

КОМПЬЮТЕРНОЕ КОНСТРУИРОВАНИЕ ЛЕКАРСТВЕННЫХ СРЕДСТВ

УДК 577.1.615.2

©Коллектив авторов

МОДЕЛИРОВАНИЕ ТРЕХМЕРНОЙ СТРУКТУРЫ ЦИТОХРОМА P450 1A2 И ПОИСК ЕГО НОВЫХ ЛИГАНДОВ

Н.В. БЕЛКИНА, В.С. СКВОРЦОВ, А.С. ИВАНОВ, А.И. АРЧАКОВ.

НИИ Биомедицинской химии РАМН, Москва, Россия, 119832, Москва,
Погодинская ул., 10, Телефон (095) 245-0768, Факс (095) 245-0857

Вещества, способные ингибировать цитохром P450 1A2 (CYP1A2), представляют собой перспективный класс новых лекарственных средств, применение которых в клинической практике может стать важным звеном в профилактике онкологических заболеваний. Данная работа посвящена компьютерному моделированию трехмерной структуры CYP1A2 и поиску новых ингибиторов с помощью скрининга баз данных низкомолекулярных соединений.

Моделирование CYP1A2 основывалось на гомологии с бактериальными цитохромами P450 с известной трехмерной структурой. С целью оптимизации структуры активного центра CYP1A2 были созданы модели его комплексов с характерными субстратами (кофеином и этоксирезурфином) с последующей их оптимизацией методами молекулярной динамики в присутствии растворителя. Далее, средствами программ DockSearch и LeapFrog были созданы модели 24 комплексов CYP1A2 с известными лигандами, для которых известны экспериментальные значения константы диссоциации (K_d). На основе этих комплексов была построена 3D-QSAR модель с хорошей предсказательной силой. На заключительном этапе был выполнен поиск лигандов CYP1A2 в базе низкомолекулярных соединений (более 23000 веществ из базы данных фирмы Maybridge и 112 известных лигандов CYP1A2 из базы данных metabolite фирмы MDL). Было получено 680 потенциальных лигандов CYP1A2 с предсказанной величиной K_d , сравнимой с известными лигандами. В это число вошли соединения из 112 известных лигандов, введенных в качестве внутреннего контроля в тестируемую базу данных.

Ключевые слова: цитохром P450 1A2, компьютерное моделирование, трехмерная структура, молекулярный докинг, 3D-QSAR.

ВВЕДЕНИЕ. Цитохром P450 1A2 (CYP1A2) является членом надсемейства гем-содержащих монооксигеназ, имеющих почти у всех видов живых существ. Цитохромы отвечают за окислительный метаболизм широкого спектра как экзогенных, так и эндогенных низкомолекулярных веществ. В частности, субстратами CYP1A2 являются многие экзогенные канцерогены и лекарства, а также эндогенные стероиды. Вещества, способные ингибировать CYP1A2, представляют собой перспективный класс новых лекарственных средств, применение которых в клинической практике может стать важным звеном в профилактике онкологических заболеваний как при лекарственной

пе
пр
ра
ег

P4
С
7-
пр
мо
эк
по
вь
ве
М
пр
73
те

ст

содер
P450
эндо
неко
спос
лекар
важн

терапии, так и при контакте с вредными химическими соединениями на производстве, в быту или при проживании в экологически неблагополучных районах.

Одним из перспективных направлений в создании новых ингибиторов является поиск новых базовых структур потенциальных ингибиторов в компьютерных банках данных, содержащих описание сотен тысяч структур низкомолекулярных веществ. Поиск может осуществляться с помощью различных методов и подходов. Методы, основанные на знании структуры места связывания лиганда в макромолекуле-мишени, являются наиболее современными и называются прямыми методами поиска. В их основе, как правило, лежат процедуры молекулярного докинга. В качестве примеров могут быть приведены такие компьютерные программы как GRID [1] и DOCK [2]. В нашей лаборатории разработана программа DockSearch [3,4], позволяющая осуществлять докинг большого числа низкомолекулярных соединений из базы данных в активный центр фермента в итерационном режиме. Фактически эта программа работает как генератор гипотез возможных комплексов, оценка и отбор которых может быть осуществлена на последующих этапах с помощью различных критериев и дополнительных методов анализа.

