

Отзыв официального оппонента
на диссертационную работу
Виноградовой Светланы Владимировны
**“Предсказание структурных элементов РНК с
использованием экспериментальных данных”**
**представленной на соискание ученой степени кандидата биологических наук
по специальности 03.01.09 – математическая биология, биоинформатика**

Основная догма молекулярной биологии отводила рибонуклеиновым кислотам вспомогательную роль посредника при переносе генетической информации. Однако впоследствии стало появляться все больше данных о разнообразии функций РНК от регуляторных до катализитических. В последние годы в процессе анализа геномов и пост-геномных данных обнаруживается масса новых типов РНК. Многие типы РНК взаимодействуют с белками, в то время как другие могут взаимодействовать с низкомолекулярными лигандами. Концепция «мира РНК», в которой предполагается, что первичные формы жизни были основаны не на белках, а на РНК приобретает все большую популярность. Функционирование РНК в большой степени определяется не только ее последовательностью (первичной структурой), но и пространственной (третичной) структурой. Принято считать, что промежуточным уровнем организации структур РНК является вторичная структура – набор спаренных оснований. Для предсказания вторичных структур РНК разработан ряд вычислительных методов. С другой стороны, есть экспериментальные методы определения элементов вторичной структуры РНК, которые в большой степени являются косвенными. В последние годы, в связи с появлением методов массового секвенирования, экспериментальные подходы определения вторичных взаимодействий в РНК приобретают все большую популярность.

Работа С.В.Виноградовой посвящена разработке подходов и методов учета экспериментальных данных при поиске структурированных сегментов РНК. В связи с развитием экспериментальных методов определения вторичных структур становится необходимым разработка вычислительных методов анализа этих данных, поэтому актуальность диссертации С.В.Виноградовой не вызывает сомнения.

Диссертация построена по традиционной схеме и содержит введение, обзор литературы, две главы, посвященные результатам автора, заключение и список литературы. Во введении обосновывается актуальность работы. Обзор литературы можно условно разбить на три части. Первая посвящена объекту исследования – структуре РНК и ее свойствам. Вторая описывает основные вычислительные методы, применяемые к анализу и предсказанию вторичных структур РНК. Наконец, в третьей части автор описывает экспериментальные методы определения вторичных структур РНК.

Глава 2 посвящена исследованию характера экспериментальных данных. Сравнение данных с известными структурами показывает достаточно большое расхождение, причем характер расхождений разный для разных типов эксперимента. Автор справедливо отмечает влияние уровня покрытия на адекватность данных. Кроме того, отмечено влияние на результаты эксперимента SHAPE положения спаренных нуклеотидов – на краю петли нуклеотиды менее доступны для SHAPE-реагента, что говорит о меньшей подвижности, что, по-видимому, связано с сохранением однонитевого стэкинга. С другой стороны, крайние нуклеотиды спиралей более доступны, чем внутренние, что подтверждает представление о динамике открытия-замыкания краевых пар спиралей. Важным представляется введение способа пересчета экспериментальных данных в универсальную меру структурированности. В конце главы проведен анализ возможного влияния транслирующих рибосом на структуру *in vivo*. Показано, что кодирующая область мРНК в эксперименте *in vivo* значительно менее структурирована, чем в эксперименте *in vitro*.

В главе 3 описаны методы поиска структурированных участков РНК. Предсказание структуры РНК *de novo* в настоящее время сталкивается с большими трудностями, обусловленными, в частности, большим количеством субоптимальных структур. Поэтому предложено искать не структуры РНК, а участки на последовательности, которые способны формировать стабильную структуру. В первом параграфе этой главы предпринята попытка поиска потенциальных консервативно структурированных участков в полных геномах рода *Drosophila*. Предложен способ учета экспериментальных данных, основанный на том, что к стандартной энергии добавляется член, ответственный за экспериментальную информацию – псевдоэнергия. При вычислении псевдоэнергии используется отношение правдоподобия, введенное в предыдущей главе. Поиск структурированных участков существенно опирается на фоновую модель генома. Показано, что использование наивных фоновых моделей, основанных на перемешивании последовательностей недостаточно адекватно. Предложена фоновая модель, основанная на случайных выборках из генома. Проведен детальный анализ метода поиска структурированных фрагментов и показано, что предложенный способ учета экспериментальных данных значительно повышает качество поиска структурированных участков. Показана необходимость учета качества экспериментальных данных, в частности уровня покрытия. Разработан веб-ресурс для поиска структурированных участков РНК.

Последний раздел «Выводы» суммирует основные результаты работы. Выводы выглядят обоснованными и адекватно отражают содержание работы. Работа описана достаточно подробно и ясно, чтобы читатель мог вникнуть в суть предмета и понять, что и как делалось. Несмотря на то, что главные результаты работы не вызывают сомнения, необходимо привести ряд замечаний и вопросов.

