

УДК 004.942, 519.876.5, 577.151.042, 577.151.43

Лизунов А.Ю.

*ФГУП «Научно–исследовательский институт прикладной акустики»;
ул. 9 Мая, д. 7А, г. Дубна Московской области, 141980; al_bio@mail.ru*

**ВЛИЯНИЕ ВЫБОРА ТРЕХМЕРНОЙ СТРУКТУРЫ БЕЛКА
НА РЕЗУЛЬТАТЫ ВИРТУАЛЬНОГО СКРИНИНГА
НА ПРИМЕРЕ БЕЛКА КАСПАЗА–7**

Получена 17 августа 2010 года

Опубликована 6 сентября 2010 года

05.13.18 Математическое моделирование, численные методы и комплексы программ

В работе рассматриваются шесть различных структур белка каспаза–7 из Protein Data Bank. Между структурами вычисляется расстояние, определяемое как среднееквадратичное расстояние между соответствующими атомами активных сайтов белковых структур после совмещения. С помощью программы Алгокомб производится виртуальный скрининг набора веществ в выбранные структуры. Показано, что несмотря на малое расстояние между выбранными структурами, результаты виртуального скрининга в эти структуры качественно отличаются.

Ключевые слова: виртуальный скрининг, влияние начальных данных, зависимость от начальных данных, докинг, влияние структуры белка, каспаза –7.

ВВЕДЕНИЕ

Один из современных подходов к созданию лекарственных средств включает в себя компьютерное моделирование и расчёт ингибирующей активности веществ в отношении интересующего белка–мишени. Общепринятым источником используемой при расчётах информации о пространственной структуре белков является открытая база данных PDB (Protein Data Bank) [3]. Однако для некоторых белков в PDB представлено по несколько различных расшифровок пространственной структуры белка, полученных в разных условиях, в виде комплексов с различными соединениями или разными методами, что неизбежно приводит к вопросу: как влияет выбор определённой структуры из PDB на результат расчётов? В данной статье исследуется влияние различий в пространственных структурах белка, представленных в PDB, на результат виртуального скрининга на примере белка каспаза–7.

МАТЕРИАЛЫ И МЕТОДЫ

Структуры каспазы-7 и их сравнение

В PDB представлено 19 структур каспазы-7. Для изучения были отобраны 6 комплексов (таблица 1) согласно следующим критериям: каспаза должна быть в активированной форме [5] и отобранные комплексы должны быть достаточно разнообразны.

Таблица 1. Используемые для анализа структуры. APO- и HOLO- – стандартные приставки для обозначения структуры белка в не связанном с лигандом состоянии и в связанном состоянии, соответственно

<i>PDB ID</i>	<i>Краткое описание</i>	<i>Год публикации</i>	<i>Разрешение снимка (ангстремы)</i>
1F1J	HOLO-структура с небольшим ковалентно связанным пептидным ингибитором.	2000	2.35
1I4O	HOLO-структура с большим (17 аминокислотных остатков) белковым ингибитором XIAP, связанным нековалентно.	2001	2.40
1I51	HOLO-структура с большим (17 аминокислотных остатков) белковым ингибитором XIAP, связанным нековалентно.	2001	2.45
3IBF	APO-структура активированной каспазы7	2009	2.50
3IBC	HOLO-структура с небольшим ковалентно связанным пептидным ингибитором.	2009	2.75
3EDR	HOLO-структура с небольшим ковалентно связанным пептидным ингибитором.	2008	2.45

Между выбранными комплексами было рассчитано «расстояние» (таблица 2). Расстояние между комплексами определялось как корень из среднего квадрата расстояния между соответствующими атомами сайтов связывания (RMSD) после совмещения белков.

Таблица 2. Расстояние между структурами, определяемое как среднеквадратичное отклонение положений атомов активного сайта белка. Оранжевым выделены структуры, при виртуальном скрининге в которые не удастся отличить активные вещества от неактивных.