Существенным ограничением использования прямых методов поиска является то, что для их осуществления необходимо знание трехмерной структуры комплекса макромолекулы-мишени как минимум с одним из известных лигандов. Наиболее благоприятной ситуацией для осуществления подобного поиска является наличие нескольких комплексов и набор известных лигандов с экспериментально установленными биохимическими характеристиками их взаимодействия с мишенью, например, константа диссоциации комплекса (K_d). К сожалению такая идеальная стартовая ситуация бывает редко. В большинстве случаев, в том числе и в случае с CYP1A2, трехмерная структура макромолекулы-мишени, так же как и ее комплексов с известными лигандами, отсутствует. Однако эта проблема может быть преодолена путем создания трехмерной компьютерной модели макромолекулы-мишени и ее комплексов с известными лигандами. При наличии хорошего набора известных лигандов с экспериментально установленными величинами K_d может быть построена 3D-QSAR модель, позволяющая компенсировать неизбежные неточности в модели активного центра путем оценки возможных величин K_d для найденных потенциальных лигандов. Данный подход пока редко используется, но на наш взгляд является наиболее перспективным.

В настоящей работе он реализован для поиска новых потенциальных ингибиторов CYP1A2. В базе данных Protein Data Bank (PDB) [5] существуют данные о четырех цитохромах P450, для которых известны трехмерные структуры, что создает предпосылку для успешного моделирования структуры CYP1A2 по гомологии с этими белками. С другой стороны, известно довольно много субстратов CYP1A2, что может позволить уточнить строение активного центра белка и получить независимую оценку результатов поиска новых лигандов путем предсказания K_d с помощью 3D-QSAR модели.

МЕТОДИКА. Молекулярные объекты, программные и вычислительные средства.

В работе были использованы следующие базы данных: белковые банки данных PDB [5] и SwissProt [6]; базы данных низкомолекулярных соединений Maybridge [7] и Metabolite [8].

Трехмерные координаты и аминокислотные последовательности четырех бактериальных P450 (CYP101, CYP102, CYP108, CYP107A1) были получены из банка PDB (2CP4, 2HPD, 1CPT, 1OXA соответственно).

Аминокислотная последовательность CYP1A2 получена из банка SwissProt (P05177). Перед моделированием трехмерной структуры CYP1A2 N-концевые

аминокислотные остатки с 1 по 48 были удалены, так как они соответствуют мембранной части CYP1A2, отсутствующей у бактериальных P450.

Из различных литературных источников были подобраны данные о 24 известных лигандах CYP1A2 (Таблица 1) с известными величинами Kd.

Таблица 1. Лиганды, для которых известны величины констант диссоциации.

Номер	Лиганд	Номер	Лиганд
1	2-Фенилпропанол	13	S-Пропранолол
2	7-Метоксирезурфин	14	Альфа-Нафтофлавон
3	7-Этоксикумарин	15	Ацетаминофен
4	7-Этоксирезурфин	16	Бета-Нафтофлавон
5	Abbott-66193	17	Бунитролол
6	R-1-(1-нафтил)этиламин	18	Дантролен
7	R-1-(4-пиридил)этанол	19	Диоксин
8	R-1-циклогексилэтиламин	20	Зилеутон (Zileuton)
9	R-Пропранолол	21	Кофеин
10	S-1-(1-нафтил)этиламин	22	Нарингенин
11	S-1-(4-пиридил)этанол	23	Теофиллин
12	S-1-циклогексилэтиламин	24	Фенацетин

Все процедуры молекулярного моделирования, энергетической оптимизации и молекулярной динамики осуществлялись с помощью программного комплекса Sybyl (Tripos, США) [9]. 3D-QSAR анализ и оценка величины Kd для лигандов CYP1A2 выполнялись средствами программ SPSS и Fiesta [10].

Скрининг баз данных был осуществлен с помощью оригинальной программы DockSearch [3,4]. Все работы были выполнены на рабочей графической станции Indigo-2 (R4400-150) фирмы Silicon Graphics (США).

Моделирование трехмерной структуры CYP1A2 и комплексов белок-лиганд.

С помощью программы COMPOSER (модуль программного комплекса Sybyl) был выполнен сравнительный анализ пространственных структур четырех бактериальных P450 и их аминокислотных последовательностей с CYP1A2. Были определены структурно-консервативные регионы (SCR) в последовательности CYP1A2 (рис.1). Трехмерные координаты для SCR в модели CYP1A2 были взяты из соответствующих участков бактериальных цитохромов P450. Для петель белковой цепи, которые не вошли в состав SCR, координаты подбирались из гомологичных участков различных белков из PDB или создавались *de novo*. Полученная структура CYP1A2 была оптимизирована с помощью процедуры минимизации энергии.

