



Олег Владимирович Колясников

Ассистент кафедры химии СУНЦ МГУ им. М.В. Ломоносова.

Биоинформатика в современной школе

Биоинформатика вообще

Сегодня биоинформатика является чрезвычайно перспективной наукой, имеющей обширные приложения в фармацевтике, медицине, криминалистике, этнографии и т. д. Относительно недавнее успешное завершение секвенирования генома человека обеспечило учёным практически неисчерпаемый источник исходных данных для сравнения и осмысления. Вслед за геномикой бурный рост переживают транскриптомика, протеомика, метаболомика и другие отрасли биоинформатики.

К сожалению, в классических курсах химии, биологии и информатики эта междисциплинарная область практически не упоминается вследствие своей относительной молодости и хронического недостатка учебного времени. В то же время основные тезисы направлений биоинформатики достаточно просты для усвоения при наличии минимального базиса знаний по упомянутым

выше дисциплинам.

Несомненным плюсом для обучения биоинформатике является то, что зачастую для исследований в этой области вполне достаточно обычного настольного компьютера с доступом в Интернет, в отличие, например, от минимальных требований к оборудованию для научной работы в области химии или классической биологии. Как показывает знакомство с материалами текущих школьных конференций, в настоящее время работа с учащимися в области биоинформатики находится в зачаточном состоянии, и нашими усилиями СУНЦ МГУ сейчас является одним из лидеров в этом процессе. Не буду говорить, что мы являемся пионерами в этой области – как пример, можно привести работу со школьниками ныне, к сожалению, покойного А.А. Недоспасова [1]. Но, так или иначе, сейчас таких групп в России очень мало.

Биоинформатика в преподавании

Наши исследования в основном сосредоточены на приложениях биоинформатики к структурной биологии. Основным орудием работы являются программы визуализации структурных данных, к которым относятся, например, SwissPDBViewer [2], Raswin [3] или Rymol [4]. Данные о десятках тысяч структур биополимеров, определённых экспериментально, содержатся в открытом виде в Интернете на сайте RCSB [5] в виде текстовых файлов. Основной объём этих файлов занимают трёхмерные координаты всех атомов за исключением водорода (так называемых «тяжёлых атомов») изучаемой молекулы. Программы визуализации преобразуют эти координаты в векторное изображение (рис. 1).

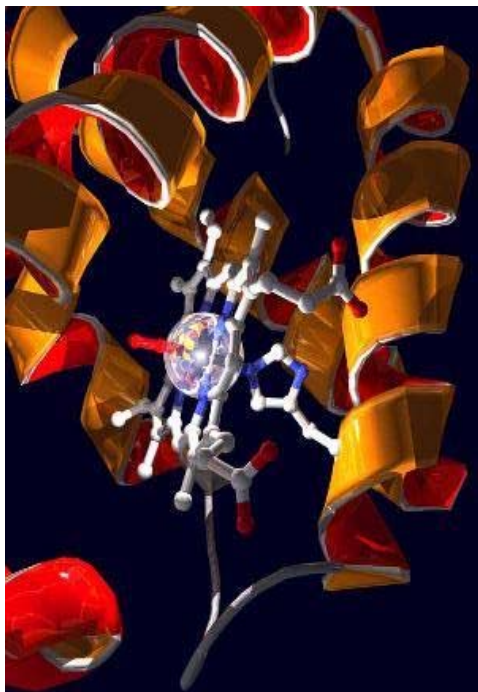


Рис. 1. Пример изображения фрагмента молекулы гемоглобина с помощью программы SwissPDBViewer

На наших уроках информатики в биологическом классе мы используем эти программы как пример программ построения векторной графики, тем самым одновременно расширяя кругозор знаний учащихся о современной науке и проводя первичный отбор кандидатов для дальнейшей научной работы. Проблемой на этом этапе является языковой барьер. К сожалению, почти все материалы и программы представлены на английском языке. Вследствие этого приходится уделять время переводу и разъяснению значения основных терминов как в рамках уроков информатики, так и сверх них.

Как говорилось выше, на уровне начала 10 класса современные школьные курсы биологии и химии содержат крайне мало материала о современной молекулярной биологии и биоорганической химии. Большую часть этих пробелов, крайне затрудняющих работу, можно закрыть дополнительными факультативными курсами. В качестве адаптированного материала для этих курсов можно рекомендовать современную тематическую научно-популярную литературу [6,7,8]. В нашей школе курс обучения составляет 2 года (10–11 классы). Поэтому читалось 2 курса – «Основы геномики» и «Основы химии и физики белка» – поочерёдно. Дополнительный плюс факультативных курсов состоит в усвоении учащимися современной научной лексики как на русском, так и на английском языках.

