

О Т З Ы В

официального оппонента д.ф.-м.н. Макеева В. Ю. о диссертационной работе Ромеро Рейес Илякай Владиславовны «Оценка аффинности комплексов белок-лиганд с применением нейронных сетей», представленной на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ.

Актуальность проблемы. Диссертация Илякай Владиславовны Ромеро Рейес посвящена важной практической проблеме — разработке математических методов и компьютерных программ для вычисления химического сродства небольших молекул к белковой мишени. Задача эта важна для разработки новых лекарственных препаратов. Действительно, достаточно часто имеется представление о возможной мишени лекарственного препарата, но доступные препараты неприменимы по каким-то причинам или просто отсутствуют. В этих случаях обычно используют экспериментальный скрининг кандидатов. Предварительная оценка возможного сродства препарата-кандидата к предполагаемой мишени позволяет отвергнуть препараты-кандидаты с низким сродством и уменьшить необходимое число кандидатов, сократив как стоимость, так и время необходимых экспериментальных исследований. Исследования диссертанта сочетают прямое молекулярное моделирование взаимодействия лиганда с мишенью, и предсказания, основанные на статистике. В рамках статистического подхода химическое сродство лиганда-кандидата к молекуле мишени вычисляется путем сопоставления связывания лиганда-кандидата с другими, схожими лигандами, для которых есть экспериментальные данные по связыванию. Экстраполяция осуществляется с помощью некоторой функции, аргументом которой являются молекулярные дескрипторы. Работа диссертанта заключается в построении гибридного алгоритма, сочетающего оба подхода, и позволяющего существенно повысить точность предсказания аффинности молекулы-кандидата к исследуемому белку.

Научная новизна. В основе работы лежит красивая и оригинальная идея. Кратко ее можно изложить следующим образом. Известно, что использование молекулярного докинга для предсказания сродства малых молекул к белкам с известными пространственными структурами очень ненадежно. Результаты докинга сильно зависят от начальных условий, в частности от ориентации лиганда, а оптимизация решений методом молекулярной динамики сильно зависит от выбранной модели и стратегии моделирования. В диссертации предложено отказаться от непосредственного предсказания химического сродства на основании вычисленных значений энергии связывания, но использовать эти вычисленные значения в качестве некоторого параметра (дескриптора), характеризующего взаимодействие. Такое вычисление проводится для ряда лигандов с известной константой взаимодействия с целевым белком; полученные значения используются для предсказания силы взаимодействия для неизвестного лиганда с помощью нелинейной регрессии. Для построения регрессии используется нейронная сеть. В качестве дополнительных дескрипторов для предсказания используются физико-химические свойства лиганда, такие как молекулярный вес, площадь поверхности и т.п. Построение алгоритма проведено в высшей степени аккуратно и хорошо описано в диссертации. Алгоритм включает предварительный анализ данных, снижение размерности методом главных компонент, контроль переобучения путем введения трех выборок (обучающей, контрольной и тестовой) выборок, и детальный статистический анализ полученных результатов. В результате удалось построить алгоритм, обеспечивающий существенный рост коэффициента детерминации (до 0.94 по сравнению со средней величиной, характерной для молекулярного моделирования, равной 0.1). Следует особо отметить, что диссертанту удалось организовать экспериментальную проверку связывания для 8 тестовых молекул. По сравнению с традиционными методами 3D QSAR использование авторского алгоритма позволяет увеличить коэффициент детерминации с 0.39 до 0.77. Такой рост качества предсказания позволяет

аккуратно ранжировать лиганды по силе связывания, что было недоступно при использовании алгоритмов 3D QSAR.

Большая работа проведена диссертантом для создания эффективного программного обеспечения, реализующего предложенный алгоритм. Алгоритм смоделирован на Matlab, и после оптимизации реализован с использованием графического процессора на NVIDIA CUDA. Созданная версия программы позволяет в реальном времени осуществлять анализ связывания для большого числа молекул, что важно для реальных приложений в области моделирования потенциальных лекарственных соединений.

Практическая значимость. Диссертант разработала новый оригинальный алгоритм предсказания силы связывания малых молекул с белками с известной структурой, работающий при условии наличия некоторого набора данных по связыванию аналогичных лигандов. Алгоритм очень быстро работает, и дает результаты, существенно превосходящие аналогичные алгоритмы по качеству предсказания. Я уверен, что идеи, использованные при создании алгоритма, найдут применение при разработке практических программ для поиска новых лекарственных соединений.

Замечания к работе. Работа блестяще подготовлена. Мне не удалось найти ни одного замечания к оформлению. В выводах (или в научной новизне) следовало бы выделить основную идею алгоритма – использование данных молекулярного моделирования не непосредственно, а в качестве дескриптора. Вызывает вопрос построение выборок. Диссертант (на стр. 57) описывает три выборки: обучающую, контрольную, и тестовую, утверждая, что разбиение на выборки происходило случайным образом, но так, чтобы точки покрывали весь диапазон изменения коэффициента Танимото. Мне кажется, что равномерность диапазона покрытия пространства структур – это не самое главное свойство при построении выборок; важнее избежать присутствия в обучающей выборке очень схожих структур,

чтобы не создавать статистических неоднородностей, а также избежать близкого сходства тестовой и обучающей выборок, во избежание переобученности.

Указанные замечания не снижают общей положительной оценки диссертации, которая представляет собой законченное исследование в рамках выдвинутых целей и имеет как научное, так и практическое значение. Автореферат и публикации полностью отражают содержание диссертационной работы. По актуальности, объёму и научному уровню она полностью соответствует требованиям Положения о присуждении ученых степеней, предъявляемым к кандидатским диссертациям, а ее автор – Ромеро Рейес И. В. заслуживает присуждения искомой степени кандидата физико-математических наук по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ, а именно пунктам 4 и 5 паспорта специальности 05.03.18 («Реализация эффективных численных методов и алгоритмов в виде комплексов проблемно-ориентированных программ для проведения вычислительного эксперимента» и Комплексные исследования научных и технических проблем с применением современной технологии математического моделирования и вычислительного эксперимента).

Официальный оппонент,
доктор физико-математических наук,
заведующий отделом вычислительной
системной биологии Института общей генетики
им. Н.И. Вавилова РАН

14.05.2014

В. Ю. Макеев

Подпись В.Ю. Макеева заверяю:
Ученый секретарь Института общей генетики
им. Н.И. Вавилова РАН
доктор биологических наук



О.А. Огаркова