1. Стр. 13 – Хорошо было бы проиллюстрировать, что такое коаксиальный и спиральный стекинг. Про коаксиальный стекинг объяснено словами, спиральный – вообще не про-комментирован.
2. Рис. 1.5 – Отсутствуют пояснения в подписи, объясняющие, что обозначено разными цветами. Что означает график сверху?
3. Стр. 38. Непонятна методика отбора транскриптов. Две последовательно написанные фразы: «Мы отобрали только транскрипты со средним покрытием, равным как минимум одно чтение на позицию. Далее мы применили второй фильтр, требуя, чтобы покрытие чтениями в обоих экспериментах было как минимум 10 чтений на позицию» кажутся противоречащими друг другу. Выше в том же абзаце написано о том, что работа проводилась с РНК, при выделении которой использовались денатурирующие условия для очистки от белков. Далее написано «В этом случае структура РНК наиболее близка к структуре РНК в клетке». Что имеется в виду? Структура РНК с удаленными РНК-связывающими белками не может быть «наиболее близкой» к структуре РНК, связанной с белками – не важно, в клетке, или вне клетки.
4. Стр. 39 – Вместо кальки с английского «псевдокаунт» лучше употреблять распространенный в русскоязычной научной литературе термин «псевдоотсчет». Почему использованные в данной формуле псевдокаунты равны⁻⁵? Не обоснован выбор данного значения.
5. Стр. 40 – есть упоминание о ROC кривых, но сами кривые находятся очень далеко по тексту. Ясно, что такое размещение определяется тем, что методы отделены от результатов, но представляется уместным наличие в данном месте ссылки на страницу с соответствующим рисунком.
6. Стр. 42 – автором определена мера «структурное значение» для нуклеотидов и фрагмента РНК. Название меры нужно писать в кавычках. Для данной меры написана формула, но, собственно, мера никак не обозначена. Это неудобно для восприятия дальнейшего изложения.
7. Стр. 43. Нигде не объяснено, что такое Vault РНК.
8. Рис. 2.2 и 2.4 – что отложено по оси Y?
9. Рис. 2.7. Вызывает недоумение способ нумерации панелей рисунка: «а, б, с, д». В русскоязычных текстах нормой считается нумерация в алфавитном порядке: «а, б, в, г». Тот же способ нумерации используется и в других рисунках.
10. Рис 2.9 - В подписи к рисунку не приведено описание панелей, что характерно и для других рисунков.

11. Что такое анализ периодичности? Обнаруженная три-периодичность, вероятно, связана с кодонами, однако хотелось бы иметь объяснение того, какое отношение кодоны могут иметь к структуре РНК.
12. Автор использует в процессе работы разнообразные компьютерные программы, написанные другими авторами, оперируя в тексте параметрами данных программ и результатами их работы, при этом сами программы либо не описаны вообще, либо описаны столь кратко, что непонятно в чем смысл их использования в работе. Это чрезвычайно затрудняет чтение и восприятие текста диссертации, поскольку требует от читателя отдельно разбираться с каждой из программ.
13. Веб-сервер RNASurface практически не описан.
14. У таблицы 3.1. нет подписи. В таблице приведены отрывочные сведения – по одному значению чувствительности метода при данной длине окна (при этом – разной для разных типов РНК) у данного типа РНК, что никоим образом не подкрепляет сделанный в тексте вывод о том, что «чувствительность метода в рамках одного класса сильно зависит от длины используемого окна и, таким образом, окно является сильным ограничением метода».
15. Рис. 3.8 на стр. 76 имеет название: «Распределений сдвигов в Z-значениях (пробинг-Z-значение минус Z-значение) для случайных сегментов мРНК и функциональных структурированных мРНК». Имеется в виду «распределения»? Что обозначено жирными точками на одном из распределений? Почему на втором распределении их нет? С помощью какой программы построены данные изображения?
16. Рис. 3.10. Для оценки увеличения предсказательной силы при использовании экспериментальных данных было бы полезно посчитать и привести в диссертации AUC – площади под полученными ROC кривыми.
17. Представляется сомнительным употребление некоторых терминов, например, «реакцион способность», стр. 25, стр. 35. Корректным термином в данном случае является «реакционная способность». Следствием прямого перевода с английского является употребление такого термина, как «пробинг данные» - стр. 68. Корректнее писать «данные, полученные методом ДМС-пробинга».
18. Имеются некоторые проблемы с употреблением введенной автором аббревиатуры «МСЭ» - минимизация свободной энергии (стр. 16). Например, в предложении «Таким образом, вместо жестких ограничений лучше использовать мягкие ограничения, которые позволяют использовать экспериментальные данные в качестве дополнительной информации при нахождении МСЭ», по-видимому, данное сокращение использовать некорректно. Имеется в виду не «минимизация», а «минимум» свободной энергии. Далее по

тексту диссертации данная аббревиатура используется очень часто, при этом подразумевается по крайней мере три варианта расшифровки – «минимизация», «минимум», а также «метод минимизации» свободной энергии. В автореферате же, на стр. 10 сокращение МСЭ расшифровывается как «минимальная свободная энергия». Таким образом, данные три буквы автор употребляет для обозначения, как минимум, четырех понятий, что делает несколько неоднозначной трактовку ряда мест диссертации.