Исследования в области биоинформатики чаще всего основаны на методе индукции, когда из сравнения многих данных мы приходим к тем или иным выводам. Таким образом, к «идеальному» объекту исследования мы имеем противоречивые

пожелания максимального разнообразия структур для увеличения значимости результатов исследования и максимальной общности устройства для упрощения сравнения отдельных представителей. По нашему мнению, наиболее удобным объектом являются антитела [9]. Антитела вырабатываются белыми клетками крови – лимфоцитами. Они помогают иммунной системе опознавать и уничтожать патогенные бактерии, вирусы и токсины. Чтобы осуществлять эти функции, антитела, с одной стороны, должны проявлять сродство к любой молекуле (так называемому антигену), с другой стороны, должны быть построены на каркасе, практически одинаковом для всех антител (рис. 2).

Как упомянуто ранее, весь материал о завершённых экспериментальных работах в области структурной биологии содержится в базе данных RSCB. Разновидность антител хотя и велика, но в общем объёме занимает небольшое место. Проблемой нашей школы (и, наверное, не только нашей) являлось то, что скорость доступа в Интернет, увы, была ограничена.

Резко сэкономить время на исследование позволило создание локальной базы структур антител

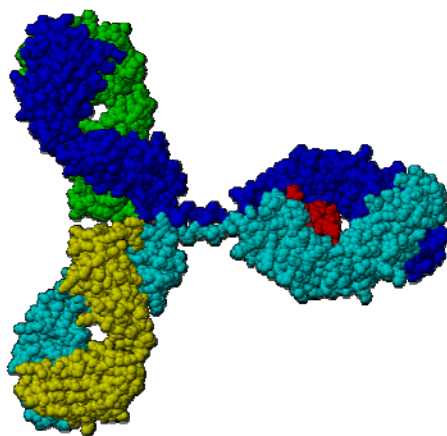


Рис. 2. Общий вид молекулы антитела, принадлежащего к наиболее распространённому классу (IgG)

учащимися нашей школы. Регулярно обновляемая русифицированная база, состоящая из полной выборки структур антител по базе RSCB (на данный момент это почти 1000 структур) предоставляет возможность найти и открыть в программе визуализации любой структурный файл в течение секунд.

Перечисленные компоненты создают необходимую базу для ключевого этапа – самостоятельной научной работы учащихся.

Биоинформатика в школьном исследовании

Рассмотрим пример нашей недавней работы в области биоинформатики. Известно, что в основе многих структур биополимеров лежит пи-стекинговое взаимодействие (от англ. stack – стопка, штабель). Это взаимодействие пи-электронных облаков, возникающее между ароматическими системами, например, между бензольными кольцами. Как правило, при нём плоскости ароматических колец взаимодействующих молекул параллельны, что упрощает обнаружение взаимодействия в

структурном файле. Широко известна роль пи-стекинга в структуре двойной спирали ДНК. Сама спираль образуется сахарофосфатным остовом молекулы, охватывающим взаимодействующие азотистые основания. Они, в свою очередь, образуют водородные связи между основаниями разных цепей ДНК и пи-стекинг между основаниями одной цепи. Сближение плоскостей азотистых оснований за счёт пи-стекинга позволяет экранировать от молекул воды область образования водород-

ных связей – ключевого взаимодействия между нитями ДНК, увеличивая точность работы систем репликации и транскрипции. И не случайно попадание в живую клетку веществ, связывающихся между плоскостями азотистых оснований и формально не влияющих на взаимодействие между цепями ДНК, чаще всего приводит к мутагенезу и канцерогенезу. Так, например, действует бензпирен – основной канцероген

табачного дыма и выхлопных газов.

В живой природе пи-стекинг встречается гораздо шире, чем в школьном курсе. Мы помним, что антитела могут связывать практически любую молекулу и, тем самым, могут служить моделью взаимодействия молекул в общем случае. Известно также, что иногда малые ароматические молекулы связываются белками с образованием пи-стекинга (рис. 3).

