

В.Б. Сулимов: Не могли бы сформулировать на каких принципах и законах работает библиотека PLIP? Меня интересуют законы для идентификации вот этих взаимодействий "Водородные связи, гидрофобные и ароматические взаимодействия определяются с помощью библиотеки PLIP".

Д.И. Баширова: Для идентификации донорных и акцепторных атомов PLIP использует программу OpenVabel. После того, как такие атомы идентифицированы, взаимодействие проверяется на соответствие геометрическим требованиям (расстояние между донором и акцептором водородной связи не должно превышать 4.1 Å, а угол в группе (D-H...A) должен быть больше 100°), для нескольких возможных водородных связей образованных от одного донора водорода остается только та связь, угол в группе (D-H...A) которой ближе к 180°. В программе PLIP атом классифицируется как гидрофобный, если он является углеродом и связан только с атомами углерода или водорода. Гидрофобные взаимодействия между всеми парами гидрофобных атомов идентифицируются на расстоянии до 4 Å. Так как нет четкой геометрии гидрофобных взаимодействий и программой выбран довольно широкий диапазон расстояний, может определяться слишком большое число гидрофобных взаимодействий, поэтому количество гидрофобных взаимодействий уменьшается: если гидрофобный атом лиганда взаимодействует с несколькими атомами белка в сайте связывания в одном и том же остатке, сохраняется только взаимодействие с ближайшим расстоянием, а также, если атом белка взаимодействует с несколькими соседними атомами лиганда, сохраняется только взаимодействие с ближайшим расстоянием. Для идентификации колец и их ароматичности PLIP использует программу OpenVabel, где идентификация производится поиском наименьшего набора наименьших колец. В том случае, если программа OpenVabel не определяет кольцо ароматичным, оно проверяется на плоскостность. Для двух ароматических колец идентифицируется ароматическое взаимодействие, если их центры находятся на расстоянии, не превышающем 5.5 Å, а угол отклонения не превышает 30° от оптимального угла 90° для T-стэкинга или 180° для P-стэкинга.

В.Б. Сулимов: PLIP свободно доступная библиотека?

Д.И. Баширова: Библиотека PLIP находится в свободном доступе (<https://github.com/pharmai/plip>).

В.Б. Сулимов: (С. 4) Какое силовое поле использовалось в молекулярной динамике? Какая вода? Какой пакет МД был использован. Шаг траектории, ансамбль и термостат?

Д.И. Баширова: (С. 4) Для генерации топологий белков мы использовали силовое поле Amber99SB-ILDN, для лигандов- Antechamber 17.3 с использованием параметров силового поля GAFF2. Для описания воды использовалась модель TIP3. Моделирование было выполнено с использованием GROMACS. После минимизации энергии каждая система была уравновешена в течение 100 пс на каждое уравновешивание NVT и NPT. После чего мы моделировали комплексы в NPT ансамбле с модифицированным термостатом Берендсена (V-rescale) и баростатом Парринелло-Рахмана при 310 K в течение 50 нс с шагом 2 фс.

В.Б. Сулимов: (С. 5) Что такое Биннинг расстояний? Что такое сигнатура фармакофора? Генерация фармакофорных моделей и/или сигнатур проводится на лигандах или на белках, или только на их комплексах?

Д.И. Баширова: (С. 5) Под биннингом расстояний подразумевается, что все расстояния между фармакофорными центрами переводятся в 'округленные' расстояния с заранее заданным шагом биннинга (например, 1 Å), такая процедура позволяет схожим фармакофорам обладать одинаковым хешем. Сигнатура фармакофора - это и есть 3D-хеш фармакофора, которые кодируют взаимное расположение фармакофорных центров и их метки. В работе генерация фармакофорных моделей осуществляется на основе комплекса, фармакофорные центры размечаются на лиганде, поэтому генерация сигнатур проводится на основе лигандов, что также позволяет использовать данный подход для генерации фармакофорных моделей на основе структур лигандов (ligand-based pharmacophore modeling).

В.Б. Сулимов: С6. Что такое 3D-хэши фармакофоров?

Поясните пожалуйста нижнюю часть слайда

"Репрезентативные фар... модели", "конформеры соединения", какой физ. смысл в дробях?

С7: Что такое соединения приманки? Это неактивные по данному белку соединения? А кто доказал их неактивность по данному белку?

Почему коэффициент обогащения больше единицы? Или это проценты? Ну вообще-то коэффициент обогащения - это площадь под кривой обогащения - так же как AUC это площадь под ROC-кривой! Или используемое Вами определение эквивалентно построению этих кривых?

Приведенные Вами 4 лиганда это что? Это нативные лиганды из соответствующих комплексов из Protein Data Bank? Я думал, что Вы вычисляете коэффициент обогащения используя набор из известных активных соединений и неактивных соединений. А что Вы делаете с этими четырьмя лигандами? Для них известны характеристики связывания с мишенью?