Были построены две модели комплексов CYP1A2 с известными субстратами (кофеином и 7-этоксирезурфином). Для этого использовали процедуру молекулярного докинга (программа DockSearch) с последующей оптимизацией активного центра методами молекулярной динамики (алгоритм Verlet) в присутствии воды (давление 1 атм, температура 300 К, шаг - 1 фс, длительность - 10 пс для комплекса с кофеином и 35 пс для комплекса с 7-этоксирезурфином). Были также построены модели комплексов CYP1A2 с 24 известными лигандами (см. таблицу 1). Оптимизация структуры комплексов была осуществлена с помощью программы LeapFrog (модуль программного комплекса Sybyl).

Создание тестируемой базы данных низкомолекулярных соединений.

Тестируемая база данных, содержащая трехмерные структуры низкомолекулярных соединений, была составлена из двух частей: около 23000 соединений (с молекулярным весом от 150 до 500) из базы Maybridge и контрольная выборка из базы Metabolite, содержащая 112 известных лигандов CYP1A2.

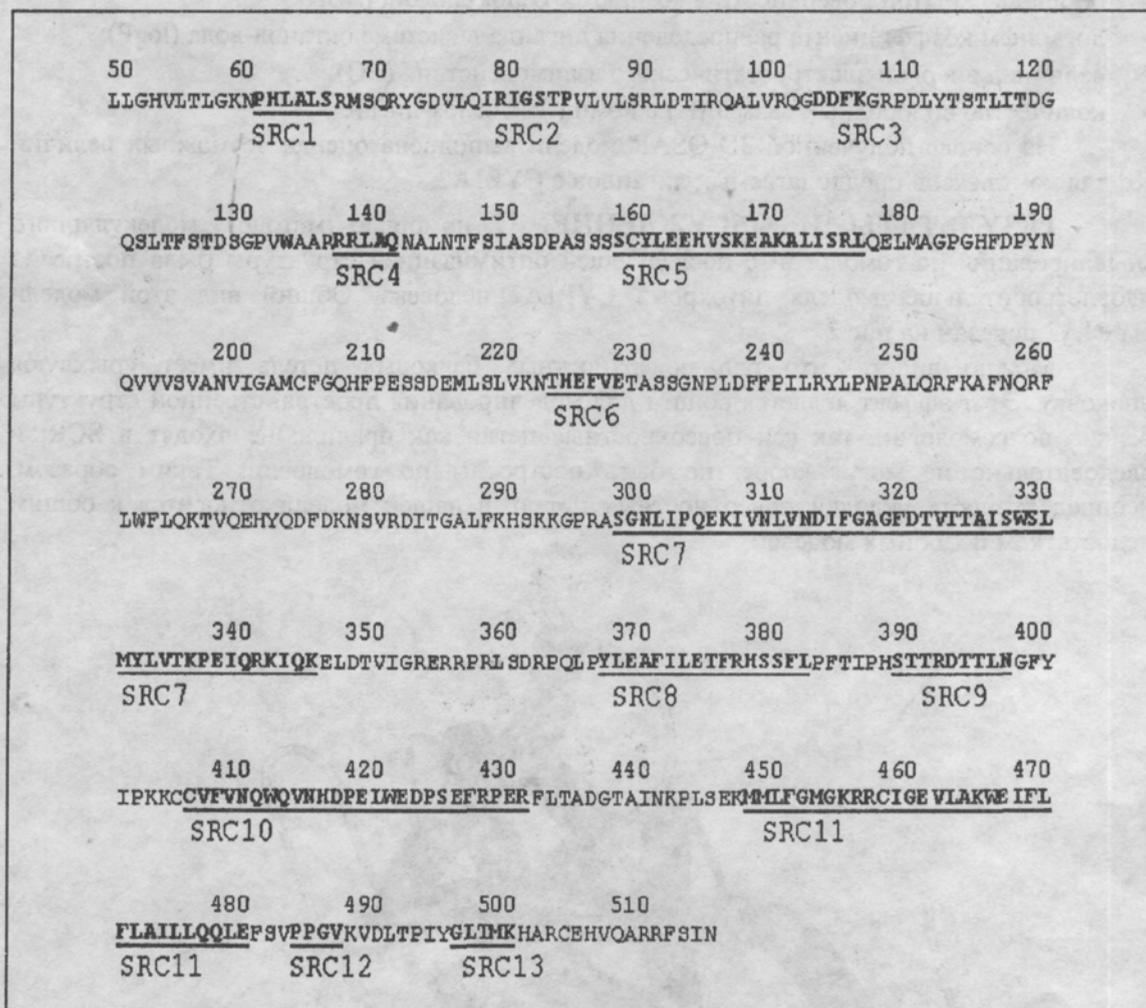


Рисунок 1.
Аминокислотная последовательность CYP1A2, где выделены структурно-консервативные регионы (SCR).

Скрининг базы данных.

Поиск новых лигандов CYP1A2 в тестируемой базе данных был осуществлен средствами программы геометрического докинга DockSearch. В качестве белка-мишени использовалась построенная на первом этапе работы модель CYP1A2. Среди полученных первичных гипотез о строении комплексов лиганд-белок были отобраны наиболее энергетически выгодные с использованием программы LeapFrog (модуль программного комплекса Sybyl).

Оценка величины Kd для лигандов CYP1A2.