	1F1J	1I4O	1I51	3EDR	3IBC	3IBF
1F1J		1.12	1.33	0.64	1.09	1.27
1I4O			0.57	1.13	1.17	0.78
1I51				1.1	1.22	0.79
3EDR					0.83	1.11
3IBC						1.16
3IBF						

Для совмещения сайтов связывания и расчёта RMSD использовалась написанная на C++ программа, реализующая алгоритм совмещения макромолекул как набора жестко связанных точек (атомов) с минимизацией RMSD на основе первичного «якоря». На первом этапе программа совмещает два белка с минимизацией RMSD для указанных пар атомов. На втором этапе программа из атомов, входящих в сайт связывания белка, выбирает такие, для которых расстояние после первого этапа оказалось менее порогового значения (1 ангстрем), и строит второе преобразование движения, совмещающее белки с минимизацией RMSD для выбранных атомов. Такой алгоритм совмещения белков отражает наличие консервативных и переменных участков в сайте связывания.

Сайт связывания определялся как совокупность атомов, находящихся на расстоянии менее порогового (7 ангстрем) от лиганда из комплекса 3IBC после первого этапа совмещения.

Виртуальный скрининг тестовых веществ

Из электронной базы данных BindingDB [4] были выбраны две группы веществ: активные и неактивные в отношении каспазы-7. В качестве активных были взяты 123 вещества, для которых в BindingDB указаны данные по связыванию с каспазой-7. В качестве неактивных были выбраны 2417 веществ,

для которых в BindingDB нет данных по связыванию с каспазой–7, но есть данные по связыванию с другими протеазами. Такой выбор неактивных веществ позволил сразу исключить вещества, очевидно не являющиеся хорошими кандидатами для связывания с белком.

Процедура поиска оптимальной позиции лиганда в сайте связывания белка называется «докинг». В каждую из шести выбранных структур каспазы–7 с помощью программы Алгокомб [1] был выполнен докинг набора из активных и неактивных веществ и рассчитана оценка качества связывания.

Программа докинга рассчитывает оценку качества связывания, которую в дальнейшем мы будем называть словом «скор» (от английского слова score). Основное требование, предъявляемое к скору, – это возможность с его помощью ранжировать вещества по свободной энергии образования комплекса белок–лиганд и таким образом отличать вещества, прочнее связывающиеся с данным белком–мишенью. В идеале скор должен равняться свободной энергии комплекса белок–лиганд, однако для правильной сортировки лигандов по энергии связывания достаточно того, чтобы скор являлся монотонной функцией от свободной энергии.

В используемой нами для докинга программе Алгокомб скор вычисляется на основе двух компонент: энергетической и «облачной». Энергетическая компонента вычисляется на основе эмпирического атом–атомного сора Тарасова–Мурышева [2] и дает оценку свободной энергии комплекса. «Облачная» составляющая сора отражает локальное сходство оцениваемого комплекса белок–лиганд с известными комплексами из PDB и не имеет энергетического смысла [1].

Для каждой из шести структур каспазы–7 была построена зависимость эмпирической плотности распределения активных и неактивных веществ по скору.