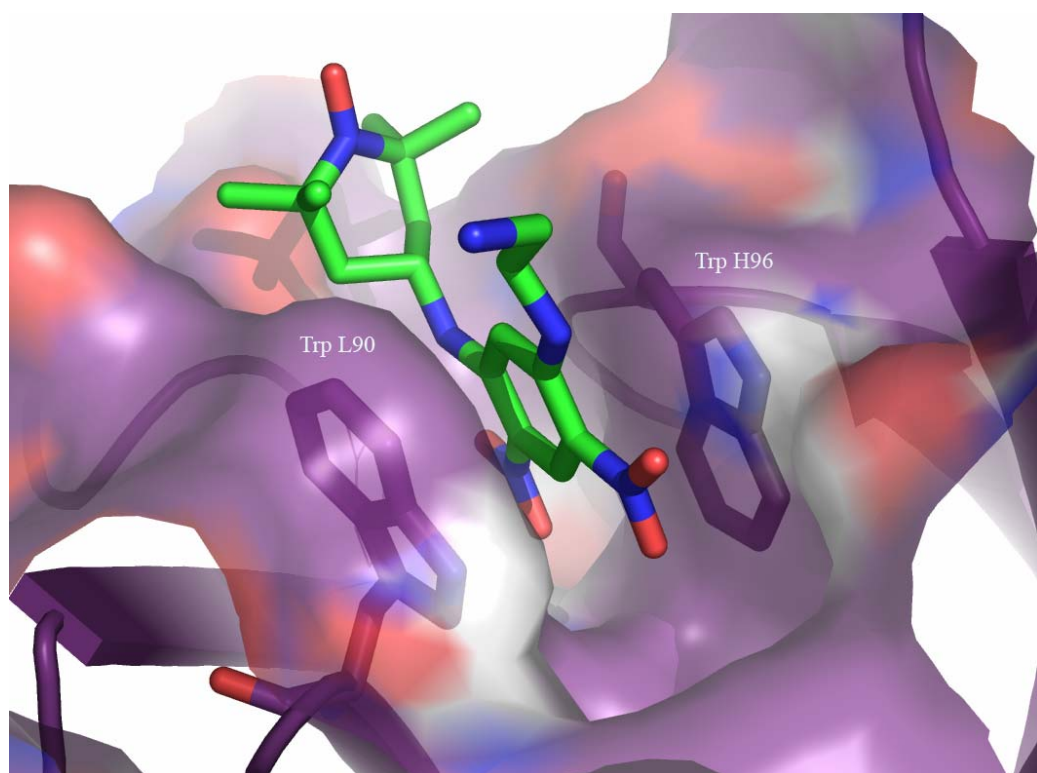


Рис. 3. Комплекс типа «сэндвич» антитела с антигеном, родственным 2,4-динитрофенолу (код PDB 1baf)

Возникает вопрос: можем ли мы на основе наших знаний о структуре антител вывести какие-либо общие условия для наблюдения пи-стекинга, по крайней мере, для случая системы антитело-антиген? Ответ на этот вопрос может быть полезным в оценке результатов моделирования комплексов белков с малыми молекулами.

Для начала, очевидно, надо задать граничные условия поиска. Минимальная стабильная ароматическая система состоит из 5 тяжёлых атомов. Порфириновое кольцо – одна из самых больших ароматических систем, встречающихся в живой природе, – состоит из 24 тяжёлых атомов. Сделав некоторый за-

пас, мы ввели критерий отбора в 50 тяжёлых атомов для связываемой молекулы антигена. Это позволило примерно на порядок сократить количество структур для более внимательного изучения. К сожалению, полностью автоматизировать этот этап на нашем уровне было сложно, поэтому пришлось прибегнуть к ручному отбору целевых структур из ранее описанной локальной базы. Это был один из самых продолжительных и рутинных этапов работы, в конечном счёте занявший несколько месяцев.

В качестве первого приближения был выдвинут тезис, что малая ароматическая молекула всегда должна связываться с антителом с участием пи-стекинга. Исключения тщательно выделялись и сортировались. Почти сразу же стало очевидно, что исключения возможно отнести к одной из трёх групп.

Во-первых, к группе структур абзимов – молекул антител, разрабатанных учёными для осуществления ферментативных реакций. Крайне редко эти реакции затрагивают ароматическое ядро. Поэтому связывание ароматической части молекулы может быть менее специфичным и может в равной мере как включать, так и не включать пи-стекинг.

Во-вторых и в-третьих, к группам структур металлокомплексов и каркасных антигенов, где ароматическая часть явно уступает по значению другим частям связываемой молекулы.