19. Имеется целый ряд замечаний, касающийся оформления диссертации: диссертация не оформлена в соответствии с правилами ГОСТа – неправильно оформлен список литературы, подписи к рисункам и таблицам, отсутствует список сокращений. Ряд сокращений, например, ДМС и PARS не расшифрованы при первом упоминании. Напротив, то, что Рибонуклеиновая кислота сокращается как РНК, написано дважды – на стр. 4 и 9. При расшифровке сокращения рiРНК (стр. 10) оно повторено дважды. Латинские названия организмов пишутся курсивом (стр. 42). При повторном использовании видовых названий организмов название рода сокращается до первой буквы с точкой – стр. 82.
20. Следует отметить, что сокращение ДМС в диссертации используется как в русскоязычном, так и в англоязычном варианте (DMS). Расшифровка данного сокращения приведена лишь на странице 26. До этого данное сокращение используется и для обозначения химического вещества (диметил сульфата) и для обозначения метода, основанного на применении данного химического вещества, усугубляя недоумение читателя по поводу данной аббревиатуры.
21. Текст диссертации местами плохо вычитан. Встречаются фразы с несогласованными окончаниями слов, например: стр. 10 - «Последовательности нуклеотидов молекулы РНК представляет собой первичную структуру РНК»; стр. 20 – «Таким образом, эволюционный подход, требующий наличие выравнивания, применим далеко не всегда»; стр. 25 «Реконструкция вторичной структуры генома ВИЧ с помощью метода SHAPE стало доказательством того, что экспериментальные методы действительно могут быть очень полезными при анализе структур РНК». Имеются опечатки. Например, на стр. 24: «Несмотря на простоту процедуры с случае коротких РНК, исследование структуры более длинных РНК является непростой задачей»; там же – «высокопроизводительному». Слово «электрофорограмма» пишется с двумя «м» - стр. 25.

Однако все вышеуказанные замечания ни в коей мере не умаляют достоинств работы. В целом работа С.В.Виноградовой выполнена на высоком методическом уровне. Автор показал хорошее владение материалом и высокий профессионализм в его анализе. Удачное включение экспериментальных данных в систему поиска структурированных участков позволило существенно повысить качество поиска структурированных сегментов РНК.

Диссертационная работа выполнена автором самостоятельно на высоком научном уровне. Полученные результаты достоверны, выводы и заключения обоснованы. Работа базируется на достаточном количестве данных, которые изложены в диссертации четко и ясно, а сама рукопись аккуратно оформлена. По каждой главе и работе в целом сделаны убедительные выводы, соответствующие поставленным задачам и полученным данным. Результаты диссертационной работы С.В.Виноградовой могут быть использованы в работе научно-исследовательских и учебных организаций.

Основные результаты диссертации С.В.Виноградовой представлены в ряде статей и неоднократно обсуждались на отечественных и международных конференциях. Автoreферат соответствует основному содержанию диссертации и полностью отражает идею работы.

Таким образом, диссертационная работа С.В.Виноградовой является самостоятельной научно-квалификационной работой, содержащей новое решение актуальной для биологии проблемы разработки подходов и методов учета экспериментальных данных при поиске структурированных сегментов РНК. По своей актуальности, научной новизне, теоретической и практической значимости, объему и методическому уровню работа Светланы Владимировны Виноградовой "Предсказание структурных элементов РНК с использованием экспериментальных данных" соответствует требованиям пункта 9 Положения о порядке присуждения ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. №842, предъявляемым к кандидатским диссертациям, а соискатель несомненно заслуживает присуждения искомой ученой степени кандидата биологических наук по специальности 03.01.09 – математическая биология, биоинформатика.

Карягина-Жулина Анна Станиславовна,
доктор биологических наук, профессор,
главный научный сотрудник лаборатории
биологически активных наноструктур
ФГБУ «Федеральный научно-исследовательский
центр эпидемиологии и микробиологии имени
почетного академика Н.Ф.Гамалеи»
Министерства здравоохранения
Российской Федерации
123098, Москва, ул. Гамалеи, д. 18;
тел. +7(926)4955132;
akaryagina@gmail.com

Подпись д.б.н., проф. Карягиной-Жулиной А.С. заверяю.
Ученый секретарь ФГБУ «ФНИЦЭМ им. Н.Ф.Гамалеи»,
Минздрава России,
к.б.н.



Кожевникова Л.К.