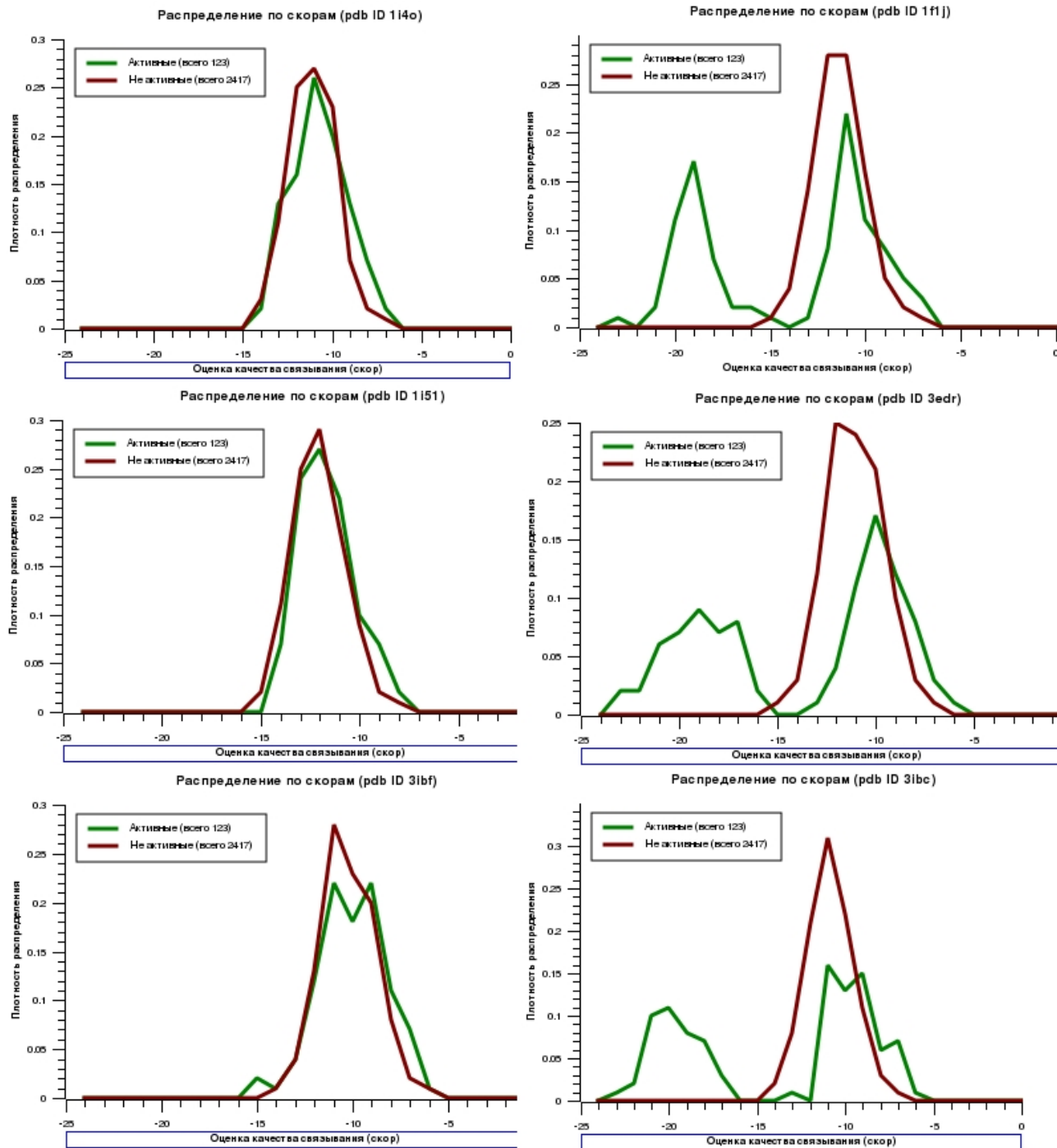


Рис. 1: Графики эмпирической плотности распределения по скорам для ингибиторов (зеленый) и неингибиторов (красный) каспазы-7 при докинге в различные структуры из PDB. Для трех структур – 1i4o, 1i51 и 3ibf — графики фактически идентичны. Для трех других структур – 1f1j, 3edr и 3ibc — у распределения ингибиторов есть второй пик в районе сора –19, что позволяет при виртуальном скрининге отличить часть ингибиторов от неингибиторов по скору

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЯ

Из приведённых графиков видно, что для комплексов 1i4o, 1i5l и 3ibf распределения активных и неактивных веществ фактически совпадают. Это означает, что отличить активные вещества от неактивных по скору при докинге в эти комплексы невозможно.

Совершенно другой тип распределений получается при докинге в комплексы 1flj, 3edr и 3ibc. Распределение неактивных веществ остается таким же, как при докинге в первые три комплекса (распределение Гауссовского типа с максимумом в районе скоры -10), в то время, как распределение активных веществ имеет два ярко выраженных максимума: в районе скоры -10 и в районе скоры -19 . Объяснить такую форму распределений активных веществ по скору можно гибкостью полипептидных цепей белка: из 133 активных веществ часть связывается с белком в конформации, похожей на представленную в структурах 1flj, 3edr и 3ibc, поэтому для них оценка качества связывания резко возрастает (по модулю), в то время как остальные вещества связываются с молекулами белка в других конформациях, и при докинге в эти структуры оценка качества связывания для них получается такой же, как для неактивных веществ.

