

ОТЗЫВ

официального оппонента на диссертационную работу

Занегиной Ольги Николаевны

на тему «Сравнительная характеристика структур ДНК-белковых комплексов»,
представленную на соискание ученой степени кандидата биологических наук по
специальности 03.01.09 – математическая биология, биоинформатика.

Диссертация Занегиной О.Н. посвящена анализу и классификации ДНК-белковых взаимодействий с позиций трехмерной организации контактов ДНК и ДНК-связывающих доменов белков. Методы на основе массового секвенирования, активно применяемые сегодня, в первую очередь предоставляют информацию о последовательности и геномной локализации участков ДНК, связываемых белком. Однако с точки зрения реалий ДНК-белкового взаимодействия, информация о последовательностях ДНК и белка является крайне ограниченной и, в силу отсутствия однозначного кода соответствия предпочтительных аминокислот и нуклеотидов в контактах белок-ДНК, дает лишь опосредованное понимание архитектуры и механизмов ДНК-белкового узнавания. В свою очередь, структурный подход, несмотря на значительно меньший объем данных, позволяет детально идентифицировать ключевые структурные особенности, что важно, в том числе, для инженеринга ДНК-белковых взаимодействий, как на уровне ДНК-сайтов, так и на уровне ДНК-связывающих белковых доменов.

В работе Занегиной О.Н. предложена новая систематическая классификация ДНК-белковых контактов и, путем анализа конкретных белковых семейств на основе предложенной классификации, сделан как ряд важных наблюдений о ДНК-белковых взаимодействиях общего значения (о принципиальной схожести типов ДНК-белковых контактов у различных, согласно общепринятой аннотации SCOP, белковых доменов), так и разбор интересных конкретных случаях, в том числе особенностей узнавания ДНК для TATA-связывающего белка и семейства эндонуклеаз LAGLIDADG. В последнем случае,

получен особенно интересный результат, а именно, показана роль неспецифических контактов, опосредованных молекулами воды. Еще одно важное систематическое наблюдение – о дополняющей роли гидрофобных кластеров между молекулами белка и ДНК, которые, с одной стороны, никогда не являются основой контакта, но, с другой стороны, вносят значительный вклад в ДНК-белковое узнавание.

Вообще, наличие актуальной и регулярно актуализируемой классификации, представленной в работе, открывает возможности для множества более детальных исследований, связанных с ДНК-белковыми взаимодействиями, в частности, становится возможным получить ответы на принципиальные вопросы о связи между типами взаимодействия и различными функциональными классами белков (факторы транскрипции, рестриктазы, гистоны, и т.д.), а также о распространении и степени сходства различных типов взаимодействий среди белков одного функционального класса; причем даже в тех случаях, когда изучаемые белки имеют существенно отличающиеся ДНК-связывающие домены.

Суммируя вышесказанное, не вызывает сомнений актуальность темы рассматриваемой диссертации, и безусловная научная и практическая ценность работы.

Структура диссертации построена по стандартной схеме, с важной особенностью, а именно, материалы и методы для каждого раздела работы вынесены в начало соответствующих глав. Более того, в ряде случаев основные понятия или определения также даны в начале разделов. Все это относится к положительным сторонам работы и заметно упрощает понимание текста.

К сожалению, текст диссертации не лишен недостатков, как стилистических, так и содержательных.

Так, в разделе «Актуальность работы» (стр. 5) автор пишет, «Развитие методов ... привело к экспоненциальному росту количества пространственных структур макромолекул». Можно было бы думать, что для структур пропущено определение «изученных» или «расшифрованных». Но в следующем же предложении сказано, что «ежегодно их <структур комплексов белков с ДНК> количество увеличивается на ~10%», что вызывает вопрос о корректности количественных оценок.

Здесь же, на стр. 5, автор пишет о том, что «не существует общепринятых подходов к предсказанию НК-белковых взаимодействий». Следовало бы явно сказать,

что речь идет о структурных подходах, поскольку в анализе последовательностей такие подходы безусловно существуют.

На странице 7 автор утверждает, что предложена «первая классификация семейств ДНК-узнающих белковых доменов». Безусловно, предложенная классификация имеет ряд важных особенностей и более полна, чем опубликованные другими авторами ранее, но, строго говоря, классификации ДНК-узнающих доменов предлагались и ранее, что отражено в обзоре литературы.

