S. MacArthur, R. Bourgon, et al., *PLoS Biol.* **6** (2), e27 (2008).
85. J. S. Margolis, M. Borowsky, C. W. Shim, and J. W. Posakony, *Dev. Biol.* **163**, 381 (1994).

86. G. Struhl, P. Johnston, and P. A. Lawrence, *Cell* **69** (2), 237 (1992).
87. J. Jaeger, *Cell. Mol. Life Sci.* **68** (2), 243 (2011).
88. R. Kraut and M. Levine, *Development* **111** (2), 611 (1991).
89. J. P. Bothma, H. G. Garcia, S. Ng, et al., *Elife* **4**, e07956 (2015).
90. E. Z. Kvon, T. Kazmar, G. Stampfel, et al., *Nature* **512** (7512), 91 (2014).
91. A. Pisarev, E. Poustelnikova, M. Samsonova, and J. Reinitz, *Nucl. Acids Res.* **37** (Database issue), D560-6 (2009).
92. A. C. Fowlkes, C. L. Luengo Hendriks, S. V. E. Keränen, et al., *Cell* **133**, 364 (2008).

Fuzzy Linguistic Modelling of Regulation of *Drosophila* Segmentation Genes

A.A. Makashov*, E.M. Myasnikova*, and A.V. Spirov**, ***

*Peter the Great St. Petersburg Polytechnic University,
ul. Polytechnicheskaya 29, St. Petersburg, 195251 Russia

**Sechenov Institute of Evolutionary Physiology and Biochemistry, Russian Academy of Sciences,
prosp. Toreza 44, St. Petersburg, 194223 Russia

*** Institute of Scientific Information for Social Sciences, Russian Academy of Sciences,
ul. Krzhizhanovskogo 15/2, Moscow, 117997 Russia

Understanding molecular mechanisms involved in gene regulation is one of the underlying issues in systems biology. Often, common approaches used to model gene regulation fail to distinguish between the model properties that adequately reflect the characteristics of a biological object and those associated with the nature of the model. This study concerns the development of a logical approach for modeling the regulation of gene expression thereby enabling further maximum use of all the available empirical data and minimum use of a priori hypotheses. Moreover, the rules for gene regulation are formulated in the form of linguistic modelling, namely, restated in the language of approach developers instead of the language of researchers. We use a fuzzy approach known as fuzzy linguistic modelling which reflects reality more accurately. In our study, for the first time, we used linguistic modelling for modeling the activity of the cis-regulatory modules and applied it to the problems of analysis of gene expression for biochemical labeling of embryos. This approach made it possible to clarify the essential details of the mechanisms of “reading out” primary morphogen gradients by two or more cis-regulatory modules of the same gene in combination.

Keywords: gene regulation, expression patterns, enhancers, enhancer grammar, logical modelling, fuzzy modelling