Такой вид распределений по скорам для активных и неактивных веществ при докинге в структуры 1flj, 3edr и 3ibc позволяет надеяться, что произвольные вещества, получающие при докинге в эти структуры скор более 16 (по модулю), будут связываться с каспазой-7. То есть при виртуальном скрининге произвольного набора веществ в конформации белков, представленные в этих комплексах, высока вероятность найти активные вещества, если в качестве кандидатов брать вещества, получившие скор более 16, что позволяет искать новые ингибиторы каспазы-7 среди, например, веществ из открытых баз данных.

Таким образом, изученные шесть комплексов каспазы–7 распадаются на две группы: 1) 1i4o, 1i51 и 3ibf – комплексы, не позволяющие отличить активные вещества на основе виртуального скрининга, и 2) 1flj, 3edr и 3ibc – комплексы, позволяющие отличить часть активных веществ.

Из таблицы 2 видно, что расстояние между всеми шестью рассмотренными структурами каспазы–7, определяемое как среднеквадратичное отклонение положений атомов сайта после совмещения, не превосходит 1,5 ангстрем. Внутри группы комплексов, не позволяющих отличить активные вещества, расстояние между комплексами не превосходит 1 ангстрем, в то время как внутри группы комплексов, позволяющих отличить активные вещества, расстояние между комплексами колеблется от 0,65 до 1,1 ангстрем. Разрешение рассматриваемых рентгеновских структур каспазы–7 колеблется от 2,35 до 2,75 ангстрем (см. таблицу 1). Таким образом, по усреднённому значению отклонения всех атомов сайта, на фоне погрешности расшифровки рентгеновских структур, значимого отличия между комплексами из первой и второй группы нет.

Более детальный анализ показывает, что резкое возрастание сора для части известных ингибиторов каспазы–7 при докинге в комплексы из второй группы связано с положением единственного аминокислотного остатка – тирозина TYR230, этот остаток поворачивается в сторону и открывает доступ к остаткам GLN184, ARG87 и ARG233, расположенным в полости внутри белка, с которыми известные ингибиторы каспазы–7 образуют водородные связи. На рисунке 2 показаны активные сайты белка с закрытым и открытым положением тирозина 230.