Каково количественное выражение этого значения, позволяющее нам судить о наличии или отсутствии пи-стекинга? Нами были проанализированы различные количественные характеристики связываемых малых молекул. Наиболее чётким параметром оказалась доля поверхности молекулы, занимаемая ароматической частью. Было выяв-

лено, что при площади ароматической части молекулы более 36% практически все структуры комплексов демонстрируют наличие пи-стекинга. Следует отметить, что на этом этапе методически крайне полезным оказалось обсуждение работы с выпускниками биокласса СУНЦ, обучающимися на факультете биоинформатики и биоинженерии МГУ.

Таким образом, в результате работы нами была получена практически важная количественная оценка, лимитирующая применение концепции пи-стекинга к описанию взаимодействия биомолекул.

Необходимо отметить, что не менее важным итогом работы явилось приобщение школьников к современной науке, освоение ими научных программных пакетов, обучение поиску тематической информации в Интернете, получение навыков презентации результатов на конференциях разного уровня, а также в виде научных публикаций. Это приближает уровень наших выпускников к студентам факультета биоинформатики и биоинженерии МГУ. Следует отметить, что все учащиеся, задействованные в работе, благополучно туда поступили.

Мы проводили сравнение результатов наших работ с другими работами, представленными на тех же конференциях. Как говорилось выше, на школьном уровне аналогичных исследований практически не наблюдается. В то же время, для работ от студенческого уровня и выше подобная проблематика весьма широко представлена. Это позволяет нам сделать вывод, что главное препятствие в приобщении учащихся к исследованиям в области биоинформатики лежит не в неспособности их к усвоению дополнительного материала, а в отсутствии опыта ведения подобных работ преподавателями.

Как пример высокой оценки наших исследований можно привести систематические выдающиеся результаты, показанные нашими участниками на Балтийском научно-техническом конкурсе, проводимом для учащихся средних школ в Санкт-Петербурге под эгидой корпорации Intel. В течение 4-х последовательных лет (2006–2009) наши работы были отмечены 2 премиями первого уровня и премией второго уровня по секции «Биология».

Впрочем, качество нашей работы позволяет быть представленным и на «взрослых» конференциях. Например, мы регулярно участвуем в Московской международной конференции по вычислительной молекулярной биологии (МССМБ), проводимой на факультете биоинформатики и биоинженерии МГУ.

Наиболее выдающиеся результаты были нами подготовлены к печати и опубликованы в международ-

ном научном тематическом журнале, где желающие могут с ними ознакомиться более подробно [10,11].

Суммируя вышесказанное, хотелось бы сказать, что изучение биоинформатики в школе, с одной стороны, приближает познания школьников к уровню современной науки и мотивирует их к углублённому изучению и пониманию ряда школьных предметов; с другой стороны, оно технически и методически просто в сравнении с классическими работами в рамках школьных программ. Это позволяет надеяться, что доля работ по биоинформатике в исследовательской деятельности учащихся будет возрастать.

Пользуясь случаем, хотелось бы ещё раз поблагодарить за активную работу ныне уже выпускников СУНЦ МГУ В. Аржаника, Д. Сви-стунову, М. Кираля, В. Ильенкова, И. Русинова, А. Самарина и М. Сая-пину, внёсших основной вклад в работу.

Литература

1. *Tatyana P. Suntsova, Natalia V. Beda and Andrey A. Nedospasov.* Structural features of proteins responsible for resistance of tryptophan residues to nitrosylation. *IUBMB Life*, Vol. 54, No. 5, 2002. 281–92.
2. <http://spdbv.vital-it.ch/>
3. <http://www.openrasmol.org/>
4. <http://www.pymol.org/>
5. <http://www.rcsb.org/>
6. Франк-Каменецкий М. Век ДНК. – М.: Книжный дом, Университет, 2004.
7. Финкельштейн А., Птицын О. Физика белка. Курс лекций. – М.: Книжный дом, Университет, 2005.
8. Тарантул В. Геном человека: энциклопедия, написанная четырьмя буквами. – М.: Языки славянской культуры, 2003.
9. Ройт А., Бростофф Дж., Мейл Д. Иммунология. – М.: Мир, 2000.
10. *Oleg V. Koliashnikov, Miroslav O. Kiral, Vitaly G. Grigorenko and Alexey M. Egorov.* Antibody CDR H3 modeling rules: extension for the case of absence of Arg H94 and Asp H101//*Journal of Bioinformatics and Computational Biology*. Vol. 4, No. 2, 2006. 415–424.
11. *Vladimir Arzhanik, Darja Svistunova, Oleg Koliashnikov and Alexey M. Egorov.* Interaction of antibodies with aromatic ligands: the role of pi-stacking//*Journal of Bioinformatics and Computational Biology*. Vol. 8, №3 2010. 471–483.