

УТВЕРЖДАЮ

И.о. генерального директора ФБУН

Государственный научный центр

вирусологии и биотехнологии «Вектор»

В.Н.. Михеев

10 ноября 2014 г.



ОТЗЫВ

ведущей организации на диссертационную работу Медведевой Ирины Вадимовны «Компьютерный анализ закономерностей кодирования функциональных сайтов белков в генах позвоночных», представленную на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – «математическая биология, биоинформатика»

Для решения многих фундаментальных и прикладных задач биологии требуется знание пространственной структуры белков. В частности, знание пространственной организации функциональных сайтов белков и их комплексов необходимо для рационального дизайна вакцинных конструкций, антигенов для тест-систем, лекарственных препаратов, а также позволяет понимать молекулярные механизмы биологических процессов. Информация о белковой структуре помогает интерпретировать результаты экспериментов и планировать дальнейшие исследования. В настоящее время разработано множество методов, осуществляющих предсказание особенностей организации сайтов связывания различных лигандов. Полученные с использованием этих методов знания позволяют предсказывать биологические свойства белков, аннотировать последовательности вновь секвенированных генов.

В связи с вышесказанным **актуальность** диссертационной работы Медведевой И.В. не вызывает сомнений, так как она посвящена изучению особенностей кодирования функциональных сайтов белков в последовательностях генов позвоночных.

Задачи, поставленные в работе, включают создание компьютерной системы, осуществляющей поиск позиций аминокислотных остатков функциональных

сайтов белков в кодирующей структуре генов, а также всесторонний анализ особенностей кодирования функциональных сайтов белков. Особое внимание в работе уделено анализу экзонов, последовательности которых кодируют аминокислотные остатки функциональных сайтов, в том числе анализу длины и расположению экзонов, а также исследованию частотных характеристик их кодонов.

Полученные автором научные результаты, безусловно, являются **новыми**. Автором впервые реализована интеграция информации о функциональных сайтах белков, их пространственных структурах и кодирующих последовательностях в рамках интегрированной базы данных SitEx. Доказано, что функциональные сайты статистически чаще кодируются более длинными, а также близлежащими экзонами. Показано, что это может быть связано с тем, что при кодировании функционального сайта на границе между экзонами чаще встречаются ненулевые фазы кодонов, что, вероятно, свидетельствует о тенденции таких экзонов к укрупнению.

Практическая ценность работы состоит в том, что созданная на основе БД SitEx компьютерная система позволяет идентифицировать позиции нуклеотидов в последовательностях генов, которые кодируют аминокислотные остатки функциональных сайтов, содержащихся в известных структурах белка, а также находить потенциальные позиции функциональных сайтов в гомологичных последовательностях и структурно сходных участках других белков. Эти возможности могут быть использованы для аннотации вновь секвенированных генов, поиска фармакологических мишеней, а также для проведения филогенетического анализа.

Содержание диссертации Медведевой И.В. изложено на 108 страницах машинописного текста и традиционно состоит из введения, обзора литературы (глава I), описания исходных данных и методов, использованных при разработке компьютерной системы SitEx (глава II), результатов исследований (глава III), обсуждения, выводов, списка литературы, включающего 130 работ, и пяти приложений. Диссертация содержит 11 таблиц и 28 рисунков.

Во введении обоснована актуальность проблемы и сформулирована цель настоящего исследования и конкретные задачи.

В главе I «Обзор литературы» подробно рассмотрены особенности структуры, функции и эволюции белков. Значительное внимание уделено современным методам проекции пространственной структуры белка на кодирующую структуру гена. Изложены современные представления о значении фазы экзонов для понимания эволюции экзонной структуры гена.

В главе II «Компьютерная система SitEx» подробно описаны используемые в данной работе геномные и белковые базы данных и их форматы, а также методы выравнивания последовательностей. Описан алгоритм и реализация разработки собственной базы данных, а также инструменты поиска потенциальных позиций функциональных сайтов в гомологичных последовательностях и аналогичных структурах белков, вводятся показатели разрывности функциональных сайтов белков, а также описывается разработанный автором веб-интерфейс, обеспечивающий доступ пользователей к базе данных SitEx и визуализацию результатов поиска. В завершении приводятся возможности применения разработанной системы для анализа конкретных структур, в том числе для анализа функциональной организации генов, а также особенностей кодирования и эволюции функциональных сайтов с учетом экзонной структуры гена.