С помощью методов регрессионного анализа и программы Fiesta [10] было выполнено построение 3D-QSAR модели, в основу которой была положена информация о структурах 24 известных лигандов CYP1A2, их пространственном взаиморасположении в модельных комплексах с CYP1A2 и экспериментальных значениях Kd. Моделирование комплексов выполнялось путем докирования каждой молекулы в полость активного центра CYP1A2, что позволило учесть особенности связывания каждого лиганда с активным центром. При построении 3D-QSAR модели в качестве предсказываемой величины использовалось значение десятичного логарифма Kd умноженное на 1000. В качестве независимых переменных использовались следующие параметры модельных комплексов:

- площадь скрытой поверхности в комплексе белок-лиганд (SURF);
- логарифм коэффициента распределения лиганда в системе октанол-вода ($\log P$);
- величина энергии электростатического взаимодействия (EQ);
- количество водородных связей (H) в комплексе белок-лиганд.

На основе полученной 3D-QSAR модели выполнена оценка возможных величин K_d для комплексов предполагаемых лигандов с CYP1A2.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ. С помощью методов молекулярного моделирования по гомологии с последующей оптимизацией структуры была построена удовлетворительная модель цитохрома CYP1A2 человека. Общий вид этой модели CYP1A2 показан на рис.2.

Можно видеть, что ряд поверхностных белковых петель имеет «рыхлую» упаковку. Этот эффект является общим для моделирования пространственной структуры белков по гомологии, так как поверхностные петли как правило не входят в SCR, и следовательно не могут корректно быть построены по гомологии. Таким образом, неопределенность укладки поверхностных петель в нашей модели относится к общим недостаткам подобных моделей.