На странице 12 идет речь о «консервативных» молекулах воды, на мой взгляд, здесь не хватает явного описания понятия консервативности, которое в деталях обсуждается гораздо позже. На страницах 16-17 приведен ряд абсолютных значений расстояний, энергий связи и конкретных названий компьютерных программ и баз данных. Фактическая информация без привязки к конкретным примерам трудна для восприятия читателя, недостаточно близко знакомого с тематикой; особенно это касается названий программ (которые приводятся без расшифровок и прямых ссылок на интернет-ресурсы). Для некоторых программ детально описаны численные параметры, но без содержательных комментариев о принципах их выбора.

На странице 22 появляется термин «мотив», но определение мотива не было введено ранее. На странице 25 читаем: «Отсутствие однозначного кода для узнавания не мешает формировать...», на мой взгляд, эта формулировка нуждается в корректировке. Далее, на странице 28, автор пишет: «Мультимеризация белка и кооперативное узнавание повышает способность к специальному узнаванию ДНК». Если закрыть глаза на стиль фразы, требует разъяснения, что автор имеет в виду.

Работа содержит и незначительное количество опечаток, например «часть существующих баз данных ... устарело» (стр. 6); «по-дробный» анализ (стр. 8), «интерфейсов» (стр. 25), «различиями во определении» (стр. 53). Есть и стилистические и терминологические недочеты, мешающие восприятию материала; так, на странице 19 читаем «Группы атомов выращиваются остаток за остатком жадным алгоритмом». Можно предположить, что эта фраза использует профессиональный жаргон, но в текущей формулировке ее смысл ускользает. На странице 33 под «неперепредставленным» списком, по всей видимости, имеется в виду «неизбыточный». Понятная, но замечательная фраза есть также в разделе 1.4: «авторы сталкивались с необходимостью описать все

разнообразие наблюдаемых взаимодействий, что приводило к возникновению классификаций».

Раздел 1.3.2.1 озаглавлен как «Специфичность ДНК-белкового узнавания с точки зрения последовательностей»; стоило явно сказать, что речь идет о белковых последовательностях, поскольку при анализе участков ДНК специфичность понимается по-другому.

Несмотря на хороший стиль изложения материала, некоторые иллюстрации нуждаются в дополнительных пояснениях. Например, на рисунке 6 есть **заметный пик** в **распределении на 4 структурах** в одной записи PDB. Это содержательное наблюдение или какая-то техническая особенность используемого набора структур из PDB?

По сравнению с основным текстом работы достаточно слабо, на мой взгляд, выглядит Глава 4, которая представляет собой описание возможностей веб-интерфейса базы данных, снабженное большим количеством копий экрана интернет-браузера. На мой взгляд, эта информация полезна в качестве инструкции пользователя, и ее следовало поместить в приложения, а не представлять в качестве отдельного раздела диссертации.

В работе недостаточно подробно изложена методическая часть реализации предложенных подходов, например, неясно насколько существующая процедура **построения классификации автоматизирована** и какой объем ручных действий (запуска программ и подбора параметров) и дальнейшей экспертной оценки требуется для поддержания классификации в актуальном состоянии.

Тем не менее, даже с учетом всех указанных замечаний, хочется еще раз отметить высокий уровень работы, а кроме того, в целом зрелый и ровный стиль изложения материала, что является несомненным достоинством диссертации и делает возможным прямое использование ее материалов в образовательных целях. Автореферат диссертации хорошо отражает ее содержание.

Учитывая все вышесказанное, диссертационная работа Занегиной Ольги Николаевны на тему «Сравнительная характеристика структур ДНК-белковых комплексов» без сомнений является законченным научно-квалифицированным исследованием и бесспорно соответствует требованиям «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства РФ от 24 сентября 2013 г. №842. Ольга Николаевна заслуживает присуждения учёной степени

кандидата биологических наук по специальности 03.01.09 «математическая биология, биоинформатика».

кандидат физико-математических наук, старший научный сотрудник
Лаборатории вычислительных методов системной биологии
Федерального государственного бюджетного учреждения науки
Института молекулярной биологии имени В.А. Энгельгардта
Российской академии наук (ИМБ РАН)
ГСП-1, 119991, г. Москва, ул. Вавилова, д. 32
ivan.kulakovskiy@eimb.ru
+7 495 135 60 00

Кулаковский / И.В. Кулаковский /
13 апреля 2016 года