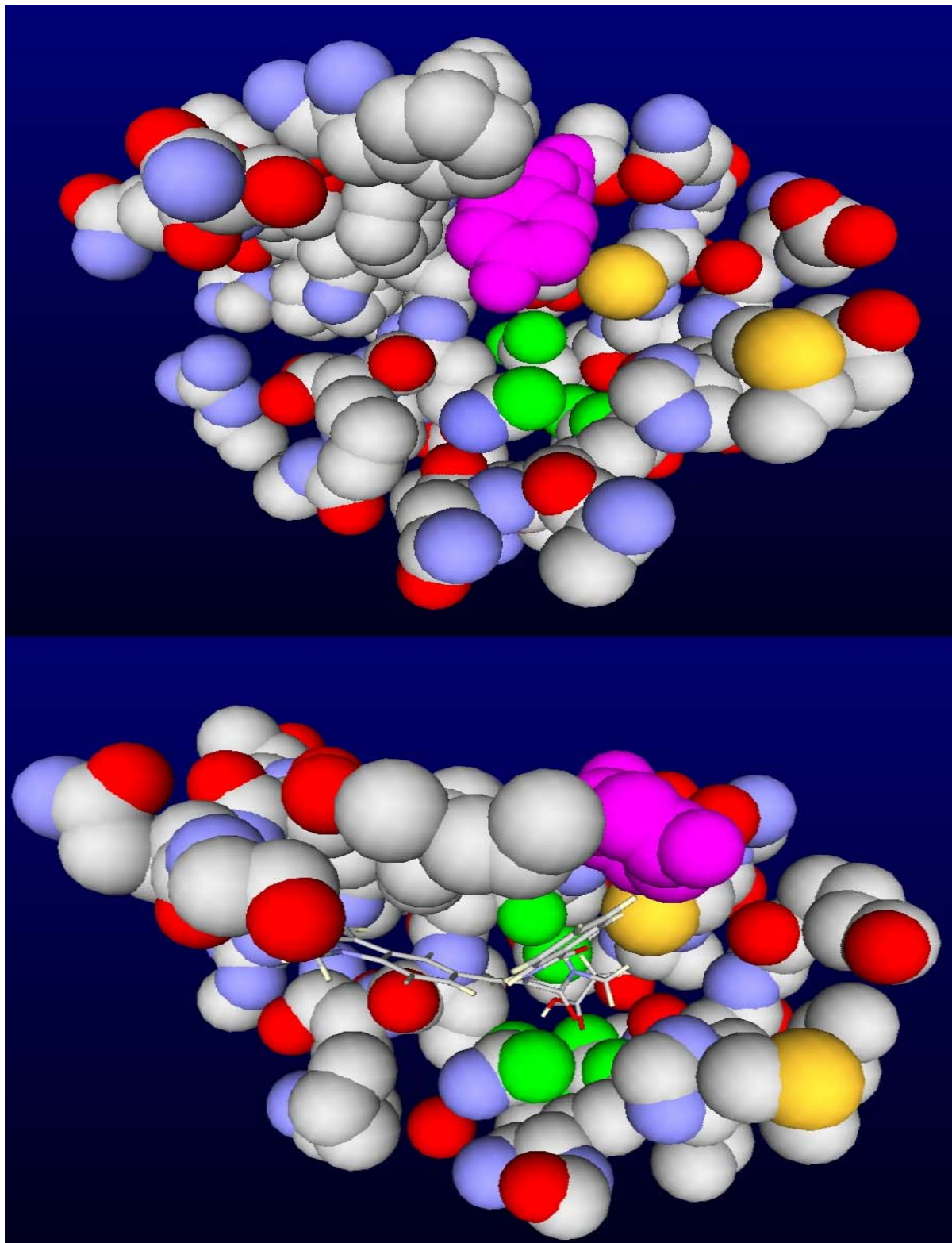


Рис.2: Активные сайты из комплексов 3ibf и 3ibc, соответственно. Фиолетовым выделен тирозин 230, зеленым — закрываемые им атомы белка, способные образовывать водородные связи. На первом рисунке тирозин 230 в «закрытом», а на втором — в «открытом» положении.

ВЫВОДЫ

Проведённое исследование показало, что малое расстояние между структурами одного и того же белка, рассчитанное как среднеквадратичное расстояние между соответствующими атомами активных сайтов после совмещения, ещё не гарантирует эквивалентных результатов виртуального скрининга веществ в эти структуры. Поэтому при практическом поиске ингибиторов расчётными методами следует крайне внимательно относиться к выбору структуры белка для докинга и по возможности проводить предварительные калибровочные запуски программы докинга с веществами, для которых уже известна ингибирующая активность.

СПИСОК ЛИТЕРАТУРЫ

1. Ramensky V., Sobol A., Zaitseva N., Rubinov A., Zosimov V. A novel approach to local similarity of protein binding sites substantially improves computational drug design results // *Proteins*. 2007. V. 69(2). P. 349–57.

2. A.E. Muryshev, D.N. Tarasov, A.V. Butygin, O.Yu. Butygina, A.B. Aleksandrov and S.M. Nikitin. A novel scoring function for molecular docking // *J. of Comp.–Aided Mol. Design*. 2003. V. 17. P. 597–605.

3. RCSB Protein Data Bank (PDB). Электронная база данных расшифрованных трехмерных структур белковых комплексов PDB. [Электронный ресурс] // [сайт the Research Collaboratory for Structural Bioinformatics (RCSB)]. [2003–2010]. URL: <http://www.rcsb.org/pdb/> (дата обращения: 15.02.2010).

4. Электронная база данных по константам связывания органических веществ с белками BindingDB. [Электронный ресурс] // [сайт the Center for Advanced Research in Biotechnology, a center of the University of Maryland Biotechnology Institute, Rockville, Maryland, USA]. [2010].

URL: <http://www.bindingdb.org>. (дата обращения: 15.02.2010).

5. Riedl S.J., Fuentes–Prior P., Renatus M., Kairies N., Krapp S., Huber R., Salvesen G.S., Bode W. Structural basis for the activation of human procaspase–7 // Proc. Natl. Acad. Sci. USA. 2001 Dec 18; 98(26). P. 14790–5.