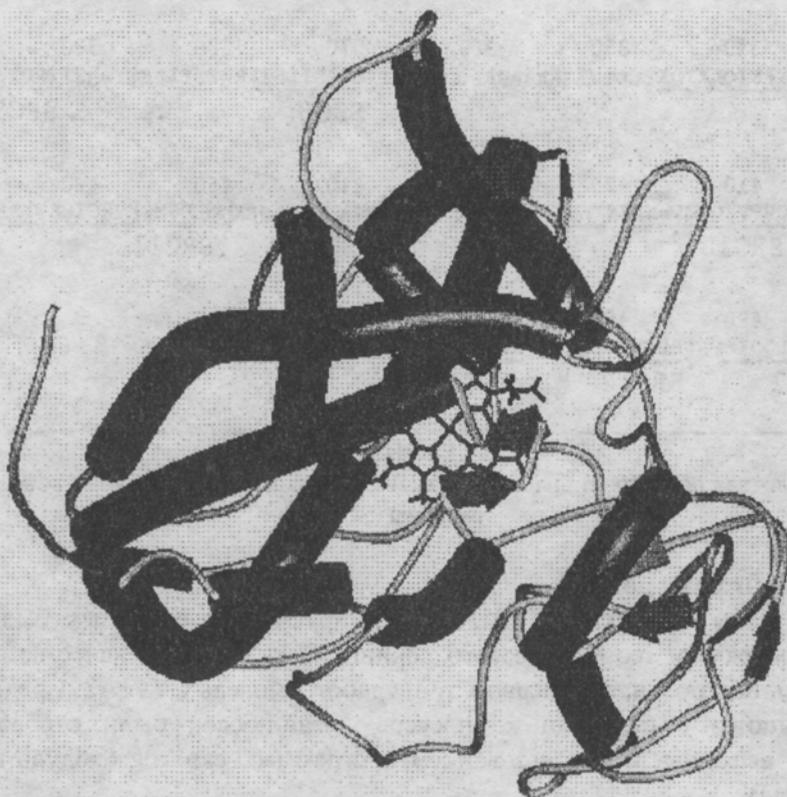


Рисунок 2

Компьютерная модель трехмерной структуры CYP1A2. α -спирали представлены в виде труб, β -слои представлены в виде плоских стрелок.

На рис.3 показана область активного центра CYP1A2. Сравнительный анализ последовательностей CYP1A2 человека и крысы показал, что ASP-320 и THR-321 в цитохроме человека соответствуют GLU-318 и THR-319 цитохрома крысы, которые находятся в активном центре фермента и принимают участие во взаимодействии с субстратом [11].

Моделирование с помощью молекулярной динамики позволило проследить изменения взаимного расположения, как молекулы субстрата, так и элементов активного

центра CYP1A2. На рис. 4а и 4б представлены изменения во времени расстояний между кислородом, связанным с атомом железа гема, и некоторыми функционально важными атомами субстратов. Результаты молекулярной динамики совпадают с экспериментальными данными о том, что основным путем метаболических превращений кофеина в организме человека является деметилирование в третьем положении ксантинового цикла. Соответственно для 7-этоксирезурфина единственно возможным путем является О-деэтилирование.

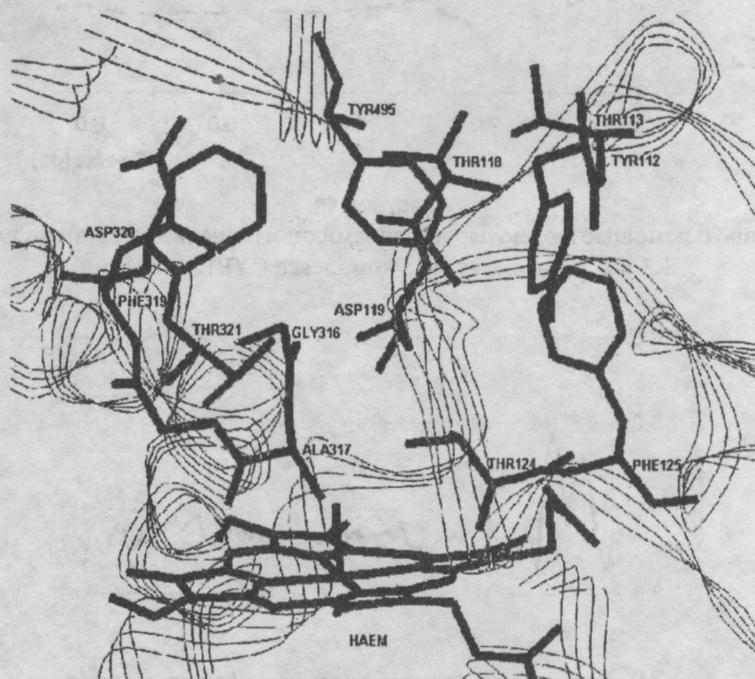
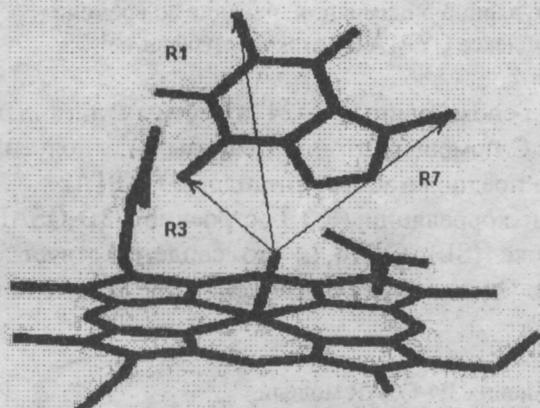
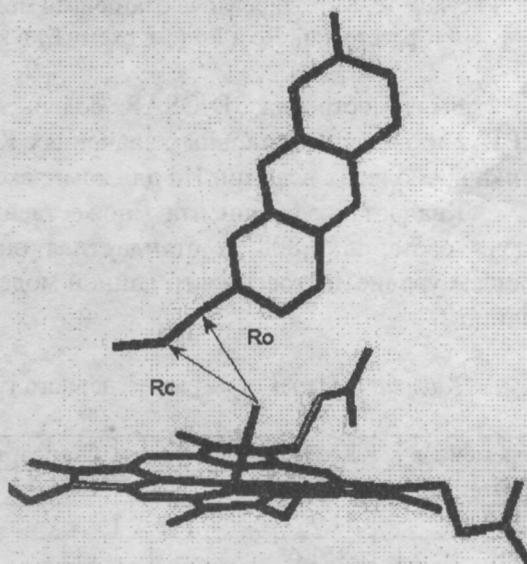