С8: Чем определяется учитываемое число фармакофорных центров для данной молекулы? Почему у Вас несколько вариантов для каждой молекулы?

Д.И. Баширова: (С. 6) На данном слайде представлены фармакофорные модели и соответствующие им 3D-хэши, кодирующие их структуру. После чего удаляются дубликаты (некоторые модели обладают идентичными хэшами), остаются фармакофорные модели только с уникальными хэшами - репрезентативные фармакофорные модели. Далее рассматривается подход общих хитов, молекула соответствует трем репрезентативным моделям из шести, согласно подходу общих хитов ее оценка (или score) будет равен $3/6 = 0.5$. Далее рассматривается подход охвата конформеров, четыре конформера молекулы из пяти соответствуют репрезентативным моделям, поэтому согласно этому подходу score молекулы будет равен $4/5 = 0.8$.

В.Б. Сулимов: С6. Что такое 3D-хэши фармакофоров? Поясните пожалуйста нижнюю часть слайда. "Репрезентативные фар... модели", "конформеры соединения", какой физ. смысл в дробях?

Д.И. Баширова: (С. 6) На данном слайде представлены фармакофорные модели и соответствующие им 3D-хэши, кодирующие их структуру. После чего удаляются дубликаты (некоторые модели обладают идентичными хэшами), остаются фармакофорные модели только с уникальными хэшами - репрезентативные фармакофорные модели. Далее рассматривается подход общих хитов, молекула соответствует трем репрезентативным моделям из шести, согласно подходу общих хитов ее оценка (или score) будет равен $3/6 = 0.5$. Далее рассматривается подход охвата конформеров, четыре конформера молекулы из пяти соответствуют репрезентативным моделям, поэтому согласно этому подходу score молекулы будет равен $4/5 = 0.8$.

В.Б. Сулимов: С7: Что такое соединения приманки? Это неактивные по данному белку соединения? А кто доказал их неактивность по данному белку? Почему коэффициент обогащения больше единицы? Или это проценты? Ну вообще-то коэффициент обогащения это площадь под кривой обогащения - так же как AUC это площадь под ROC-кривой! Или используемое Вами определение эквивалентно построению этих кривых? Приведенные Вами 4 лиганда это что? Это нативные лиганды из соответствующих комплексов из Protein Data Bank? Я думал, что Вы вычисляете коэффициент обогащения используя набор из известных активных соединений и неактивных соединений. А что Вы делаете с этими четырьмя лигандами? Для них известны характеристики связывания с мишенью?

Д.И. Баширова: (С. 7) Соединения-приманки - это известные неактивные соединения по отношению к данному белку, соединения-приманки и известные ингибиторы были взяты из базы данных DUD-E, которая обеспечивается лабораторией Shoichet (UCSF). Enrichment factor = precision / baseline precision, для используемого набора данных baseline precision = 0,0167, в результате чего коэффициент обогащения может быть больше единицы. В работе исследуется четыре комплекса одного белка с четырьмя лианами, обладающими высокой аффинностью к данной мишени, структуры комплексов взяты из базы данных PDB. Коэффициент обогащения вычисляется на основе известных активных и неактивных соединений из базы данных DUD-E. Четыре разных лиганда были отобраны для более подробного изучения нашего подхода.

В.Б. Сулимов: С8: Чем определяется учитываемое число фармакофорных центров для данной молекулы? Почему у Вас несколько вариантов для каждой молекулы?

Д.И. Баширова: (С. 8) На данном слайде приводится анализ оптимального выбора минимального количества фармакофорных центров в модели фармакофора, таким образом, по результатам

скрининга можно сделать вывод, что более сложные модели приводят к более высокой точности виртуального скрининга, но уменьшает общее количество хитов.

В.Б. Сулимов: С9: вообще нельзя понять. Поясните суть.

Д.И. Баширова: (С. 9) На данном слайде представлено сравнение подходов ранжирования: подход охвата конформеров (указан пунктирной линией), подход общих хитов (указан сплошной линией). Чем выше коэффициент обогащения, тем лучше подход ранжирования, подход охвата конформеров продемонстрировал более высокие показатели коэффициента обогащения.

В.Б. Сулимов: С11: п.1 "В работе были продемонстрированы преимущества использования фармакофоров, выявленных из молекулярно-динамических траекторий". Преимущества по сравнению с чем?

Д.И. Баширова: (С. 11) В работе были продемонстрированы преимущества использования фармакофоров, выявленных из молекулярно-динамических траекторий по сравнению с фармакофорами, извлеченными из комплекса белок-лиганд, структура которых взята из базы данных PDB (сравнение представлено на слайде 10): фармакофоры PDB являются либо очень специфичными, либо слишком простыми, что затрудняет их применение.