Результаты исследования изложены в главе III «Статистический анализ закономерностей кодирования функциональных сайтов белков в генах позвоночных». Вычислительные эксперименты, проведенные с использованием репрезентативных выборок и применением разных схем анализа, позволили автору получить ряд новых статистически значимых результатов. Показано, что (а) длина экзонов, кодирующих функциональные сайты превышает среднюю длину некодирующих функциональные сайты экзонов; (б) функциональные сайты белков статистически чаще кодируются одним или близко расположенными экзонами; (в) фрагменты белков, кодируемые отдельно взятыми экзонами, проявляют большее структурное сходство в случае, если они содержат аминокислотные остатки функциональных сайтов; (г) распределение частот представленности различных фаз кодонов в районах 5'-концов экзонов, статистически значимо отличается между кодонами, кодирующими и не кодирующими аминокислоты в позициях функционального сайта белка.

Сделанные автором выводы статистически обоснованы и полностью соответствуют полученным результатам.

Диссертационная работа И.В. Медведевой производит хорошее впечатление. Тем не менее, некоторые замечания общего плана хотелось бы все-таки высказать:

- 1) На наш взгляд в диссертации отсутствует должное обсуждение полученных результатов. Представленный вариант обсуждения изложен на двух страницах, содержит две ссылки и фактически является общим заключением, содержащим оценку полученных результатов. Автору следовало более обстоятельно обсудить полученные результаты и сделанные на их основе предположения, используя ссылки из обзора литературы.
- 2) Из данных, представленных на рис. 3.6 автор делает вывод, что среди кодонов, кодирующих пролин, наибольшая частота соответствует кодону ССА. Однако фактически представленные данные показывают, что частоты всех кодонов для пролина равны.
- 3) Для сопоставления границ экзона в гене и проекции домена кодируемого белка на нуклеотидную последовательность автор в одних случаях использует показатель «фазы экзонов», а в других – «фазы кодонов». Из контекста следует, что это фактически один показатель, поэтому для него нужно было использовать одно обозначение.
- 4) Автору нужно было пояснить, почему для расчета частоты кодонов вблизи границ экзонов рассматривали участки, ограниченные только пятью кодонами на 5'-конце и 3'-конце экзона.
- 5) В разделах 2.1.2 и 2.2.4, когда речь идет о об информации, касающейся описания БД PDB, автор отсылает читателя к раздел 1.1.3 вместо раздела 1.1.5.
- 6) В оглавлении и соответственно в тексте диссертации разделы «Влияние мутаций на структуру и функцию белка» и «Эволюция структуры и функции белков» имеют одинаковый номер 1.3, а два других раздела «Базы данных, посвященные функциональным сайтам белков» и «Проекция пространственной структуры белка на структуру кодирующего гена» имеют одинаковый номер 1.4.

Отмеченные недостатки не являются принципиальными и не умоляют значимости работы. Они носят рекомендательный характер и ни в коей мере не ставят под сомнения выводы работы.

В целом, следует высоко оценить диссертационную работу И.В. Медведевой. Она является законченной научно-квалификационной работой, в которой содержится решение задач, имеющих существенное значение для понимания особенностей кодирования структурно-функциональной организации белков, а именно для установления связи между структурой доменов и функциональных сайтов белков, с одной стороны, и экзон-интронной организацией кодирующих их генов – с другой.

Новизна и значимость представленной работы не вызывают сомнений. Выводы диссертации являются обоснованными, автореферат адекватно и полно отражает содержание работы. Материал диссертации соответствует указанной специальности. Диссертация апробирована на шести международных научных конференциях. Все основные результаты опубликованы в работах соискателя, в том числе в трех журналах из списка ВАК. Кроме того, получено свидетельство о государственной регистрации БД SitEx.

Таким образом, диссертационная работа И.В. Медведевой удовлетворяет всем требованиям ВАК России, предъявляемым к диссертациям на соискание ученой степени кандидата биологических наук по специальности 03.01.09 – «математическая биология, биоинформатика», а ее автор, несомненно, достоин присуждения искомой степени.

Отзыв обсужден и одобрен на семинаре теоретического отдела ФБУН ГНЦ ВБ «Вектор».

Заведующий теоретическим отделом

ФБУН ГНЦ ВБ "Вектор"

д.б.н.

С.И. Бажан

Подпись заведующего отделом д.б.н. С.И. Бажана заверяю:

Ученый секретарь ФБУН ГНЦ ВБ «Вектор»

к.б.н., доцент



О.А. Плясунова

10 ноября 2014 г.