Рисунок 3.

Модель структуры активного центра CYP1A2.



А. Схема взаиморасположения гема и молекулы кофеина в комплексе CYP1A2/кофеин.



Б. Схема взаиморасположения гема и молекулы 7-этоксирезурфина в комплексе CYP1A2/7-этоксирезурфин.

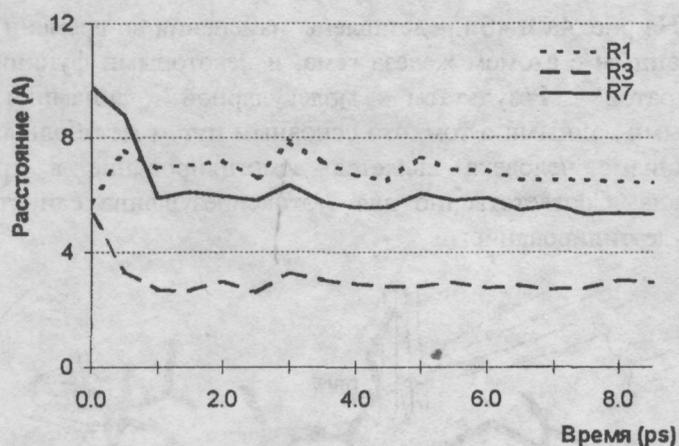


Рисунок 4а.

Моделирование с помощью молекулярной динамики: изменение во времени расстояний R1, R3 и R7 (см. схему А) в комплексе CYP1A2/кофеин.

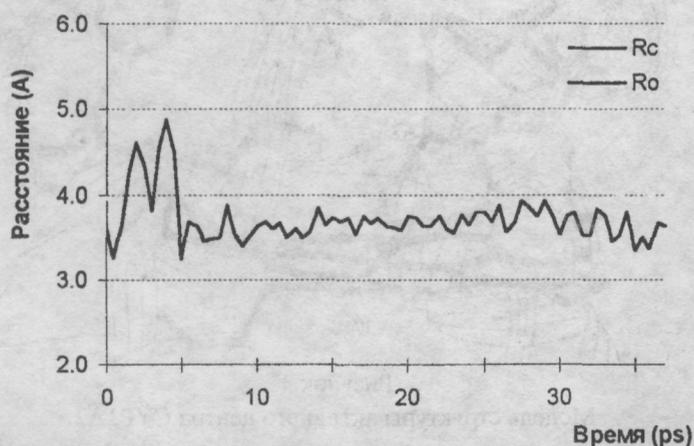


Рисунок 4б. Моделирование с помощью молекулярной динамики: изменение во времени расстояний Rc и Ro (см. схему Б) в комплексе CYP1A2/7-этоксирезурфин.

Была построена 3D-QSAR модель на основе данных о 24 известных лигандах CYP1A2 и экспериментальных значениях Kd. С помощью этой модели была выполнена оценка возможных величин Kd для комплексов предполагаемых лигандов с CYP1A2.

Квадрат коэффициента множественной корреляции (R^2) построенной 3D-QSAR модели составил 0,808, а стандартная ошибка (SE) – 775,4, что свидетельствует о хорошем уровне достоверности данной модели. Окончательное уравнение представлено в таблице 2.

Таблица 2. Параметры регрессионного уравнения 3D-QSAR модели.

Переменная	Коэффициент (?)	Стандартная ошибка (SE)	Уровень значимости (p)
Константа	8454,9	1034,3	0,000
SURF	54,9	7,8	0,000
LogP	-451,6	139,2	0,004
EQ	94,4	26,5	0,002
H	724,9	257,9	0,011

На рис.5 показана корреляционная зависимость между реальными и рассчитанными по регрессионному уравнению величинами K_d для известных лигандов СУР1А2, на основе которых была построена 3D-QSAR модель.

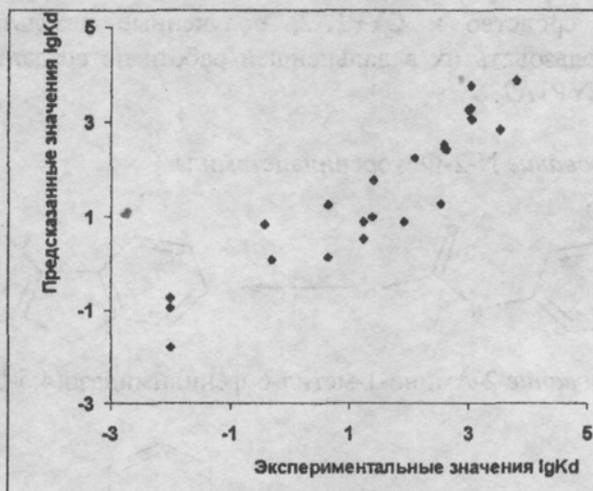


Рисунок 5.

Корреляционная зависимость между предсказанными и экспериментально определенными значениями K_d (мкМ) для 3D-QSAR модели.

В результате скрининга базы данных методом геометрического докинга с последующей энергетической минимизацией было отобрано 1.410 предполагаемых лигандов и для каждого из них были предсказаны значения K_d . Распределение всех 1.410 лигандов в зависимости от величины предсказанной K_d представлено на рис.6.



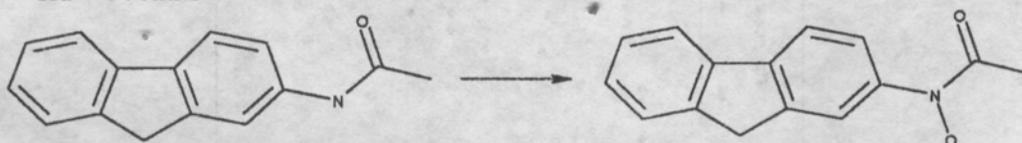
Рисунок 6.

Распределение 1.410 выбранных комплексов, основанное на величине предсказанной K_d (мкМ).

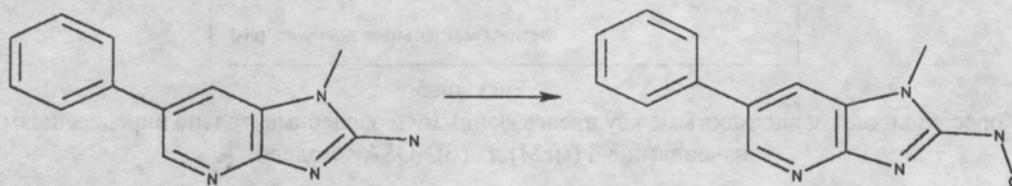
Из полученной выборки было отобрано 680 потенциальных лиганда СУР1А2 с величиной K_d , сравнимой с известными лигандами ($K_d < 70$ мкМ). В это число вошли 73 соединения из 112 известных лигандов, введенных в качестве внутреннего контроля в тестируемую базу данных. На рис.7 в качестве примера показаны три подобных

соединения с соответствующими предсказанными значениями K_d . Таким образом, использование для поиска новых лигандов процедуры докирования в модель активного центра позволило выявить определенное количество потенциальных лигандов СУР1А2. Учитывая, что более двух третей известных лигандов были определены как имеющие достаточно высокое сродство к СУР1А2, полученные результаты можно считать достоверными и использовать их в дальнейшей работе по созданию базовых структур новых ингибиторов СУР1А2.

N-Гидроксилирование *N*-2-Флуоренилацетамида
 $K_d \sim 64 \text{ мкМ}$



N-Гидроксилирование 2-Амино-1-метил-6-фенилимидазо(4,5-*b*)пиридина
 $K_d \sim 1 \text{ мкМ}$



Эпоксидация 6-Аминокризена
 $K_d \sim 26 \text{ мкМ}$

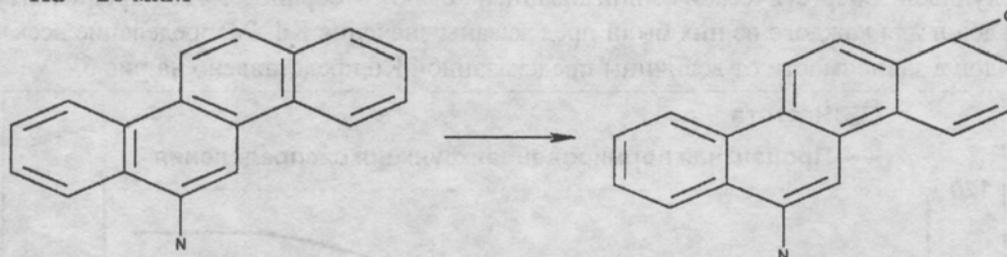


Рисунок 7.

Пример трех известных субстратов СУР1А2, реакции, осуществляемые СУР1А2, и соответствующие предсказанные значения K_d .

Данная работа выполнена при поддержке Российского Фонда Фундаментальных Исследований (грант № 96-07-89360). Авторы также благодарны фирмам Tripos GmbH (Мюнхен, Германия) и MDL Information Systems (Сан Леандро, США) за научную и техническую поддержку.

ЛИТЕРАТУРА

1. Goodford, P.J. J., (1985), *Med. Chem.*, **28**, 849-857.
2. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Landridge and Ferrin, T.E. J., (1982), *Mol. Biol.* **161**, 269-288.
3. DockSearch (<http://lmgdd.ibmh.msk.su/docksearch>).
4. Ivanov A.S., Skvortsov V.S., Archakov A.I., (1997), *FASEB J.*, **11**(9), A785.
5. Bernstein C., Koetzle T.F., Williams G.J.B., Meyer E.F.Jr., Brice M.D., Rodgers J.R., Kennard O., Shimanouchi T., Tasumi M. (1977), *J. Mol. Biol.* **112**, 535-542.

6. *Bairoch A., Apweiler R.*, (1997), *J. Mol. Med.* **75**, 312-316.
7. Maybridge Catalog Database. Ryan Scientific. Columbia, USA.
8. Metabolite ISIS™/Base, Molecular Design Limited Information Systems, Inc., 14600 Catalina Street, San Leandro, California, 94577, USA.
9. Sybyl 6.4, Tripos Inc., 1699 South Hanley Road, St Louis, Missouri, 63144, USA.
10. *Sklenar H., Eisenhaber F., Poncin M., Lavery R.* (1990), *Theoretical Biochemistry and Molecular Biophysics*. (Eds. D.L.Beveridge & R. Lavery). Adenine Press. pp.317-335.
11. *Tuck S.F., Hire K., Shimizu T., Hatano M., Ortiz de Montellano P.R.* (1993), *Biochemistry*, **32**. p.2548-2553

MODELLING OF 3-D STRUCTURE OF CYTOCHROME P450 1A2 AND SEARCHING FOR ITS NEW LIGANDS

N.V. BELKINA, V.S. SKVORTSOV, A.S. IVANOV, A.I. ARCHAKOV

Institute of Biomedical Chemistry, 10 Pogodinskaya St., 119832, Moscow, Russia
Telephone (095) 245-0768, Fax (095) 245-0857

The substances inhibiting cytochrome P450 1A2 (CYP1A2) represent a perspective class of new drugs, which application in clinical practice can become the important part in preventive maintenance in oncology. The present work is devoted to computer modelling of 3-D structure of CYP1A2 and searching of new inhibitors by database mining.

The modelling of CYP1A2 was done based on homology with 4 bacterial cytochromes P450 with known 3-D structure. For optimization of CYP1A2 active site structure the models of its complexes with characteristic substrates (caffeine and 7-ethoxyresorufin) were designed. These complexes were optimized by molecular dynamics simulation in water. The models of 24 complexes of CYP1A2 with known ligands with known K_d were designed by means of DockSearch and LeapFrog programs. 3D-QSAR model with good predictive force was created based on these complexes. On a final stage the search of new CYP1A2 ligands in testing database (more than 23.000 substances from database Maybridge and 112 known CYP1A2 ligands from database Metabolite, MDL) was executed. 680 potential ligands of CYP1A2 with K_d values, comparable with known ones were obtained. This number has included 73 compounds from 112 known ligands, introduced in tested database as the internal control.

Key words: cytochrome P450 1A2, computer modelling, 3-D structure, molecular docking, 3D-QSAR.